nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01785-5

Exome sequencing identifies genes associated with sleep-related traits

Received: 3 June 2023

Accepted: 15 November 2023

Published online: 04 January 2024

Check for updates

Chen-Jie Fei[1,6], Ze-Yu Li[2,3,6], Jing Ning[1,6], Liu Yang[1,6], Bang-Sheng Wu[1], Ju-Jiao Kang[2,3], Wei-Shi Liu[1], Xiao-Yu He[1], Jia You[2,3], Shi-Dong Chen[1], Huan Yu[1], Zhi-Li Huang[4], Jian-Feng Feng[2,3,5], Jin-Tai Yu[1] ✉ & Wei Cheng[1,2,3] ✉

Sleep is vital for human health and has a moderate heritability. Previous genome-wide association studies have limitations in capturing the role of rare genetic variants in sleep-related traits. Here we conducted a large-scale exome-wide association study of eight sleep-related traits (sleep duration, insomnia symptoms, chronotype, daytime sleepiness, daytime napping, ease of getting up in the morning, snoring and sleep apnoea) among 450,000 participants from UK Biobank. We identified 22 new genes associated with chronotype (*ADGRL4*, *COL6A3*, *CLK4* and *KRTAP3-3*), daytime sleepiness (*ST3GAL1* and *ANKRD12*), daytime napping (*PLEKHM1*, *ANKRD12* and *ZBTB21*), snoring (*WDR59*) and sleep apnoea (13 genes). Notably, 20 of these genes were confirmed to be significantly associated with sleep disorders in the FinnGen cohort. Enrichment analysis revealed that these discovered genes were enriched in circadian rhythm and central nervous system neurons. Phenotypic association analysis showed that *ANKRD12* was associated with cognition and inflammatory traits. Our results demonstrate the value of large-scale whole-exome analysis in understanding the genetic architecture of sleep-related traits and potential biological mechanisms.

Sleep is a vital biological process regulated by the circadian system. Chronic sleep disturbances, affecting approximately 30% of adults worldwide[1], have been linked to a broad spectrum of negative consequences, including cardiometabolic diseases, neuropsychiatric disorders, immunological dysfunctions and all-cause mortality[2–5]. However, the molecular underpinning of sleep regulation in overall health remains unclear. Family studies have indicated that self-reported sleep behaviours have a heritability of 10–45%[6,7]; therefore, conducting genetics studies might offer valuable insights into the physiology of sleep and its association with chronic diseases.

Over the past decade, genome-wide association studies (GWAS) have identified hundreds of independent loci associated with self-reported or device-measured sleep-related traits, including sleep duration[8–11], insomnia[12], excessive daytime sleepiness[13], daytime napping[14], chronotype[15–17] and snoring[18]. The most consistently observed associations with longer sleep duration include single nucleotide polymorphisms (SNPs) located near the *PAX8* and *VRK2* genes, which have been replicated in several studies[8–11,19]. SNPs near clock genes (*PER2*, *PER3* and *PATJ*), which are involved in the sleep/wake cycle, have been identified to be associated with chronotype[9,20]. These studies have laid

[1]Department of Neurology and National Center for Neurological Disorders, Huashan Hospital, State Key Laboratory of Medical Neurobiology and MOE Frontiers Center for Brain Science, Shanghai Medical College, Fudan University, Shanghai, China. [2]Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [3]Key Laboratory of Computational Neuroscience and Brain-Inspired Intelligence, Fudan University, Ministry of Education, Shanghai, China. [4]Department of Pharmacology, School of Basic Medical Sciences, State Key Laboratory of Medical Neurobiology, Institutes of Brain Science and Collaborative Innovation Center for Brain Science, Fudan University, Shanghai, China. [5]Department of Computer Science, University of Warwick, Coventry, UK. [6]These authors contributed equally: Chen-Jie Fei, Ze-Yu Li, Jing Ning, Liu Yang. ✉e-mail: jintai_yu@fudan.edu.cn; wcheng@fudan.edu.cn

the foundation for unravelling the underlying genetic mechanisms involved in sleep regulation.

However, previous GWAS have primarily focused on common variants (minor allele frequency (MAF) > 1%), which tend to have smaller effect sizes. Additionally, many of the identified loci from GWAS are mapped to non-coding regions of the genome, posing challenges in exploring the underlying mechanisms. In contrast, whole-exome sequencing (WES) analysis could overcome these limitations. WES data concentrate on protein-coding regions of the genome, providing the opportunity to identify rare and ultrarare variants that may not be genotyped or may be genotyped and imputed inaccurately[21]. Currently, there is a lack of studies using large-scale WES data to investigate sleep-related phenotypes. Such analysis may therefore provide a more comprehensive understanding of the genetic architecture underlying sleep-related traits and guide health improvement strategies.

The WES analysis of sleep we conducted in this study is, to the best of our knowledge, the largest analysis of this type performed thus far. Through single-variant association tests and gene-based collapsing analysis of the WES data from approximately 450,000 UK Biobank participants[22], we investigated the associations between genetic variants and eight sleep-related traits, including sleep duration, insomnia symptoms, chronotype, ease of getting up in the morning, daytime sleepiness, daytime napping, snoring and sleep apnoea. Sixty-eight genes, of which 22 have not been reported in previous studies, were identified to be associated with sleep-related traits at exome-wide significance, and 20 genes were validated for their associations with sleep disorders in the FinnGen cohort. We also provided biological insights into the identified genes and estimated their genetic burden heritability. Additionally, we performed pathway enrichment analysis and a phenome-wide association study (PheWAS) to uncover the underlying mechanisms and potential impacts on other traits. This study advances our understanding of the genetic factors influencing sleep and may guide personalized approaches for improving sleep health.

Results

Description of the study population and data

We used phenotypic and genetic data from UK Biobank, including exome sequencing data, sleep-related traits (Supplementary Table 1) and 249 phenotypes used in the phenome-wide association analysis (Supplementary Table 2). The exome sequencing data were put through a series of quality control steps (Methods) to remove low-quality variants and samples. In the main analysis, about 294,000 white British individuals with sleep-related traits were used to perform the association tests, with ages from 38 to 72 at enrolment, of which 54.7% were female. The sleep-related traits included sleep duration, insomnia symptoms, chronotype, ease of getting up in the morning, daytime sleepiness, daytime napping, snoring and sleep apnoea. The demographic characteristics of sleep-related traits after exclusion and genetic quality control (Methods) are provided in Supplementary Table 3. Finally, we obtained a total of 13,553,257 distinct autosomal genetic variants, comprising 100,081 common variants (MAF > 1%) and 13,453,176 rare variants (MAF < 1%). Our overall study design is shown in Fig. 1.

Exome-wide association analysis for sleep-related traits

To identify the degree to which different types of mutations would impact sleep, we assessed the associations between sleep-related traits and five mutation groups and found that the loss-of-function (LoF) variant burden had a significant influence on insomnia symptoms (regression coefficient (β) = 0.15; $P = 7.4 \times 10^{-5}$), daytime sleepiness ($\beta = 0.01$, $P = 9.1 \times 10^{-5}$) and daytime napping ($\beta = 0.02$, $P = 2.7 \times 10^{-7}$) (Supplementary Fig. 1 and Supplementary Table 4). To further explore the relationship between sleep-related traits and coding variants, we used SAIGE[23] to calculate associations at both the single-variant and gene-based levels.

First, we performed a single-variant exome-wide association analysis to identify genetic loci associated with sleep, adjusting for age, sex, body mass index (BMI) and ten principal components. We identified 75 genetic associations across seven of the sleep-related traits at the significance threshold ($P < 1 \times 10^{-8}$) (Fig. 2 and Supplementary Tables 5 and 6). The corresponding quantile–quantile plots are shown in Supplementary Fig. 2. The identified significant genomic loci mapped to 60 genes, and 16 of them have not been reported in previous studies (Table 1). For insomnia symptoms, two associations were observed at the *VPS8* locus (rs6766721T, $\beta = 0.037$, $P = 6.3 \times 10^{-9}$) and the *CNNM2* locus (rs943037T, $\beta = -0.067$, $P = 9.0 \times 10^{-9}$). For chronotype, the strongest association was observed at rs1144566T, a missense variant in *RGS16* ($\beta = 0.078$, $P = 1.3 \times 10^{-31}$), which contributed to morning sleep. For ease of getting up in the morning, the strongest association was rs2653349A, a missense variant in *HCRTR2* ($\beta = 0.028$, $P = 1.7 \times 10^{-26}$). We identified six genes associated with chronotype and ease of getting up in the morning, which suggests a close relationship between the two phenotypes. The *PATJ* locus (rs12140153T, $\beta = -0.023$, $P = 1.5 \times 10^{-14}$) was most closely associated with daytime sleepiness, and rs1876831T in *CRHR1* ($\beta = -0.027$, $P = 1.0 \times 10^{-31}$) was most closely associated with daytime napping, contributing to less daytime sleepiness and less napping during the day, respectively. Only one locus at *MSRB3* (rs61921506T, $\beta = 0.037$, $P = 2.5 \times 10^{-10}$) was associated with snoring. Rs1287581425A in *CGN* ($\beta = 3.32$, $P = 8.9 \times 10^{-17}$) was associated with sleep apnoea and manifested the largest coefficient. In addition, to validate the significance of the 16 genes that were not reported previously, we queried the FinnGen online GWAS summary statistics[24], and 14 of them were found to have significant associations with sleep disorders (Table 1).

Second, we performed the gene-based collapsing test. We used LoF and missense variants of each gene to perform the analysis, with two MAF thresholds (<1% and <0.1%). In total, we found that seven genes were significantly associated with sleep-related traits after Bonferroni correction ($P < 2.5 \times 10^{-6}$) (Fig. 3, Table 2 and Supplementary Table 7). The corresponding quantile–quantile plots are shown in Supplementary Fig. 3. Rare variants in *PER3* ($\beta = 4.9 \times 10^{-3}$, $P = 6.7 \times 10^{-17}$), *PER2* ($\beta = 2.0 \times 10^{-2}$, $P = 3.3 \times 10^{-11}$) and *MTNR1B* ($\beta = 4.0 \times 10^{-3}$, $P = 1.1 \times 10^{-6}$) were associated with the morning chronotype. *PER3* ($\beta = 2.8 \times 10^{-3}$, $P = 2.9 \times 10^{-8}$) and *PER2* ($\beta = 1.7 \times 10^{-2}$, $P = 1.7 \times 10^{-9}$) were also associated with getting up in the morning easily. *ST3GAL1* ($\beta = 7.7 \times 10^{-3}$, $P = 1.4 \times 10^{-6}$) and *ENSG00000258603* ($\beta = 1.5 \times 10^{-1}$, $P = 7.6 \times 10^{-9}$, not formally named and not considered in subsequent analysis) were associated with more daytime sleepiness. *ZBTB21* ($\beta = 1.8 \times 10^{-2}$, $P = 1.1 \times 10^{-6}$) was associated with more napping during the day. *WDR59* ($\beta = 2.6 \times 10^{-2}$, $P = 4.2 \times 10^{-7}$) was associated with snoring. In addition, three genes were identified after false discovery rate (FDR) correction, including *CLK4* (associated with the morning chronotype), *KRTAP3-3* (associated with the evening chronotype) and *ANKRD12* (associated with more daytime sleepiness and daytime napping). Among these associations, those of *PER3*, *PER2* and *MTNR1B* with the morning chronotype have been previously reported by other GWAS[17] (Supplementary Fig. 4). In addition, all of the genes identified in the gene-based test were found to be significantly associated with sleep disorders in the FinnGen study (Table 2). Most single variants in genes identified by the gene-based collapsing test showed a consistency in effect directions, forming long allelic series (Supplementary Table 8). Moreover, the burden of the rare variants had the same direction as each variant accumulated. In addition, the pleiotropy of sleep-related signals is shown in Supplementary Tables 9 and 10, and shared genetic determinants were found for chronotype and ease of getting up in the morning.

