Genome-wide analysis in over 1.6 million participants uncovers 147 loci associated with obstructive sleep apnoea

Luis M. García-Marín[1,2], Zuriel Ceja[1,2], Abishna Parasuraman[3], Jia Wen Xu[4], Santiago Díaz-Torres[1,2], Victor Flores-Ocampo[1,2], Asma M. Aman[2,3], Mateo Maya-Martínez[5,6], Xueyan Huang[1,2], Camilla Pasquali[1,7], Aura Aguilar-Roldán[1,8], Bade Uckac[1,9], Fangyuan Cao[1,2], Natalia S. Ogonowski[1,2,3], Nicholas G. Martin[1], Stuart MacGregor[2,3], Xianjun Dong[10,11] , Sarah J. Lewis[12], Mathias Seviiri[2,3,9], Jiao Wang[4,13], and Miguel E. Rentería[1,2,9,*]

1. Brain and Mental Health Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
2. School of Biomedical Sciences, Faculty of Health, Medicine and Behavioural Sciences, The University of Queensland, Brisbane, QLD, Australia
3. Population Health Program, QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia
4. School of Public Health, Sun Yat-sen University, Guangzhou, Guangdong, China
5. McLean Hospital, Belmont MA, USA
6. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
7. Department of Bioscience, Universitá degli Studi di Milano, Milan, Italy
8. Escuela Nacional de Estudios Superiores Unidad Juriquilla, Universidad Nacional Autónoma de México (UNAM), Querétaro, México
9. School of Biomedical Sciences, Faculty of Health, Queensland University of Technology, Brisbane, QLD, Australia
10. Stephen & Denise Adams Center for Parkinson's Disease Research of Yale School of Medicine, New Haven, CT, USA
11. Department of Neurology, Yale School of Medicine, Yale University, New Haven, CT, USA
12. Department of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
13. Greater Bay Area Public Health Research Collaboration, China

* Correspondence: Miguel E. Rentería (miguel.renteria@qimrb.edu.au)

Key words: Sleep, sleep apnoea, genetics, GWAS, meta-analysis, UK Biobank, CLSA, MGB Biobank, Guangzhou Biobank, AGDS.

ABSTRACT

We conducted the largest GWAS meta-analysis for obstructive sleep apnoea (OSA; N_{cases}= 230,657; $N_{controls}$= 1,377,442) using European ancestry genetic data from five countries. We identified 147 independent loci associated with OSA, and estimated SNP-based heritability at 16%. We report six independent loci in a separate African population meta-analysis (N_{cases}= 46,834; $N_{controls}$= 149,192). We observed spatially resolved gene enrichment involving GABAergic and glutamate pathways, synaptic transmission, and cytoskeletal remodelling. OSA-derived polygenic risk scores showed predictive ability for clinician□ascertained OSA status, Fitbit-derived sleep features, and self-reported sleep traits in participants of diverse ancestral backgrounds. We identified putative causal relationships with ADHD, depression, multisite chronic pain, body mass index, and schizophrenia, among others. Our findings demonstrate a robust genetic component underlying OSA risk, independent of body mass index, implicating distinct neurobiological pathways related to synaptic function and corticothalamic feedback loops.

INTRODUCTION

Obstructive sleep apnoea (OSA) is characterised by pharyngeal narrowing and upper airway collapse during sleep, resulting in halted breathing and fragmented non-restorative sleep[1]. Its most prominent symptoms include loud snoring, gasping for air, and waking with a dry mouth[1]. Although frequently underdiagnosed, OSA is a common sleep-related disturbance affecting an estimated 1 billion people worldwide. Its prevalence is notably higher in males (25-30%) than in females (9-17%)[2–4].

The aetiology of OSA is multifactorial, with neuromuscular and lifestyle factors, reduced ventilatory drive during sleep, and individual anatomy (e.g., tongue size and position) all considered important contributors[1]. Among these, obesity, typically measured by the body mass index (BMI), is a well-established major risk factor[5], with genetic studies suggesting the genetic risk for OSA is partly mediated by obesity[6–8]. While twin and family studies indicate that OSA is highly heritable (25-87%)[6,7,9], and genome-wide association studies (GWAS) have identified up to 49 risk loci[6,7], a comprehensive understanding of its genetic architecture independent of obesity remains incomplete.

In the present study, we analysed individual- and summary-level genetic data for OSA from 1.6 million participants across eight international cohorts. Our analytical approach involved a large-scale GWAS meta-analysis in individuals of European ancestry, followed by comprehensive post-GWAS analyses. These included fine-mapping, gene-based tests, expression quantitative trait loci (eQTL) mapping, and the integration of spatial transcriptomics data to investigate the putative role of specific brain structures in OSA[10]. We also evaluated the predictive utility of OSA-derived polygenic risk scores (PRS) in individuals from diverse ancestral backgrounds. Finally, we explored the genetic overlap and investigated putative causal relationships between OSA and a range of physical and neuropsychiatric phenotypes, sleep disturbances, lifestyle factors, and structural brain neuroimaging measures. This work advances our understanding of the molecular underpinnings of OSA, elucidating its relationships with physical and mental health while providing evidence for its genetic components beyond the effects of BMI.

RESULTS

Discovery Genome-Wide Association Analyses in Europeans

Our meta-analysis of European-ancestry GWAS summary data for OSA identified 147 independent genome-wide significant loci ($P < 5 \times 10^{-8}$) (**Figure 1** and **Supplementary Tables 1 - 3**), with the strongest association observed for SNP rs1421085 in chromosome 16 within the *FTO* gene. The single-nucleotide polymorphism (SNP)-based heritability on the liability scale was estimated at 16.26% (SE = 0.006). The linkage disequilibrium (LD) score regression intercept was 1.004 (SE = 0.01), with an attenuation ratio of 0.005 (SE = 0.01), suggesting that confounding from population stratification was negligible. As a conditional analysis, we generated GWAS summary statistics for OSA after adjusting for the genetic effect of BMI ($OSA_{noBMI\text{-}effect}$). In this secondary analysis, we identified 39 independent genome-wide significant loci, of which 27 loci were within a 500kb window of SNPs identified in our primary OSA meta-analysis and 12 were novel (**Supplementary Table 4**). Manhattan and quantile–quantile plots for individual cohort and meta-analysis are available in **Supplementary Figures 1-18**.

GWAS in African Ancestry

Next, we meta-analysed African-ancestry GWAS summary data for OSA, identifying six independent genome-wide significant loci ($P < 5 \times 10^{-8}$) (**Supplementary Table 5**; **Supplementary Figures 19 and 20**). The SNP-based heritability on the liability scale for this analysis was estimated at 5.35% (SE = 0.007). The trans-ethnic genome-wide genetic correlation between OSA in the African and European cohorts was high at 0.91 (SE = 0.05, $P = 0.08$), indicating a largely shared genetic architecture (*P*-value tests for deviation from 1, thus not rejecting the hypothesis of perfect genetic sharing). Manhattan and quantile–quantile plots for individual cohort and meta-analysis are available in **Supplementary Figures 21-24**.

Functional Annotation and Gene Prioritisation

To translate genetic loci into biological function, we applied a suite of post-GWAS analyses to the European-ancestry summary statistics.

Gene discovery and pathway analysis. Gene-based association tests using MAGMA and fastBAT yielded consistent findings (MAGMA $P < 2.65 \times 10^{-06}$; fastBAT $P < 2.04 \times 10^{-06}$; *see Methods*), implicating hundreds of genes after multiple testing correction (e.g., *CADM2, FTO, ETV5, AUTS2, SCAPER, APOE, ATXN2L*; see **Supplementary Tables 6 and 7** for full lists). Subsequent gene-set analysis revealed enrichment in 17 pathways, including neurogenesis, axon development, and neuron development (**Supplementary Tables 8 and 9**).

Tissue and cell-type specificity. We next performed three complementary analyses, MAGMA tissue-expression analysis, LDSC-SEG, and eQTL mapping, to investigate tissue specificity. These analyses converged to show significant enrichment of OSA-associated gene expression in the brain, particularly in the cerebellum, frontal cortex, and subcortical basal ganglia structures (e.g., nucleus accumbens, caudate nucleus) ($P < 9.26 \times 10^{-04}$; *see Methods*; **Supplementary Tables 10-12**). Integrating the OSA GWAS with GTEx v8 brain eQTL data further supported these findings, identifying numerous genes whose expression in these brain regions is associated with OSA. Across all 13 brain tissues tested, six genes were ubiquitously implicated: *LRRC37A2*, *AS3MT*, *TTC12*, *LRRC37A*, *GTF2IRD2*, and *ZSCAN31* (**Supplementary Tables 13 and 14**).

Causal gene prioritisation and spatial mapping. Summary-based Mendelian Randomisation (SMR) identified five potential causal genes in the prioritised brain tissues, including *GTF2IRD2* and *SCAPER* ($P<7.56\times10-07$; **Supplementary Table 15**). We also observed potential causal gene candidates of nominal statistical significance ($P < 0.05$). These findings included *AS3MT* across all 13 brain tissues; *ZSCAN31* in all tissues except the cerebellar hemisphere; *CADM2* in the nucleus accumbens, caudate nucleus, putamen, substantia nigra, cervical spinal cord, and hippocampus; and *MAPT* in the cerebellum, among others (**Supplementary Table 16**).

To further explore the neural circuitry, we leveraged spatial human cortical brain tissue data through the gsMap method. This revealed that 17 prioritised genes were robustly enriched in excitatory neurons of layers 2/3, 4, and 6, and in specific inhibitory interneurons and astrocytes. Genes such as *NPEPPS* and *ETV5* showed the strongest enrichment in layer 6, while *NRXN1* and *CACNB2* were restricted to layers 2/3 (**Supplementary Tables 17-19**). Protein-protein interaction network analysis highlighted modules involved in synaptic organisation and neurotransmitter secretion (**Figure 2**). This complements SMR findings and spatial enrichment of genes in excitatory and inhibitory neurons, together indicating that the genetic risk for OSA may converge on disruption of synaptic signalling pathways within specific cortical microcircuits.

Conditional analysis removing BMI effects. As a conditional analysis, we repeated the functional annotation for the $OSA_{noBMI\text{-}effect}$ GWAS. This resulted in a marked reduction in the number of associated genes (e.g., 41 genes in MAGMA vs. 249 in the primary analysis). Notably, the significance for the well-known obesity-related gene *FTO* was attenuated, and associations for *CADM2*, *ETV5*, and *AUTS2* were no longer significant after correction, suggesting their effects on OSA are at least partially mediated by BMI (**Supplementary Tables 20-24**).

Polygenic Risk Score Prediction Across Diverse Ancestries

We evaluated the predictive utility of a polygenic risk score (PRS) derived from our European GWAS in individuals of diverse ancestries (European, African, Admixed American, South Asian, East Asian, and Middle Eastern) from the *All of Us* cohort, and East Asian (Chinese) participants from the *Guangzhou Biobank Cohort Study* (GBCS) cohort.