Sensitivity analysis, leave-one-variant-out analysis and conditional analysis

For all identified significant associations, we performed sex-specific and ancestry-specific analysis. In the sex-specific analysis, the β values of the associations were very similar to the original values, and

https://doi.org/10.1038/s41562-023-01785-5

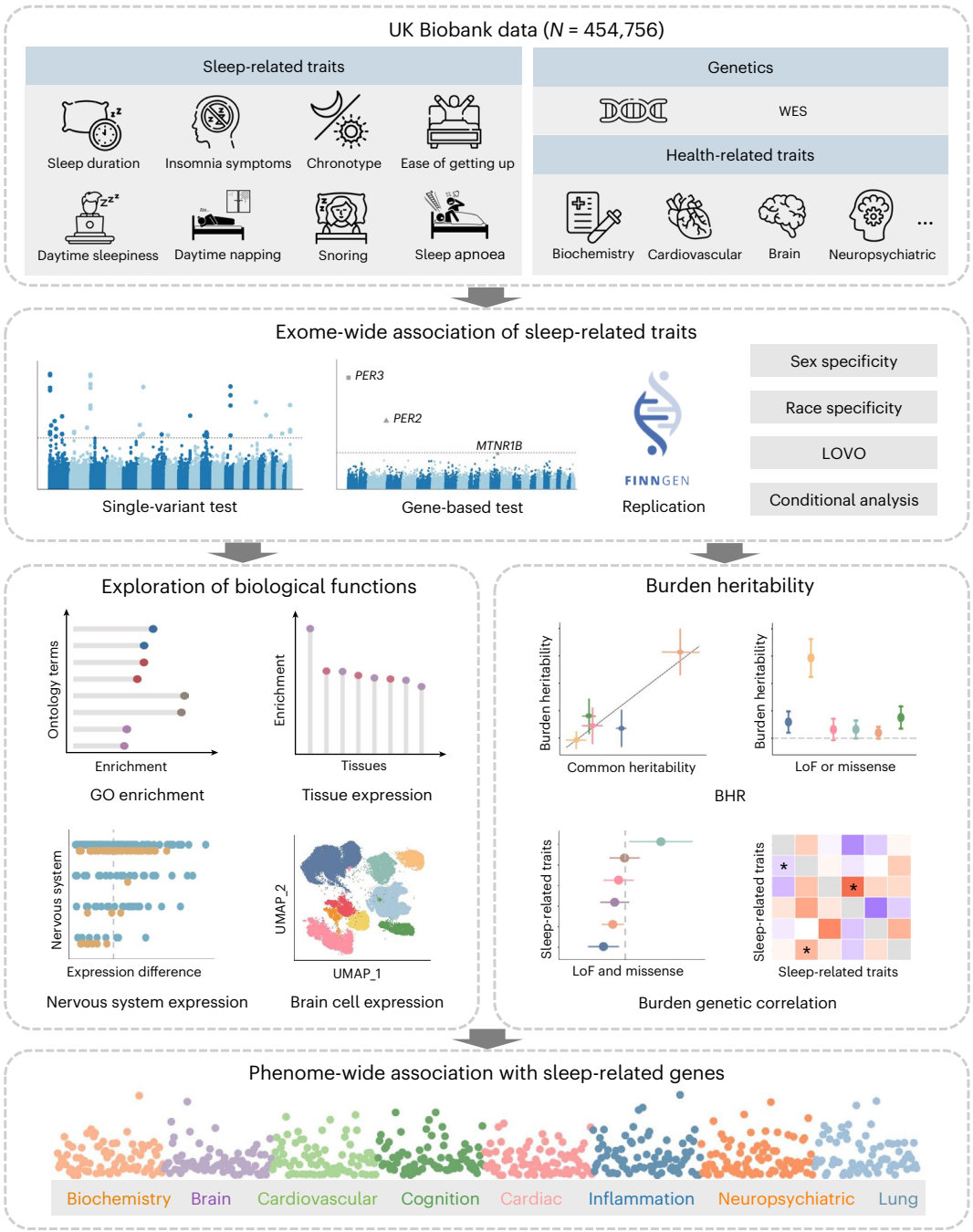

Fig. 1 | Design of the study. Top, data used in the analyses, including sleep-related traits, exome sequencing data and health-related traits. Second row, exome-wide association analysis of sleep-related traits, including single-variant association tests, gene-based collapsing tests, replication analysis, sex and race specificity analysis, LOVO analysis and conditional analysis. Third row left, biological function analyses of genes identified in the exome-wide analyses, including GO enrichment, tissue expression, nervous system expression and brain cell expression analyses. Third row right, BHR and burden genetic correlation of sleep-related traits. Bottom, phenome-wide association analysis of the identified genes.

all genes and variants showed associations with sleep-related traits (Supplementary Fig. 5 and Supplementary Tables 11 and 12). In individuals of white (non-British) ancestry, two associations (*PER3* and *PER2* with chronotype) were nominally significant ($P = 1.5 \times 10^{-2}$ and $P = 3.0 \times 10^{-2}$, respectively). Two associations (*ANKRD12* with daytime napping and *PER2* with ease of getting up in the morning) were identified in individuals of Asian ancestry ($P = 1.2 \times 10^{-2}$ and $P = 2.2 \times 10^{-3}$, respectively), and one association (*PER3* with ease of getting up in the morning) was identified in individuals of Black ancestry ($P = 1.7 \times 10^{-2}$) (Supplementary Fig. 6 and Supplementary Tables 13 and 14).

To further assess the stability of the identified associations in the gene-based collapsing test, we performed a leave-one-variant-out (LOVO) analysis. For 8 of 12 associations, the removal of one variant did not affect the overall associations, which implies that these associations were driven by a burden of multiple contributing variants. After we removed the locus chr1:g.7809893[C/G] from the gene *PER3*, the association with the morning chronotype and ease of getting up became less significant ($P = 8.4 \times 10^{-3}$ and $P = 2.4 \times 10^{-2}$, respectively). After we removed the locus chr5:g.178608454[T/A] from the gene *CLK4* and the locus chr17:g.40994058[C/T] from the gene *KRTAP3-3*,

https://doi.org/10.1038/s41562-023-01785-5

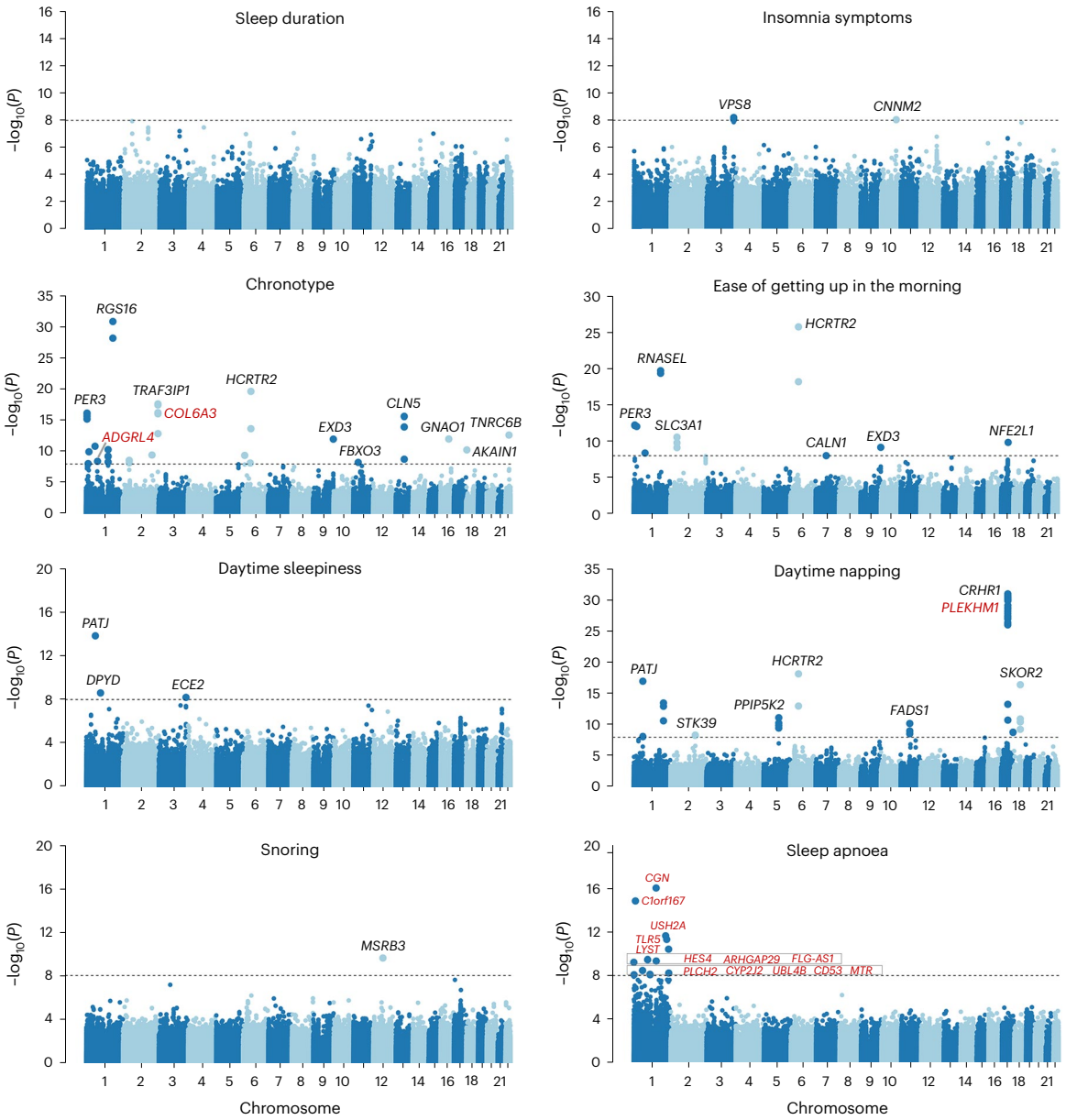

Fig. 2 | Exome-wide single-variant tests for sleep-related traits. Manhattan plots showing the results of the tests for single variants associated with sleep-related traits, including sleep duration, insomnia symptoms, chronotype, ease of getting up in the morning, daytime sleepiness, daytime napping, snoring and sleep apnoea. The *x* axis indicates the positions of the single variants on 22 chromosomes, and the *y* axis indicates the $-\log_{10}$ of the *P* value for each association. The grey dashed line indicates the threshold for significant associations ($P < 1 \times 10^{-8}$). The genes corresponding to the variants significantly associated with sleep-related traits are marked on the plots, with red representing new identified genes. Only the top signals and new genes in each chromosome are tagged for chronotype, ease of getting up in morning and daytime napping. Single-variant association analyses were performed using SAIGE-GENE+ software. The *P* values shown are two-sided and unadjusted for multiple testing. All models were adjusted for age, sex, BMI and the top ten ancestral principal components.

the association with chronotype became less significant ($P = 1.2 \times 10^{-1}$ and $P = 2.0 \times 10^{-1}$, respectively). (Supplementary Figs. 7–11 and Supplementary Table 15). This implies that, in certain cases, a single variant might play an important role in the relationship between genes and sleep-related traits.

We then performed the conditional analysis to evaluate whether the identified genes were independent of nearby common variants. First, we ran the common variant (MAF > 0.5%) association analysis within the region ±500 kb of the gene. We reran the gene-based collapsing analysis for the identified genes, adding the clumped common variants as covariates (Methods). After conditioning on nearby common variants, the *P* values of the associations were not substantially attenuated, and the effect sizes remained similar. All associations examined were significant after undergoing conditional analysis (Supplementary Table 16).