In *All Of Us*, OSA PRS was significantly associated with clinician-ascertained OSA status across all six ancestral groups analysed ($N_{Europeans}$=191,199; $N_{African}$=74,120; $N_{Admixed\ American}$=66,896; $N_{Middle\ Eastern}$=8,727; $N_{South\ Asian}$=4,751; $N_{East\ Asian}$=1,361; **Figure 3a**; **Supplementary Tables 25 and 26**). A clear dose-response relationship was observed; for instance, in Europeans, individuals in the top PRS decile had a 2.64-fold increase odds of OSA compared to the bottom decile (**Figure 3b**). Similar gradients were observed in all other ancestries. Per-standard-deviation odds ratios (ORs) ranged from 1.14 (95% CI 1.11–1.16; $P=2.55\times10^{-27}$; AUC=0.65) in African-ancestry to 1.47 (95% CI 1.16–1.87; $P=1.49\times10^{-3}$; AUC=0.72) in Middle Eastern–ancestry participants explaining up to 12% of Nagelkerke R-squared, with similar strong effects in European (OR 1.36; $P=1.00\times10^{-300}$; AUC=0.65), Admixed American (OR 1.33; $P=1.59\times10^{-83}$; AUC=0.71), East Asian (OR 1.26; $P=2.25\times10^{-6}$; AUC=0.70), and South Asian (OR 1.21; $P=6.95\times10^{-3}$; AUC=0.75) groups. These associations were attenuated but remained largely significant after adjusting for BMI. The PRS also correlated with objective, FitBit-derived sleep metrics (e.g., fragmentation index, number of awakenings, REM-sleep percentage, sleep efficiency, and wake after sleep onset; *see Methods* and **Supplementary Table 27**). In Europeans ($N_{Europeans}$=8,779), OSA PRS correlated with more frequent awakenings ($P = 2.46 \times 10^{-4}$) and wake after sleep onset (WASO) ($P = 2.87 \times 10^{-2}$). Associations for other ancestry groups ($N_{Admixed\ American}$=771; $N_{African}$=617; $N_{East\ Asian}$=340; $N_{South\ Asian}$=112; $N_{Middle\ Eastern}$=27) were largely non-significant ($P > 0.05$) yet showed similar r-squared values.

In the East Asian GBCS cohort, each standard deviation increase in the European-derived OSA PRS was associated with a 10% higher odds of having a phenotypic OSA risk score ≥ 2 based on the (*see Methods*; $N_{TOTAL} = 2{,}023$; 863 cases) the 4-item STOP questionnaire (OR = 1.10; 95% CI 1.00–1.20; $P = 0.01$; $R^2 = 0.002$), independent of age and sex. This effect persisted (OR = 1.10; $P = 0.01$; $R^2 = 0.002$) after further adjustment for BMI. Similar results were observed when defining OSA risk with a STOP score ≥ 3 (OR = 1.10; 95% CI 1.00–1.20; $P = 0.01$; $R^2 = 0.002$), both before and after BMI adjustment. For self-reported snoring ($N_{TOTAL} = 2{,}023$; 891 cases), OSA PRS conferred a 10% increase in odds per standard deviation

(OR = 1.10; 95% CI 1.00–1.20; P = 0.008; $R^2 \approx 0.003$), even after including BMI as a covariate in the model (**Supplementary Table 28**).

Genetic Overlap and Causality with Other Complex Traits

Genome-wide genetic correlations. We estimated the genetic correlation (r_G) between OSA and 114 complex traits (**Figure 4; Supplementary Table 29**). After Bonferroni correction ($P < 2.19 \times 10^{-04}$; *see Methods*), OSA showed significant positive genetic correlations with neuropsychiatric disorders (attention-deficit/hyperactivity disorder (ADHD), anxiety, major depressive disorder) cardiometabolic traits (BMI, stroke, type 2 diabetes), and other sleep-related phenotypes (snoring, napping, long/short sleep duration). We also observed significant negative correlations ($P < 2.19 \times 10^{-04}$) with anorexia nervosa, schizophrenia, and HDL cholesterol.

As a conditional analysis, we investigated if and to what extent removing the effect of BMI influenced the genetic correlations of OSA with complex human traits. The genetic correlation between OSA and $OSA_{noBMI\text{-}effect}$ was 0.82 (SE = 0.01, $P < 1 \times 10^{-300}$) while the estimate for $OSA_{noBMI\text{-}effect}$ and BMI was 0.06 (SE = 0.02, $P < 6.28 \times 10^{-04}$). $OSA_{noBMI\text{-}effect}$ results are, for the most part, consistent with OSA's, with slight reductions in the magnitude of the genetic correlations. When compared to OSA results, although most of the following genetic correlations maintained the same direction albeit a reduction in their magnitude, we observed the absence of nominal genetic correlations ($P < 0.05$) with autism spectrum disorder, Alzheimer's disease, birthweight, cannabis use disorder, subcortical brainstem and globus pallidus volumes, and some brain cortical measurements. Similarly, genetic correlations with anorexia nervosa and schizophrenia were no longer statistically significant after multiple testing corrections.

Colocalisation. Further analysis with GWAS-PW identified 544 unique genomic segments jointly influencing OSA and at least one other phenotype (**Supplementary Table 30**). Cardiometabolic traits, such as BMI (n = 344), type 2 diabetes (n = 198), and HDL cholesterol (n = 100), shared the largest number of genomic segments (n) with OSA, with 49 genomic segments being common for the aforementioned traits. Similarly, neuropsychiatric traits, including major depressive disorder (n = 87), multisite chronic pain (n = 50), neuroticism score (n = 42), ADHD (n = 39), and schizophrenia (n = 34), shared a considerable number of genomic segments with OSA.

Mendelian randomisation. To investigate potential causal relationships, we used Latent Heritable Confounder Mendelian Randomisation (LHC-MR). After multiple testing corrections ($P < 1.47 \times 10^{-03}$; **Supplementary Table 31**), we found evidence that genetic liability for ADHD and

depression increased the risk for OSA. Conversely, genetic liability for OSA appeared to increase the risk for multisite chronic pain, type 2 diabetes, and snoring. We also observed bidirectional relationships between OSA and daytime dozing, napping, HDL cholesterol, schizophrenia, and, notably, with BMI, where the influence of BMI on OSA risk showed stronger statistical evidence (i.e., smaller *P*) yet smaller a effect magnitude than the reverse.

DISCUSSION

In the largest GWAS meta-analysis of obstructive sleep apnoea (OSA) to date, including over 1.6 million individuals from 5 countries (Australia, USA, UK, Canada, and Finland), we identified 147 independent genome-wide significant loci ($P < 5 \times 10^{-8}$), of which 141 had not been reported previously[6,7]. LD score regression intercepts suggest that the elevated lambda and inflation in the quantile plots are most likely due to polygenicity rather than population stratification. Similarly, attenuation ratios indicate correct genomic control. Our SNP-based heritability estimate of 16% is substantially higher than previous reports (8-13%)[6,7]. Crucially, our findings provide significant new insights into the genetic basis of OSA beyond the well-established effects of BMI. Through in-depth functional annotation, we prioritised genes and implicated specific biological pathways and tissues. We demonstrated that a European-derived polygenic risk score (PRS) for OSA is robustly associated with clinician-ascertained OSA status, Fitbit sleep features, and self-reported sleep traits, with its predictive power varying across ancestries. We also clarified the genetic relationships between OSA and a range of complex health traits.

Our gene-level analyses uncovered associations with metabolic, neurodevelopmental, and neural signalling pathways. For instance, associations with *FTO* and *GPD2* affirm links to adiposity, energy homeostasis and sleep-disordered breathing[11–13], while genes regulating synaptic connectivity and neurodevelopment, such as *CADM2*, *ETV5*, and *AUTS2*, implicate brain-related processes in sleep disruption[14–17]. Furthermore, the identification of *APOE* and *MSRB3* suggests a potential role for lipid metabolism and oxidative stress, which have been previously linked to respiratory muscle function and inflammation in OSA[18–20]. The enrichment of neurogenesis and axon-development pathways further implies that perturbations in neural development and brain circuitry could contribute to OSA susceptibility, a hypothesis supported by our tissue- and cell-specific analyses (*see discussion on tissue- and cell-specific enrichment and spatial mapping below*).

Our tissue-specificity investigations converged to show that OSA-associated genetic variants are enriched in multiple brain structures. This broad enrichment may indicate a general brain-wide signal, consistent with the central neural regulation of breathing and arousal[21]. More specifically, enrichment in subcortical structures like the basal ganglia implicates circuits responsible for respiratory rhythm, stress-arousal responses, and autonomic tone[22]. Likewise, signals in the hypothalamus and brainstem point towards key regulators of sleep-wake transitions, homeostatic drives, and chemosensation.[23,24,25]. Enrichment in cortical areas, such as the frontal and parietal lobes, suggests a role for structures involved in voluntary respiratory control and sensory

feedback[26,27]. These observations collectively support the hypothesis that OSA susceptibility could be mediated not only by upper airway anatomy but also by genetic factors acting on neural circuits that regulate breathing, arousal, and autonomic balance.

Fine-grained spatial gene analyses revealed layer- and cell-type–specific expression patterns in the human cortex, with signals concentrated in layer 6 excitatory neurons, and other signals in layers 2/3 and 4, inhibitory interneuron subtypes (Inh.1–3), and astrocyte subtype 2. Functionally, the implicated genes converge on fundamental processes of synaptic transmission, cytoskeletal remodelling, and ion homeostasis, which represent critical pathways for sleep–wake regulation. Layer 6 is a key regulator of corticothalamic feedback, which governs slow oscillations and sleep spindles essential for sleep-stage regulation and memory consolidation[28,29]. The *ETV5* gene is implicated in neuroprotection[30], body weight regulation[31], and sleep[7], but it also promotes the generation of GABAergic neurons through the suppression of *NEUROG2*[15], which in turn promotes the formation of excitatory neurons[32]. Thus, suggesting a potential shift in the deep-cortical excitatory–inhibitory balance that could alter arousal thresholds and sleep stability[33].

Complementing this, we identified enrichment of genes involved in the glutamate pathway in superficial cortical layers (2/3), including *NRXN1*, essential for the specification and release of glutamate at excitatory synapses[34], and *CACNB2*, which controls presynaptic Ca^{2} influx to trigger glutamate vesicle fusion[35]. Similarly, *RAB11A*, a regulator of AMPA receptor recycling[36], and *DOCK3*, which contributes to dendritic spine architecture via Rac1 signaling[37], could point to layer-specific remodelling of glutamatergic synapses. Other identified genes, such as *SLC8A1*, responsible for Na/Ca^{2} exchange and astrocytic regulation of extracellular glutamate[38]; and *PSMD2*, a proteasome subunit linked to activity-dependent protein turnover[39], could suggest a role of homeostatic clearance and receptor turnover in sleep. Our findings suggest that genetic susceptibility for OSA is enriched in cortical circuitry, implicating dysregulated peptide processing, synaptic receptor turnover, and excitatory–inhibitory balance in corticothalamic feedback loops as putative partial contributors to sleep fragmentation and impaired restorative sleep observed in OSA. These observations provide testable hypotheses for future functional investigations using animal or cellular models.

We demonstrated that European-derived polygenic risk scores for OSA predict clinician-ascertained OSA status, some Fitbit-derived sleep features, and self-reported sleep traits when applied to participants of diverse ancestral backgrounds, even when including BMI as a covariate in the predictive models. We show that each SD increase in PRS confers 14 - 47% higher

odds of clinically ascertained OSA in European, African, Admixed American, South Asian, East Asian, and Middle Eastern participants. When stratifying each ancestry‐specific PRS distribution into deciles, we observed that individuals in the top PRS decile face a 2–3‐fold greater OSA risk when compared to the bottom decile. Moreover, PRS correlations with continuous Fitbit metrics were largely non-statistically significant for non-Europeans. This finding is most likely explained by the substantial differences in sample sizes across ancestries, as only a small fraction of the All of Us participants have Fitbit data available (*see Methods*). Related to this, in the independent Chinese cohort, OSA PRS modestly predicted STOP‐questionnaire OSA risk scores and self‐reported snoring, confirming that European‐derived PRS retained predictive power across ancestries.