Burden heritability of sleep-related traits

Burden heritability regression (BHR) is a method to quantify the heritability explained by the burden of rare coding variants. Weiner et al. showed that rare coding variants explain 1.3% of the phenotypic variance on average[25]. We first used BHR to estimate the burden heritability of eight sleep-related traits. As recommended in BHR, variants were stratified into bins according to allele frequency and functional categories. For ultrarare LoF variants, all sleep-related traits except sleep

https://doi.org/10.1038/s41562-023-01785-5

Table 1 | New exome-wide significant variants associated with sleep-related traits ($P<1\times10^{-8}$)

Trait	Chr	rsID	A0	A1	A1 freq (%)	β	s.e.	P	Gene	Function	P (FinnGen)
Quantitative											
Chronotype	1	rs985277	G	T	40.0	−0.0138	0.0023	3.72×10^{-9}	*ADGRL4*	Intronic	5.3×10^{-1}
	2	rs2077061420	C	T	4.3	−0.0471	0.0057	8.15×10^{-17}	*COL6A3*	Missense	1.5×10^{-3}
Daytime napping	17	rs1879581	T	C	18.6	−0.0273	0.0024	6.58×10^{-29}	*PLEKHM1*	Synonymous	8.9×10^{-1}
Binary											
Sleep apnoea	1	rs369900342	T	A	0.007	2.4627	0.3978	5.96×10^{-10}	*HES4*	Missense	4.7×10^{-4}
	1	rs774775589	G	A	0.005	2.6723	0.4638	8.33×10^{-9}	*PLCH2*	Missense	5.3×10^{-5}
	1	rs1284009203	CA	C	0.006	2.9943	0.3749	1.39×10^{-15}	*C1orf167*	Frameshift	1.5×10^{-4}
	1	rs759605633	T	C	0.005	2.5755	0.4357	3.41×10^{-9}	*CYP2J2*	Intronic	6.5×10^{-3}
	1	rs752253132	G	A	0.013	2.1973	0.3501	3.45×10^{-10}	*ARHGAP29*	Splice region variant	2.7×10^{-4}
	1	rs769644497	G	A	0.005	2.6392	0.4568	7.60×10^{-9}	*UBL4B*	Missense	4.9×10^{-3}
	1	rs188506897	G	A	0.008	2.4122	0.4188	8.42×10^{-9}	*CD53*	Intronic	3.9×10^{-4}
	1	rs1287581425	G	A	0.004	3.3212	0.3992	8.87×10^{-17}	*CGN*	Missense	3.2×10^{-4}
	1	rs1218957616	A	T	0.004	2.9595	0.4747	4.54×10^{-10}	*FLG-AS1*	Intronic, non-coding transcript	2.8×10^{-4}
	1	rs199885325	C	T	0.004	3.1108	0.4433	2.27×10^{-12}	*USH2A*	Missense	1.3×10^{-4}
	1	rs200954588	G	A	0.004	2.9723	0.4303	4.92×10^{-12}	*TLR5*	Missense	3.3×10^{-3}
	1	rs766823455	A	G	0.004	2.9138	0.4410	3.92×10^{-11}	*LYST*	Missense	1.1×10^{-3}
	1	rs371286986	C	T	0.004	2.7215	0.4676	5.86×10^{-9}	*MTR*	Intronic	6.6×10^{-5}

The association results for the corresponding genes for 'sleep disorders (combined)' or 'sleep apnoea' in FinnGen are also shown. Single-variant association analyses were performed using SAIGE-GENE+ software. The *P* values shown are two-sided and unadjusted for multiple testing. Positive β values mean that mutation carriers tend to have longer sleep duration, the morning chronotype, more daytime sleepiness and more daytime napping; get up in the morning easily; and have a higher risk of insomnia symptoms, snoring and sleep apnoea. Chr, chromosome; A0, allele 0; A1, allele 1; A1 freq, allele 1 frequency in the analysed sample.

duration had significant non-zero burden heritability ranging from 0.10% to 0.42% (Fig. 4a and Supplementary Table 17). For rare LoF variants, sleep duration (h^2 = 0.14%, s.e. = 0.05%), chronotype (h^2 = 0.10%, s.e. = 0.05%) and sleep apnoea (h^2 = 0.26%, s.e. = 0.06%) had significant non-zero burden heritability (Fig. 4b and Supplementary Table 17). Damaging missense variants explained less burden heritability than LoF variants for both ultrarare and rare bins. We aggregated ultrarare and rare LoF burden heritability together as total burden heritability and compared it with the common-variant heritability of each trait (reported in refs. 9,13,14,17,18,26, using linkage disequilibrium score regression[27]). Insomnia symptoms had the highest total burden heritability (h^2 = 0.73%, s.e. = 0.08%) and common-variant heritability (h^2 = 20.6%, s.e. = 1.1%). Burden heritability and common-variant heritability of sleep-related traits were correlated, while common variants explained a much larger fraction of phenotypic variance (mean, 11.5% versus 0.4%) (Fig. 4c and Supplementary Table 18).

We then used BHR to compute the genetic correlations between LoF and missense variants across sleep-related traits (Fig. 4d,e and Supplementary Table 19). LoF–missense genetic correlations were low in sleep apnoea, suggesting that LoF and missense variants in the same genes might have divergent phenotypic effects. We also estimated the burden genetic correlations across sleep-related traits from ultrarare and rare LoF variants (Fig. 4f and Supplementary Table 20). The pair of chronotype and ease of getting up in the morning showed a large genetic correlation (for ultrarare LoF variants, r_g = 64.7%, s.e. = 25.5%; for rare LoF variants, r_g = 62.0%, s.e. = 32.3%), indicating shared genetic architecture between the two traits, which is consistent with the findings in the gene-based collapsing analysis.

Biological functions of sleep-related genes

To characterize the biological properties of the identified genes, we first conducted a pathway enrichment analysis. As expected, Gene Ontology (GO) enrichment analysis revealed that 68 significant genes in single-variant and gene-based tests were enriched in the sleep-related ontologies, such as regulation of circadian sleep/wake cycle ($P = 5.1\times10^{-5}$), circadian sleep/wake cycle process ($P = 6.7\times10^{-5}$) and circadian sleep/wake cycle ($P = 9.6\times10^{-5}$) (Fig. 5a and Supplementary Table 21).

We then explored the expression of the identified genes in different tissues and cell types. First, we performed enrichment analysis on 68 significant genes in 54 different tissues from GTEx[28], and the top three significantly enriched tissues were pancreas ($P = 9.4\times10^{-10}$), brain hippocampus ($P = 1.0\times10^{-6}$) and adrenal gland ($P = 1.1\times10^{-6}$) (Fig. 5b and Supplementary Table 22). Next, we used a single-cell atlas of the mouse nervous system[29] and performed a one-sided Wilcoxon rank sum test comparing the expression of the 68 genes against all remaining genes. We found that signals in four parts of the nervous system were higher in neuronal cells and the central nervous system than those in non-neuronal cells and the peripheral nervous system (Fig. 5c and Supplementary Table 23). In addition, brain single-cell RNA sequencing (scRNA-seq) data[30] were used to analyse the expression of each gene in different neuronal cells, revealing high expression levels of *PER3* and *ANKRD12* in neuronal cells (Fig. 5d–f and Supplementary Fig. 12).

Phenotypic association with sleep-related genes

To explore the association between sleep genes and a large range of phenotypes, we conducted a phenotypic association analysis on 25 sleep-associated genes (16 significant new genes in the single-variant test and 9 significant genes in the gene-based test). The phenotypes we used included eight categories (biochemistry, brain structure, cardiac function, cardiovascular diseases, cognition, inflammation, lung function and neuropsychiatric diseases). The results for genes identified in the gene-based test are shown in Fig. 6 and Supplementary Table 24. We found that *ANKRD12* had associations with the widest range of traits—for example, fluid intelligence ($P = 2.2\times10^{-10}$), total protein ($P = 1.1\times10^{-11}$), white blood cell count or percentage (for example,

https://doi.org/10.1038/s41562-023-01785-5

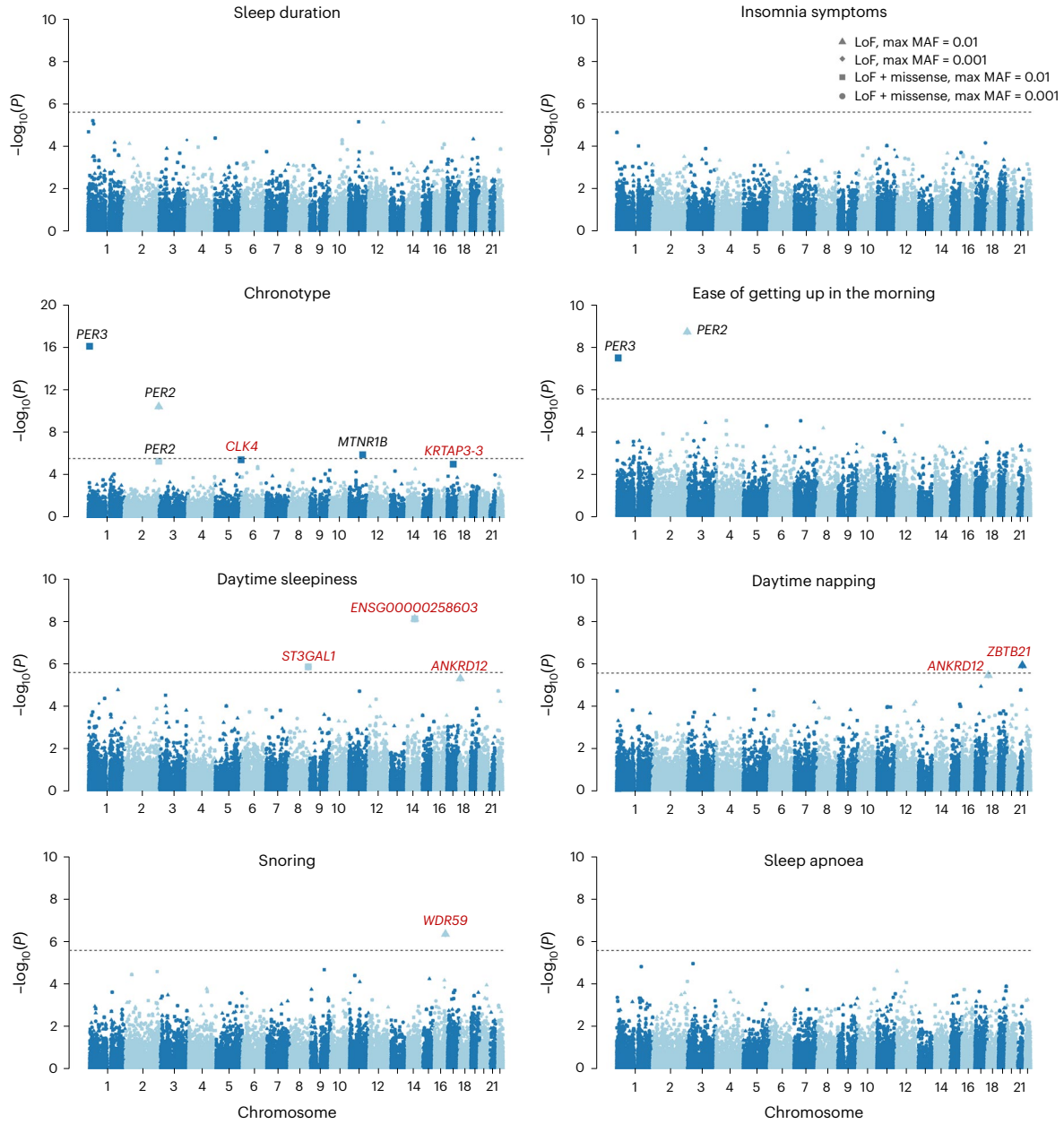

Fig. 3 | Exome-wide gene-based tests for sleep-related traits. Manhattan plots showing the results of the tests for genes associated with sleep-related traits, including sleep duration, insomnia symptoms, chronotype, ease of getting up in the morning, daytime sleepiness, daytime napping, snoring and sleep apnoea. Rare variants used in the gene-based collapsing test have two different annotation groups (LoF and LoF + missense) and two different maximum MAF cut-offs (1% and 0.1%). The *x* axis indicates the positions of the genes on 22 chromosomes, and the *y* axis indicates the $-\log_{10}$ of the *P* value for each association. The grey dashed line indicates the threshold for significant associations using Bonferroni correction ($P < 2.5 \times 10^{-6}$). The genes significantly associated with sleep-related traits, using FDR correction, are marked on the plots, with red representing new identified genes. Gene-based collapsing analyses were performed using SKAT-O tests provided by SAIGE-GENE+ software. The *P* values shown are two-sided and unadjusted for multiple testing. All models were adjusted for age, sex, BMI and the top ten ancestral principal components.

neutrophil percentage, $P = 3.1 \times 10^{-14}$) and lung function. *PER2* was associated with multiple brain regions, from frontal lobe (left middle frontal gyrus, $P = 3.2 \times 10^{-4}$; right dorsolateral superior frontal gyrus, $P = 2.8 \times 10^{-4}$) to supplementary motor area (left, $P = 5.8 \times 10^{-5}$; right, $P = 9.2 \times 10^{-6}$). *CLK4* was associated with orbitofrontal cortex, such as right posterior orbital gyrus ($P = 2.4 \times 10^{-6}$) and anterior orbital gyrus (left, $P = 1.5 \times 10^{-4}$; right, $P = 4.9 \times 10^{-5}$). In addition, *WDR59* was associated with triglycerides ($P = 1.1 \times 10^{-4}$), which may suggest that obese people are more likely to have snoring symptoms. The results for genes identified in the single-variant test are shown in Supplementary Table 25.

Discussion

In this extensive WES analysis, we identified 68 genes significantly associated ($P < 1 \times 10^{-8}$ for single-variant tests and FDR $Q < 0.05$ for gene-based tests) with eight sleep-related traits including sleep duration, insomnia symptoms, chronotype, ease of getting up in the morning, daytime sleepiness, daytime napping, snoring and sleep apnoea. In addition to confirming known sleep-related genes, we identified 22 genes that have not been reported in previous studies, including *ADGRL4*, *COL6A3*, *CLK4* and *KRTAP3-3* associated with chronotype; *ST3GAL1* and *ANKRD12* associated with daytime sleepiness; *PLEKHM1*, *ZBTB21* and *ANKRD12* associated with daytime napping; *WDR59*

https://doi.org/10.1038/s41562-023-01785-5

Table 2 | Genes associated with sleep-related traits at FDR Q<0.05

Trait	Chr	Gene	Group	Max MAF	MAC	β (95% CI)	P	P for sleep disorders (FinnGen)
Quantitative								
Chronotype	1	*PER3*	LoF+missense	0.01	3,850	0.005 (0.004, 0.006)	6.66×10^{-17}	7.5×10^{-4}
	2	*PER2*	LoF	0.001	142	0.020 (0.014, 0.026)	3.33×10^{-11}	4.7×10^{-4}
	5	***CLK4***	LoF+missense	0.01	3,637	0.003 (0.002, 0.004)	3.25×10^{-6}	6.6×10^{-4}
	11	*MTNR1B*	LoF+missense	0.01	2,109	0.004 (0.002, 0.006)	1.09×10^{-6}	2.8×10^{-3}
	17	***KRTAP3-3***	LoF+missense	0.01	286	−0.006 (−0.010, −0.002)	8.46×10^{-6}	2.4×10^{-4}
Daytime sleepiness	8	***ST3GAL1***	LoF+missense	0.001	244	0.008 (0.004, 0.011)	1.35×10^{-6}	5.5×10^{-4}
	18	***ANKRD12***	LoF	0.001	186	0.009 (0.005, 0.013)	4.80×10^{-6}	2.1×10^{-4}
Ease of getting up in the morning	1	*PER3*	LoF+missense	0.01	3,858	0.003 (0.002, 0.004)	2.93×10^{-8}	7.5×10^{-4}
	2	*PER2*	LoF	0.001	143	0.017 (0.011, 0.022)	1.74×10^{-9}	4.7×10^{-4}
Daytime napping	18	***ANKRD12***	LoF	0.001	188	0.010 (0.006, 0.014)	3.19×10^{-6}	2.1×10^{-4}
	21	***ZBTB21***	LoF	0.001	64	0.018 (0.011, 0.025)	1.10×10^{-6}	1.2×10^{-3}
Binary								
Snoring	16	***WDR59***	LoF	0.001	291	0.026 (0.016, 0.035)	4.22×10^{-7}	3.9×10^{-5}

The association results for the corresponding genes for 'sleep disorders (combined)' in FinnGen are also shown. Gene-based collapsing analyses were performed using SKAT-O tests provided by SAIGE-GENE+ software. The *P* values shown are two-sided and unadjusted for multiple testing. Positive β values mean that mutation carriers tend to have longer sleep duration, morning chronotype, more daytime sleepiness and more daytime napping; get up in the morning easily; and have a higher risk of insomnia symptoms, snoring and sleep apnoea. Gene names in boldface indicate that the gene has not been previously reported in sleep GWAS. MAC, minor allele count; CI, confidence interval.

associated with snoring; and *HES4*, *PLCH2*, *C1orf167*, *CYP2J2*, *ARHGAP29*, *UBL4B*, *CD53*, *CGN*, *FLG-AS1*, *USH2A*, *TLR5*, *LYST* and *MTR* associated with sleep apnoea. In the validation analysis, we confirmed the associations of 20 genes with sleep disorders in the FinnGen cohort. In addition, pathway enrichment analysis revealed that these genes exhibited high expression in neurons and were involved in circadian rhythm pathways. Through PheWAS analysis, we demonstrated significant correlations between *ANKRD12* and cognitive decline and inflammatory markers, enhancing our understanding of the biological mechanisms underlying the circadian cycle, sleep regulation and related disorders.

Our study validated 47 genes previously identified in GWAS analysis, and most showed consistent effect directions across multiple cohorts, including *EXD3*, *HCRTR2*, *PER3* and *PER2* with chronotype[20], *PATJ* with daytime napping[14] and *MSRB3* with snoring[18]. The BHR analysis revealed that the gene-wise burden of rare and ultrarare LoF variants explained 0.3% of the phenotypic variance on average across sleep-related traits. Additionally, in the conditional analysis, we demonstrated a direct relationship between these identified rare variants and sleep-related traits, independent of nearby (±500 kb) common variants. These findings highlight that the identification of rare variants and genes may contribute to a more comprehensive genetic landscape of sleep[31].

In the single-variant analysis, we have expanded our knowledge of the contribution of rare variants to sleep-related traits, going beyond the scope of previous GWAS, which mainly covered common variants. *ADGRL4* encodes a key regulator of angiogenesis, and the silence of *ADGRL4* affects the Notch pathway[32], subsequently leading to disturbances in the sleep circadian rhythm and resulting in a late chronotype[33]. In the current study, we identified a significant association between the missense variant rs2077061420 in *COL6A3* and the late chronotype. *COL6A3* encodes the α-3 chain of collagen VI, and mutations of the gene have been implicated in a continuum of skeletal muscle phenotypes, ranging from mild Bethlem myopathy (Mendelian Inheritance in Man (MIM) 158810) to severe Ullrich congenital muscular dystrophy (MIM 254090)[34,35]. Recent studies have provided evidence suggesting that collagen VI plays a role in mitigating central neuron apoptosis and maintaining the dopamine circuitry function[36,37], thereby contributing to the regulation of circadian rhythm homeostasis[38]. Our findings suggest that a lack of collagen VI may manifest as myopathy and circadian clock disruption. Thirteen rare variants were found to be correlated with sleep apnoea, indicating that obesity and lipid metabolism[39,40] (*CYP2J2* and *LYST*) and deficiency in craniofacial development[41] (*ARHGAP29*) are risk factors in sleep apnoea.

To identify the effect of the burden of rare genetic variants with MAF < 1% on sleep-related traits, we conducted a gene-based collapsing analysis and identified six previously unreported sleep-related genes. In the present study, we identified significant associations between *ST3GAL1* and excessive daytime sleepiness. *ST3GAL1* encodes β-galactoside α2-3 sialyltransferase, which is involved in terminal sialylation in ganglioside biosynthesis[42], and deficiency of gangliosides in the central nervous system contributes to neuropsychiatric diseases and narcolepsy[43]. *ANKRD12* was another gene significantly associated with daytime sleepiness and daytime napping, and previous GWAS or proteomics studies have revealed that *ST3GAL1* and *ANKRD12* are risk genes for bipolar disorder[44] and the negative symptoms of schizophrenia[45,46], suggesting a genetic correlation between daytime dozing and neuropsychiatric diseases[9]. In the PheWAS analysis, we also found that *ANKRD12* was significantly associated with cognitive decline and serum inflammatory markers, providing us insights into targets for improving daytime sleepiness symptoms. For chronotype, *KRTAP3-3* (associated with the late chronotype) and *CLK4* (associated with the early chronotype) could be potential clock genes; however, the underlying mechanisms remain to be further explored. We identified a significant link between snoring and *WDR59*, which encodes a component of the GATOR2 complex, recognized for its role in activating mTORC1[47], and mTORC1 orchestrates lipid biosynthesis through its interaction with the sterol responsive element binding protein transcription factor[48]. In the PheWAS analysis, we found a connection between *WDR59* and increased triglyceride levels, which implies its relationship with obesity and offers insights into the genetic connection between snoring and lipid metabolism or obesity[18]. However, the effect of *WDR59* on snoring remained significant after adjusting for BMI, suggesting the existence of other mechanisms that require further investigation. These findings highlight the value of exome sequencing in identifying rare and coding genes with crucial roles in sleep-related traits, as well as facilitating potential mechanistic exploration.