Previous studies have aimed to investigate the genetic overlap of OSA with complex human traits[6,7]. In the present study, we identified genetic correlations between OSA and 17 complex human traits, including psychiatric, anthropometric, cardiometabolic, and other sleep-related phenotypes. Further interrogation of these associations uncovered potential causal genetic effects in which genetic liability to ADHD and depression tends to increase OSA risk, mirroring clinical observations where these disorders often co‐occur and may share neurobiological pathways[40,41]. Conversely, genetic susceptibility to OSA may contribute to the development of cardiometabolic and pain-related phenotypes, elevating the risk for multisite chronic pain, type 2 diabetes, and snoring, which is consistent with systemic effects of intermittent hypoxia and sleep fragmentation on inflammation[42,43], reduced insulin sensitivity[44], and autonomic regulation[45]. Our bidirectional LHC‐MR results with BMI underscore a self‐reinforcing obesity–OSA cycle, in which adiposity mechanically promotes airway collapse while OSA could reciprocally worsen energy balance via hormonal dysregulation[46]. The relationships between OSA and HDL cholesterol point to lipid metabolism as both a contributor and a consequence of OSA pathophysiology. Similarly, findings of nominal statistical significance for anxiety, stroke, ventral diencephalon volume, and anorexia nervosa could provide hints at additional neurovascular and neurodevelopmental axes influencing OSA susceptibility that warrant further investigation. Our observations highlight the relevance of comorbid psychiatric conditions in the aetiology of OSA, which may inform screening, diagnosis and management.

It is well-established that increased adiposity is a risk factor for OSA, partly through mechanisms such as upper airway fat deposition and reduced lung volume.[12] However, the genetic architecture of OSA is not solely driven by BMI-related effects[6,7], and while adipokine signalling has been proposed as a link,[47,48] associations with related genes (e.g., *LEP*, *ADIPOQ*, *RETN*) were not prominent in our results. Our analysis conditioning on BMI (i.e., $OSA_{noBMI\text{-}effect}$) confirmed a

substantial BMI-independent genetic component. As expected, this analysis reduced the number of significant loci from 147 to 39, and attenuated the signal for key metabolic genes, including *FTO*. Crucially, this still represents a major advance over previous BMI-conditioned analyses, which identified far fewer loci. Most importantly, while metabolic signals were reduced, biological pathways implicating neurogenesis-related processes persisted, and eQTL mapping continued to implicate the same brain structures (albeit with fewer and more specific candidate genes). Furthermore, and in contrast to previous studies[7], the genetic correlations profile for $OSA_{noBMI\text{-}effect}$ largely mirrored our primary results. Taken together, these findings provide compelling evidence that while BMI is a major contributor to OSA aetiology, a distinct and prominent neurobiological architecture independently contributes to individual risk.

We acknowledge this study has several limitations. First, OSA is frequently underdiagnosed[2,3], creating a risk of misclassifying cases within control groups. Second, reliance on self-report questionnaires in some cohorts, as opposed to clinical ICD-10 or SNOMED codes, could lead to the inadvertent inclusion of a small number of central sleep apnoea cases. To mitigate both issues, we applied stringent, multi-item criteria for case-control definitions in all self-report cohorts (see *Methods*), which minimises these potential sources of misclassification. Furthermore, given that the population prevalence of central sleep apnoea is low (<1%) and typically skews older than our study's predominant age group[49], any minor inclusion of central sleep apnoea is unlikely to have substantially influenced our findings.

In the discovery phase, our GWAS meta-analysis only included participants of European ancestry. Thus, the genome-wide loci reported here are only representative for individuals of European ancestry until confirmed in samples of other ancestral backgrounds, with the exception of those genetic loci identified in the independent meta-analysis for individuals of African ancestry, which doubled those reported in a previous African GWAS[7]. Regarding the estimation of potential causal genetic effects, we detected slight evidence of horizontal pleiotropy for some cardiometabolic phenotypes (**Supplementary Table 31**). However, LHC-MR models both sample overlap and correlated pleiotropy, providing putative causal estimates considering these factors. While not all genetic overlap between OSA and another complex human phenotype may be causal, the strength and robustness of the remaining causal components suggest it is unlikely that our findings are explained by pleiotropy alone. Similarly, we note that while European-derived OSA PRS showed encouraging predictive power across ancestries, the clinical utility of these scores may ultimately be greatest when derived from ancestry-specific GWAS data. These findings highlight the relevance and potential of genomics for clinical screening of patients at high risk for OSA.

In conclusion, this study fundamentally reframes our understanding of the genetic architecture of OSA, establishing a complex polygenic basis that extends beyond adiposity to implicate distinct neurobiological pathways related to synaptic function and corticothalamic feedback loops. The predictive utility of our polygenic risk score across diverse ancestries marks a tangible step towards genomic medicine, offering a potential tool for early risk stratification in clinical settings. Critically, our findings provide a rich atlas of testable hypotheses for the field; future research must now focus on functionally validating these novel genes and pathways to mechanistically dissect their roles in sleep fragmentation and arousal. Expanding discovery efforts in non-European populations will be paramount to ensuring that the benefits of genomic discovery are equitable. Ultimately, this work paves the way for future therapeutic strategies aimed at the molecular roots of OSA, offering new hope for the millions affected by this debilitating condition.

ACKNOWLEDGMENTS

This project used data from the UK-Biobank under application number 25331. We want to acknowledge Mass General Brigham Biobank for providing samples, genomic data, and health information. We thank the National Institutes of Health's All of Us Research Program for making the participant level data available for our study. We also thank the participants in the All of US Research Program for making this study possible. 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, as well as the following provinces, Newfoundland, Nova Scotia, Quebec, Ontario, Manitoba, Alberta, and British Columbia. This research has been conducted using the CLSA dataset [Baseline Comprehensive Dataset version 4.0, Follow-up 1 Comprehensive Dataset version 1.0, Genetic data version 3.0], under Application Number 190225. The CLSA is led by Drs. Parminder Raina, Christina Wolfson and Susan Kirkland. The time and commitment of the participants to the CLSA study platform is gratefully acknowledged, without whom this research would not be possible. 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. 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). ZC is supported by funding from the National Institute of Mental Health (R01MH129356 awarded to ACE). VFO expresses thanks for the PhD Scholarship support from the Global Parkinson's Genetic Program (GP2). BU is supported by the Higher Degree Research (HDR) Tuition Fee Scholarship (TFS) from Queensland University of Technology (QUT) and a QIMR Computational Neurogenomics Laboratory scholarship. FC is supported by scholarships from the University of Queensland and the QIMR Computational Neurogenomics Laboratory. MER thanks the support from the Rebecca L. Cooper Medical Research Foundation (F20231230). We thank all the investigators of the Guangzhou Biobank Cohort Study (GBCS) for their generous contribution of genetic and longitudinal data that made this work possible. SM is supported by a research fellowship from the Australian NHMRC.

AUTHOR CONTRIBUTIONS

Core analysis and writing group: LMGM, ZC, AP, JX, MS, JW, MER

Contributed to the editing of the manuscript: LMGM, ZC, AP, JX, SDT, VFO, MMM, NSO, CP, AA, XH, AAR, BU, FC, NGM, SM, XD, SL, MS, JW, MER

Genetic data analysis: LMGM, ZC, AP, JX, SDT, VFO, MMM, NSO, CP, AA, XH, AAR, BU, FC, MS, JW, MER

FIGURE LEGENDS

Figure 1. Obstructive sleep apnoea GWAS meta-analysis results.

Manhattan plot illustrating loci associated with obstructive sleep apnoea. The common genome-wide significance threshold ($P < 5 \times 10\text{-}8$) is represented by the red dashed line. The p-values referenced here correspond to a two-tailed Z-test as implemented in METAL.

Figure 2. Protein-protein interaction (PPI) network of genes generated using STRING. Only medium and high-confidence interactions (score ≥ 0.4) are shown for PPI network for prioritised genes surviving multiple testing correction in the discovery phase and replicating in spatial transcriptomics analyses on cortical brain tissue. Edges in black represent co-expression evidence, green indicates associations supported by text mining, light blue represents associations supported by curated databases, and dark blue indicates gene co-occurrence, while purple represents protein homology.

Figure 3. Polygenic prediction in populations from diverse ancestral backgrounds in the All Of Us cohort. Clinician‐ascertained obstructive sleep apnoea (OSA) status polygenic risk scores (PRS) were derived using *SBayesRC*. Odds ratios (ORs) were estimated with logistic regressions. **a)** Per standard‐deviation ORs for clinician‐ascertained OSA in each ancestry, including European (N = 224,180), African (N = 77,064), Admixed American (N = 74,036), East Asian (N = 9,743), South Asian (N = 5,362), and Middle Eastern (N = 1,473). **b)** Risk stratification by PRS decile: ORs in deciles 2–10 (vs. decile 1) for each ancestry. In both a) and b), orange markers show models without body mass index (BMI) adjustment; blue markers show models with BMI as a covariate.

Figure 4. Genetic overlap with other complex human traits

Heatmap depicting genetic correlations (rG) of obstructive sleep apnoea (OSA) and OSA without the effect of BMI with complex human phenotypes. *p-value < 0.05; **p-value significant after Bonferroni multiple testing correction (0.05/228 (total number of genetic correlation tests) = 2.19×10^{-04}). Genetic correlations were estimated using LD score regression. P-values correspond to chi-squared tests with one degree of freedom as implemented in LD score regression.

REFERENCES

1. Slowik, J. M., Sankari, A. & Collen, J. F. Obstructive Sleep Apnea. in *StatPearls [Internet]* (StatPearls Publishing, 2024).

2. Benjafield, A. V. *et al.* Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *The Lancet. Respiratory medicine* **7**, 687 (2019).

3. Malhotra, A. *et al.* Metrics of sleep apnea severity: beyond the apnea-hypopnea index. *Sleep* **44**, zsab030 (2021).

4. Geer, J. H. & Hilbert, J. Gender Issues in Obstructive Sleep Apnea. *The Yale Journal of Biology and Medicine* **94**, 487 (2021).

5. Romero-Corral, A., Caples, S. M., Lopez-Jimenez, F. & Somers, V. K. Interactions Between Obesity and Obstructive Sleep Apnea: Implications for Treatment. *Chest* **137**, 711 (2010).

6. Campos, A. I. *et al.* Discovery of genomic loci associated with sleep apnea risk through multi-trait GWAS analysis with snoring. *Sleep* **46**, (2023).

7. Sofer, T. *et al.* Genome-wide association study of obstructive sleep apnoea in the Million Veteran Program uncovers genetic heterogeneity by sex. *EBioMedicine* **90**, 104536 (2023).

8. Campos, A. I. *et al.* Insights into the aetiology of snoring from observational and genetic investigations in the UK Biobank. *Nature Communications* **11**, 1–12 (2020).

9. Szily, M. *et al.* Genetic influences on the onset of obstructive sleep apnoea and daytime sleepiness: a twin study. *Respiratory Research* **20**, 1–6 (2019).

10. Structural and functional neural adaptations in obstructive sleep apnea: An activation likelihood estimation meta-analysis. *Neuroscience & Biobehavioral Reviews* **65**, 142–156 (2016).