https://doi.org/10.1038/s41562-023-01785-5

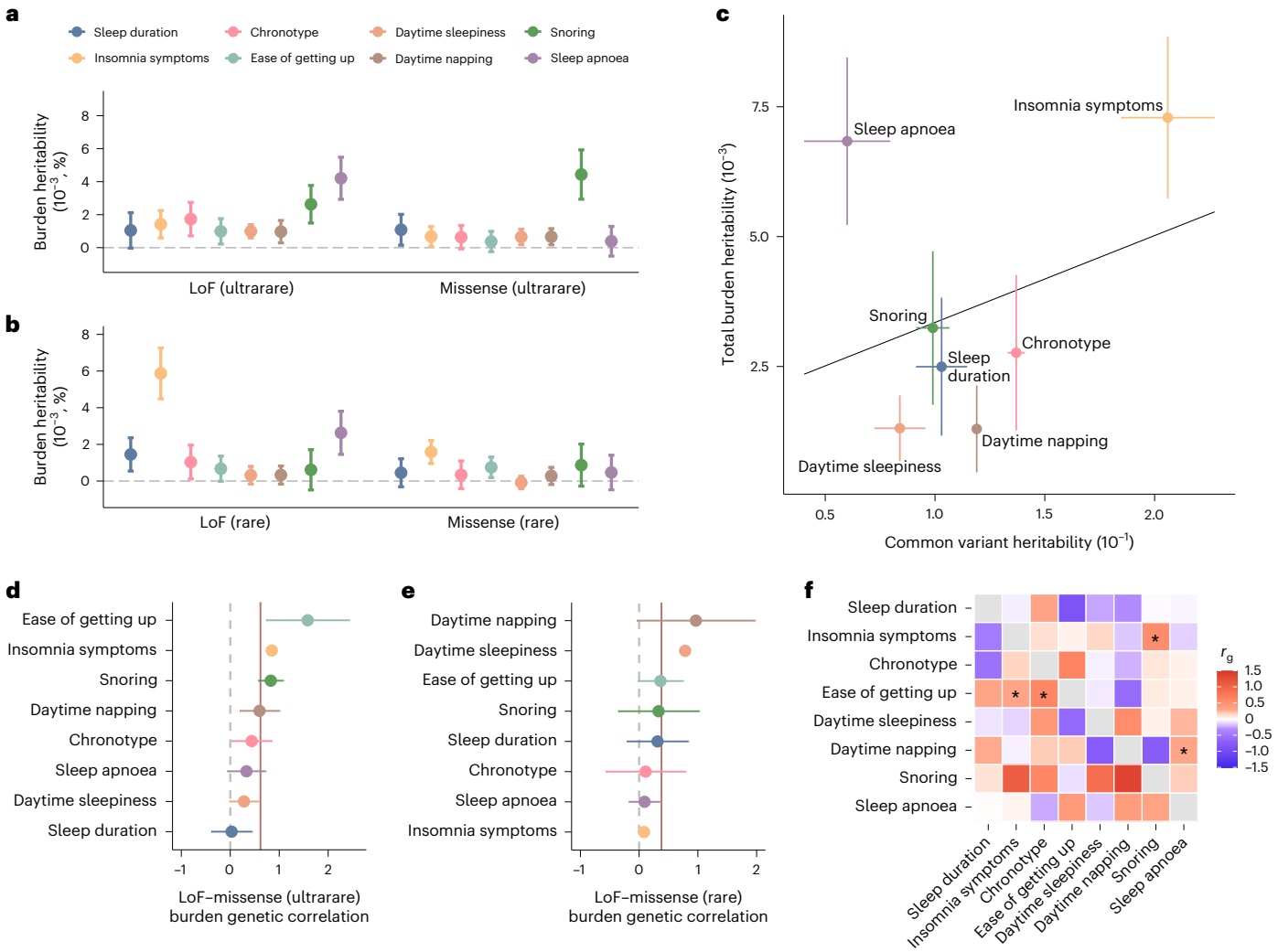

Fig. 4 | Burden heritability of sleep-related traits. a, Comparison of the total burden heritability (ultrarare and rare variants) with the common-variant heritability of each sleep trait (estimated using linkage disequilibrium score regression). Common-variant and total burden heritability estimates and 95% CIs are plotted. The linear-regression line shows the correlation between burden heritability and common-variant heritability. **b**, Burden heritability estimates and 95% CIs for only ultrarare variants (MAF $< 1 \times 10^{-5}$) in the LoF and missense groups. **c**, Burden heritability estimates and 95% CIs for only rare variants ($1 \times 10^{-5} \leq$ MAF $< 1 \times 10^{-3}$) in the LoF and missense groups. **d**, Burden genetic correlation estimates and 95% CIs between LoF and missense variants, for only ultrarare variants. The dashed line indicates a correlation of 0, and the solid line indicates the mean correlation across all sleep-related traits. **e**, Burden genetic correlation estimates and 95% CIs between LoF and missense variants, for only rare variants. The dashed line indicates a correlation of 0, and the solid line indicates the mean correlation across all sleep-related traits. **f**, Burden genetic correlations between sleep-related traits (*$P < 0.05$). The tests conducted in the analyses were two-sided, and no adjustments were made for multiple comparisons. The top triangle represents genetic correlations using ultrarare variants, and the bottom triangle represents genetic correlations using rare variants.

Our study also explored the biomedical mechanisms of these sleep-related genes. GO pathway enrichment analysis validated the significant association of the identified genes with circadian sleep/wake cycle and rhythmic behaviour[49]. Furthermore, these genes exhibited a significant correlation with magnesium ion homeostasis, aligning with previous animal research findings that reveal the crucial role of brain interstitial magnesium ion concentration in regulating the sleep/wake cycle[50]. A previous clinical study demonstrated that magnesium supplementation effectively enhances sleep quality[51], and our findings provide promising insights for advancing targeted therapeutic interventions aimed at addressing sleep disorders from a genetic perspective. In addition, tissue and single-cell expression analysis were conducted on 68 sleep-associated genes, revealing high expression in central nervous system neurons. The elevated expression of sleep-related genes in the hippocampus provides a genetic foundation supporting synaptic remodelling in the hippocampus during sleep and the association between weakened circadian rhythms and increased dementia risk in elder individuals[52–54]. Our research provides genetic insights for future studies exploring the mechanisms underlying the hypothesized neural network involved in the sleep/wake cycle. Furthermore, our results demonstrate that these genes are enriched in the pancreas. This finding substantiates previous research from a genetic perspective, which has shown that heightened GLP-1 and insulin secretion in the pancreas directly triggers the upregulation of the *PER1* and *PER2* genes in the liver, consequently instigating the synchronization of the circadian clock[55,56].

In addition, our study used gene-based collapsing analysis and BHR to explore the shared genetic architecture between sleep-related traits and the identified relationships between snoring, shorter sleep duration and insomnia symptoms, consistent with previous genetic and phenotypic correlation findings[9]. We also observed a strong genetic

https://doi.org/10.1038/s41562-023-01785-5

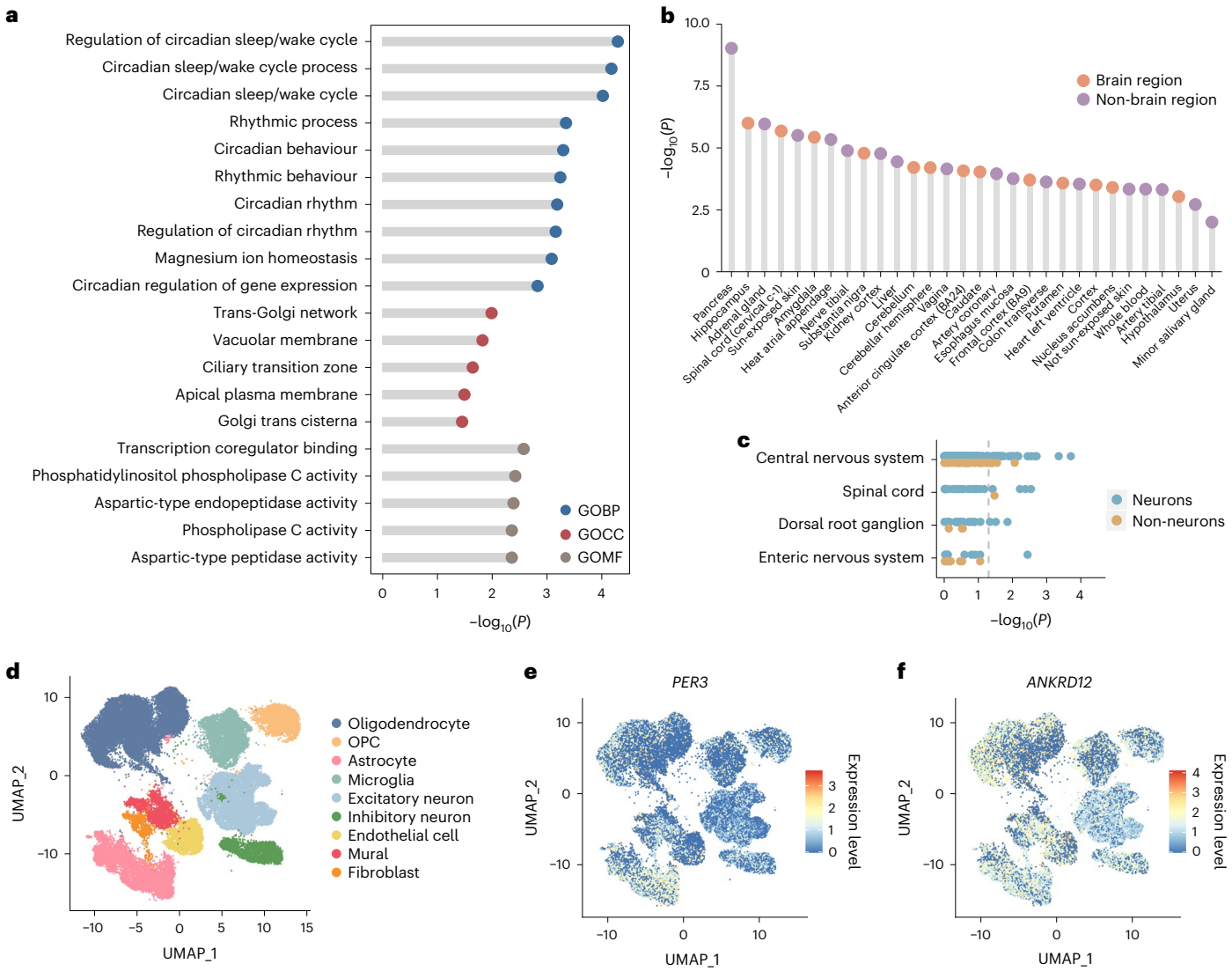

Fig. 5 | Biological functions of genes associated with sleep-related traits. **a**, Enrichment analysis of 68 significant genes identified in the single-variant test and the gene-based test, using the GO database. The *P* values shown are two-sided and unadjusted for multiple testing. GOBP, GO terms from the Biological Process domain; GOCC, GO terms from the Cellular Component domain; GOMF, GO terms from the Molecular Function domain. **b**, Enrichment analysis of 68 genes in 54 tissues in GTEx using FUMA. The top 30 tissues are shown in the plot. The *P* values shown are two-sided and unadjusted for multiple testing. **c**, Relative expression of sleep-associated genes in single-cell data from the mouse central nervous system. Each dot shows the *P* value from a one-sided Wilcoxon rank sum test comparing normalized expression values between sleep-associated genes and all other genes in a single cell type. The dots are stratified by four anatomical locations. **d**, Uniform manifold approximation and projection (UMAP) visualization of scRNA-seq data described previously[30]. The colour of each dot indicates the cell type. OPC, oligodendrocyte progenitor cell. **e**, Feature plot showing the expression level of *PER3* in different cell types. **f**, Feature plot showing the expression level of *ANKRD12* in different cell types.

correlation between the early chronotype and getting up easier in the morning, indicating a shared genetic basis involving genes such as *EXD3*, *EIF4G3*, *HCRTR2*, *PER3*, *PER2*, *PREPL* and *SLC3A1*, revealing shared biological mechanisms underlying the regulation of circadian rhythm for these two traits.