11. Pugliese, G. *et al.* Sleep Apnea, Obesity, and Disturbed Glucose Homeostasis: Epidemiologic Evidence, Biologic Insights, and Therapeutic Strategies. *Current Obesity Reports* **9**, 30–38 (2020).

12. Deng, H. *et al.* Association of adiposity with risk of obstructive sleep apnea: a population-based study. *BMC Public Health* **23**, 1–12 (2023).

13. Loos, R. J. F. & Yeo, G. S. H. The bigger picture of FTO—the first GWAS-identified obesity gene. *Nature Reviews Endocrinology* **10**, 51–61 (2013).

14. Pasman, J. A. *et al.* The CADM2 Gene and Behavior: A Phenome-Wide Scan in UK-Biobank. *Behavior Genetics* **52**, 306–314 (2022).

15. Liu, Y. & Zhang, Y. ETV5 is Essential for Neuronal Differentiation of Human Neural Progenitor Cells by Repressing NEUROG2 Expression. *Stem cell reviews and reports* **15**, (2019).

16. Loberti, L. *et al.* AUTS2-related syndrome: Insights from a large European cohort. *Genetics in medicine : official journal of the American College of Medical Genetics* **27**, (2025).

17. Stein, M. B. *et al.* Genome-wide analysis of insomnia disorder. *Molecular Psychiatry* **23**, 2238–2250 (2018).

18. Lavie, L. Oxidative stress in obstructive sleep apnea and intermittent hypoxia--revisited--the bad ugly and good: implications to the heart and brain. *Sleep medicine reviews* **20**, (2015).

19. Pau, M. C. *et al.* Evaluation of Inflammation and Oxidative Stress Markers in Patients with Obstructive Sleep Apnea (OSA). *Journal of Clinical Medicine* **12**, 3935 (2023).

20. Huang, Y. & Mahley, R. W. Apolipoprotein E: Structure and Function in Lipid Metabolism, Neurobiology, and Alzheimer's Diseases. *Neurobiology of disease* **72PA**, 3 (2014).

21. Chamberlin, N. L. Brain circuitry mediating arousal from obstructive sleep apnea. *Current opinion in neurobiology* **23**, (2013).

22. Lazarus, M., Chen, J.-F., Urade, Y. & Huang, Z.-L. Role of the basal ganglia in the control of sleep and wakefulness. *Current opinion in neurobiology* **23**, 780 (2013).

23. Adamantidis, A. R. & de Lecea, L. Sleep and the hypothalamus. *Science* (2023) doi:10.1126/science.adh8285.

24. Nattie, E. & Li, A. Central Chemoreceptors: Locations and Functions. *Comprehensive Physiology* **2**, 221 (2012).

25. Smith, J. C., Abdala, A. P., Borgmann, A., Rybak, I. A. & Paton, J. F. Brainstem respiratory networks: building blocks and microcircuits. *Trends in neurosciences* **36**, (2013).

26. McKay, L. C., Evans, K. C., Frackowiak, R. S. J. & Corfield, D. R. Neural correlates of voluntary breathing in humans. *Journal of Applied Physiology* (2003) doi:10.1152/japplphysiol.00641.2002.

27. Holton, P. *et al.* Differential responses to breath-holding, voluntary deep breathing and hypercapnia in left and right dorsal anterior cingulate. *Experimental physiology* **106**, (2021).

28. Herrera, C. G. & Tarokh, L. A Thalamocortical Perspective on Sleep Spindle Alterations in Neurodevelopmental Disorders. *Current Sleep Medicine Reports* **10**, 103 (2024).

29. Mölle, M., Bergmann, T. O., Marshall, L. & Born, J. Fast and Slow Spindles during the Sleep Slow Oscillation: Disparate Coalescence and Engagement in Memory Processing. *Sleep* **34**, 1411–1421 (2011).

30. Pisanu, C. *et al.* Evidence that genes involved in hedgehog signaling are associated with both bipolar disorder and high BMI. *Translational Psychiatry* **9**, 1–13 (2019).

31. Gutierrez-Aguilar, R. *et al.* The obesity-associated transcription factor ETV5 modulates circulating glucocorticoids. *Physiology & behavior* **150**, 38 (2015).

32. Hulme, A. J., Maksour, S., Glover, M. S.-C., Miellet, S. & Dottori, M. Making neurons, made easy: The use of Neurogenin-2 in neuronal differentiation. *Stem Cell Reports* **17**, 14 (2021).

33. Kim, J., Matney, C. J., Blankenship, A., Hestrin, S. & Brown, S. P. Layer 6 Corticothalamic

Neurons Activate a Cortical Output Layer, Layer 5a. *J. Neurosci.* **34**, 9656–9664 (2014).

34. Boxer, E. E. & Aoto, J. Neurexins and their ligands at inhibitory synapses. *Frontiers in Synaptic Neuroscience* **14**, 1087238 (2022).

35. Modrell, M. S. *et al.* Insights into electrosensory organ development, physiology and evolution from a lateral line-enriched transcriptome. (2017) doi:10.7554/eLife.24197.

36. Moretto, E. & Passafaro, M. Recent Findings on AMPA Receptor Recycling. *Frontiers in Cellular Neuroscience* **12**, 286 (2018).

37. Wu, Q. *et al.* Electroacupuncture inhibits dendritic spine remodeling through the srGAP3-Rac1 signaling pathway in rats with SNL. *Biological Research* **56**, 1–18 (2023).

38. Rose, C. R., Ziemens, D. & Verkhratsky, A. On the special role of NCX in astrocytes: Translating Na+-transients into intracellular Ca2+ signals. *Cell calcium* **86**, (2020).

39. Sun, L. *et al.* Acetylation-dependent regulation of core spliceosome modulates hepatocellular carcinoma cassette exons and sensitivity to PARP inhibitors. *Nature Communications* **15**, 1–17 (2024).

40. Oğuztürk, Ö., Ekici, M., Çimen, D., Ekici, A. & Senturk, E. Attention Deficit/Hyperactivity Disorder in Adults with Sleep Apnea. *Journal of Clinical Psychology in Medical Settings* **20**, 234 (2012).

41. Jehan, S. *et al.* Depression, Obstructive Sleep Apnea and Psychosocial Health. *Sleep medicine and disorders : international journal* **1**, 00012 (2017).

42. Cubillos-Zapata, C. *et al.* Differential effect of intermittent hypoxia and sleep fragmentation on PD-1/PD-L1 upregulation. *Sleep* **43**, zsz285 (2019).

43. Sforza, E. & Roche, F. Chronic intermittent hypoxia and obstructive sleep apnea: an experimental and clinical approach. *Hypoxia* **4**, 99 (2016).

44. Ip, M. S. *et al.* Obstructive sleep apnea is independently associated with insulin resistance. *American journal of respiratory and critical care medicine* **165**, (2002).

45. Leung, R. S. Sleep-disordered breathing: autonomic mechanisms and arrhythmias. *Progress in cardiovascular diseases* **51**, (2009).

46. Shechter, A. Obstructive sleep apnea and energy balance regulation: A systematic review. *Sleep medicine reviews* **34**, (2017).

47. Schwartz, A. R. *et al.* Obesity and Obstructive Sleep Apnea: Pathogenic Mechanisms and Therapeutic Approaches. *Proceedings of the American Thoracic Society* **5**, 185 (2008).

48. Cowan, D. C. & Livingston, E. Obstructive Sleep Apnoea Syndrome and Weight Loss: Review. *Sleep Disorders* **2012**, 163296 (2012).

49. Rana, A. M. & Sankari, A. Central Sleep Apnea. in *StatPearls [Internet]* (StatPearls Publishing, 2023).

50. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nature Genetics* **50**, 1335–1341 (2018).

51. Mbatchou, J. *et al.* Computationally efficient whole-genome regression for quantitative and binary traits. *Nature Genetics* **53**, 1097–1103 (2021).

52. Genomic data in the All of Us Research Program. *Nature* **627**, (2024).

53. Vuokko, R., Vakkuri, A. & Palojoki, S. Systematized Nomenclature of Medicine–Clinical Terminology (SNOMED CT) Clinical Use Cases in the Context of Electronic Health Record Systems: Systematic Literature Review. *JMIR Medical Informatics* **11**, e43750 (2023).

54. Lind, P. A. *et al.* Preliminary results from the Australian Genetics of Bipolar Disorder Study: A nation-wide cohort. *The Australian and New Zealand Journal of Psychiatry* **57**, 1428 (2023).

55. Byrne, E. M. *et al.* Cohort profile: the Australian genetics of depression study. *BMJ Open* **10**, e032580 (2020).

56. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).

57. Raina, P. *et al.* Cohort Profile: The Canadian Longitudinal Study on Aging (CLSA). *International Journal of Epidemiology* **48**, 1752 (2019).

58. Raina, P. S. *et al.* The Canadian longitudinal study on aging (CLSA). *Canadian journal on aging = La revue canadienne du vieillissement* **28**, (2009).

59. Forgetta, V. *et al.* Cohort profile: genomic data for 26 622 individuals from the Canadian Longitudinal Study on Aging (CLSA). *BMJ Open* **12**, e059021 (2022).

60. Sudlow, C. *et al.* UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

61. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

62. Kurki, M. I. *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, (2023).

63. Verma, A. *et al.* Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program. *Science* (2024) doi:10.1126/science.adj1182.

64. Karlson, E. W., Boutin, N. T., Hoffnagle, A. G. & Allen, N. L. Building the Partners HealthCare Biobank at Partners Personalized Medicine: Informed Consent, Return of Research Results, Recruitment Lessons and Operational Considerations. *Journal of personalized medicine* **6**, (2016).

65. Boutin, N. T. *et al.* The Evolution of a Large Biobank at Mass General Brigham. *Journal of Personalized Medicine* **12**, 1323 (2022).

66. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics* **81**, 559 (2007).

67. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190 (2010).

68. Huang, J. *et al.* Genomics and phenomics of body mass index reveals a complex disease network. *Nature Communications* **13**, 1–10 (2022).

69. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nature Communications* **9**, 1–12 (2018).

70. Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *American Journal of Human Genetics* **99**, 76 (2016).

71. de Leeuw, C. A., Mooij, J. M. & Heskes, T. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology* **11**, (2015).

72. Watanabe, K., Taskesen, E. & van Bochoven, A. Functional mapping and annotation of genetic associations with FUMA. *Nature communications* **8**, (2017).

73. Bakshi, A. *et al.* Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Scientific Reports* **6**, 1–9 (2016).

74. dbGaP Study. https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v8.p2.

75. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature genetics* **50**, 621 (2018).

76. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics* **48**, 481–487 (2016).

77. Song, L., Chen, W., Hou, J., Guo, M. & Yang, J. Spatially resolved mapping of cells associated with human complex traits. *Nature* 1–10 (2025).

78. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

79. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, (2015).

80. Zasadzińska-Stempniak, K., Zajączkiewicz, H. & Kukwa, A. Prevalence of Obstructive Sleep Apnea in the Young Adult Population: A Systematic Review. *Journal of clinical medicine* **13**, (2024).

81. García-Marín, L. M. *et al.* Genomic analysis of intracranial and subcortical brain volumes yields polygenic scores accounting for variation across ancestries. *Nature Genetics* **56**, 2333–2344

(2024).

82. Wu, K. *et al.* Obstructive sleep apnea and structural and functional brain alterations: a brain-wide investigation from clinical association to genetic causality. *BMC Medicine* **23**, 1–11 (2025).

83. Joo, E. Y., Jeon, S., Kim, S. T., Lee, J.-M. & Hong, S. B. Localized Cortical Thinning in Patients with Obstructive Sleep Apnea Syndrome. *Sleep* **36**, 1153 (2013).

84. Grasby, K. L. *et al.* The genetic architecture of the human cerebral cortex. *Science* (2020) doi:10.1126/science.aay6690.

85. Sullivan, P. F. *et al.* Psychiatric Genomics: An Update and an Agenda. *The American journal of psychiatry* **175**, (2018).

86. Middeldorp, C. M., Felix, J. F., Mahajan, A. & McCarthy, M. I. The Early Growth Genetics (EGG) and EArly Genetics and Lifecourse Epidemiology (EAGLE) consortia: design, results and future prospects. *European journal of epidemiology* **34**, (2019).

87. Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics* **48**, (2016).

88. García-Marín, L. M. *et al.* Shared molecular genetic factors influence subcortical brain morphometry and Parkinson's disease risk. *npj Parkinson's Disease* **9**, 1–10 (2023).

89. Reyes-Pérez, P. *et al.* Investigating the Shared Genetic Etiology Between Parkinson's Disease and Depression. *Journal of Parkinson's Disease* (2024) doi:10.3233/JPD-230176.

90. García-Marín, L. M. *et al.* Investigating the genetic relationship of intracranial and subcortical brain volumes with depression and other psychiatric disorders. *Imaging Neuroscience* (2024) doi:10.1162/imag_a_00291.

91. Darrous, L., Mounier, N. & Kutalik, Z. Simultaneous estimation of bi-directional causal effects and heritable confounding from GWAS summary statistics. *Nature communications* **12**, (2021).

92. Burgess, S., Foley, C. N., Allara, E., Staley, J. R. & Howson, J. M. M. A robust and efficient method for Mendelian randomization with hundreds of genetic variants. *Nature Communications* **11**, 1–11 (2020).

93. Jiang, C. Q. *et al.* An overview of the Guangzhou biobank cohort study-cardiovascular disease subcohort (GBCS-CVD): a platform for multidisciplinary collaboration. *Journal of human hypertension* **24**, (2010).

94. Zheng, Z. *et al.* Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries. *Nature genetics* **56**, (2024).

95. Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics* **50**, 746–753 (2018).

96. Adekolu, O. & Zinchuk, A. Sleep Deficiency in Obstructive Sleep Apnea. *Clinics in chest medicine* **43**, 353 (2022).

97. Park, M., Senel, G. B., Modi, H., Jain, V. & DelRosso, L. M. Combined impact of obstructive sleep apnea and periodic limb movements on sleep parameters. *Sleep medicine* **129**, (2025).

98. Krakow, B., Romero, E., Ulibarri, V. A. & Kikta, S. Prospective Assessment of Nocturnal Awakenings in a Case Series of Treatment-Seeking Chronic Insomnia Patients: A Pilot Study of Subjective and Objective Causes. *Sleep* **35**, 1685 (2012).

99. Bonsignore, M. R. *et al.* REM sleep obstructive sleep apnoea. *European Respiratory Review* **33**, 230166 (2024).

100. Rapid eye movement related obstructive sleep apnea: Where do we stand? *Respiratory Investigation* **59**, 589–595 (2021).

101. Staats, R. *et al.* The Importance of Sleep Fragmentation on the Hemodynamic Dipping in Obstructive Sleep Apnea Patients. *Frontiers in Physiology* **11**, 104 (2020).

102. Patel, D. *et al.* Validation of the STOP questionnaire as a screening tool for OSA among different populations: a systematic review and meta-regression analysis. *Journal of Clinical Sleep Medicine : JCSM : Official Publication of the American Academy of Sleep Medicine* **18**, 1441 (2022).

METHODS

Statistics

In this study we performed several statistical approaches, including linear regression, linear mixed-effects associations, genome-wide association studies, LD-score regression, and METAL-based meta-analysis of GWAS summary statistics. Each approach is described in detail below. GWAS analyses for individual cohorts in the present study were conducted with either SAIGE[50] or REGENIE[51]. We used the same parameters and quality control procedures as the GWAS summary data that were leveraged from public repositories (e.g. FinnGen and Million Veteran Program) to ensure consistency. Details for each individual cohort and the meta-analyses are provided below.

Cohorts and GWAS for European populations

All of Us. The All of Us research program is a longitudinal cohort study seeking to recruit at least one million individuals from diverse ancestral backgrounds across the USA by 2027[52]. We performed GWAS for OSA (32,618 cases and 192,980 controls) using whole-genome sequencing data for participants of European ancestry from All of Us version 8. The whole genome sequencing data were processed using the Illumina DRAGEN platform[52]. Ancestry was defined using the genetically inferred ancestry provided by the All of Us research program, computed through Rye[52]. OSA was defined based on the (Systematized Nomenclature of Medicine) SNOMED[53] vocabulary concept ID 78275009. Participants who self-reported snoring were removed from the controls in the OSA GWAS due to the possibility of representing undiagnosed OSA. GWAS were conducted using REGINIE[51] including age, sex, and 10 principal components as covariates. During quality control, we excluded variants with a low minor allele count (<20) and minor allele frequency (<0.01) from the analysis.

Australian cohorts. We leveraged data from participants of European ancestry from the Australian Genetics of Bipolar Disorder Study (AGBD)[54] and the Australian Genetics of Depression Study (AGDS)[55] to conduct GWAS for OSA. A thorough description for each of these cohorts is available elsewhere[54,55]. Briefly, AGBD is a nationwide cohort of adults living with bipolar disorder. So far, AGBD includes ~7,000 participants, of which 70% have been genotyped[54]. Similarly, AGDS is a nation-wide cohort study of individuals who have been diagnosed with depression at some point in their lifetime[55]. AGDS includes ~20,000 participants, of which ~18,000 have been genotyped[55]. Both cohort studies were genotyped with the Illumina Global Screening Array v2, leveraged the Michigan imputation server with the Haplotype Reference Consortium reference panel[56] to impute the genotypes and performed similar cohort quality control procedures[54,55]. In addition, both studies

included the same items in their self-reported questionnaires to assess OSA. Therefore, we were able to conduct GWAS for OSA using the samples from both cohorts as a larger single sample.

In the two Australian cohorts, OSA was defined using two items: *'During the last month, on how many nights or days per week have you had or been told you had snorting or gasping?'* and '*During the last month, on how many nights or days per week have you had or been told you had your breathing stop or you choked or struggled for breath?'* Participants were classified as OSA cases (N = 4,616) if they reported a frequency anywhere between 1 to 7 times per week for either item. In addition to this, for an individual to be considered a case for OSA, the individual must have been deemed as a snorer. The OSA control group (N = 12,692) included individuals who did not snore (*Rarely, less than once a week* or *Never*) or have OSA. The OSA control group also excluded those with snoring but no OSA. GWAS for OSA was performed using SAIGE[50] and a generalised linear mixed model. The GWAS was adjusted for age and sex, and 10 principal components. We excluded variants with a low minor allele frequency (<0.01) or a low-quality imputation score (<0.60) from the analysis.

Canadian Longitudinal Study on Aging (CLSA). The Canadian Longitudinal Study of Ageing (CLSA) seeks to understand and address the needs of an ageing population. It includes ~50,000 participants AGED 45 to 85 years at the time of recruitment, of which ~20,000 have been genotyped via Affymetrix Axiom array and imputed through the Michigan Imputation Server using the TOPMed reference panel[57]. A thorough description of this cohort is available elsewhere[57–59]. In this cohort, OSA prevalence (4,151 cases and 11,436 controls) was defined via the categories "SNO_STOPBREATH_MCQ" and "SNO_STOPBREATH_COF1" from baseline and follow-up surveys, respectively. These categories include the item *'stop breathing in sleep'*, with '*Yes'*, '*No'*, and *'Don't know'* as possible responses. At baseline, 14.7% responded '*Yes'*, while 83.2% responded *'No'*. In the follow-up, 16.8% responded '*Yes'*, and 78.3% '*No'*. In this analysis, we excluded participants who responded '*Don't know'*. In addition, if a participant reported *'Yes'* at baseline but 'No' in the follow-up survey or vice versa, they were deemed as cases since they reported OSA at least once. Participants who self-reported snoring were also removed from the controls in the OSA GWAS due to the possibility of representing undiagnosed OSA. The GWAS included individuals of European ancestry for OSA and was conducted using REGINIE(v2.2.4)[51], including age, sex, and 10 principal components as covariates. We excluded variants with a low minor allele frequency (<0.01) or a low-quality imputation score (<0.60) from the analysis. Also, the Regenie '--firth' flag was used to perform Firth correction to provide a reduction in bias due to the relatively small sample size[51].

UK Biobank (UKB). We conducted GWAS for OSA (5,959 cases and 402,925 controls) using data for participants of European ancestry from the UK Biobank[60]. Genotyping details, which included using the UK BiLEVE Axiom Array for the first ~50,000 participants and the UK Biobank Axiom Array for the remaining ~450,000, as well as phenotyping details for this cohort have been described elsewhere[61]. Briefly, for consistence with previous studies[6,8], we defined OSA through the ICD-10 sleep apnoea diagnosis code (code G47.3 in the 41270 UKB data field), and self-reported sleep apnoea (code 1123 in the 20002 UKB data field). Participants with only *'Yes'* or *'No'* responses were included, leaving out those who answered '*I don't know*' or *'Prefer not to answer'*. Participants who self-reported snoring (defined using data field 1210) were removed from the controls in the OSA GWAS due to the possibility of representing undiagnosed OSA. We conducted the GWAS using SAIGE[50] and a mixed generalised linear model, and included age, sex, genotyping array, and 10 principal components as covariates to adjust further for population stratification. We excluded variants with a low minor allele frequency (<0.01) or a low-quality imputation score (<0.60) from the analysis.

FinnGen. We leveraged publicly available GWAS summary statistics, including participants of European ancestry for OSA from the FinnGen cohort - release 6 (G6)[62]. A detailed description of this cohort is available elsewhere[62]. OSA was defined based on the ICD-10 sleep apnoea diagnosis code (G47)[62]. GWAS summary data were derived using a linear mixed-model implemented in SAIGE[50], with sex, age, genotyping batch and 10 principal components as covariates[62]. The analysis included a total of 56,885 cases and 441,137 controls, and data was filtered to include variants with INFO >0.60[62]. In addition, variants with a low minor allele frequency (<0.01) were excluded.

Million Veteran Program (MVP). We obtained publicly available GWAS summary statistics, including participants of European ancestry for OSA from MVP, genotyped via the MVP 1.0 Axiom array and imputed using the TOPMed reference panel. The summary data included 126,695 cases and 303,630 controls and were derived as part of the MVP genome-wide PheWAS project[63] in a collaboration between the US Department of Veterans Affairs and the US Department of Energy[63]. OSA was defined via electronic health records using a multimodal automated phenotyping procedure, which analyses ICD codes and natural language processing of clinical notes. GWAS were conducted using a linear mixed model approach in SAIGE[50] adjusting for age, sex, and the first 10 principal components[63]. Further details regarding GWAS summary statistics from MVP are available elsewhere[63]. We performed quality control procedures on the GWAS

summary data and excluded variants with a low minor allele frequency (<0.01) or a low-quality imputation score (<0.60).