The strength of our study is its large-scale WES analysis of sleep-related traits, applying both single-variant and gene-based analysis and enhancing the completeness of the genetic landscape underlying sleep regulation. Moreover, through pathway enrichment and PheWAS analysis, we have gained better insights into the biological mechanisms underlying sleep regulation and the shared genetic correlations with other phenotypes, providing a solid foundation for the treatment of sleep disorders. However, the current study has several potential limitations. First, our primary study included exome data from white British participants in UK Biobank, and the ancestry-specific analysis showed insignificant gene–phenotype associations in other ethnic groups due to smaller sample sizes. Additionally, the FinnGen validation cohort employed a metric based on sleep disorders, distinct from the subjective sleep-related traits explored in our current study, and rare variants were not captured in GWAS. Future analysis could use large-scale whole-exome data from diverse ethnic populations encompassing subjective sleep characteristics, to validate and extend the findings. Second, self-reported sleep characteristics may cause recall bias and not fully represent the objective sleep conditions. Subsequent research could perform whole-exome association analysis on device-measured sleep parameters[57]. Third, the UK Biobank population predominantly consists of middle-aged and relatively healthy individuals[58–60], which potentially inflates genetic effects for healthy sleep characteristics and attenuates those for sleep disturbances.

https://doi.org/10.1038/s41562-023-01785-5

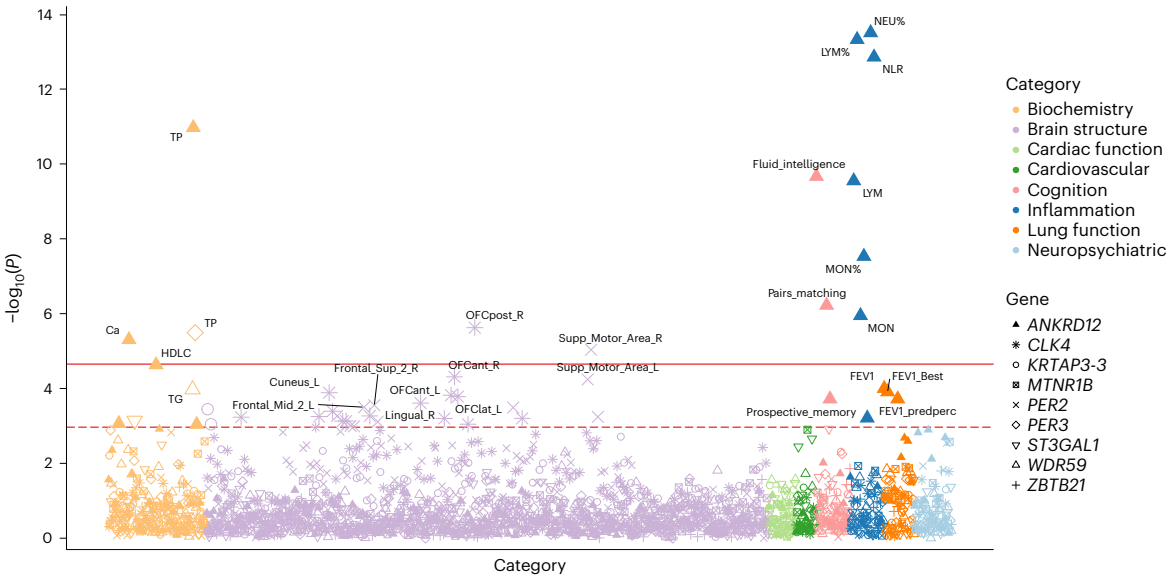

Fig. 6 | Phenome-wide association of sleep-related genes. Associations between genes identified in gene-based tests and a wide range of phenotypes, including eight categories (biochemistry, brain structure, cardiac function, cardiovascular diseases, cognition, inflammation, lung function and neuropsychiatric diseases). The *y* axis indicates the $-\log_{10}$ of the *P* value for each association. The *P* values shown are two-sided and unadjusted for multiple testing. The red solid line indicates the threshold for significant associations using Bonferroni correction ($P < 2.2 \times 10^{-5}$), and the red dashed line indicates the threshold using FDR correction ($P < 1.5 \times 10^{-3}$). The colour of each point represents the category, and the shape represents the gene. SKAT-O tests were used for gene-based analyses, and the models were adjusted for age, sex and the top ten ancestral principal components. Cuneus_L, left cuneus; FEV1, forced expiratory volume in 1 second; FEV1_Best, best measure of FEV1; FEV1_predperc, predicted percentage of FEV1TP, total protein; Frontal_Mid_2_L, left middle frontal gyrus; Frontal_Sup_2_R, right superior frontal gyrus (dorsolateral); HDLC, high-density lipoprotein cholesterol; Lingual_R, right lingual gyrus; LYM%, lymphocyte percentage; LYM, lymphocyte count; MON%, monocyte percentage; MON, monocyte count; NEU%, neutrophil percentage; NLR, neutrophil lymphocyte ratio; OFCant_L, left anterior orbital gyrus; OFCant_R, right anterior orbital gyrus; OFClat_L, left lateral orbital gyrus; OFCpost_R, right posterior orbital gyrus; Supp_Motor_Area_L, left supplementary motor area; Supp_Motor_Area_R, right supplementary motor area; TG, triglycerides.

In summary, our study identified 22 newly discovered genes significantly associated with sleep-related traits and demonstrated their significant associations with psychiatric, cognitive, metabolic and inflammatory traits. The results advance our understanding of the underlying biological mechanisms involved in sleep regulation and its profound impact on human health, offering promising avenues for precision medicine.

Methods

Study population and phenotypes

UK Biobank is a large prospective study with phenotypic and genetic data on approximately 500,000 participants, aged 38–72 years at recruitment[61]. The UK Biobank cohort was approved by the North West Multi-centre Research Ethics Committee (https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics), which provided oversight for this study. Written informed consent was obtained from all participants. The data used in the analysis contained demographic characteristics, sleep-related traits, biochemistry, brain structure, cardiovascular diseases, cognition, cardiac function, inflammation, neuropsychiatric diseases and lung function. Sleep-related traits used in the study included sleep duration, insomnia symptoms, chronotype (morning or evening person), ease of getting up in the morning, daytime sleepiness, daytime napping, snoring and sleep apnoea (Supplementary Table 1). This study was conducted under project application number 19542.

For sleep duration, we removed individuals whose duration was less than 3 h or greater than 18 h. For all sleep-related traits, we excluded sleep medication users and shift workers (Supplementary Methods). The demographic characteristics of the participants with sleep-related traits after exclusion and genetic quality control (Supplementary Methods) are provided in Supplementary Table 3. Details about other phenotypic data, which were used in the phenome-wide association analysis, are provided in Supplementary Table 2.

Exome sequencing and quality control

WES was conducted on 454,756 participants from UK Biobank[22]. The IDT xGen Exome Research Panel v.1.0 was used to capture the exomes, and the sequencing protocols have been described in detail elsewhere[21]. In addition to the quality control that was performed centrally, we applied extensive genotype-level, variant-level and sample-level quality control procedures, similar to the previous study[62], to ensure a high-quality dataset, for which the details are provided in the Supplementary Methods. Briefly, we split multi-allelic sites into bi-allelic sites in our analysis and removed the calls that had a low genotype quality and extremely low or high genotype depth. We next removed monomorphic variants in the final dataset and those that failed the filters (call rate, ≤90%; $P \leq 10^{-15}$ in the liberal Hardy–Weinberg equilibrium test). For the sample quality control, we removed the samples that had withdrawn their consent, were duplicates, had discordance between self-reported and genetically inferred sex, and had irrational call rates or additional metrics. We also excluded any individuals related at the third degree or closer on the basis of the KING-robust algorithm (kinship coefficient threshold, 0.0442; Supplementary Methods). Finally, we included 326,788 white British individuals in the main exome-wide association analysis.

Variant annotation

Variants were filtered to identify those that were rare (MAF < 1%), and they were annotated using SnpEff[63]. The most severe consequence for each gene transcript was retained. We then grouped variants annotated as stop gained, start lost, splice donor, splice acceptor, stop lost or frameshift into LoF variants. Likely deleterious missense variants were defined as those predicted consistently as deleteriousness in SIFT[64], PolyPhen2 HDIV, PolyPhen2 HVAR[65], LRT[66] and MutationTaster[67]. Variants were collapsed for each gene for the gene-based collapsing test. Each LoF and missense variant gained a score from the pLI[68,69] or REVEL[70] quantitative algorithm accordingly.

Effects of genetic variants on sleep-related traits

To explore whether mutations of different types or different severity within the same type can impact sleep traits to different degrees, we calculated the associations between the number of mutations in different groups and sleep traits. First, we selected genes with a pLI score greater than 0.9. The pLI score ranges from 0 to 1 for the most tolerant to the most intolerant genes, and a pLI score of >0.9 indicates that a gene is very likely to be intolerant to LoF[68,69]. We then grouped the missense mutations in these genes into four categories according to the REVEL score (<0.25, 0.25–0.5, 0.5–0.75, ≥0.75). The REVEL score for an individual missense variant ranges from 0 to 1, representing the likelihood that the variant is disease-causing[70]. Finally, we calculated the total number of mutations in LoF and four missense variant groups in these genes to reflect the burden of variants and employed linear and logistic models to test the associations between the burden and sleep traits.

Exome-wide association analysis

To identify genetic variation relevant to sleep-related traits, we performed single-variant association tests and gene-based collapsing tests, using SAIGE-GENE+ software[23]. SAIGE-GENE+ can test rare variants and ultra-rare variants (ultra-rare variants are collapsed to a pseudo marker) together, which reduces the data sparsity due to the effects of ultra-rare variants. In the single-variant association test, we utilized all variants with minor allele count ≥20. In the gene-based collapsing test, we applied two different maximum MAF cut-offs (1% and 0.1%) and two different variant annotation groups (LoF and LoF + missense) to perform SKAT-O tests. The relative coefficient cut-off of 0.05 for the sparse genetic relationship matrix (GRM) for the variance ratio estimation was used in the collapsing test. For quantitative traits, we used the inverse normalization. In both the single-variant test and the gene-based collapsing test, all models were adjusted for age, sex, BMI and the top ten ancestral principal components. $P < 1 \times 10^{-8}$ was considered significant in the single-variant association test, and $P < 2.5 \times 10^{-6}$ was considered significant in the gene-based collapsing test.

Sensitivity analysis for sex and ancestry

For all identified significant associations, we ran a number of sensitivity analyses. First, we performed the single-variant association and gene-based collapsing tests in different sexes to evaluate whether the results were impacted by sex. Second, we performed analyses on individuals of white (non-British), mixed, Asian and Black ancestry (UKB Field 21000) to evaluate whether significant associations were consistent across ethnicities.

LOVO analysis

To evaluate the robustness of significant associations in the gene-based collapsing test, we performed the LOVO analysis. Within each iteration, we performed the collapsing test upon removal of one variant. This procedure was repeated iteratively throughout the analysis and obtained multiple *P* values. When the *P* value obtained from the LOVO analysis differed significantly from the original *P* value, it was demonstrated that the deleted variant had a large impact on the association result—in other words, the variant was more important.

Conditional analysis adjusting for nearby common variation

To explore whether the identified rare signals were independent of nearby common variants, we reran the gene-based collapsing analysis of the significant genes, adjusting for the common variants in the region. First, we ran the common variant (MAF > 0.5%) association analysis in the genomic regions, which were 500 kb upstream and downstream of the significant genes. In the analysis, we used the software PLINK v.2.0 (https://www.cog-genomics.org/plink/2.0/) to perform genome-wide association analysis, and genotype data from the UK Biobank v.3 imputation (Supplementary Methods). We then clumped the results with cut-offs of $P < 1 \times 10^{-5}$ and $r^2 < 0.01$. Finally, we reran the gene-based collapsing analysis of the identified genes with the adjustment of the clumped common variants.

Replication of identified signals

To replicate our identified signals in the analysis, we used the FinnGen cohort to verify the associations. The FinnGen cohort is a large Finnish biobank, including 342,499 individuals with genotype data and digital health record data[24]. The details of the imputation and quality control for the genotype data were described previously[24]. We used genome-wide association analysis summary results, which were publicly available (https://r9.finngen.fi/), to find associations between the identified genes and sleep. GWAS summary statistics on 'sleep disorders (combined)' were used for the validation of seven self-reported sleep-related traits and 'sleep apnoea' (G6_SLEEPAPNO) for sleep apnoea. The neurological endpoint 'sleep disorders (combined)' in FinnGen included sleep disorders (F5_SLEEP), sleep apnoea (G6_SLEEPAPNO), narcolepsy and cataplexy (G6_NARCOCATA), and other sleep disorders (G6_SLEEPDISOTH). A detailed description of each endpoint can be accessed through https://risteys.finregistry.fi/endpoints/. The *P* value of another locus in the same gene was used for validation if a rare locus was not captured in the GWAS of FinnGen.

Burden heritability and burden genetic correlation

GWAS have identified hundreds of independent loci associated with sleep disturbance traits, and heritability explained by common variants across sleep traits has also been estimated. However, the contribution of rare coding variants to heritability is unclear. We used BHR[25] (https://github.com/ajaynadig/bhr) to estimate the heritability explained by the gene-wise burden of rare coding variants across sleep traits. As recommended in BHR, the variants were stratified into bins according to allele frequency and functional categories. Ultrarare was defined as MAF $< 1 \times 10^{-5}$, and rare was defined as $1 \times 10^{-5} \leq$ MAF $< 1 \times 10^{-3}$. Functions of the variants were defined in the variant annotation section. All BHR analyses were run with the baseline file provided by ref. 25. The variant-level summary statistics derived from SAIGE-GENE+ outputs were used, and effect sizes of quantitative traits were obtained directly from SAIGE-GENE+ outputs, while effect sizes of binary traits were calculated using allele frequency among cases and controls.