Mass General Brigham Biobank. We leveraged GWAS summary statistics for OSA (3,102 cases and 16,945 controls)[6], including participants of European ancestry from the Mass General Brigham Biobank (formerly the Partners Healthcare Biobank)[64,65], generated in a previous study[6]. Full details about the Mass General Brigham Biobank are available elsewhere[64,65]. Briefly, this Biobank focuses on electronic health records and survey data, including ~135,000 participants, for which ~65,000 have genomic data available[64,65]. Data was genotyped using the Illumina Multi-Ethnic Global Array (MEGA) (Illumina, Inc., San Diego, CA) GWAS array[64,65], and imputation was performed on the Michigan Imputation server[6]. OSA was defined using electronic health records data and the ICD-10 sleep apnoea diagnosis code (G47.3)[6]. The GWAS was conducted using PLINK[66] v2.00 and a logistic regression with age, sex, genetic principal components, and genotype batch as covariates[6]. The GWAS excluded variants with a low minor allele frequency (<0.01)[6].

GWAS meta-analysis in European populations

METAL. We used METAL[67] (v20200505) to perform a sample-size weighted (p-value based) meta-analysis for OSA using data from the cohorts described previously (N_{cases}= 230,657; $N_{controls}$= 1,377,442; N_{total}= 1,608,099; **Supplementary Table 32**). Following the well-established recommendations for case-control imbalance, cohorts were weighted according to their effective sample size as described by the equation: Neff = 4/(1/Ncases + 1/Ncontrols)[67]. Independent loci for OSA were determined by performing a conservative clumping procedure in PLINK 1.9[66] ($P\_1 = 1 \times 10^{-8}$, $P\_2 = 1 \times 10^{-5}$, $r2 = 1 \times 10^{-3}$, kb = 1,000). Independent genome-wide loci not reported in previous studies are claimed based on a comparison of the independent unique loci identified in the present study against independent genome-wide significant loci for OSA[6,7] reported in previous studies based on exact rs numbers. We considered LD information in the definition of the independent genome-wide loci not reported in previous studies by performing a clumping procedure using PLINK 1.9[66] ($P\_1 = 1 \times 10^{-8}$, $P\_2 = 1 \times 10^{-5}$, $r2 = 1 \times 10^{-3}$, kb = 1,000).

Conditional analysis

Body mass index effect removal.

Body mass index (BMI) is an established risk factor for OSA, and genetic variants that influence BMI are also likely to contribute to OSA risk. However, previous studies have suggested that genetic factors may influence the aetiology of OSA independently of BMI[6–8]. In the present study,

we sought to further interrogate the effect of BMI on OSA. Therefore, we leveraged a set of publicly available GWAS summary statistics for BMI, including ~1.1 million participants[68] and multi-trait-based conditional and joint analysis (mtCOJO)[69] using GCTA (v1.91.7), to adjust for BMI while avoiding biases due to collider bias (i.e., the emergence of a spurious association between a pair of variables when a common outcome is modeled as a covariate). With this approach, we aimed to estimate, for each SNP, the association with OSA that was independent of the SNP's association with BMI. In addition, to estimate independent genome-wide loci that where common for both the $OSA_{noBMI\text{-}effect}$ and our original OSA meta-analyses, we compared SNPs identified via clumping (performed as described above) in both datasets. Specifically, common independent genome-wide loci were defined either by exact rsID match or by genomic proximity, where a SNP from $OSA_{noBMI\text{-}effect}$ was considered common with the original OSA meta-analysis if it fell within ±500 kb of any independent genome-wide SNP from the original OSA analysis.

Cohorts and GWAS for African populations

All of Us. We performed GWAS for OSA (8,822 cases and 71,463 controls) using whole genome sequencing data for participants of African ancestry from All of Us version 8. OSA was defined based on the SNOMED[53] vocabulary concept ID 78275009. Participants who self-reported snoring were removed from the controls in the OSA GWAS due to the possibility of representing undiagnosed OSA. GWAS were conducted using REGINIE(v3.2.6),[51] including age, sex, and 10 principal components as covariates. During quality control, we excluded variants with a low minor allele count (<20) and minor allele frequency (<0.05) from the analysis.

Million Veteran Program (MVP). We obtained publicly available GWAS summary statistics, including participants of African ancestry for OSA from MVP. The summary data included 38,012 cases and 77,729 controls. GWAS were conducted using a linear mixed model approach in SAIGE[50] adjusting for age, sex, and the first 10 principal components[63]. Further details regarding GWAS summary statistics from MVP are available elsewhere[63]. We performed quality control procedures on the GWAS summary data and excluded variants with a low minor allele frequency (<0.01) or a low-quality imputation score (<0.60).

GWAS meta-analysis in African populations

METAL. We used METAL[67] (v20200505) to perform a sample-size weighted (p-value based) meta-analysis for OSA using data from the cohorts described previously (N_{cases}= 46,834; $N_{controls}$= 149,192; N_{total}= 196,026). Following the well-established recommendations for case-control imbalance, cohorts were weighted according to their effective sample size as described by the

equation: Neff = 4/(1/Ncases + 1/Ncontrols)[67]. Independent loci for OSA were determined by performing a conservative clumping procedure in PLINK 1.9[66] ($P\_1 = 1 \times 10^{-8}$, $P\_2 = 1 \times 10^{-5}$, $r2 = 0.05$, kb = 1,000).

Transethnic genetic correlation

Popcorn. We leveraged GWAS summary statistics for OSA in individuals of European and African ancestries to estimate the transethnic genetic correlation using Popcorn(v1.0)[70]. Popcorn models the covariance of SNP effect sizes across two populations by regressing the product of Z scores on cross-population LD scores, adjusting for differences in linkage disequilibrium and allele frequency. By comparing this covariance to the heritability within each ancestry, we derived a trans ethnic genetic correlation that quantifies the proportion of shared genetic architecture between the traits in European and African cohorts[70]. The p-value reported by Popcorn tests whether the genetic correlation is significantly less than 1. A large p-value (e.g., >= 0.05) implies not rejecting the hypothesis of perfect genetic sharing between populations[70].

Functional annotation and gene prioritisation

We performed functional annotation and gene prioritisation analyses. Specifically, these analyses included MAGMA, fastBAT, LDSC-SEG, eQTL mapping, Summary-based Mendelian Randomization (SMR), and spatial mapping of cells in cortical brain tissue (gsMap).

We conducted MAGMA[71] gene-based and gene-set (v1.08) analyses as implemented in FUMA (v1.7.0)[72] (https://fuma.ctglab.nl/). Briefly, MAGMA aggregates association *P* values considering all the genetic variants mapping to a given gene and its corresponding regulatory region[71]. We applied a Bonferroni multiple-testing correction (0.05/18,871 (total number of tested genes) $= 2.65 \times 10^{-06}$). As a complementary approach to MAGMA's gene-based test, we performed fastBAT (v1.94.1)[73] as part of our initial gene discovery phase and applied a Bonferroni multiple-testing correction (0.05/24,568 (total number of tested genes) $= 2.04 \times 10^{-06}$). MAGMA gene-set tests assess predefined biological pathways and GO terms for enrichment of OSA-associated genes, highlighting potential dysregulated processes.

We used MAGMA's Tissue Expression Analysis using GTEx v8[74] as implemented in FUMA to identify tissue specificity of genes associated with OSA. We implemented a Bonferroni multiple-testing correction for general (0.05/30 (total number of general tissues) $= 1.67 \times 10^{-03}$) and specific tissues (0.05/54 (total number of specific tissues) $= 9.26 \times 10^{-04}$). Furthermore, we investigated tissue-specific enrichment (for those identified with MAGMA's Tissue Expression

Analysis) with LDSC-SEG (v1.0.1)[75], a method that integrates GWAS summary data with gene expression patterns via stratified LD score regression to identify tissues or cell types where OSA heritability is enriched. We implemented a Bonferroni multiple-testing correction (0.05/24 (total number of tested tissues) = 2.08 × 10^{-03}). In addition, we performed eQTL mapping in FUMA to map OSA-associated SNPs to potential target genes via the expression quantitative trait loci based on 13 GTEx (v8) brain tissues. We applied Bonferroni multiple testing correction and accounted for the number of brain tissues in our analysis (0.05 / 15,757 [average number of annotations per brain volume] * 13 [number of tested brain tissues] = $2.44x10^{-07}$). Then, we followed up on brain tissues identified in our tissue-specificity analyses by performing SMR (v1.3.1) analyses, which integrate GWAS and eQTL data in enriched tissues (GTEx v8) to test for causal gene candidates in the prioritised tissues[76]. We applied a Bonferroni multiple-testing correction and accounted for the number of brain tissues in our analysis (0.05 / 5087 [average number of annotations per brain volume] * 13 [number of tested brain tissues] = $7.56x10^{-07}$).

Considering our gene discovery and tissue enrichment findings, we further investigated the genes prioritised in the gene discovery phase via spatial transcriptomics analyses using gsMap (v1.71.2)[77]. We leveraged human cortical brain tissues to investigate fine-grained spatial gene expression patterns underlying OSA's neural circuitry based on the results from our tissue expression analyses described above. We filtered our results for those involving genes that were identified in MAGMA or fastBAT analyses. We implemented a Bonferroni multiple-testing comparison based on the total number of tested cells (0.05 / 14 = $3.57x10^{-03}$), and a strict Pearson correlation coefficient (PCC) between the modelled expression and true expression (PCC > 0.70).

We performed protein–protein interaction (PPI) network analysis using the STRING database v12.0[78] to examine potential biological interactions among protein-coding genes mapped via MAGMA and fastBAT, replicating in our cell-specific gene expression analysis on spatial transcriptomics analyses on cortical brain tissue. We retained medium and high-confidence interactions with a score ≥ 0.4. We also applied the Markov Cluster Algorithm (MCL) with an inflation parameter of 3 to identify potential functional modules. Network enrichment was assessed by comparing the observed number of edges with the expected number given the size of the input set, using STRING's built-in enrichment test.

As a conditional post-GWAS analysis, to further understand the influence of BMI in OSA, we sought to identify potential differences in OSA with and without the effect of BMI. Thus, we performed gene-based and gene-set MAGMA (v1.08) and eQTL mapping as implemented in FUMA for $OSA_{noBMI\text{-}effect}$. We implemented a Bonferroni multiple-testing correction, considering

the total number of associations (MAGMA gene-based test P = 0.05/18,784 [total number of tested genes] = 2.66×10^{-06}; eQTL mapping P = 0.05 / 728 [average number of annotations per brain volume] * 13 [number of tested brain tissues] = 5.28×10^{-06}). We also conducted fastBAT analyses with a Bonferroni multiple-testing correction based on the total number of tests (P = 0.05/24,278 (total number of tested genes) = 2.06×10^{-06}).

SNP-based heritability and genetic correlations

We estimated the heritability for OSA using LDSC(v1.0.1)[79], with heritability on the liability scale based on a 16% population prevalence[80]. LDSC leverages the expected association between LD variant tags and the expected degree of association with a phenotype. Therefore, variants that are in higher LD with others are expected to capture more of the true polygenic signal, and LDSC leverages this relationship to distinguish polygenic heritability from confounding biases, such as population stratification. Using the munge function from LDSC, we processed the GWAS meta-analysis for OSA and conducted LDSC to estimate the variance explained by the SNPs in the GWAS summary statistics.

The genetic correlation between two complex human traits denotes the relationship between the direction and magnitude of genetic effect between the two. We leveraged LDSC to conduct genetic correlation analyses between OSA and several complex human phenotypes. These traits included neuropsychiatric disorders, anthropometric measurements, sleep-related traits, lifestyle factors, cardiometabolic traits, cortical brain measurements, as well as intracranial and subcortical brain volumes. Further details for the GWAS summary statistics for these complex human traits are available in **Supplementary Table 33**. It is important to note that these traits were selected based on the following criteria: (i) The relationship between OSA[6,7] and a trait that has been investigated before at a genetic[81,82] or observational[10,83] level, and (ii) The trait has publicly available and well-powered (i.e., large enough sample size to detect genome-wide loci) GWAS summary statistics that can be leveraged to perform LDSC analyses. Thus, GWAS summary data for the other complex human traits included in these analyses were primarily retrieved from international consortia, such as ENIGMA[81,84], the Psychiatric Genetics Consortium (PGC)[85], and the Early Growth Genetics Consortium (EGG)[86], among others. We accounted for multiple testing using Bonferroni correction (0.05/228 (total number of genetic correlation tests) = 2.19×10^{-04}).

As a conditional analysis, we sought to identify potential differences in OSA with and without the effect of BMI. Therefore, we estimated the genetic correlation between OSA and $OSA_{noBMI\text{-}effect}$. In addition, we also estimated genetic correlations between OSA, removing the effect of BMI, and

complex human traits, ultimately providing genetic correlation results for OSA and $OSA_{noBMI\text{-}effect}$ with complex human phenotypes. We accounted for multiple testing using Bonferroni correction (0.05/228 (total number of genetic correlation tests) = 2.19 x 10^{-04}).

Colocalisation

We leveraged the GWAS-PW (v.0.3.6) method[87] to identify segments of the genome with genetic variants influencing the etiology of OSA and also another human complex trait. For this analysis we leveraged traits with a statistically significant genetic correlation based on our LDSC results to conduct GWAS-PW analyses. Specifically, the GWAS-PW method estimates the posterior probability of association for four different models. In the first model, the segment of the genome is only associated with phenotype A (i.e., OSA in the present study)[87]. For model two, the genomic segment is only influencing the aetiology of phenotype B (i.e., a complex human trait in the present study)[87]. Moreover, in model three, the genomic segment is influencing both phenotypes via the same genetic variants, while in model four the genomic segment influences both phenotypes, but via different genetic variants[87]. In the present study, we provide findings for genomic segments where model three showed a PPA > 0.5 since this threshold has been used in previous studies[81,88–90].

Potential causal genetic effects

We leveraged the Latent Heritable Confounding Mendelian Randomisation (LHC-MR; v0.0.0.9000)[91] method in R (v4.3.1) to investigate potential causal genetic effects between OSA and the complex human phenotypes with a statistically significant genetic correlation after Bonferroni multiple-testing correction. We selected LHC-MR, as it represents an alternative to traditional Mendelian randomisation methods, which assume no sample overlap between phenotypes[92]. In addition, LHC-MR, unlike traditional Mendelian randomisation methods, leverages the entire set of GWAS summary data (not only genome-wide independent loci) to estimate putative causal genetic effects between two traits. Therefore, LHC-MR has been reported to improve statistical power to estimate bidirectional putative causal genetic effects, direct heritabilities, and confounder effects while accounting for sample overlap[91]. Further details for LHC-MR are available elsewhere[91]. We implemented a Bonferroni multiple testing correction (0.05/34 (total number of LHC-MR tests in the present study) = 1.47 X 10^{-03} 2.19 x 10^{-04}).

Polygenic scores estimation and association analyses

We leveraged GWAS summary statistics for OSA to derive PRS for OSA and tested their predictive capability on related outcomes in samples of diverse ancestral backgrounds, including All Of Us[52] and the Guangzhou Biobank Cohort Study (GBCS)[93], with full details provided below. Across

cohorts, we derived OSA PRS using *SBayesRC (v2.2)*[94] with the Genome-wide Complex Trait Bayesian (GCTB; v2.0) software tool[95]. PRS were calculated through the multiplication of the multivariate effect size (obtained from *SBayesRC*) times the allelic dosage of the effect allele and summing across all loci for each participant. In addition, to derive the PRS, we only included SNPs passing quality control (MAF > 0.01; call rate > 0.90 ; and for GBCS an imputation score > 0.60). In the All Of Us cohort, ancestry was defined as having at least 60% composition or greater of a specific ancestry.

Specifically, we leveraged our European ancestry GWAS summary statistics to derive OSA PRS to predict OSA and five features derived from Fitbit data (sleep efficiency[96], wake after sleep onset (WASO)[97], number of awakenings[98], REM sleep percentage[97,99,100], and fragmentation index[101]) in participants from the All Of Us cohort across ancestries. We note that to evaluate the predictive ability of OSA PRS in European ancestry participants in All Of Us as the target sample, we derived the PRS using a modified version of our original GWAS OSA meta-analysis, in which we left out the All Of Us European ancestry participants from original the meta-analysis. Sample sizes for clinically diagnosed OSA included up to 224,108 European, 77,064 African, 74,036 Admixed American, 9,743 East Asian, 5,362 South Asian, and 1,473 Middle Eastern participants in All Of Us. In contrast, for Fitbit derived measures, sample sizes were substantially reduced, as only a fraction of the participants have these data available. Sample sizes for Fitbit data included up to 10,528 Europeans, 895 Admixed Americans, 713 Africans, 400 East Asians, 135 South Asians, and 32 Middle Eastern participants in All Of Us. Ancestry groups were determined based on individuals who showed 60% or greater ancestry composition for a specific group. A full description of the demographic characteristics and sample sizes for each ancestry is available in **Supplementary Tables 34** and **35**.

We also used OSA PRS derived from the European ancestry GWAS to predict snoring and risk for OSA based on the self-reported 4-item STOP questionnaire[102] in up to ~3,000 participants of East Asian, specifically of Chinese ancestry, from GBCS. The STOP questionnaire was not available in any other cohort leveraged in the present study, so we used it to complement our previous cross-ancestry analyses. In particular, two risk phenotypes for OSA were evaluated for GBCS, (i) having a score of at least two, where one of the points in the score indicates that the patient snores; (ii) having a score of at least three, where one of the points in the score indicates that the patient snores. If a participant had a score greater than two or three, but none of the points represented the presence of snoring, the participant was defined as a control.

Once we derived OSA PRS, we implemented multivariate regressions (logistic for binary outcomes and linear for continuous outcomes) in R for two models. The first one with sex, age, OSA-PRS and the first 10 principal components as covariates to adjust for population stratification. The second one with sex, age, OSA PRS, the first 10 principal components, and BMI as covariates to assess the potential influence of BMI on the predictive ability of the OSA PRS. The BMI covariate used in these models corresponds to a physically examined measure, not a self-reported item. We created figures showing our results in R (v4.3.1) using the *dplyr* package.

ETHICS STATEMENT

Our study is based on previously published GWAS summary data and publicly available data for which appropriate site-specific institutional review boards and ethical reviews at local institutions were approved for the use of these data. All participants included in the present study provided written informed consent, and the investigators in the participating studies obtained approval from their institutional review board or equivalent organization.

DATA AVAILABILITY

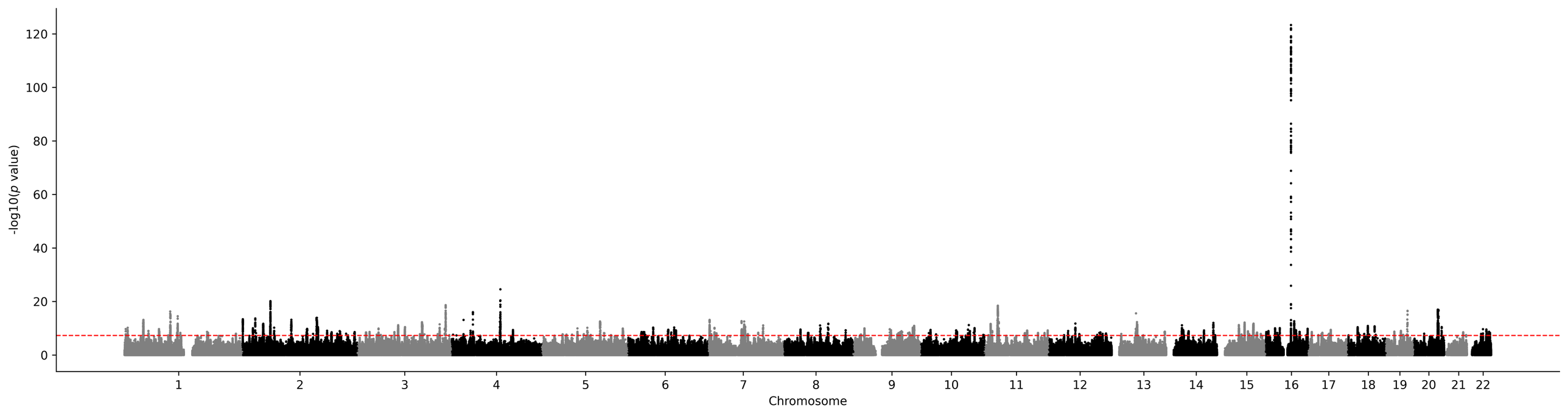
Full GWAS summary statistics for OSA and $OSA_{noBMI\text{-}effect}$ generated with METAL are available through the GWAS Catalogue (https://www.ebi.ac.uk/gwas/downloads/summary-statistics). Researchers can access individual-level data from the individual cohorts leveraged in the present study following the corresponding data application procedures and data transfer agreements. Work performed using UKB data was done under application 25331. 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. This project used data from the National Institutes of Health's All of Us Research Program's Controlled Tier Dataset version 8, accessible to authorized researchers on the Researcher Workbench.

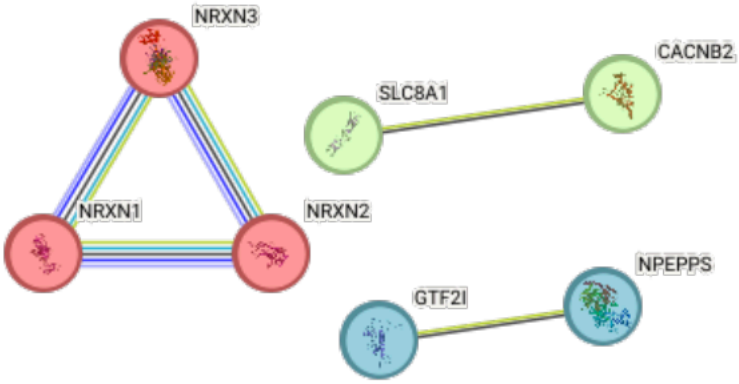
CODE AVAILABILITY

No custom code was used in this study. Publicly available software tools were used to perform genetic analyses and are referenced throughout the manuscript.

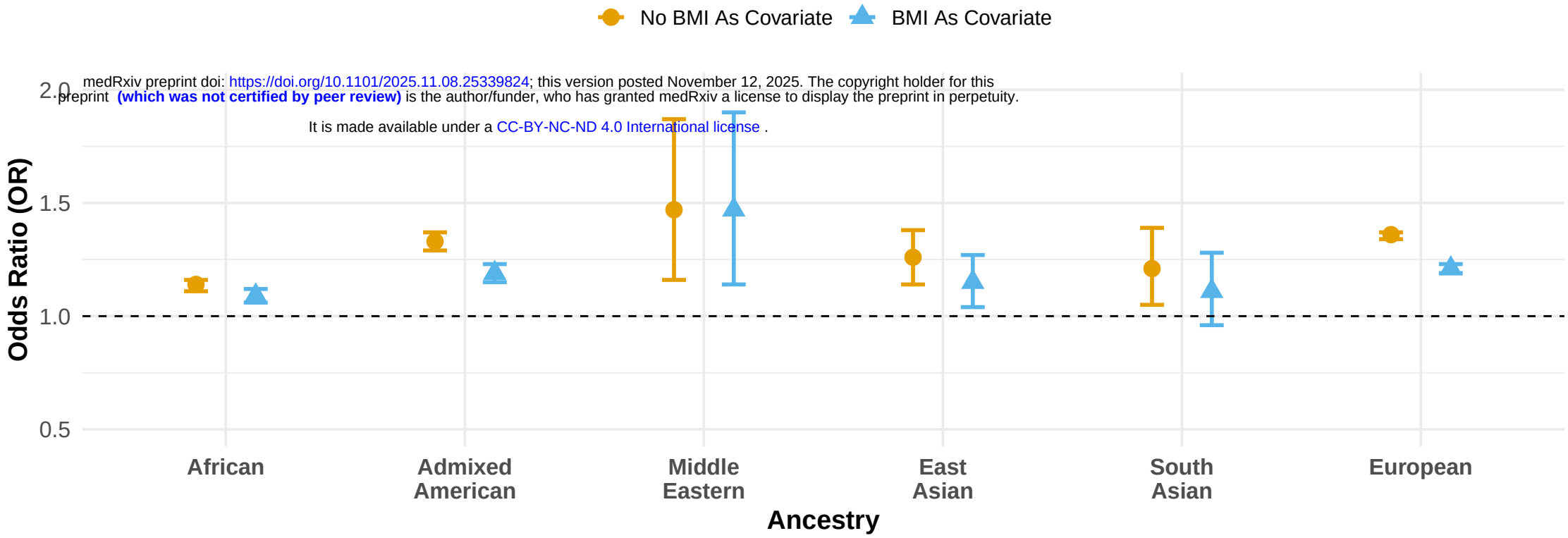
COMPETING INTERESTS

All authors declare no competing interests.

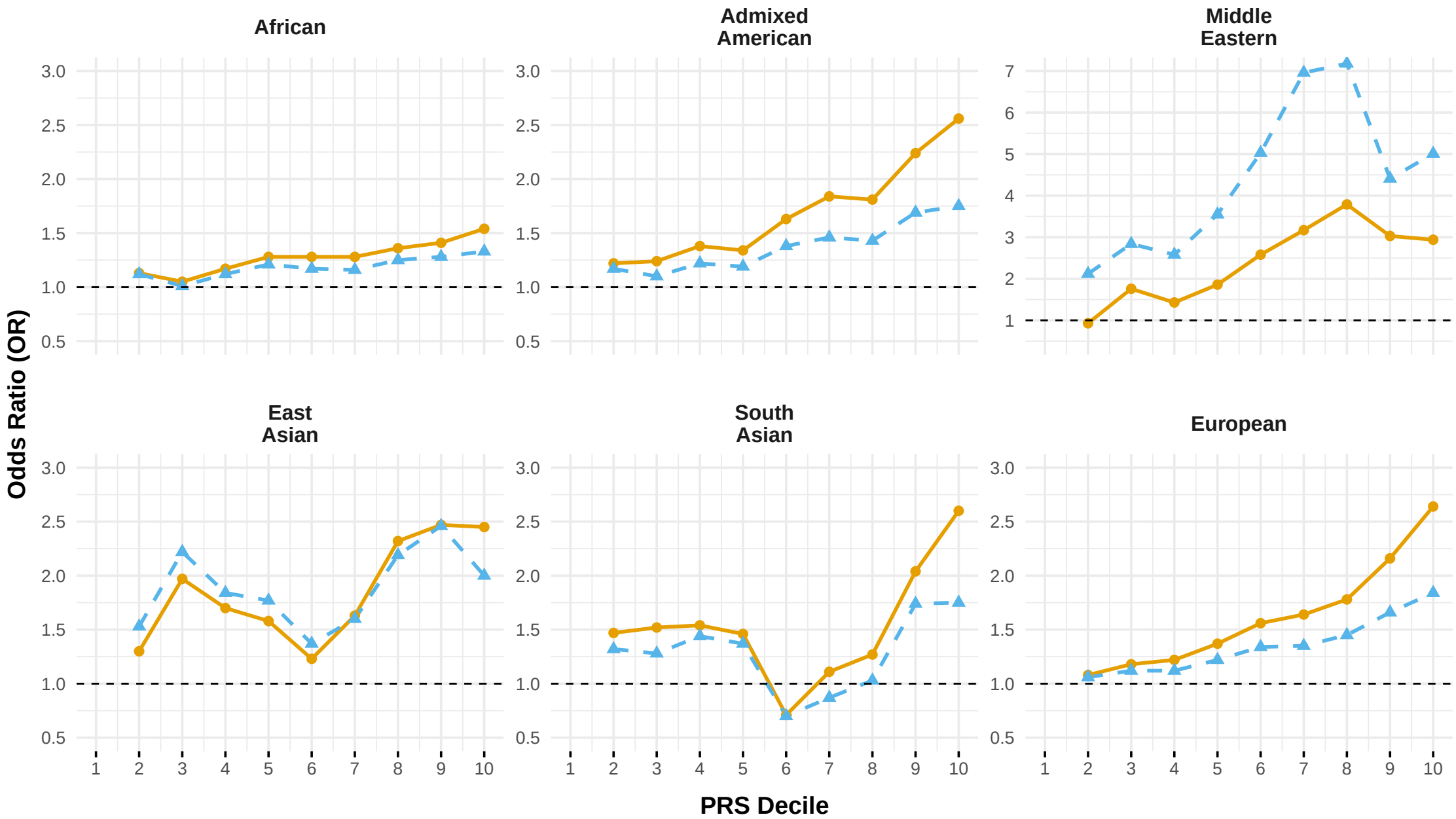

-log10(p value)
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome

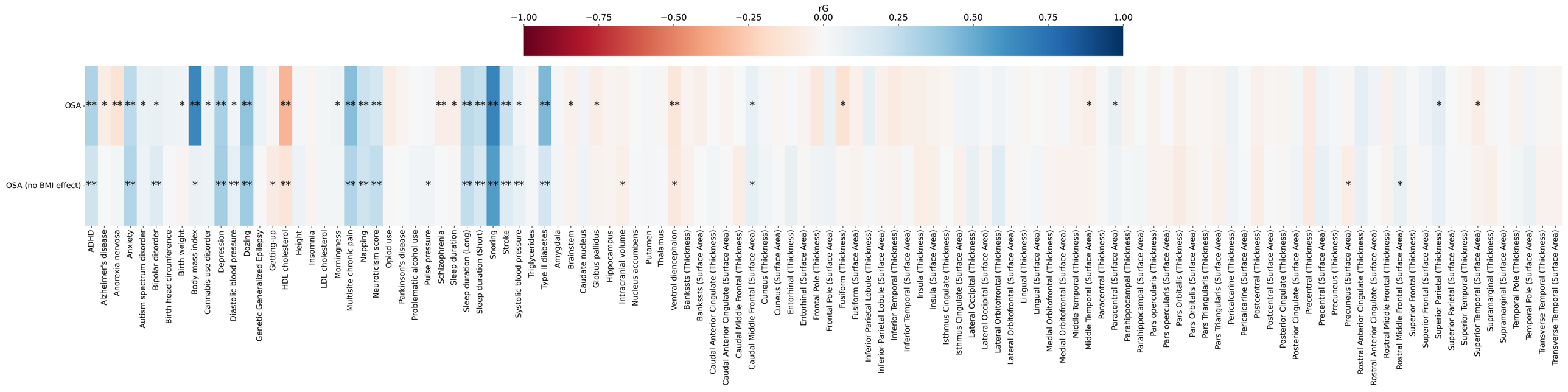


A



B




rG
−1.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
OSA
OSA (no BMI effect)
ADHD
Alzheimer's disease
Anorexia nervosa
Anxiety
Autism spectrum disorder
Bipolar disorder
Birth head circunference
Birth weight
Body mass index
Cannabis use disorder
Depression
Diastolic blood pressure
Dozing
Genetic Generalized Epilepsy
Getting-up
HDL cholesterol
Height
Insomnia
LDL cholesterol
Morningness
Multisite chronic pain
Napping
Neuroticism score
Opiod use
Parkinson's disease
Problematic alcohol use
Pulse pressure
Schizophrenia
Sleep duration
Sleep duration (Long)
Sleep duration (Short)
Snoring
Stroke
Systolic blood pressure
Triglycerides
Type II diabetes
Amygdala
Brainstem
Caudate nucleus
Globus pallidus
Hippocampus
Intracranial volume
Nucleus accumbens
Putamen
Thalamus
Ventral diencephalon
Bankssts (Thickness)
Bankssts (Surface Area)
Caudal Anterior Cingulate (Thickness)
Caudal Anterior Cingulate (Surface Area)
Caudal Middle Frontal (Thickness)
Caudal Middle Frontal (Surface Area)
Cuneus (Thickness)
Cuneus (Surface Area)
Entorhinal (Thickness)
Entorhinal (Surface Area)
Frontal Pole (Thickness)
Frontal Pole (Surface Area)
Fusiform (Thickness)
Fusiform (Surface Area)
Inferior Parietal Lobule (Thickness)
Inferior Parietal Lobule (Surface Area)
Inferior Temporal (Thickness)
Inferior Temporal (Surface Area)
Insula (Thickness)
Insula (Surface Area)
Isthmus Cingulate (Thickness)
Isthmus Cingulate (Surface Area)
Lateral Occipital (Thickness)
Lateral Occipital (Surface Area)
Lateral Orbitofrontal (Thickness)
Lateral Orbitofrontal (Surface Area)
Lingual (Thickness)
Lingual (Surface Area)
Medial Orbitofrontal (Thickness)
Medial Orbitofrontal (Surface Area)
Middle Temporal (Thickness)
Middle Temporal (Surface Area)
Paracentral (Thickness)
Paracentral (Surface Area)
Parahippocampal (Thickness)
Parahippocampal (Surface Area)
Pars opercularis (Thickness)
Pars opercularis (Surface Area)
Pars Orbitalis (Thickness)
Pars Orbitalis (Surface Area)
Pars Triangularis (Thickness)
Pars Triangularis (Surface Area)
Pericalcarine (Thickness)
Pericalcarine (Surface Area)
Postcentral (Thickness)
Postcentral (Surface Area)
Posterior Cingulate (Thickness)
Posterior Cingulate (Surface Area)
Precentral (Thickness)
Precentral (Surface Area)
Precuneus (Thickness)
Precuneus (Surface Area)
Rostral Anterior Cingulate (Thickness)
Rostral Anterior Cingulate (Surface Area)
Rostral Middle Frontal (Thickness)
Rostral Middle Frontal (Surface Area)
Superior Frontal (Thickness)
Superior Frontal (Surface Area)
Superior Parietal (Thickness)
Superior Parietal (Surface Area)
Superior Temporal (Thickness)
Superior Temporal (Surface Area)
Supramarginal (Thickness)
Supramarginal (Surface Area)
Temporal Pole (Thickness)
Temporal Pole (Surface Area)
Transverse Temporal (Thickness)
Transverse Temporal (Surface Area)