Univariate BHR analysis was performed to estimate the burden h^2 of each sleep trait, and we aggregated ultrarare and rare LoF burden heritability together as total burden heritability and compared it with the common-variant heritability of each trait, which was obtained using linkage disequilibrium score regression[27]. Bivariate BHR analysis was then used to compute the genetic correlation r_g. The burden genetic correlation between LoF and missense variants within each sleep trait was estimated. Burden genetic correlations across sleep traits using ultrarare and rare LoF variants were also obtained.

Functional enrichment analysis and tissue expression

We used the R package clusterProfiler[71,72] to perform the enrichment analysis of the 68 genes that were identified in the single-variant association tests and the gene-based collapsing tests. GO was selected as the gene set database to perform the enrichment analysis. GO terms were categorized as Biological Process, Cellular Component and Molecular Function. To gain more insight into how genes may influence sleep-related traits, we examined the enrichment of 68 genes in 54 different tissues from GTEx[28] using FUMA[73].

Single-cell expression

For the single-cell expression analysis, we first used data from a single-cell atlas of the mouse nervous system based on the RNA sequencing of half a million cells[29] to prioritize cell types as possibly involved in sleep-related traits. Cells from the mouse nervous system were clustered into 265 cell types. Similar to ref. 74, we downloaded the summary expression data, and we rescaled the expression of each

gene into *Z*-scores, which had a mean of 0 and a standard error of 1, to enable comparisons of genes between clusters. Analysis was restricted to 17,434 mouse genes that could be mapped to equivalent human protein-coding genes. For each cell type, we performed a one-sided Wilcoxon rank sum test comparing the expression *Z*-scores of the 68 sleep-associated genes against the values of all remaining genes. We report the *P* values of these 265 tests in Supplementary Table 23. We then subdivided these cell types into neurons and non-neurons and by region (central nervous system and peripheral nervous system) to show the enriched expressions of sleep-associated genes in neurons.

Second, we used brain scRNA-seq data from Garcia et al. in the Gene Expression Omnibus database with the accession ID GSE173731 (ref. 30). The scRNA-seq data were generated with brain temporal cortex tissues of individuals with refractory epilepsy, and all cell types within the brain were sequenced[30]. We used the R package Seurat to perform the main analysis and visualization[75]. The clustering and annotation of brain scRNA-seq data were conducted using the metadata files provided by Garcia et al.[30].

Phenome-wide association analysis

We performed PheWAS between the identified genes and various phenotypes (Supplementary Table 2). In our analysis of diseases, 11 neuropsychiatric diseases and 6 cardiovascular diseases were analysed. In our analysis of continuous phenotypes, 10 cognition tasks, 10 inflammatory traits, 29 blood biochemistry traits, 166 grey matter measures, 8 cardiac function measures and 9 lung function measures were analysed. PheWAS of the genes identified in the single-variant and gene-based collapsing tests were analysed. For genes in the single-variant test, we used linear or logistic models to find associations between significant variants and phenotypes. For genes in the collapsing test, we performed SKAT-O tests between significant genes and phenotypes. All models were adjusted for age, sex and the top ten ancestral principal components.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The main data, including the individual-level phenotypic and genetic data used in this study, were accessed from UK Biobank under application number 19542 and are available through UK Biobank (https://www.ukbiobank.ac.uk/). Summary GWAS statistics of FinnGen were obtained through https://r9.finngen.fi/. The single-cell sequencing data from the human brain were obtained from the Gene Expression Omnibus database (GSE173731). The single-cell sequencing data from the mouse central nervous system were obtained from the Mouse Brain Atlas (http://mousebrain.org/adolescent/).

Code availability

The code used for single-variant and gene-based analysis is an adaptation of the R package SAIGE-GENE+ v.1.1.6.2 (https://github.com/saigegit/SAIGE/). Quality control of individual-level data was performed using Hail v.0.2 (https://hail.is) and PLINK v.2.0 (https://www.cog-genomics.org/plink/2.0/). Variant annotation was performed using SnpEff v.5.1 (https://pcingola.github.io/SnpEff/). Burden heritability estimation was performed using BHR v.0.1.0 (https://github.com/ajaynadig/bhr/). The analysis and visualization of scRNA-seq data were performed using Seurat v.4.3.0 (https://github.com/satijalab/seurat/). GO enrichment analysis was performed using clusterProfiler v.4.2.2 (https://github.com/YuLab-SMU/clusterProfiler/). Tissue expression enrichment analysis was performed using FUMA v.1.5.6 (https://fuma.ctglab.nl/). Custom scripts for the analyses in this paper are available at https://github.com/cjfei18/sleep_wes.

References

1. Stranges, S., Tigbe, W., Gomez-Olive, F. X., Thorogood, M. & Kandala, N. B. Sleep problems: an emerging global epidemic? Findings from the INDEPTH WHO–SAGE study among more than 40,000 older adults from 8 countries across Africa and Asia. *Sleep* **35**, 1173–1181 (2012).
2. Leng, Y., Musiek, E. S., Hu, K., Cappuccio, F. P. & Yaffe, K. Association between circadian rhythms and neurodegenerative diseases. *Lancet Neurol.* **18**, 307–318 (2019).
3. Irwin, M. R. Sleep and inflammation: partners in sickness and in health. *Nat. Rev. Immunol.* **19**, 702–715 (2019).
4. Fan, M. et al. Sleep patterns, genetic susceptibility, and incident cardiovascular disease: a prospective study of 385 292 UK Biobank participants. *Eur. Heart J.* **41**, 1182–1189 (2020).
5. Cappuccio, F. P., D'Elia, L., Strazzullo, P. & Miller, M. A. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep* **33**, 585–592 (2010).
6. Lane, J. M. et al. Genetics of circadian rhythms and sleep in human health and disease. *Nat. Rev. Genet.* **24**, 4–20 (2023).
7. Ambrosius, U. et al. Heritability of sleep electroencephalogram. *Biol. Psychiatry* **64**, 344–348 (2008).
8. Gottlieb, D. J. et al. Novel loci associated with usual sleep duration: the CHARGE Consortium Genome-Wide Association Study. *Mol. Psychiatry* **20**, 1232–1239 (2015).
9. Lane, J. M. et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.* **49**, 274–281 (2017).
10. Doherty, A. et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat. Commun.* **9**, 5257 (2018).
11. Dashti, H. S. et al. Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat. Commun.* **10**, 1100 (2019).
12. Hammerschlag, A. R. et al. Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. *Nat. Genet.* **49**, 1584–1592 (2017).
13. Wang, H. et al. Genome-wide association analysis of self-reported daytime sleepiness identifies 42 loci that suggest biological subtypes. *Nat. Commun.* **10**, 3503 (2019).
14. Dashti, H. S. et al. Genetic determinants of daytime napping and effects on cardiometabolic health. *Nat. Commun.* **12**, 900 (2021).
15. Hu, Y. et al. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* **7**, 10448 (2016).
16. Lane, J. M. et al. Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nat. Commun.* **7**, 10889 (2016).
17. Jones, S. E. et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat. Commun.* **10**, 343 (2019).
18. Campos, A. I. et al. Insights into the aetiology of snoring from observational and genetic investigations in the UK Biobank. *Nat. Commun.* **11**, 817 (2020).
19. Jones, S. E. et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat. Commun.* **10**, 1585 (2019).
20. Jones, S. E. et al. Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS Genet.* **12**, e1006125 (2016).
21. Van Hout, C. V. et al. Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749–756 (2020).

22. Backman, J. D. et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature* **599**, 628–634 (2021).
23. Zhou, W. et al. SAIGE-GENE+ improves the efficiency and accuracy of set-based rare variant association tests. *Nat. Genet.* **54**, 1466–1469 (2022).
24. Kurki, M. I. et al. FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).
25. Weiner, D. J. et al. Polygenic architecture of rare coding variation across 394,783 exomes. *Nature* **614**, 492–499 (2023).
26. Strausz, S. et al. Genetic analysis of obstructive sleep apnoea discovers a strong association with cardiometabolic health. *Eur. Respir. J.* **57**, 2003091 (2021).
27. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
28. GTEx Consortium. The Genotype–Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
29. Zeisel, A. et al. Molecular architecture of the mouse nervous system. *Cell* **174**, 999–1014.e22 (2018).
30. Garcia, F. J. et al. Single-cell dissection of the human brain vasculature. *Nature* **603**, 893–899 (2022).
31. van Dijk, E. L., Jaszczyszyn, Y., Naquin, D. & Thermes, C. The third revolution in sequencing technology. *Trends Genet.* **34**, 666–681 (2018).
32. Favara, D. M. et al. ADGRL4/ELTD1 silencing in endothelial cells induces ACLY and SLC25A1 and alters the cellular metabolic profile. *Metabolites* **9**, 287 (2019).
33. Seugnet, L. et al. Notch signaling modulates sleep homeostasis and learning after sleep deprivation in *Drosophila*. *Curr. Biol.* **21**, 835–840 (2011).
34. Lampe, A. K. et al. Automated genomic sequence analysis of the three collagen VI genes: applications to Ullrich congenital muscular dystrophy and Bethlem myopathy. *J. Med. Genet.* **42**, 108–120 (2005).
35. Nadeau, A. et al. Natural history of Ullrich congenital muscular dystrophy. *Neurology* **73**, 25–31 (2009).
36. Cescon, M., Chen, P., Castagnaro, S., Gregorio, I. & Bonaldo, P. Lack of collagen VI promotes neurodegeneration by impairing autophagy and inducing apoptosis during aging. *Aging (Albany NY)* **8**, 1083–1101 (2016).
37. Gregorio, I. et al. Collagen VI deficiency causes behavioral abnormalities and cortical dopaminergic dysfunction. *Dis. Models Mech.* **15**, dmm049481 (2022).
38. Majcin Dorcikova, M., Duret, L. C., Pottié, E. & Nagoshi, E. Circadian clock disruption promotes the degeneration of dopaminergic neurons in male *Drosophila*. *Nat. Commun.* **14**, 5908 (2023).
39. Yang, Y. et al. Endothelium-specific CYP2J2 overexpression attenuates age-related insulin resistance. *Aging Cell* **17**, e12718 (2018).
40. Petrovan, R. J., Yuan, Y. & Curtiss, L. K. Expression of the Lyst(beige) mutation is atheroprotective in chow-fed apolipoprotein E-deficient mice. *J. Lipid Res.* **49**, 429–437 (2008).
41. Leslie, E. J. et al. Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. *Birth Defects Res. A* **94**, 934–942 (2012).
42. Schneider, J. S. Altered expression of genes involved in ganglioside biosynthesis in substantia nigra neurons in Parkinson's disease. *PLoS ONE* **13**, e0199189 (2018).
43. Cutillo, G., Saariaho, A.-H. & Meri, S. Physiology of gangliosides and the role of antiganglioside antibodies in human diseases. *Cell. Mol. Immunol.* **17**, 313–322 (2020).
44. Zhang, P. et al. Family-based association analysis to finemap bipolar linkage peak on chromosome 8q24 using 2,500 genotyped SNPs and 15,000 imputed SNPs. *Bipolar Disord.* **12**, 786–792 (2010).
45. Xu, C. et al. BCL9 and C9orf5 are associated with negative symptoms in schizophrenia: meta-analysis of two genome-wide association studies. *PLoS ONE* **8**, e51674 (2013).
46. Smirnova, L. et al. The difference in serum proteomes in schizophrenia and bipolar disorder. *BMC Genomics* **20**, 535 (2019).
47. Valenstein, M. L. et al. Structure of the nutrient-sensing hub GATOR2. *Nature* **607**, 610–616 (2022).
48. He, A. et al. Acetyl-CoA derived from hepatic peroxisomal β-oxidation inhibits autophagy and promotes steatosis via mTORC1 activation. *Mol. Cell* **79**, 30–42.e4 (2020).
49. Patke, A., Young, M. W. & Axelrod, S. Molecular mechanisms and physiological importance of circadian rhythms. *Nat. Rev. Mol. Cell Biol.* **21**, 67–84 (2020).
50. Ding, F. et al. Changes in the composition of brain interstitial ions control the sleep–wake cycle. *Science* **352**, 550–555 (2016).
51. Nielsen, F. H., Johnson, L. K. & Zeng, H. Magnesium supplementation improves indicators of low magnesium status and inflammatory stress in adults older than 51 years with poor quality sleep. *Magnes. Res.* **23**, 158–168 (2010).
52. Cullell, N. et al. Sleep/wake cycle alterations as a cause of neurodegenerative diseases: a Mendelian randomization study. *Neurobiol. Aging* **106**, 320.e1–320.e12 (2021).
53. Cox, S. R. et al. Sleep and cognitive aging in the eighth decade of life. *Sleep* **42**, zsz019 (2019).
54. Spano, G. M. et al. Sleep deprivation by exposure to novel objects increases synapse density and axon–spine interface in the hippocampal CA1 region of adolescent mice. *J. Neurosci.* **39**, 6613–6625 (2019).
55. Sato, T. & Sato, S. Circadian regulation of metabolism: commitment to health and diseases. *Endocrinology* **164**, bqad086 (2023).
56. Tahara, Y., Otsuka, M., Fuse, Y., Hirao, A. & Shibata, S. Refeeding after fasting elicits insulin-dependent regulation of Per2 and Rev-erbα with shifts in the liver clock. *J. Biol. Rhythms* **26**, 230–240 (2011).
57. Rowlands, A. V. et al. Accelerometer-assessed physical activity in epidemiology: are monitors equivalent? *Med. Sci. Sports Exerc.* **50**, 257–265 (2018).
58. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
59. Batty, G. D., Gale, C. R., Kivimäki, M., Deary, I. J. & Bell, S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *Br. Med. J.* **368**, m131 (2020).
60. Schoeler, T. et al. Participation bias in the UK Biobank distorts genetic associations and downstream analyses. *Nat. Hum. Behav.* **7**, 1216–1227 (2023).
61. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
62. Jurgens, S. J. et al. Analysis of rare genetic variation underlying cardiometabolic diseases and traits among 200,000 individuals in the UK Biobank. *Nat. Genet.* **54**, 240–250 (2022).
63. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* **6**, 80–92 (2012).
64. Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M. & Ng, P. C. SIFT missense predictions for genomes. *Nat. Protoc.* **11**, 1–9 (2016).
65. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **76**, 7.20.1–7.20.41 (2013).
66. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).

67. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat. Methods* **7**, 575–576 (2010).
68. Kosmicki, J. A. et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49**, 504–510 (2017).
69. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
70. Ioannidis, N. M. et al. REVEL: an ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.* **99**, 877–885 (2016).
71. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innov. (Camb.)* **2**, 100141 (2021).
72. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**, 284–287 (2012).
73. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
74. Singh, T. et al. Rare coding variants in ten genes confer substantial risk for schizophrenia. *Nature* **604**, 509–516 (2022).
75. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

Acknowledgements

We thank all UK Biobank participants for their time and UK Biobank team members for collating the data. We thank the participants and investigators of the FinnGen study. Z.-L.H. was supported by grants from the Science and Technology Innovation 2030 Major Projects (no. 2021ZD0203400). J.-T.Y. was supported by grants from the Science and Technology Innovation 2030 Major Projects (no. 2022ZD0211600), the National Natural Science Foundation of China (nos 82071201, 81971032 and 92249305), the Shanghai Municipal Science and Technology Major Project (no. 2018SHZDZX01), the Research Start-Up Fund of Huashan Hospital (no. 2022QD002), the Excellence 2025 Talent Cultivation Program at Fudan University (no. 3030277001), Shanghai Talent Development Funding for the Project (no. 2019074), and the Zhangjiang Lab, Tianqiao and Chrissy Chen Institute, and the State Key Laboratory of Neurobiology and Frontiers Center for Brain Science of Ministry of Education, Fudan University. W.C. was supported by grants from the National Natural Science Foundation of China (no. 82071997) and the Shanghai Rising-Star Program (no. 21QA1408700). J.-F.F. was supported by the National Key R&D Program of China (nos 2018YFC1312904 and 2019YFA0709502), the Shanghai Municipal Science and Technology Major Project (no. 2018SHZDZX01), the 111 Project (no. B18015), the Shanghai Center for Brain Science and Brain-Inspired Technology and the Zhangjiang Lab. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

W.C. and J.-T.Y. designed the study. C.-J.F., Z.-Y.L., J.N. and L.Y. conducted the main analysis. C.-J.F., Z.-Y.L., J.N., L.Y., J.Y. and S.-D.C. drafted the manuscript. B.-S.W., J.-J.K., W.-S.L. and X.-Y.H. contributed to the data collection and analysis. W.C., J.-T.Y., J.-F.F., Z.-L.H. and H.Y. critically revised the manuscript. All authors reviewed and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01785-5.

Correspondence and requests for materials should be addressed to Jin-Tai Yu or Wei Cheng.

Peer review information *Nature Human Behaviour* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Wei Cheng, Jin-Tai Yu

Last updated by author(s): Nov 10, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was involved in data collection (data used is all directly available from UKB and FinnGen as described in detail in the paper).
Data analysis	The code used for single-variant and gene-based analysis is an adaptation of the R package SAIGE-GENE+ v.1.1.6.2 (https://github.com/saigegit/SAIGE/). Quality control of individual-level data was performed using Hail v.0.2 (https://hail.is) and PLINK v.2.0 (https://www.cog-genomics.org/plink/2.0/). Variant annotation was performed using SnpEff v.5.1 (https://pcingola.github.io/SnpEff/). Burden heritability estimation was performed using BHR v.0.1.0 (https://github.com/ajaynadig/bhr/). Analysis and visualization of single-cell RNA sequencing data was performed using Seurat v.4.3.0 (https://github.com/satijalab/seurat/). Gene Ontology enrichment analysis was performed using clusterProfiler v.4.2.2 (https://github.com/YuLab-SMU/clusterProfiler/). Tissue expression enrichment analysis was performed using FUMA v.1.5.6 (https://fuma.ctglab.nl/). Custom scripts for the analyses in this paper are available at https://github.com/cjfei18/sleep_wes. Details of specific software and references can be found within text in the relevant Methods and Supplementary Information sections.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The individual-level phenotypic and genetic data used in the present study are available from UKB with restrictions applied. Data were used under the application number 19542 and are thus not publicly available. Access to the UKB data can be requested through a standard protocol (https://www.ukbiobank.ac.uk/enable-your-research). Summary GWAS statistics of FinnGen were obtained through https://r9.finngen.fi/. The single-cell sequencing data of human brain were obtained from GEO database (GSE173731). The single-cell sequencing data of mouse central nervous system were obtained from Mouse Brain Atlas (http://mousebrain.org/adolescent/).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	We took sex into consideration in our study and our findings could apply to both males and females. Sex (UKB Field ID 31) in the UK Biobank was acquired from central registry at recruitment, and all included participants gave written informed consent for sharing of individual-level data.
Reporting on race, ethnicity, or other socially relevant groupings	Ethnic background (UKB Field ID 21000) was used to include the white British participants in the main analysis. Participants with White (non-British), Mixed, Asian, and Black ancestry were included in the sensitivity analysis.
Population characteristics	In the main analysis, about 294,000 white-British individuals with sleep-related traits were used to perform association tests, with ages from 38 to 72 at enrollment, of which 54.7% were female. The baseline demographic data of participants was shown in Supplementary Table 3.
Recruitment	The UK Biobank enrolled the participants aged 38-72 years between 2006 and 2010 for baseline assessments in 22 centers across the UK. The assessment visits comprised interviews and questionnaires covering lifestyles and health conditions, physical measures, biological samples, imaging, and genotyping. The database is linked to national health datasets, including primary care, hospital inpatient, death, and cancer registration data.
Ethics oversight	The UK Biobank has approval from the North West Multi-centre Research Ethics Committee (https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics) as a Research Tissue Bank approval and provides oversight for this study. Written informed consent was obtained from all participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample sizes, and eligible participants with both genetic and phenotypic data were included as much as possible.
Data exclusions	Participants without whole-exome sequencing data and sleep traits data, and those who failed to pass quality control were excluded. Sleep medication users and shift workers were also excluded. Details of data quality control were provided in Methods.
Replication	The GWAS summary statistics data from the FinnGen study were used for external replication. GWAS summary statistics of "sleep disorders (combined)" were used for validation of seven self-reported sleep-related traits and "sleep apnoea" for sleep apnoea. Among 22 genes that have not been reported in previous studies, 20 of them are associated with sleep disorders in FinnGen (see Tables 1 and 2). Rs985277 and rs1879581 were two exceptions that did not show a significant association with sleep disorders in FinnGen, possibly due to differences between the phenotype "sleep disorders (combined)" and sleep-related traits or due to heterogeneity across different populations.
Randomization	Covariates including age, sex, BMI, and the first 10 genetic principal components (calculated with whole-exome sequencing data) were adjusted in the study.

Blinding | Blinding was not applicable to this study as this study is observational.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Plants

Seed stocks: *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes: *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication: *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

Magnetic resonance imaging

Experimental design

Design type: T1-weighted structural MRI data was obtained from UK Biobank (application 19542).

Design specifications: UK Biobank designed the imaging acquisition protocols including 6 modalities, covering structural, diffusion and functional imaging. In the current study, T1-weighted structural image was used and the image was acquired using straight sagittal orientation for 5 minutes.

Behavioral performance measures: N/A

Acquisition

Imaging type(s): T1-weighted structural imaging

Field strength: 3T

Sequence & imaging parameters: The EPI-based acquisitions utilize simultaneous multi-slice (multiband) acceleration. UK Biobank uses pulse sequences and reconstruction code from the Center for Magnetic Resonance Research (CMRR), University of Minnesota https://www.cmrr.umn.edu/multiband. The resolution is 1x1x1 mm and field of view is 208x256x256 matrix. Straight sagittal orientation is used. TR and TE are 2000ms and 2.01ms respectively. The flip angle is 8 deg. Detailed sequence and imaging parameters are openly available here: https://biobank.ndph.ox.ac.uk/showcase/showcase/docs/brain_mri.pdf.

Area of acquisition: Whole brain

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software: We used the pipeline with the Statistical Parametric Mapping software version 12 (http://www.fil.ion.ucl.ac.uk/spm) using the CAT12 toolbox (http://dbm.neuro.uni-jena.de/cat) with default settings, to preprocess the structural MRI data, which

	contained the usage of high-dimensional spatial normalization with an already integrated Dartel template in Montreal Neurological Institute (MNI) space. All images were subjected to nonlinear modulations and corrected for each individual head size. Images were then smoothed with an 8 mm full-width at half-maximum Gaussian kernel with the resulting voxel size of 1.5 mm3. The automated anatomical labeling 3 (AAL3) atlas, which partitioned the brain into 166 regions of interest, was employed to obtain the region-wise gray matter volume.
Normalization	See above
Normalization template	See above
Noise and artifact removal	See above
Volume censoring	See above

Statistical modeling & inference

Model type and settings	Mass univariate
Effect(s) tested	Associations of sleep-related genes and gray matter measures were tested using SAIGE-GENE+.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☒ Both

Anatomical location(s)	All imaging derived phenotypes were extracted based on the automated anatomical labeling 3 (AAL3) atlas.
Statistic type for inference (See Eklund et al. 2016)	N/A (see methods on phenome-wide association analysis)
Correction	We applied Bonferroni correction and false discovery rate (FDR) correction for testing of multiple genes and traits.

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis