

# UCLA

## UCLA Previously Published Works

### Title

Genome-wide association analysis of composite sleep health scores in 413,904 individuals

### Permalink

<https://escholarship.org/uc/item/3b7621dj>

### Journal

Communications Biology, 8(1)

### ISSN

2399-3642

### Authors

Goodman, Matthew O

Faquih, Tariq

Paz, Valentina

et al.

### Publication Date

2025

### DOI

10.1038/s42003-025-07514-0

Peer reviewed

<https://doi.org/10.1038/s42003-025-07514-0>

# Genome-wide association analysis of composite sleep health scores in 413,904 individuals

Check for updates

Matthew O. Goodman<sup>1,2,3</sup>, Tariq Faquih<sup>1,2,3</sup>, Valentina Paz<sup>4,5,6</sup>, Pavithra Nagarajan<sup>1</sup>, Jacqueline M. Lane<sup>1,2,3,6</sup>, Brian Spitzer<sup>7</sup>, Matthew Maher<sup>6</sup>, Joon Chung<sup>8</sup>, Brian E. Cade<sup>1,2,3</sup>, Shaun M. Purcell<sup>1,2,3,9</sup>, Xiaofeng Zhu<sup>10</sup>, Raymond Noordam<sup>11</sup>, Andrew J. K. Phillips<sup>12</sup>, Simon D. Kyle<sup>13</sup>, Kai Spiegelhalter<sup>14</sup>, Michael N. Weedon<sup>15</sup>, Deborah A. Lawlor<sup>16,17</sup>, Jerome I. Rotter<sup>18</sup>, Kent D. Taylor<sup>18</sup>, Carmen R. Isasi<sup>19</sup>, Tamar Sofer<sup>1,2,7,20</sup>, Hassan S. Dashti<sup>3,6,21</sup>, Martin K. Rutter<sup>22,23</sup>, Susan Redline<sup>1,2</sup>, Richa Saxena<sup>3,6,21</sup> & Heming Wang<sup>1,2</sup> ✉

Recent genome-wide association studies (GWASs) of several individual sleep traits have identified hundreds of genetic loci, suggesting diverse mechanisms. Moreover, sleep traits are moderately correlated, so together may provide a more complete picture of sleep health, while illuminating distinct domains. Here we construct novel sleep health scores (SHSs) incorporating five core self-report measures: sleep duration, insomnia symptoms, chronotype, snoring, and daytime sleepiness, using additive (SHS-ADD) and five principal components-based (SHS-PCs) approaches. GWASs of these six SHSs identify 28 significant novel loci adjusting for multiple testing on six traits ( $p < 8.3e-9$ ), along with 341 previously reported loci ( $p < 5e-08$ ). The heritability of the first three SHS-PCs equals or exceeds that of SHS-ADD (SNP- $h^2 = 0.094$ ), while revealing sleep-domain-specific genetic discoveries. Significant loci enrich in multiple brain tissues and in metabolic and neuronal pathways. Post-GWAS analyses uncover novel genetic mechanisms underlying sleep health and reveal connections (including potential causal links) to behavioral, psychological, and cardiometabolic traits.

Sleep is an essential biological process, orchestrated by interrelated neurologic and physiologic regulatory processes, responding to individual, social, and environmental influences<sup>1-3</sup>. Positive sleep traits have been associated with lower rates of cardiometabolic and neuropsychiatric diseases, as well as higher productivity and well-being<sup>4</sup>. Moreover, general sleep health has come into recent focus as a consequential and modifiable health factor, with the combined presence of multiple healthy sleep factors frequently being a stronger predictor of positive health outcomes<sup>5,6</sup>. As a composite, sleep health is recognized to involve multiple domains, including regularity, satisfaction, alertness, timing, efficiency, and duration<sup>1</sup>.

Several recent studies leveraging biobank-scale data have resulted in well-powered genome-wide association studies (GWASs) of sleep phenotypes, capturing several aspects of sleep health, including self-reported sleep duration<sup>7</sup>, insomnia<sup>8,9</sup>, sleepiness<sup>10</sup>, snoring<sup>11</sup>, and chronotype<sup>12</sup>. These GWASs have begun to elucidate the genetic architecture of sleep, while revealing the presence of widespread genetic correlations across both sleep and related neuropsychiatric and cardiometabolic traits<sup>2,3</sup>. Associated genomic loci and pathways are often shared across multiple sleep traits,

suggesting a shared genetic basis and co-regulated processes. Therefore, a more complete and robust understanding of sleep may be achieved by describing patterns across multiple traits, pointing toward underlying domains, highlighting the potential utility in analyzing composite sleep health scores (SHSs).

Recently, an additive sleep health score (SHS-ADD) consisting of five self-reported sleep behaviors was studied in unrelated individuals in the UK Biobank (UKB), yielding new genetic findings<sup>13</sup>. However, additive scores compress data across multiple domains to a single metric, resulting in potential information loss, increased genetic heterogeneity, and weaker signal for genetic analyses.

Prior data-driven phenotyping of complex traits and correlated disease conditions have often relied on principal components analysis (PCA). The PCA algorithm is well-recognized for its ability to sensibly restructure multi-dimensional phenotypes, by constructing statistically independent linear composites, or principal components (PCs), sequentially prioritized in terms of variance explained. The construction of PCs is informed by underlying relationships across traits, with correlated traits tending to be combined,

A full list of affiliations appears at the end of the paper. ✉ e-mail: [hwang@bwh.harvard.edu](mailto:hwang@bwh.harvard.edu)

given the independence constraint, since a composite of correlated traits will explain more of the remaining total variance. In the broader field of genomics, PCA has been effectively applied as a dimension reduction technique to derive composite phenotypes and to refine and improve investigations of the genetic architecture of complex traits, with applications to both subject-level phenotypic data and GWAS summary statistics. Multivariate analyses with PCA have been used for various target phenotypes in genetic analyses, including recently in cardiac imaging<sup>14</sup>, anthropometric traits reflecting body shape<sup>15</sup>, metabolic profiling<sup>16</sup> and lipid profiling<sup>17</sup>.

In this study, we expanded the prior UKB SHS-ADD GWAS to a larger UKB sample and constructed five novel PC-derived SHSs as linear combinations of the same five underlying traits. We conjectured that, when compared with SHS-ADD, SHS-PCs would create more precisely targeted phenotypes, resulting in potentially greater heritability, as well as distinct and interpretable domain-specific associations in secondary analyses.

## Results

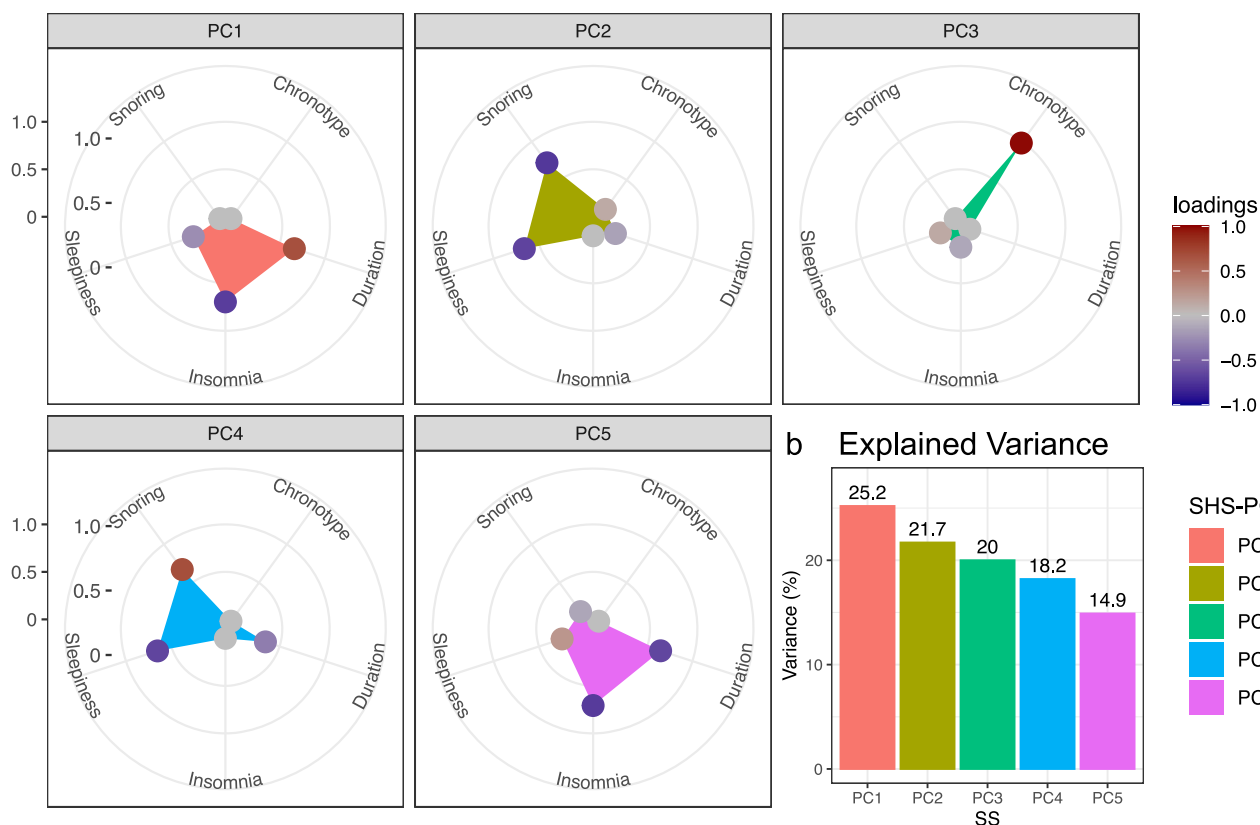
### Sleep health score construction in UKB

The study population consisted of 413,904 UKB participants of European ancestry with complete sleep and genomics data (“Methods”). Sample characteristics are provided in Supplementary Data 1. Self-reported sleep traits (sleep duration, insomnia, chronotype, snoring, and daytime sleepiness) derived from the baseline questionnaire were used to construct SHS traits (“Methods”). Briefly, SHS-ADD was operationalized as in previous UKB studies<sup>5,13</sup>, defined as the sum of five dichotomized positive sleep health characteristics: sleep duration of 7–8 h, morning chronotype preference, no snoring, infrequent insomnia symptoms and infrequent daytime sleepiness

(“Methods”). SHS-PCs were extracted from the original integer-scale same underlying sleep traits, treated as linear continuous measures (after being mean-centered and variance-standardized), and then oriented to positively correlate with self-assessed overall health.

Several findings emerged in constructing the Sleep Health Score Principal Components (SHS-PCs) in the UK Biobank, providing context and guiding interpretation. SHS-PCs 1–5 individually explained from 25.2% to 14.9% of the phenotypic variation (Fig. 1 and Supplementary Data 2). Based on their PC loadings (Fig. 1 and Supplementary Data 2), higher scores on SHS-PCs are interpreted as follows – SHS-PC1: longer sleep with less-frequent insomnia symptoms and sleepiness; SHS-PC2: healthier sleep with less-frequent sleepiness and without snoring (i.e., without symptoms of sleep apnea syndrome); SHS-PC3: morningness chronotype; SHS-PC4: snoring with less frequent sleepiness; SHS-PC5: shorter sleep duration with less-frequent insomnia symptoms. Substantial non-normality was observed for SHS-PC2, SHS-PC3, and SHS-PC4 (Supplementary Fig. 1), resulting from the underlying trait distributions (“Methods”). The direction and loadings of the PCs follow from underlying covariance and corresponding Pearson correlations among the self-report sleep traits (Supplementary Data 3): SHS-PC1 was driven primarily by correlations between sleep duration and insomnia ( $r = -0.24$ ) and between insomnia and sleepiness ( $r = 0.09$ ); SHS-PC2 was driven by the correlation between sleepiness and snoring ( $r = 0.08$ ); SHS-PC3 loaded on chronotype, which was independent of the other underlying traits; whereas SHS-PC4 and SHS-PC5 appear to be driven largely by the PC independence constraint, such that they loaded on both positive and negative sleep attributes, and does not imply these combinations constitute clusters in the data.

### a PC loadings



**Fig. 1 | Loadings and variance explained for the five principal component (PC)-based composite sleep health scores (SHS).** **a** SHS-PC loadings on self-reported sleep phenotypes. Interpretation of higher SHS-PC scores, based on loadings, are: PC1—longer sleep with less-frequent insomnia and lower sleepiness; PC2—absence of snoring and lower sleepiness; PC3—morningness chronotype; PC4—presence of

snoring and less frequent sleepiness; PC5—shorter sleep and less-frequent insomnia. Radar plots display loading magnitudes (radial distance). Red dots: positive loadings; Blue dots: negative loadings. **b** Percent of the phenotypic variance explained by each sleep health score. SHS sleep health score, PC principal component.

The interpretation of the SHSs was further clarified by their Spearman rank correlations ( $r_s$ ) with objective traits not used in their construction (Supplementary Figs. 2a and 3a). For example, SHS-PC1 was the SHS trait most positively correlated with accelerometry-based sleep duration ( $r_s = 0.09$ ) and sleep efficiency ( $r_s = 0.08$ ) metrics. SHS-PC2 was positively correlated with higher sleep efficiency ( $r_s = 0.06$ ), lower daytime inactivity ( $|r_s| = 0.10$ ), lower BMI ( $|r_s| = 0.19$ ), and being female ( $|r_s| = 0.17$ ), features which suggest the absence of sleep apnea. SHS-PC3 correlated with earlier accelerometry-based measures of maximum activity timing ( $|r_s| = 0.29$ ) and sleep midpoint ( $|r_s| = 0.27$ ), (measurements related to circadian timing). Notably, SHS-PC4 showed no significant correlations with self-assessed overall health ( $p > 0.05$ ) but was nonetheless correlated with other measures of overall health, including lower levels of self-reported disability ( $|r_s| = 0.06$ ) and lower numbers of treatments/medications taken ( $|r_s| = 0.05$ ). Notably, SHS-ADD had the strongest association with self-reported overall health ( $|r_s| = 0.21$ ), as well as with the accelerometry-derived sleep regularity index ( $|r_s| = 0.10$ ).

### Genome-wide association analysis

We performed GWAS for the six SHS traits using linear mixed regression models, adjusting for age, sex, genotyping array, ten genetic PCs and genetic relatedness matrix (“Methods”). We identified 31,188 genome-wide significant (GWS) SNPs ( $p < 5e-8$ ), resulting in 45 loci for SHS-ADD (SNP- $h^2 = 0.094$ ), and 91, 48, 166, 26, and 24 loci for SHS-PC1-5 (SNP- $h^2 = 0.117$ , 0.093, 0.153, 0.070, and 0.068), respectively (Fig. 2a, Table 1, Supplementary Figs. 4 and 5, and Supplementary Data 4 and 5; “Methods”). Function annotation of all SNPs in linkage disequilibrium (LD;  $r^2 \geq 0.6$ ) with the lead SNPs in the risk loci was performed using FUMA<sup>18</sup> (“Methods”; Fig. 2e, h).

To determine novelty, we compared the 400 unique GWS loci against those found to be GWS in prior biobank GWASs of individual sleep traits and the previously developed SHS-ADD in the UKB (“Methods”; Fig. 2c, d; Supplementary Data 5). This identified 59 unreported SHS-associated GWS loci with a lead variant at least 500 kb away from previously reported sleep variants (Supplementary Data 4). Of these unreported loci, 28 passed a stricter significance threshold ( $p < 8.3e-9$ ) accounting for multiple testing on six traits, which are defined as novel loci (Table 1). Locus zoom plots of the 28 novel loci are shown in Supplementary Fig. 6. Among the 28 novel loci, two were independent ( $r^2 < 0.1$ ) but within 50 kb of one another: rs201449027 (associated with SHS-PC1) and rs11494758 (with SHS-PC3), both in the *KLRG1* locus (a gene with reported immune system function<sup>19</sup>). The other 26 loci were distinctly associated with one SHS trait. In addition, there were two functional variants among the lead SNPs: First, SHS-PC1 associated with rs8074498, a missense variant in *ASPSERI*, a gene regulating *GLUT4* in glucose sequestration and transportation in response to insulin. Second, SHS-PC5 associated with rs138572890 in the 3' UTR of *PAPD5*, a non-canonical poly(A) polymerase involved in the surveillance and degradation of aberrant RNAs, including the glucose transporter *GLUT1*. Functional annotation for the lead SNPs at the 28 novel and 31 additional unreported loci is provided in Supplementary Data 6 and 7.

Among all 400 GWS loci, there were 62 loci colocalizing SHS traits (Fig. 2b), including 18 loci colocalizing multiple SHS-PCs, suggesting the presence of instances of pleiotropy across SHS-PC traits (“Methods”; Fig. 2b and Supplementary Data 8). Several variants colocalized across multiple SHS and were previously reported in multiple sleep GWAS, for example: Rs113851554 at *MEIS1* (colocalizing SHS-ADD, SHS-PC1, SHS-PC3, and SHS-PC5) was reported in GWASs of insomnia, chronotype, and restless leg syndrome. Rs2863957 at *PAX8* (colocalizing SHS-PC1, SHS-PC2, and SHS-PC5) was reported for sleep duration and insomnia. Rs1421085 at *FTO* (colocalizing SHS-PC3 and SHS-PC4), a widely recognized obesity gene, was reported for sleep duration, chronotype, snoring, and OSA. In addition, SHS-PC4 and SHS-PC5 colocalized at rs576981040 in a novel locus containing *TNFRSF14*, a gene involved in T-cell activation and signaling. The infrequent colocalization of loci associated with SHS-PC traits is consistent with their being largely genetic and phenotypically statistically independent (Supplementary Figs. 2b and 3b).

### Genetic overlap with individual sleep traits

Genetic correlations, between SHS traits with individual sleep and accelerometry traits, were qualitatively similar to the analogous phenotypic correlations described above, while often stronger (Supplementary Fig. 2). GWS SHS loci and their corresponding genetic risk scores (GRS; “Methods”) were associated with underlying self-reported and accelerometry derived sleep traits, largely in keeping with expectation, based on the construction of the SHS traits (Supplementary Data 9 and 10, Supplementary Note 1). Conversely, approximately 50% of 1039 GWS loci previously reported for SHS-ADD or individual sleep traits were associated with one or more SHS traits ( $p < 5e-8$ ; Supplementary Data 11 and 12).

### Sensitivity analysis

We performed 22 distinct sensitivity analyses (“Methods”), for each of the 400 GWS loci across the six SHS traits in unrelated individuals ( $n = 308,902$ ), adjusting for various factors or restricting to one of three subsets (males-only,  $n = 145,186$ ; females-only,  $n = 163,716$ ; healthy-only,  $n = 115,297$ ). Specific covariate adjustments led to modest average attenuation (<15%) in SHS genetic effects across the GWS loci (Supplementary Data 13 and Supplementary Fig. 7). For example, adjustment of adiposity measures in SHS-PC2 (9.3%) and SHS-ADD (7.9%), mood variables in SHS-ADD (14.1%) and SHS-PC1 (9.6%), and in a healthy subset without chronic diseases in SHS- SHS-PC1 (13.5%). Nearly all individual loci remained nominally significant ( $p < 0.05$ ) with similar effect size (standard error [SE] change < 2) after additional adjustment.

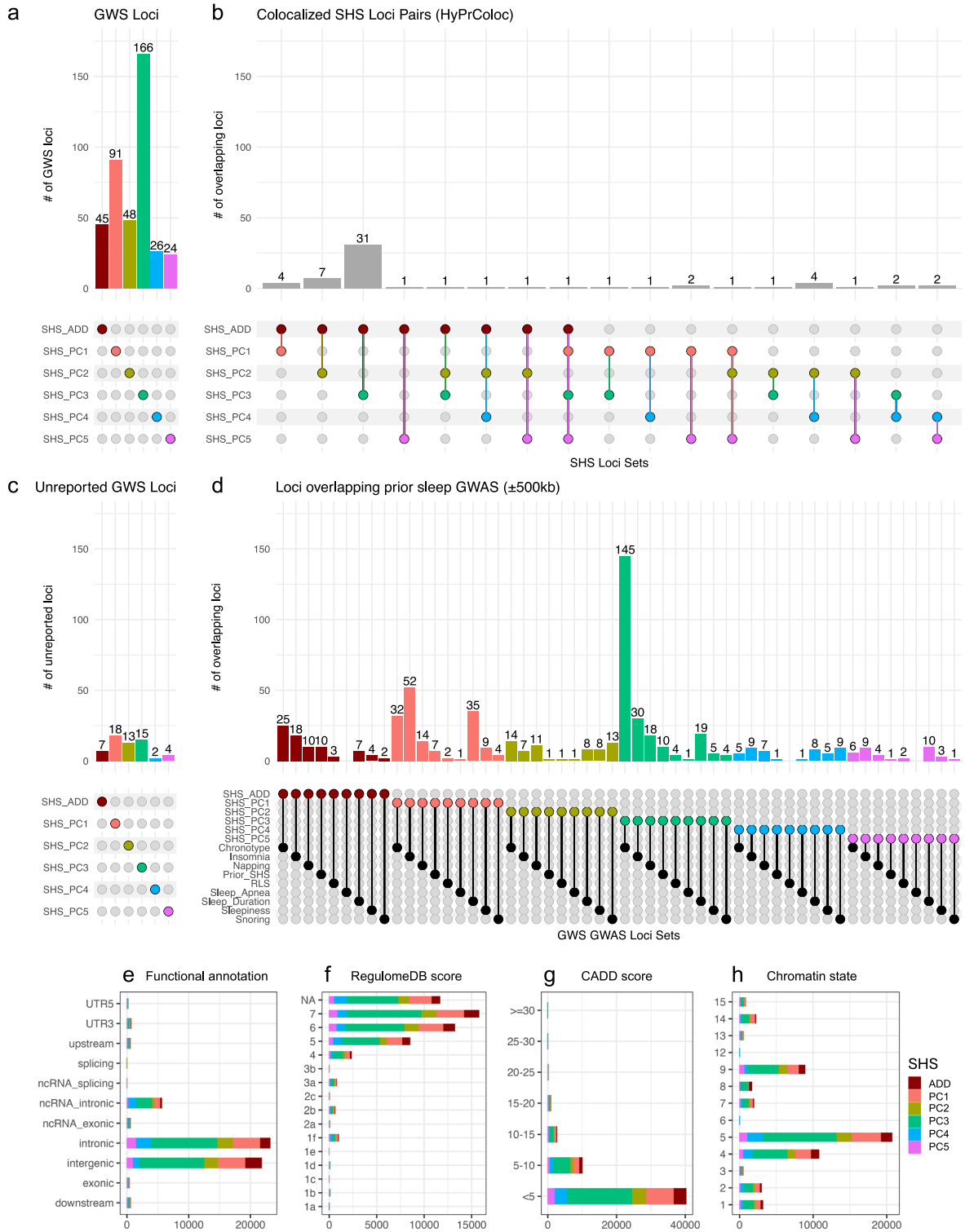
### Replication and validation analyses

We tested the replication of the 400 GWS loci in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL;  $n = 11,144$ ) using the same SHS PC loadings in UKB. Notably, HCHS/SOL differs in sample size and continental ancestry, compared with UKB, as well as in the distribution of key variables, such as age and sleep duration (“Methods”; Supplementary Data 14). Chronotype was only collected in a subset of the HCHS/SOL cohort and was imputed to the entire sample (“Methods”). Of the 400 loci tested, 12 demonstrated nominal significance ( $P < 0.05$ ) and consistent directional effects compared to UKB (Supplementary Data 15). However, we were able to validate the polygenic risk score (PRS) constructed using genome-wide summary statistics for each SHS in HCHS/SOL ( $P < 0.008$ ), except for SHS-PC3, which showed nominal significance ( $P < 0.05$ ) (“Methods”; Table 2). We further examined the PRS associations with sleep phenocodes in the MGB biobank (“Methods”; Table 2). The PRSs of both SHS-ADD and SHS-PC1 were associated with lower odds of 6 of the 13 sleep phenocodes: insomnia, obstructive sleep apnea, restless legs syndrome, sleep disorders (unspecified), organic or persistent insomnia, and sleep apnea (unspecified). PRS for SHS-PC2, SHS-PC4, and SHS-PC5 were associated with 2 sleep phenocodes: obstructive sleep apnea and sleep apnea. Whereas the SHS-PC4 PRS was associated with higher odds for sleep apnea disorders (likely reflecting its positive association with snoring), increases in PRS for SHS-PC2 and SHS-PC5 were associated with lower odds for sleep apnea.

### Implicated genes

We prioritized the genes at GWS loci using three mapping methods (position, eQTL, and Chromatin Interaction [CI]) as well as MAGMA<sup>20</sup> positional gene-based analysis in FUMA<sup>18</sup> (“Methods”; Supplementary Data 16–19; for gene-based GWAS Manhattan and QQ plots see Supplementary Figs. 8 and 9). Hundreds of genes were implicated. Here we highlight only those with the strongest evidence, as prioritized by all mapping methods. The five novel SHS-ADD loci were mapped to 39 genes, with five genes in two loci supported by all four mapping methods. The latter included *FNIP2*, which binds to AMP-activated protein kinase (AMPK) and plays a crucial role in mTORC1 signaling and the regulation of heat shock protein-90 (Hsp90)<sup>21</sup>, which has been previously linked to sleep homeostasis and behavioral rhythms<sup>22,23</sup>; *NRBF2*, which is involved in circadian rhythm via control of autophagy, and nutrient and cellular homeostasis<sup>24</sup>; and *JMJD1C*, which plays a role in DNA repair<sup>25</sup> and has been associated with

## GWAS Results



Rett syndrome (OMIM 312750), which co-occurs with epilepsy and sleep disturbance<sup>26</sup>. Nine novel SHS-PC1 loci were mapped to 140 genes, with five genes in three loci mapped by all four methods. Of these, *NMB* encodes the neuromedin B neuropeptide linked to the endocrine and exocrine systems, body temperature, and blood pressure<sup>27</sup>, while at the same locus, *WDR73* is highly expressed in cerebellar Purkinje neurons, and *ZNF592* has been implicated in cerebellar atrophy; *ANKFY1* is also involved in the

maintenance of cerebellar Purkinje cells that play a role in sleep-wake regulation<sup>28,29</sup>; *MAPK1* is part of the mitogen-activated protein kinase (MAPK)/extracellular signal-regulated kinase (ERK) signaling pathway that is linked to mental health and the circadian system<sup>30</sup>. Five novel SHS-PC2 loci were mapped to 97 genes (nine genes in four loci mapped by all methods). Of these, *SEMA3D* encodes a member of the class III semaphorins that are involved in axon guidance during neuronal development<sup>31</sup>

**Fig. 2 | Genome-wide significant SNPs associated with SHS.** **a** Number of Genome-wide significant (GWS) SNPs ( $p < 5e-08$ ) and risk loci (“Methods”). **b** Number of risk variants that colocalized for each pair of SHS (HyPrColoc; “Methods”). **c** Number of loci not reported by previous sleep GWASs in biobanks (GWS loci with a lead variant at least 500 kb from any of the previously published GWS sleep variants; “Methods”). **d** Number of loci reported by previous sleep GWASs in biobanks (lead variant within 500 kb of published GWS sleep variants; “Methods”). **e** Distribution of Functional consequences (FUMA<sup>18</sup>; “Methods”) of all annotated SNPs in LD with independent GWS SNPs by SHS; 3.2% of the annotated SNPs were in functional regions (exon, UTR, and splice site). **f** Regulome DB score

distribution (FUMA; “Methods”) of all annotated SNPs in LD with independent GWS SNPs by SHS; 3.4% of the annotated SNPs were in regulatory regions with Regulome DB score  $< 2$ . **g** CADD score distribution (FUMA; “Methods”) of all annotated SNPs in LD with independent GWS SNPs by SHS; 6.8% of the annotated SNPs likely deleterious effect with CADD score  $> 10$ . **h** Chromatin state distribution (FUMA; “Methods”) of all annotated SNPs in LD with independent GWS SNPs by SHS; 74% of the annotated SNPs were in open chromatin regions with a minimum chromatin state between 1 and 7. OSA obstructive sleep apnea, RLS restless leg syndrome, SHS sleep health score, UTR5 5' untranslated region: UTR3 3' untranslated region, ncRNA non-coding RNA.

**Table 1 | Significant novel loci associated with SHS with a lead variant at least 500 kb from any of the previously published genome-wide significant sleep variants**

SHS	Loc.	SNP	Chr.	Position (GRCh37)	Nearest gene(s)	Alleles (E/A)	EAF	INFO	BETA	SE	P	
ADD	1	rs75607302	4	159677732	<i>PPID/FNIP2</i>	A/AT	0.428	0.977	-0.013	0.002	2.8E - 09	
	2	rs10808575	8	130062388	<i>AC068570.1/LINC00977</i>	T/C	0.733	0.997	0.014	0.002	2.8E - 09	
	3	rs12257317	10	19457469	<i>UBE2V2P1/MALRD1</i>	A/G	0.685	0.997	0.013	0.002	5.7E - 09	
	4	rs7924036	10	65191645	<i>JMJD1C</i>	G/T	0.497	1.000	-0.012	0.002	7.4E - 09	
	5	rs72896891	18	42632654	<i>SETBP1</i>	A/T	0.833	0.993	-0.018	0.003	6.8E - 10	
PC1	6	rs12759956	1	18432831	<i>RP11-174G17.2/IGSF21</i>	T/A	0.713	0.995	0.017	0.003	4.0E - 10	
	7	rs12470733	2	200968215	<i>C2orf47/SPATS2L</i>	C/A	0.798	0.993	-0.019	0.003	4.6E - 10	
	8	rs1571582	9	103663962	<i>RP11-394D2.1/ RP11-62L10.1</i>	T/C	0.498	0.997	-0.016	0.002	3.4E - 11	
	9	rs201449027	12	9142784	<i>KLRG1/ RP11-259O18.4</i>	TG/T	0.561	0.965	0.016	0.002	1.4E - 10	
	10	rs4559781	13	28303803	<i>NPM1P4/GSX1</i>	C/T	0.153	0.994	0.020	0.003	2.1E - 09	
	11	rs139221256	15	85357857	<i>ZNF592/ALPK3</i>	T/TA	0.749	0.991	0.016	0.003	5.8E - 09	
	12	rs12601771	17	4108822	<i>ANKFY1</i>	G/A	0.433	0.995	0.015	0.002	5.6E - 10	
	13	rs8074498 <sup>a</sup>	17	79954544	<i>ASPSCR1</i>	T/A	0.418	0.983	0.016	0.002	1.2E - 10	
	14	rs9610500	22	22221167	<i>MAPK1</i>	A/G	0.634	0.983	-0.015	0.003	4.8E - 09	
	15	rs2821226	1	203517292	<i>OPTC/ATP2B4</i>	A/G	0.473	0.984	-0.013	0.002	5.2E - 09	
	PC2	16	rs17559978	7	84677860	<i>SEMA3D</i>	G/A	0.688	0.990	0.015	0.002	2.3E - 10
		17	rs111111069	12	102271962	<i>DRAM1</i>	C/G	0.791	0.995	-0.016	0.003	2.4E - 09
18		rs113851179	16	1733479	<i>LA16c-431H6.6/HN1L</i>	C/CT	0.925	0.987	0.026	0.004	8.6E - 10	
19		rs12979056	19	17862131	<i>FCHO1</i>	G/A	0.542	0.992	-0.014	0.002	1.5E - 09	
20		rs3788337	22	23412017	<i>RTDR1</i>	G/A	0.647	0.993	-0.014	0.002	6.5E - 09	
PC3		21	rs56049037	7	32947201	<i>AVL9</i>	G/A	0.713	0.988	0.016	0.002	1.3E - 11
	22	8:11053467-GA:G	8	11053467	<i>XKR6</i>	GA/G	0.555	0.914	0.014	0.002	1.6E - 09	
	23	rs11494758	12	9116542	<i>KLRG1</i>	C/T	0.622	0.997	-0.014	0.002	3.4E - 10	
	24	rs71272625	15	78166843	<i>LINGO1/CSGP4P13</i>	C/CT	0.331	0.931	0.014	0.002	8.8E - 10	
	25	rs11373181	22	42705672	<i>TCF20</i>	A/AC	0.482	0.990	-0.013	0.002	3.6E - 10	
PC4	26	13:58551593-GA:G	13	58551593	<i>PCDH17/RNA5SP30</i>	GA/G	0.763	0.993	-0.016	0.002	2.4E - 11	
PC5	27	rs138572890 <sup>b</sup>	16	50264953	<i>PAPD5</i>	C/CTTTA	0.929	0.968	0.023	0.004	7.7E - 10	
	28	19:59007970-CA:C	19	59007970	<i>SLC27A5</i>	CA/C	0.366	0.948	-0.013	0.002	7.0E - 11	

The significance threshold was adjusted for multiple testing across six traits ( $p < 8.3 \times 10^{-9}$ ).

Loc. locus, Chr. chromosome, E/A effect and alternative alleles, EAF effect allele frequency, INFO imputation quality score.

<sup>a</sup>Missense variant.

<sup>b</sup>3' UTR variant.

(two more class III semaphorins at this locus, *SEMA3A* and *SEMA3E*, also mapped by CI); *DRAM1* is a regulator of autophagy in the context of mitochondrial dysfunction, implicated in neurodegeneration<sup>32</sup>; *MAPK8IP3* is another MAPK/ERK gene essential for the function and maintenance of neurons, with links to neurodevelopmental disorders<sup>33</sup>; *FCHO1* is involved in clathrin-coat assembly and clathrin-mediated endocytosis and has been implicated in immune deficiency<sup>34,35</sup>. Three novel SHS-PC3 loci were mapped to 62 genes, with five genes in three loci mapped by all four methods. These include *PDE1C* a phosphodiesterase bound by calmodulin that regulates proliferation of vascular smooth muscle cells and may play a

pathological role in cardiac remodeling and dysfunction<sup>36</sup> and *TCF20*, involved in neurodevelopmental diseases and sleep disturbances<sup>37,38</sup>. One novel SHS-PC4 locus was mapped to three genes, including *PCDH17* (mapped by position, eQTL and CI in hippocampal and neural progenitor cells) involved in forming and maintaining neuronal synapses<sup>39</sup>. The two novel SHS-PC5 loci mapped to 71 genes, of which two genes were mapped by all four methods. These include *SLC27A5*, involved in bile acid synthesis and metabolism<sup>40</sup>, which has also been implicated in brain health and neural development<sup>41</sup>. In addition, fifty genes in 28 novel loci have shown drug interactions (Supplementary Data 16). Implicated genes for additional

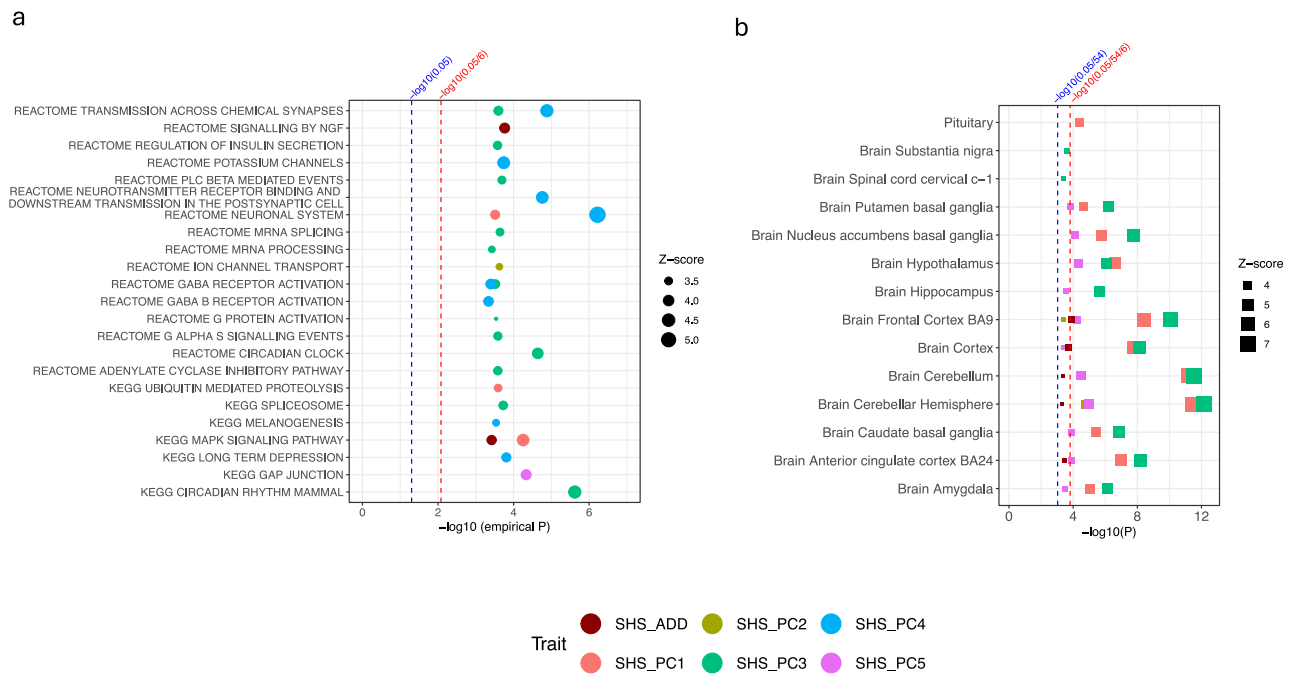
**Table 2 | Novel SHS Polygenic Risk Score (SHS-PRS) associations in independent studies**

Study	Phenotype	Phe -code	N°	SHS-ADD		SHS-PC1		SHS-PC2		SHS-PC3		SHS-PC4		SHS-PC5	
				Beta (SE)/ OR (95% CI)	P*	Beta (SE)/ OR (95% CI)	P*	Beta (SE)/ OR (95% CI)	P*	Beta (SE)/ OR (95% CI)	P*	Beta (SE)/ OR (95% CI)	P*	Beta (SE)/ OR (95% CI)	P*
HCHS/SOL	Corresponding SHS <sup>b</sup>		11,144	0.063 (0.010)	<b>3.86</b> E – 10	0.086 (0.009)	<b>9.22</b> E – 20	0.082 (0.009)	<b>3.86</b> E – 18	0.024 (0.009)	1.16 E – 02	0.047 (0.009)	<b>5.41</b> E – 07	0.094 (0.009)	<b>3.04</b> E – 23
MGB biobank	Insomnia	327.4	5748/ 30226	0.91 (0.88, 0.94)	<b>5.87</b> E – 10	0.9 (0.87, 0.92)	<b>1.15</b> E – 12	0.99 (0.96, 1.02)	0.416	0.97 (0.94, 1)	0.026	1.01 (0.98, 1.04)	0.372	0.96 (0.93, 0.98)	0.002
	Obstructive sleep apnea	327.32	6722/ 30226	0.86 (0.84, 0.88)	<b>6.73</b> E – 25	0.93 (0.9, 0.95)	<b>7.65</b> E – 08	0.86 (0.84, 0.89)	<b>3.88</b> E – 26	1.00 (0.97, 1.03)	0.998	1.08 (1.06, 1.11)	<b>5.21</b> E – 09	0.94 (0.92, 0.97)	<b>1.82</b> E – 05
	Restless legs syndrome	327.71	1281/ 30226	0.86 (0.81, 0.91)	<b>3.06</b> E – 07	0.86 (0.81, 0.91)	<b>1.94</b> E – 07	0.93 (0.88, 0.99)	0.015	0.96 (0.91, 1.02)	0.179	1.00 (0.94, 1.06)	0.982	0.91 (0.86, 0.97)	0.002
	Sleep disorders	327	3048/ 30226	0.88 (0.85, 0.92)	<b>1.30</b> E – 09	0.9 (0.87, 0.94)	<b>2.80</b> E – 07	0.94 (0.91, 0.98)	0.003	0.96 (0.93, 1)	0.053	1.01 (0.97, 1.05)	0.564	0.98 (0.94, 1.02)	0.239
	Organic or persistent insomnia	327.41	1343/ 30226	0.87 (0.82, 0.92)	<b>2.10</b> E – 06	0.87 (0.82, 0.92)	<b>3.27</b> E – 06	0.97 (0.92, 1.03)	0.325	0.98 (0.93, 1.03)	0.419	1.02 (0.96, 1.08)	0.530	0.93 (0.88, 0.98)	0.008
	Sleep apnea	327.3	4161/ 30226	0.85 (0.82, 0.88)	<b>1.06</b> E – 19	0.93 (0.9, 0.96)	<b>2.61</b> E – 05	0.84 (0.81, 0.87)	<b>2.03</b> E – 23	0.99 (0.95, 1.02)	0.403	1.07 (1.04, 1.11)	<b>3.84</b> E – 05	0.93 (0.9, 0.97)	<b>6.77</b> E – 05
	Sleep-related movement disorders	327.7	445/ 30226	0.9 (0.81, 0.99)	0.036	0.89 (0.81, 0.99)	0.024	0.99 (0.89, 1.08)	0.763	0.95 (0.86, 1.04)	0.252	1.01 (0.92, 1.11)	0.782	0.99 (0.9, 1.09)	0.774
	Circadian rhythm sleep disorder	327.6	302/ 30226	0.85 (0.76, 0.96)	0.010	0.88 (0.78, 0.99)	0.033	0.88 (0.78, 0.99)	0.034	0.94 (0.83, 1.05)	0.246	0.97 (0.86, 1.08)	0.559	0.99 (0.88, 1.11)	0.882
	Parasomnia	327.5	497/ 30226	0.94 (0.85, 1.03)	0.171	0.92 (0.83, 1)	0.061	0.99 (0.9, 1.09)	0.842	0.95 (0.87, 1.04)	0.277	1.01 (0.93, 1.11)	0.784	1.01 (0.93, 1.11)	0.771
	Hypersomnia	327.1	989/ 30226	0.89 (0.84, 0.96)	0.001	0.94 (0.88, 1)	0.064	0.9 (0.84, 0.96)	0.001	0.94 (0.88, 1)	0.049	1.02 (0.96, 1.09)	0.554	0.92 (0.87, 0.99)	0.017
	Central/nonobstructive sleep apnea	327.31	555/ 30226	0.91 (0.83, 1)	0.046	1.06 (0.97, 1.15)	0.232	0.93 (0.85, 1.01)	0.089	0.96 (0.89, 1.05)	0.407	1.07 (0.98, 1.16)	0.117	0.88 (0.8, 0.96)	0.003
	Sleep-related leg cramps	327.72	223/ 30226	1.00 (0.86, 1.15)	0.949	0.95 (0.83, 1.1)	0.511	1.04 (0.91, 1.19)	0.563	1.07 (0.94, 1.22)	0.318	1.07 (0.94, 1.22)	0.305	1.00 (0.87, 1.14)	0.990
	Cataplexy and narcolepsy	347	325/ 30860	0.88 (0.78, 0.99)	0.032	0.97 (0.86, 1.08)	0.558	0.88 (0.79, 0.98)	0.025	0.98 (0.88, 1.09)	0.660	1.01 (0.91, 1.13)	0.842	0.96 (0.86, 1.07)	0.432

<sup>a</sup>Total sample size is provided for continuous outcome in HCHS/SOL, sample sizes in cases vs controls are provided for binary sleep phenotypes in MGB biobank.

<sup>b</sup>SHS in HCHS/SOL were calculated from the SHS-ADD definition and UKB SHS-PC loadings.

<sup>c</sup>Significant P-values of PRS associations with corresponding SHS in HCHS/SOL accounting for 6 SHS traits (P < 0.008) and with sleep phenotype in MGB biobank accounting for the number of phenotype and SHS traits (P < 0.0006) shown in bold.



**Fig. 3 | Pathway and tissue enrichment analysis.** **a** Pathway gene sets significantly enriched for SHS genes (PASCAL<sup>43</sup>, “Methods”). **b** Tissue-specific expression gene sets enriched for SHS genes (MAGMA<sup>20</sup> via FUMA<sup>18</sup>, “Methods”).

unreported and reported variants are summarized in Supplementary Data 17 and 18. For completeness, all significant genes ( $p < 2.64e-6$ ) in gene-based analysis are reported in Supplementary Data 19. Note that genes implicated by fewer mapping methods may also merit prioritization. One intriguing example is that two GWS unreported SHS-PC2-associated variants, rs2821226 (*OPTC/ATP2B4*;  $5.20E-09$ ) and rs1610263 (*COL8A1*;  $4.10E-08$ ), both implicate collagen-pathway genes that may play a key role in the physiology of obstructive sleep apnea (OSA)<sup>42</sup>. Lack of causal tissues (e.g., specific brain and connective tissues) may result in incorrect mappings. Any definitive conclusions based on mapped genes will require functional follow-up.

### Gene set enrichment analyses

We performed pathway enrichment analysis, applying PASCAL<sup>43</sup> to SHS GWAS summary statistics, and identified significant enrichments of SHS-ADD variants in MAPK and NGF signaling pathways; SHS-PC1 variants in neuronal system, ubiquitin-mediated proteolysis, and MAPK signaling pathways; SHS-PC2 in ion transport; SHS-PC3 in circadian, mRNA processing and splicing, G-protein, and metabolic pathways; SHS-PC4 in neuronal system, neurotransmission at synapses, GABA receptor, and long term depression pathways; and SHS-PC5 in the gap junction pathway (Empirical  $p < 5e-4$ ; Fig. 3a and Supplementary Data 20).

We also performed gene-set enrichment analyses<sup>20</sup> (tissue, trait, cell type, and pathway) on genes mapped by FUMA using MAGMA (Supplementary Data 21–23). Tissue enrichment analyses identified multiple brain tissues for all SHS traits except SHS-PC4, with the highest enrichments in cerebellum, hypothalamus, and frontal cortex for SHS-PC1 and frontal cortex, cerebellum, and nucleus accumbens for SHS-PC3. All enriched tissue findings were brain tissues, except the pituitary tissue for SHS-PC1 ( $p < 0.05/54/6 = 1.5e-4$ ; Fig. 3b and Supplementary Data 21), which has both neural and endocrine functions, the posterior pituitary containing distal axons of hypothalamic neurons. Cell type enrichment analysis identified multiple brain cell types for each SHS, and especially for SHS-PC1, including GABAergic neurons in the human midbrain (Supplementary Fig. 10).

Enrichment of SHS-associated genes with phenotype-associated gene sets from the GWAS catalog (Supplementary Data 22) revealed associations with psychological traits (intelligence, neuroticism, impulsivity/risk-taking,

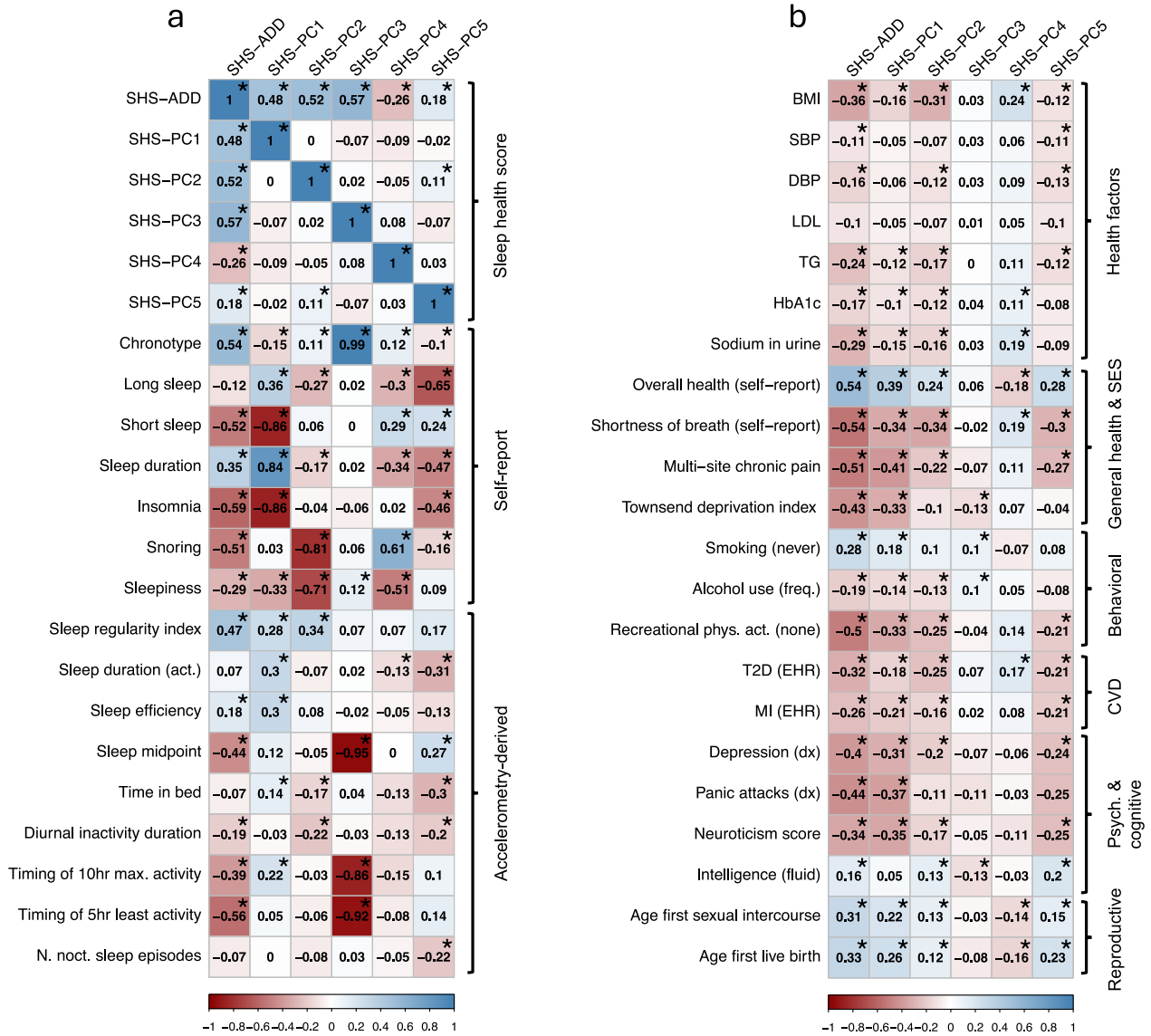
mood, and psychiatric disorders), behavior (e.g., regular activity patterns, alcohol consumption), inflammatory markers and diseases (C-reactive protein, IgG, and inflammatory bowel diseases), blood pressure, adiposity (especially in SHS-PC2 and PC4), and reproductive aging. Gene sets for Alzheimer’s disease and related biomarkers (cerebrospinal fluid tau and amyloid  $\beta$ ) were enriched in SHS-PC1, SHS-PC4, and most strongly SHS-PC5. A hippocampal volume gene set was enriched in SHS-PC2 and SHS-PC5. Dendrite gyrus brain volume and kidney disease gene sets were enriched in SHS-PC2. Gene sets associated with intracranial and subcortical brain region volumes, craniofacial microsomia, idiopathic pulmonary fibrosis, and aortic root size were enriched in SHS-PC4. An iron biomarker gene set was enriched in SHS-PC5.

### Genetic and causal relationships between SHS and other common complex traits

LD score regression (LSDC)<sup>44</sup> revealed numerous phenotypes genetically correlated to SHS traits. Among 375 phenome-wide representative heritable traits (“Methods”), 256 traits were genetically correlated with at least one SHS ( $p < 0.05/375/6 = 2.2e-5$ ; Fig. 4 and Supplementary Data 24). Genetic correlations with SHS were strongest (magnitude  $\sim 0.3$  to  $0.5$ ) with physical and mental health, and (inversely) with socio-economic status (SES), stress, pain, mental and emotional distress, and recognized health conditions and risk factors. However, these genetic correlations had discernable patterns, unique to each SHS trait. Compared with the SHS-PCs, SHS-ADD had stronger genetic correlations with non-specific health markers, e.g., traits related to overall health, physical conditioning, markers of socio-economic status, healthy lifestyle factors, as well as stronger inverse genetic relationships with pain, activities interfering with sleep, and depression.

Compared with the other SHSs, SHS-PC1 had stronger inverse genetic correlations with anxiety traits, alcohol addiction, and self-harm behavior; SHS-PC2 had comparatively stronger genetic correlations with daytime napping, diagnosed sleep disorders, and more moderate but, relative to other SHSs, still comparatively stronger inverse genetic correlations with metabolic and adiposity traits. Association patterns for SHS-PC3 differed markedly from other SHS, having lower correlations with overall health and most individual health conditions, inverse association with educational attainment. Compared with the other SHS, genetic correlations involving SHS-





**Fig. 4 | SHS genetic correlations with selected phenotypes.** **a** SHS genetic correlations (LDSC<sup>44</sup>, Methods) with sleep traits ( $*p < 0.05/132$ ). **b** SHS genetic correlations (LDSC<sup>44</sup>, Methods) with selected health outcomes ( $*p < 0.05/2250$ ). SHS sleep health score, act. derived from actigraphy, max. maximum, N number, noct. nocturnal, BMI body mass index, SBP systolic blood pressure, DBP diastolic blood

pressure, LDL low-density lipoprotein cholesterol, TG triglycerides, HbA1c glycosylated hemoglobin, phys. act. physical activity, freq. frequency, T2D type-II diabetes, MI myocardial infarction, dx self-report of physician diagnosis to trained interviewee, EHR electronic health record-derived, SES socioeconomic status, CVD cardiovascular and metabolic disease, psych. psychological.

PC4 were often inverted, including a (non-significant) inverse association with overall health rating (despite a weak positive phenotypic correlation) as well as positive associations with BMI, and cardiovascular traits, congruent with its positive loading on snoring. SHS-PC5 was preferentially genetically correlated with fluid intelligence, educational attainment, and mother’s age.

We further conducted bidirectional Mendelian randomization (MR) analyses to investigate potentially causal links between sleep health and 50 selected traits (Methods). We identified potential causal effects, estimated via inverse variance weighted (IVW) method, of lower SHS-PC1 on codeine or tramadol medication use ( $\beta_{IVW} = 0.47$ ;  $p = 4.1E - 07$ ); lower SHS-PC5 on Bipolar disorder ( $\beta_{IVW} = 1.66$ ;  $p = 1.6E - 05$ ); and lower SHS-ADD on smoking initiation ( $\beta_{IVW} = -0.42$ ;  $p = 6.3E - 05$ ) and years of schooling ( $\beta_{IVW} = 0.34$ ;  $p = 1.1E - 04$ ) (Supplementary Data 25). In the reverse direction, MR identified potential causal effects of greater years of education on higher SHS-ADD ( $\beta_{IVW} = 0.15$ ;  $p = 1.2E - 07$ ) and SHS-PC5 ( $\beta_{IVW} = 0.10$ ;  $p = 2.7E - 06$ ); lower BMI on higher SHS-PC2 ( $\beta_{IVW} = -0.10$ ;  $p = 7.2E - 07$ ); and Alzheimer’s disease risk on higher SHS-PC5 ( $\beta_{IVW} = 0.012$ ;  $p = 1.5E - 05$ ) (Supplementary Data 26).

**Comparison with composite sleep health scores constructed using genetic correlations in addition to phenotypic correlations**

As a sensitivity analysis we constructed composite sleep scores informed by genetic correlations rather than phenotypic correlations, using linear combinations of the same underlying sleep phenotypes as used in the PC-based approach. We constructed these phenotypes using the ‘MaxH’ maximally heritable approach<sup>45</sup>, which provides linear combinations of phenotypes of maximum heritability, under an independence constraint. The SHS-MaxH phenotypes were largely phenotypically concordant with SHS-PCs, with similar heritabilities (Supplementary Data 2 and 27). SHS-MaxH1 correlated phenotypically with SHS-PC3; SHS-MaxH2 with SHS-PC1; SHS-MaxH3 with SHS-PC2 ( $r > 0.9$ ). Less concordantly, SHS-MaxH4 and MaxH5 correlated with both SHS-PC4 and PC5 ( $|r| > 0.5$ ) but were no greater in heritability.

**Discussion**

We performed the first large-scale sleep health GWAS investigating five novel PC-based SHSs and compared these with an updated GWAS of

SHS-ADD by including related individuals in UKB. Each SHS was based on five underlying self-reported sleep traits: sleep duration, insomnia symptom frequency, daytime sleepiness frequency, chronotype, and snoring, resulting in distinct sleep health composites interpretable via their loadings. The SHS approach emphasizes the co-occurrence of multiple sleep traits, aligning with the multi-dimensional view of sleep health, in which individual components of sleep do not confer sleep health in isolation.

We identified 28 novel significant ( $p < 8.3e-9$ ) loci and 31 additional GWS ( $p < 5e-8$ ) loci that were not reported by previous sleep GWASs. Our findings were supported by sensitivity analysis and PRS validation and associations with clinical sleep outcomes in independent studies. These loci mapped to genes implicated in neurodevelopment, synaptic signaling, ion channel transportation, cellular energy production, and metabolic processes. The findings collectively suggest that studying SHSs has advanced genetic discovery by linking to plausible biological mechanisms and aligning with established sleep health domains, thereby uncovering novel insights.

Findings for SHS-ADD, combining 5 positive binary sleep traits, suggest a phenotype that captures global sleep health by integrating multiple independent regulatory signals and sleep domains. Notably, SHS-ADD was the SHS most strongly associated with accelerometry-derived sleep regularity index, both phenotypically and genetically, suggesting the conjunction of multiple independent sleep health traits may be a prerequisite for sleep regularity. Genetic correlations were strongest between SHS-ADD and overall health and SES, with MR indicating a potential bidirectional causal relationship with SES. SHS-ADD was sensitive to several adjustment factors, including health behaviors and health status, as well as both psychological factors (with SHS-PC1) and BMI (with PC2). Moreover, for SHS-ADD, evidence of enrichments, particularly in neuronal tissues and cell types, was not as strong as for other traits such as SHS-PC1 and SHS-PC3. Together these findings suggest SHS-ADD to be a broad sleep health phenotype, that retains significant heterogeneity, combining multiple independent sleep domains and/or sleep disorders with distinct etiologies. It nevertheless highlights the connections between global sleep health, sleep regularity, and overall health and well-being.

SHS-PC1, longer sleep with less frequent insomnia and sleepiness, reflects correlated self-reported sleep traits across the domains of satisfaction, duration, and alertness, while demonstrating higher heritability than SHS-ADD and each of its underlying sleep traits. SHS-PC1 was the SHS most strongly genetically correlated with accelerometry-based sleep efficiency and duration, suggesting shared genetics with both longer and more efficient objective sleep. Additional genetic correlation and MR analyses suggest negative relationships with chronic pain (potentially causal), anxiety, and neuroticism. Enrichment analyses identified the ubiquitin-proteasome system pathway, important for circadian rhythm regulation and sleep homeostasis<sup>46</sup>, as well as the GABAergic neuronal cell-type, central to neural orchestration of sleep. Overall, these findings suggest a phenotype capturing neurobiological sleep regulation, and support a bidirectional relationship with anxiety<sup>47</sup> via shared GABAergic regulation<sup>48</sup>, while pointing to further mechanisms involving presence or perception of distressing stimuli, including chronic pain.

SHS-PC2, interpreted as healthy sleep characterized by absent snoring and sleepiness, cardinal symptoms of sleep apnea syndrome, may serve as a surrogate indicating absence of under-diagnosed and poorly captured clinical sleep-disordered breathing conditions. Correspondingly, SHS-PC2 showed strong inverse (causal) associations with adiposity-related measures like BMI, consistent with the strong association between OSA and obesity. Likewise, SHS-PC2 was the SHS most sensitive to BMI adjustment, which nevertheless only modestly attenuated estimated GWAS effects. SHS-PC2 had inverse genetic correlations with cardiometabolic traits, notably type 2 diabetes, which has previously been linked bidirectionally to OSA<sup>49</sup>. Enrichment analyses also link SHS-PC2 with neuronal pathways and hippocampal volume, suggesting neurological involvement in OSA, and with connective-tissue genes (collagen pathway) and traits (adolescent idiopathic scoliosis, aortic root size), consistent with a role for connective tissue in pharyngeal collapsibility related to sleep-disordered breathing<sup>50</sup>.

SHS-PC3 largely recapitulates chronotype and confirmed known associations with circadian genes and pathways. It is notable that chronotype was largely independent not only from other SHS-PCs but was also more weakly genetically correlated with phenome-wide health outcomes, while being moderately genetically correlated with healthy lifestyle behaviors, such as physical activity and time spent outdoors in summer.

Though SHS-PC4 and SHS-PC5 have more complex interpretations due to both positive and negative loadings on healthy sleep traits, they were nonetheless roughly as heritable as the least heritable individual trait (sleepiness) and contributed novel genetic findings. SHS-PC4, interpreted as snoring without sleepiness, showed genetic enrichment in neurotransmission pathways and craniofacial structure, suggesting mechanisms that could lead to sleepiness without snoring, and/or snoring without sleepiness. The latter would be consistent with a sleep-disordered breathing phenotype resulting from reduced craniofacial dimensions that cause pharyngeal narrowing and turbulent airflow (or snoring) without the severe airway collapsibility, sleep fragmentation, and inflammation characteristic of obstructive sleep apnea syndrome<sup>51,52</sup>.

SHS-PC5, characterized by short sleep without insomnia, was genetically correlated with later objective sleep midpoint and fewer accelerometry-measured sleep episodes, as well as positive genetic correlations with markers of cognitive decline, including memory loss, cerebrospinal fluid t-tau levels, and Alzheimer's disease, and with enrichment of genes in the gap junction pathway, implicated in amyloid- $\beta$  clearance by astrocytes in Alzheimer's disease (AD)<sup>53</sup>. Additionally, MR showed a potential reverse causal association between AD and SHS-PC5. This suggests that SHS-PC5 may characterize shorter, delayed sleep, without insomnia symptoms, indicative of accelerated brain aging, rather than a natural short sleep or 'super-sleeper' phenotype<sup>3</sup>. The low likelihood of insomnia in this subtype is consistent with lack of consistent data implicating insomnia in cognitive decline, potentially due to the heterogeneity of conditions underlying insomnia.

Colocalization analyses revealed selective cases of shared regulation, but also pointed to sometimes differing relationships with underlying components, particularly in the special cases of PC2 vs PC4, and PC1 vs PC5, which share underlying traits with opposite loadings. For example, *DLEU7*, proximal to rs592333, colocalized opposite associations with PC2 and PC4, consistent with a role in snoring (which loaded positively in PC4 and negatively in PC2) and consistent with prior associations of the *DLEU7* locus with adiposity<sup>54</sup>. Conversely, rs1846644 colocalized positive associations with both PC2 and PC4 at the *KSR2* locus, a gene highly expressed in cerebellar Purkinje neurons<sup>55</sup>, suggesting a link to sleepiness regardless of the presence of snoring, in keeping with a role for Purkinje neurons in sleep-wake transition<sup>56</sup>. Similarly, for SHS-PC1 and PC5, *PAX8* colocalized with opposite directions, in keeping with opposite loadings on sleep duration, while *MEIS1* colocalized with similar direction in keeping with consistent loading on insomnia.

Several additional findings pointed to mechanisms shared across SHS indicating broad-based involvement in sleep health. For example, PC2 and PC4 at the *KSR2* locus, a gene highly expressed in cerebellar Purkinje neurons, which were again implicated in PC1 by two novel loci, *ANKFY1* and *WDR73/ZNF592*, suggesting cerebellar regulation of sleep maintenance efficiency. Additional findings consistently reinforced the role of neural development, as well as consistently implicating neurotransmitters and synaptic signaling. The association of SHS-PC2 with *FCHO1* at rs12979056 gives further support for a role of clathrin-coat vesicle transport in sleep health, presumably in synaptic function, a mechanism previously implicated by the *STON1-GTF2A1L*, *TOR1A*, *TOR1B*, *AP2B1* (PC3) and *AP3B2* (PC1) loci. Pathways and genes related to MAPK, GAP-junction, immune signaling, and energy metabolism processes were found to associate across multiple SHS. Interestingly, the two identified novel functional variants, rs8074498 in PC1 and rs138572890 in PC5, the PC phenotypes loading on duration/insomnia, were connected to glucose transporters (respectively *GLUT4* and *GLUT1*). Moreover, enrichment analysis for individual SHS all highlighted gene expression in brain tissues and cells, and associations with metabolic, inflammatory, and psychiatric traits, which reinforce critical roles for central nervous system, metabolic, and immune system function on

sleep regulation. Multi-trait genetic correlation analyses further suggested more broadly interrelatedness of sleep health with overall health, lifestyle, behavioral and psychological traits, as well as pain, physical frailty, and deconditioning. However, some caution is warranted as some apparent genetic relationships could be induced by factors relating to the subjective rating of sleep health, remaining heterogeneity within SHS traits, or unadjusted confounders, including remaining population stratification<sup>57</sup>.

This study has several limitations. First, SHS-PCs incorporated sleep traits in a linear fashion, which may not reflect the complexity of their contribution to sleep health, e.g., the U-shaped contribution of sleep duration as modeled in SHS-ADD. Second, SHS-PCs only focus on five self-reported sleep questions to maximize sample size and to be more comparable to SHS-ADD as previously published on in UKB by using the same underlying sleep phenotypes. This could limit our ability to fully capture a comprehensive sleep health composite, due to limitations of subjective assessment or the ability of these questions to fully capture well-being in sleep health. Third, the data-driven PC approach may limit generalizability across different studies and populations. We were able to identify HCHS/SOL as having similar questionnaire data, however, PC loading generalization to this cohort was challenging due to population and age differences. A potentially productive approach to future validation studies would be to average PC loadings across different studies in a meta-analysis<sup>58</sup>. Nonetheless, using the same PC loading in UKB, we did not replicate individual GWS loci but validated the PRSs constructed by genome-wide variants. Lastly, we note that the SHS-PCs were based on phenotypic correlations, which could limit the ability to derive heritable composite phenotypes. However, in a sensitivity analysis we constructed sleep scores informed by genetic correlations using the ‘maxH’ maximally heritable approach<sup>45</sup>, for which the loadings and heritability of the derived phenotypes were numerically and qualitatively similar. Nonetheless, genetic correlation information could be valuable in future research integrating additional objective sleep-related phenotypes.

In summary, this study introduces a novel approach to understanding sleep genetics using PC-based SHS, which effectively distinguishes differing mechanisms of distinct, domain-specific sleep-related traits. In keeping with the large influence of insomnia (related to SHS-PC1) and sleep apnea (related to PC2) on sleep health in the population, along with the independent role of chronotype (PC3), our approach appears to have enhanced genetic discovery by separately targeting these domain-specific sleep health scales. Future research involving SHSs incorporating accelerometry data and other objectively measured sleep and activity data may provide enhanced targeting of psychosocial domains and neuroregulatory sleep mechanisms, resulting in further genetic discovery.

## Methods

### Population and study design

The discovery analysis was conducted on participants of European ancestry from the UK Biobank study<sup>59</sup>. The UK Biobank is a prospective study that has enrolled over 500,000 people aged 40–69 living in the United Kingdom. Baseline measures collected between 2006–2010, including self-reported health questionnaire and anthropometric assessments, were used in this analysis. Participants taking any self-reported sleep medication (described before<sup>7,8,10</sup>) were excluded. The UK Biobank study was approved by the National Health Service National Research Ethics Service (ref. 11/NW/0382), and all participants provided written informed consent to participate in the UK Biobank study.

### Genotype

DNA samples of 502,631 participants in the UK Biobank were genotyped on two arrays: UK BiLEVE (807,411 markers) and UKB Axiom (825,927 markers). 488,377 samples and 805,426 genotyped markers passed standard QC<sup>60</sup> and were available in the full data release. 452,071 individuals of European ancestry (based on K-means clusters on genomics PCs) were studied with available phenotypes and genotyping passing quality control. SNPs were imputed to a combined Haplotype Reference Consortium (HRC)

and 1000 Genome panel. SNPs with minor allele frequency (MAF) > 0.001, BGEN imputation score > 0.3, maximum per SNP missingness of 10%, and samples with a per-sample missingness of 40% were kept in the GWAS.

### Sleep trait assessment

The UK Biobank baseline questionnaire assessed chronotype, sleep duration, insomnia symptoms, snoring, and excessive daytime sleepiness via self-report responses. Self-reported sleep duration was recorded as an integer-valued variable via responses to the question “About how many hours sleep do you get in every 24 h? (please include naps).” The remaining questions had ordinal responses, which were assigned to an integer scale as follows. Chronotype (morningness): “Do you consider yourself to be:” -2. Definitely an ‘evening’ person; -1. More an ‘evening’ than a ‘morning’ person; 0. Do not know; 1. More a ‘morning’ than an ‘evening’ person; 2. Definitely a ‘morning’ person; NA. Prefer not to answer. Insomnia Symptoms: “Do you have trouble falling asleep at night, or wake up in the middle of the night?” 1. Never/rarely; 2. Sometimes; 3. Usually; NA. Prefer not to answer. Snoring: “Does your partner or a close relative or friend complain about your snoring?” 1. Yes; 0. No; NA. Do not know or Prefer not to answer. Subjective daytime sleepiness: “How likely are you to doze off or fall asleep during the daytime when you don’t mean to? (e.g. when working, reading, or driving?)” 0. Never/rarely; 1. Sometimes; 2. Often; 3. All of the time; NA. Do not know or Prefer not to answer. Individuals with any missingness (NA) from any sleep questionnaires were excluded from the analysis.

### Sleep health score construction

For the UK Biobank additive sleep health score (SHS-ADD), consistent with previous studies<sup>5,13</sup>, we assigned one point to each of five positive sleep traits, as follows: Chronotype: More a ‘morning’ than an ‘evening’ person or Definitely a ‘morning’ person; Sleep Duration: from 7 to 8 h (inclusive); Insomnia Symptoms: Never/rarely; Snoring: No; Subjective daytime sleepiness: 0. Never/rarely. No subjects were excluded; those who did not report the positive attributes were coded as zero for that trait. The final SHS-ADD rating is the total number of positive sleep traits for each individual on a scale of 0–5. We performed principal component analysis of the above-described integer-scale self-report sleep question responses, after centering and scaling to standardize each response to mean zero, variance one, to compute SHS-PC1 through SHS-PC5. SHS-PCs were oriented so that they were positively correlated with self-assessed overall health (UKB Data-Field 2178: Overall health rating).

### Covariate measurements

Covariates used in the sensitivity analyses included potential confounders (depression, socio-economic deprivation based on residential area, alcohol intake frequency, smoking status, caffeine intake, employment status, marital status, neurodegenerative disorders, and use of psychiatric medications). Depression was recorded as a binary variable (yes/no) corresponding to question “Ever depressed for a whole week?”. Social economic status was measured by the Townsend Deprivation Index based on aggregated data from national census output areas in the UK. Alcohol intake frequency was coded as a continuous variable corresponding to “daily or almost daily”, “three or four times a week”, “once or twice a week”, “once to three times a month”, “special occasions only”, and “never” drinking alcohol. Smoking status was categorized as “current”, “past”, or “never” smoked. Caffeine intake was coded continuously corresponding to self-reported cups of tea/coffee per day. Employment status was categorized as “employed”, “retired”, “looking after home and/or family”, “unable to work because of sickness or disability”, “unemployed”, “doing unpaid or voluntary work”, or “full or part-time student”. Neurodegenerative disorder cases ( $N = 517$ ) were identified as a union of International Classification of Diseases (ICD)-10 coded Parkinson’s disease (G20–G21), Alzheimer’s disease (G30), and other degenerative diseases of nervous system (G23, G31–G32).

### Accelerometry data

Accelerometry data were collected using Axivity AX3 wrist-worn triaxial accelerometers in 103,711 individuals from the UK Biobank for up to 7 days,

3–10 years after baseline<sup>61</sup>. Sleep period time (SPT)-window and activity levels were extracted using a heuristic algorithm using the R package GGIR<sup>62</sup>. Briefly, for each individual, a 5-min rolling median of the absolute change in z-angle (representing the dorsal-ventral direction when the wrist is in the anatomical position) across a 24-h period. The 10th percentile of the output was used to construct an individual's threshold, distinguishing periods with movement from non-movement. Inactivity bouts were defined as inactivity of at least 30 min duration. Inactivity bouts with gaps of less than 60 min were combined into blocks. The SPT-window was defined as the longest inactivity block, with sleep onset as the start of the block and waking time as the end of the block. This algorithm provides comparable estimates of sleep onset time, waking time, SPT-window duration, and sleep duration within the SPT-window with polysomnography-derived metrics<sup>62</sup>. After quality control based on missingness, wear time, and calibration, eight metrics were generated and analyzed in this study, namely: M10 (midpoint of the 10 consecutive hours of maximum activity), L5 (midpoint of the 5 consecutive hours of minimum activity), sleep midpoint, sleep duration, sleep efficiency, diurnal inactivity, number of nocturnal sleep episodes, and sleep regularity index (accounting for wake after sleep onset and daytime napping<sup>63</sup>).

### Genome-wide association analysis

We performed a genome-wide association analysis (GWAS) of six SHS traits as continuous variables using linear mixed regression models adjusting for age, sex, genotyping array, 10 PCs, and genetic relatedness matrix in BOLT-LMM. Association testing was based on *p*-value from the non-infinite mixed model association test provided in the BOLT-LMM output. Reference 1000 genome European-ancestry (EUR) LD scores and genetic map (hg19) were utilized in this analysis. X-chromosome data were imputed and analyzed separately (with males coded as 0/2 and female genotypes coded as 0/1/2) using the same analytical approach in BOLT-LMM as was done for analysis of autosomes. FUMA was used to annotate the genome-wide significant (GWS) risk loci ( $p < 5e-8$ ), lead independent SNPs ( $r^2 < 0.1$ ), and all SNPs in LD with independent SNPs ( $r^2 \geq 0.6$ ) within a genomic region (including ANNOVAR functional consequence, CADD score, RegulomeDB score, as well as 15 chromatin states, and GWAS Catalog associations). LD is derived from the built-in "UKB/release2b EUR\_10k" reference panel. GWAS loci were defined in FUMA<sup>18</sup> by merging LD blocks within 250 kb of one another. We compared the GWS loci to loci reported by biobank-based GWAS of sleep traits published by June 2022, including insomnia<sup>8,9</sup>, sleep duration<sup>7</sup>, daytime sleepiness<sup>10</sup>, chronotype<sup>12</sup>, snoring<sup>11</sup>, daytime napping<sup>64</sup>, obstructive sleep apnea (OSA)<sup>65</sup>, restless legs syndrome (RLS)<sup>66</sup>, and prior SHS-ADD in UKB unrelated individuals<sup>13</sup>. GWS loci with a lead variant at least 500 kb from any of the previously published GWS sleep variants were noted as unreported. To account for performing six simultaneous GWAS, we report significant novel loci defined as the unreported loci passing a stricter significant threshold correcting for six traits ( $p < 8.3e-9$ ).

The constructed SHS traits, as sums of ordinal and integer-valued variables, were somewhat non-normal, which has the potential to affect Type-I error. Consistent with prior UKB-GWAS of ordinal sleep phenotypes, conducting GWAS using linear mixed models implemented in BOLT is expected to mostly ameliorate this concern. BOLT has previously been shown to preserve Type-I error adequately for non-normal ordinal and binary phenotypes, as long as the sample sizes in different groups not markedly imbalanced<sup>67</sup>.

### Gene mapping and gene-based analysis

We used FUMA to map SNPs with  $r^2 \geq 0.6$  with the lead-independent SNP in each locus to genes using three methods: positional mapping ( $\leq 10$  kb), cis-eQTL ( $\leq 1$  Mb) in GTEx v8 tissues (FDR  $< 0.05$ ), and 3D Chromatin interaction (CI) in 127 tissue/cell types (FDR  $< 1e-6$ ). We also looked up ensemble phenotypes using R biomaRt package and drug interaction evidence using DGIdb for mapped genes. We next performed a supplementary gene-based association analysis using genome-wide summary statistics using MAGMA<sup>20</sup> in FUMA. Input SNPs were mapped to 18,931 protein-coding genes.

Genome-wide significance level was defined as  $p < 0.05/18,931 = 2.641e-6$ . Gene-level association results were used to aid prioritization of genes at loci identified as significant in the primary SNP-level multi-GWAS analysis.

### Gene set enrichment analysis

We performed pathway enrichment analysis using PASCAL, which estimated a combined association *p*-value from the summary statistics of multiple SNPs in a gene<sup>43</sup>. Significant KEGG, Reactome, and BIOCARTA pathways for each SHS trait were identified using empirical  $p < 0.05$ . We also performed gene set enrichment analysis on positional, eQTL and CI-mapped genes in FUMA MAGMA<sup>20</sup> adjusted for gene size. Significant tissues were identified using  $p < 0.05/54/6$  accounting for 54 tissues in GTEx v8 and six traits. Significant pathways (KEGG, Reactome, and GO pathways in MSigDB), and GWAS gene set (GWAS Catalog) were identified using adjusted  $p < 0.05$ . PASCAL pathway enrichment incorporates effect sizes and LD at individual loci, whereas FUMA MAGMA gene set enrichment is based only on the overlap in sets of associated genes and does not account for LD, impacting interpretation.

### Colocalization analysis

We performed colocalization analysis across SHS traits using HyPrColoc<sup>68</sup> and genome-wide summary statistics to assess the shared genetic risk factors. Given the independence of the SHS-PCs, we expected less colocalization and for colocalized loci to reflect pleiotropic effect and play a central regulatory role.

### Sensitivity analyses

Sensitivity analyses of all GWS loci were performed additionally adjusting for potential confounders (including adiposity, socio-economic status, alcohol intake frequency, smoking status, caffeine intake, employment status, marital status, and psychiatric problems) individually in 308,902 unrelated individuals using PLINK in addition to adjusting for age, sex, genotyping array and 10 PCs in PLINK 1.9. We used a hard-call genotype threshold of 0.1, SNP imputation quality threshold of 0.80, and a MAF threshold of 0.001. We also performed the analysis excluding shift workers and individuals with chronic health or psychiatric illnesses ( $N = 115,297$ ) and in males ( $N = 145,196$ ) and females ( $N = 163,716$ ) (without adjusting for sex).

### Genetic risk score analysis

We constructed a weighted Genetic Risk Score (GRS) comprised of GWS loci for each SHS and tested for associations with other self-reported sleep traits (sleep duration, long sleep duration, short sleep duration, insomnia, chronotype, and snoring), and 7-day accelerometry traits in the UK Biobank. Association testing was conducted in a sample overlapping with sample which was used for the SHS GWAS, consisting of the available unrelated complete-case subset of UKB for each tested outcome. Weighted GRS analyses were performed by summing the products or risk allele count multiplied by the effect estimate reported in the SHS GWASs using R package gtx (<http://cran.nexr.com/web/packages/gtx/gtx.pdf>). We also tested the GRSs of reported loci for sleep traits using the same approach.

### Genetic correlation analysis

We estimated genetic correlations among SHS and with other self-reported and accelerometry sleep traits using LDSC<sup>44</sup> and genome-wide SNPs mapped to the HapMap3 reference panel. To understand the genetic overlap with a range of common health problems, we selected 381 representative UKB traits by choosing the 232 most heritable traits using UKB hierarchical phenotype categories (selecting at most one 'level' per phenotype, at most 5 phenotypes per phenotype category, and at most 25 phenotypes per category group), as well as all 195 (partly overlapping) traits in the PanUKBB maximally independent set of phenotypes (<https://pan.ukbb.broadinstitute.org/blog/2022/04/11/h2-qc-updated-sumstats/index.html>). Of the 381 selected traits, 375 passed heritability thresholds and were carried on for LDSC analyses. Traits passed heritability QC if stratified LD score regression in the UKB Europeans defined in the Pan UKBB GWAS (sldsc\_25bin\_h2\_pval\_EUR) was significant with  $p < 0.05/381$ . Multiple-testing-corrected significance level for genetic correlation was defined as  $p < 0.05/375/6 = 2.2e-5$ .

## Mendelian randomization analysis

Bidirectional mendelian randomization (MR) analyses, as implemented in the *TwoSampleMR* R package<sup>69</sup>, were conducted to investigate potentially causal links between sleep health and 50 representative traits from across the phenome, selected (prior to performing MR) based on their relationships with sleep traits, either in prior literature, or based on results of interest from the genetic correlation analysis. The final list of traits (below) was also determined by manual review of availability of non-UKB GWAS summary statistics in the IEU open GWAS project database (non-overlapping samples required for valid inference). We report associations that were significant correcting for testing 50 phenotypes and 6 sleep health traits ( $p < 0.05/300$ ), under 2-sample inverse variance weighted (IVW) methodology, while also requiring effects estimated under MR-Egger to be consistent in direction, and that the putative causal direction was not invalidated in a Steiger directionality test, such that the selected instruments for the exposure had a greater  $R^2$  variance explained in the exposure phenotype than the outcome phenotype (neither the MR-Egger nor Steiger tests were required to be statistically significant). SNP heterogeneity was evaluated by Cochran's  $Q$  (for IVW)<sup>70</sup> and Rucker's  $Q'$  (for MR-Egger)<sup>71</sup> statistics. Horizontal pleiotropy was assessed by MR-Egger intercept.

The following GWAS traits were selected for MR from the IEU open GWAS project database: finn-b-R18\_COUGH: Cough, finn-b-RX\_CODEINE\_TRAMADOL: Codeine or tramadol medication, ieu-a-73: Waist-to-hip ratio, ieu-b-103: HbA1C, finn-b-F5\_NEUROTIC: Neurotic, stress-related and somatoform disorders, ieu-a-832: Rheumatoid arthritis, ieu-a-962: Ever vs never smoked, finn-b-G6\_PARKINSON\_EXMORE: Parkinson's disease (more controls excluded), ieu-a-44: Asthma, ieu-b-18: multiple sclerosis, ieu-a-89: Height, finn-b-KRA\_PSY\_ANXIETY: Anxiety disorders, ebi-a-GCST002216: Triglycerides, finn-b-R18\_PAIN\_THROAT\_CHEST: Pain in throat and chest, ieu-b-2: Alzheimer's disease, finn-b-ANTIDEPRESSANTS: Depression medications, finn-b-Z21\_PERSONS\_W\_POTEN\_HEALTH\_HAZARDS\_RELATED\_SOCIO\_PSYCHOSO\_CIRCUMSTANC: Persons with potential health hazards related to socioeconomic and psychosocial circumstances, finn-b-K11\_DIVERTIC: Diverticular disease of intestine, ieu-a-113: Neo-agreeableness, ieu-a-115: Neo-extraversion, ieu-a-1009: Subjective well-being, finn-b-PAIN: Pain (limb, back, neck, head abdominally), finn-b-ALCOHOL\_RELATED: Alcohol related diseases and deaths, all endpoints, ieu-b-4855: FEV1/FVC, ieu-a-294: Inflammatory bowel disease, ieu-a-16: Childhood intelligence, finn-b-F5\_SUBSNOALCO: Substance use, excluding alcohol, finn-b-F5\_PANIC: Panic disorder, ieu-a-114: Neo-conscientiousness, ieu-b-4859: Physical activity, ieu-a-116: Neo-neuroticism, finn-b-F5\_GAD: Generalized anxiety disorder, finn-b-I9\_IHD: Ischemic heart disease, wide definition, finn-b-F5\_DEPRESSIO: Depression, ieu-b-4820: Age at first birth, ebi-a-GCST002223: HDL cholesterol, ieu-a-1001: Years of schooling, finn-b-RX\_PARACETAMOL\_NSAID: Paracetamol or NSAID medication, ebi-a-GCST003116: Coronary artery disease, ieu-b-38: systolic blood pressure, ieu-b-41: bipolar disorder, finn-b-M13\_ENTESOPATHYLOW: Enthesopathies of lower limb, excluding foot, ieu-b-73: Alcoholic drinks per week, finn-b-K11\_REFLUX: Gastro-esophageal reflux disease, finn-b-E4\_DIABETES: Diabetes mellitus, ieu-a-1095: Age at menarche, ieu-a-835: Body mass index, ieu-a-117: Neo-openness to experience, ieu-b-39: diastolic blood pressure, ebi-a-GCST002222: LDL cholesterol

## Replication analysis

HCHS/SOL is a community-based study in the USA, which includes 16,415 adults aged 18–74 with self-identified Hispanic/Latino background<sup>72,73</sup>. Individuals were recruited from randomly selected households near four centers in Miami, San Diego, Chicago and the Bronx area of New York. Self-reported sleep duration, insomnia (assessed by Women's Health Initiative Insomnia Rating Scale [WHIIRS]), daytime sleepiness (assessed by Epworth Sleepiness Scale [ESS]), and snoring were collected in 13,268 individuals at baseline. Chronotype was only available in 1855 individuals enrolled in the Sueño sleep ancillary study.

We converted the sleep data in HCHS/SOL to UKB scale. For insomnia, we used two questions that were part of the WHIIRS questionnaire in the HCHS/SOL: (1) "Did you have trouble falling asleep?" (2) "Did you wake up several times at night?". Each question provided 5 choices: 1. No, not in the past four weeks; 2. Yes, less than once a week; 3. Yes, 1 or 2 times a week; 4. Yes, 3 or 4 times a week; 5. Yes, 5 or more times a week. We converted the sum score of the two questions (2–10) to UKB scale as: 2–4 = "Never/rarely"; 5–8 = "Sometimes"; 9–10 = "Usually". For sleepiness, we converted the ESS score (0–24) to UKB scale as: 0–10 = "Never/Rarely"; 11–14 = "Sometimes"; 15–18 = "Often"; 19–24 = "All the time". For snoring, we converted the 4-level answers for the question "How often do you snore now?" to UKB scale as: 0 if the answer was "Never" or "Rarely" and 1 if the answer was "Sometimes" or "Always". Chronotype was collected in Sueño using the same questionnaire to UKB. We performed multiple imputation to calculate and impute the missing chronotype in the rest of the samples in HCHS/SOL ( $N = 11,413$ ) using chained equations method with linear regression on relevant variables, specifically, sex, age, BMI, and four sleep timing questions "What time do you usually go to bed in the weekday?", "What time do you usually go to bed in the weekend?", "What time do you usually wake up in the weekday?", and "What time do you usually wake up in the weekend?". We found the post-imputation distribution of the chronotype responses matched those observed in Sueño. We then constructed SHS in HCHS/SOL using the same loadings from UKB.

Of the 13,268 individuals with imputed phenotype data, 11,144 individuals with genotype data and consented to genetic research are available for replication. Genotyping was conducted using an Illumina Omni2.5 M SNP array with additional customized content, including 2,536,661 SNPs and imputed to TOPMed reference panel using TOPMed imputation server. Genetic association analysis for the 400 GWS loci were performed using linear mixed model in R Genesis software adjusting for age, sex, study center, sampling weights, five principal components of genetics representing ancestry, with random effects corresponding to kinship, household, and block unit.

We construct Polygenic risk score (PRS) for each SHS trait in the HCHS/SOL data using Polygenic Risk Score–Continuous Shrinkage (PRS-CS)<sup>74</sup> with the UKB SHS genome-wide summary statistics and the 1000 Genome European linkage disequilibrium (LD) reference panel. PLINK was used to compute the PRS scores by summing the product of the effect allele count and the effect size across all SNPs in the PRS.

## Validation of SHS PRS on clinical phenotypes in a clinical biobank

The Mass General Brigham (MGB) Biobank is a clinical biobank enriched for disease states supplemented with genetic data from the MGB healthcare network in Massachusetts. Since 2009, patients have been recruited through online channels or in person from various MGB community-based primary care facilities and specialty tertiary care centers. Among the enrolled patients, a subset ( $n = 64,639$ ) provided blood samples for genotyping. DNA extracted from samples was genotyped using the Infinium Global Screening Array-24 version 2.0 (Illumina). Imputation was carried out through the Michigan Imputation server with the Trans-Omics for Precision Medicine (TOPMed) (version r2) reference panel, and haplotype phasing was performed using Eagle version 2.3. Low-quality genetic markers and samples were excluded<sup>75</sup>. Pairs of related individuals (kinship  $> 0.0625$ ) were identified, and one sample from each related pair was excluded<sup>75</sup>. To correct for the population substructure, principal components of ancestry were computed using TRACE and the Human Genome Diversity Project<sup>76,77</sup>. Individuals with non-European ancestry were excluded to limit genetic heterogeneity in the present analysis.

Among 47,082 adult patients included in the present analysis (mean age =  $60.4 \pm 17.0$ ; 53.8% female), PRS for each SHS trait were generated using PRS-CS<sup>74</sup>. Case ascertainment for sleep disorders was based on clinical phenotypes identified from ICD-9/-10 billing codes and mapped to PheWAS codes (i.e., "phecodes") based on clinical similarity generated by the PheWAS R package<sup>78</sup>. A total of 13 sleep disorders were considered in the analysis. For each disorder, participants with at least two codes were set as

cases, and those with no relevant codes were set as controls. Associations between the PRSs and each disorder were tested via the Wald test using logistic regressions adjusted for age, sex, genotyping array, batch, and PCs of ancestry. Significance was determined using Bonferroni-adjusted  $P$  values for the total number of tests ( $0.05/(6 \text{ PRSs} \times 13 \text{ phenotypes}) = 6.4e-4$ ).

### Comparison with composite sleep health scores constructed using genetic correlations in addition to phenotypic correlations

We applied the MaxH method<sup>45</sup> to our phenotypic and genetic data to construct linear combinations of the self-reported sleep phenotypes that maximize heritability of the composite, subject to an independence constraint. From BOLT-REML analyses conducted with the regression model corresponding to the GWAS described above, we obtained the residual phenotypic covariance in the analytic sample for sleep duration, insomnia, snoring, sleepiness, and chronotype, as well as the corresponding genetic covariance. The heritability-optimized MaxH phenotype loadings were derived from eigenvectors of the Rayleigh quotient matrix, following the MaxH method. Calculations were done in base R, from BOLT-REML outputs.

### Statistics and reproducibility

Statistic used are as described above.  $P$ -values calculated from the described regression models are based on Wald tests of the appropriate parameter, unless otherwise specified. Reproducibility was maintained by use of versioned scripts and package management systems, such as the R package *renv*.

### Data availability

Summary GWAS statistics are publicly available at the Sleep Disorder Knowledge Portal (<http://sleepdisordergenetics.org/>).

### Code availability

The code used in this study is customized from standard genetic software (listed below) and is available for sharing upon reasonable request. BOLT-LMM, FUMA, HyPrColoc (<https://github.com/cnfoley/hyprcoloc>), LDSC, PASCAL, PLINK 1.9 (<https://www.cog-genomics.org/plink/>), PRS-CS (<https://github.com/getian107/PRSs>), R package bioMart (<https://bioconductor.org/packages/release/bioc/html/biomaRt.html>), R package GENESIS, R package PheWAS (<https://github.com/PheWAS/PheWAS>), R package TwoSampleMR (<https://mrcieu.github.io/TwoSampleMR>).

Received: 11 March 2024; Accepted: 10 January 2025;

Published online: 24 January 2025

### References

1. Buysse, D. J. Sleep health: can we define it? Does it matter? *Sleep* **37**, 9–17 (2014).
2. Jan, M., O'Hara, B. F. & Franken, P. Recent advances in understanding the genetics of sleep. *F1000Res* **9**, F1000 Faculty Rev-214 (2020).
3. Lane, J. M. et al. Genetics of circadian rhythms and sleep in human health and disease. *Nat. Rev. Genet.* **24**, 4–20 (2023).
4. Lim, D. C. et al. The need to promote sleep health in public health agendas across the globe. *Lancet Public Health* **8**, e820–e826 (2023).
5. 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).
6. Makarem, N. et al. Multidimensional sleep health is associated with cardiovascular disease prevalence and cardiometabolic health in US adults. *Int. J. Environ. Res. Public Health* **19**, 10749 (2022).
7. 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).
8. Lane, J. M. et al. Biological and clinical insights from genetics of insomnia symptoms. *Nat. Genet.* **51**, 387–393 (2019).
9. Jansen, P. R. et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat. Genet.* **51**, 394–403 (2019).
10. 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).
11. 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).
12. 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).
13. Yao, Y. et al. Genome-wide association study and genetic correlation scan provide insights into its genetic architecture of sleep health score in the UK Biobank Cohort. *Nat. Sci. Sleep* **14**, 1–12 (2022).
14. Bonazzola, R. et al. Unsupervised ensemble-based phenotyping enhances discoverability of genes related to left-ventricular morphology. *Nat. Mach. Intell.* **6**, 291–306 (2024).
15. Ried, J. S. et al. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. *Nat. Commun.* **7**, 1–11 (2016).
16. Avery, C. L. et al. A phenomics-based strategy identifies loci on APOC1, BRAP, and PLCG1 associated with metabolic syndrome phenotype domains. *PLoS Genet.* **7**, e1002322 (2011).
17. Harshfield, E. L. et al. An unbiased lipid phenotyping approach to study the genetic determinants of lipids and their association with coronary heart disease risk factors. *J. Proteome Res.* **18**, 2397–2410 (2019).
18. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
19. Henson, S. M. & Akbar, A. N. KLRG1—more than a marker for T cell senescence. *Age* **31**, 285–291 (2009).
20. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
21. Ramirez Reyes, J. M. J., Cuesta, R. & Pause, A. Folliculin: a regulator of transcription through AMPK and mTOR signaling pathways. *Front. Cell Dev. Biol.* **9**, 667311 (2021).
22. Chikahisa, S., Fujiki, N., Kitaoka, K., Shimizu, N. & Sei, H. Central AMPK contributes to sleep homeostasis in mice. *Neuropharmacology* **57**, 369–374 (2009).
23. Hung, H. C., Kay, S. A. & Weber, F. HSP90, a capacitor of behavioral variation. *J. Biol. Rhythms* **24**, 183–192 (2009).
24. Wang, X. et al. Rheostatic balance of circadian rhythm and autophagy in metabolism and disease. *Front. Cell Dev. Biol.* **8**, 616434 (2020).
25. Lu, J. & Matunis, M. J. A mediator methylation mystery: JMJD1C demethylates MDC1 to regulate DNA repair. *Nat. Struct. Mol. Biol.* **20**, 1346–1348 (2013).
26. Saez, M. A. et al. Mutations in JMJD1C are involved in Rett syndrome and intellectual disability. *Genet. Med.* **18**, 378–385 (2016).
27. Bouchard, L. et al. Neuromedin beta: a strong candidate gene linking eating behaviors and susceptibility to obesity. *Am. J. Clin. Nutr.* **80**, 1478–1486 (2004).
28. Chang, L., Soomro, S. H., Zhang, H. & Fu, H. Ankfy1 is involved in the maintenance of cerebellar purkinje cells. *Front. Cell. Neurosci.* **15**, 648801 (2021).
29. Zhang, L. B. et al. Neuronal activity in the cerebellum during the sleep-wakefulness transition in mice. *Neurosci. Bull.* **36**, 919–931 (2020).
30. Wang, X. L. et al. Regulation of circadian genes by the MAPK pathway: implications for rapid antidepressant action. *Neurosci. Bull.* **36**, 66–76 (2020).
31. Koncina, E., Roth, L., Gonthier, B. & Bagnard, D. Role of semaphorins during axon growth and guidance. *Adv. Exp. Med. Biol.* **621**, 50–64 (2007).
32. Li, Y.-Y., Qin, Z.-H. & Sheng, R. The multiple roles of autophagy in neural function and diseases. *Neurosci. Bull.* **40**, 362–382 (2024).
33. Platzer, K. et al. De novo variants in MAPK8IP3 cause intellectual disability with variable brain anomalies. *Am. J. Hum. Genet.* **104**, 203–212 (2019).

34. Umasankar, P. K. et al. Distinct and separable activities of the endocytic clathrin-coat components Fcho1/2 and AP-2 in developmental patterning. *Nat. Cell Biol.* **14**, 488–501 (2012).
35. Calzoni, E. et al. F-BAR domain only protein 1 (FCHO1) deficiency is a novel cause of combined immune deficiency in human subjects. *J. Allergy Clin. Immunol.* **143**, 2317–2321.e12 (2019).
36. Knight, W. E. et al. PDE1C deficiency antagonizes pathological cardiac remodeling and dysfunction. *Proc. Natl Acad. Sci. USA* **113**, E7116–E7125 (2016).
37. Vetrini, F. et al. De novo and inherited TCF20 pathogenic variants are associated with intellectual disability, dysmorphic features, hypotonia, and neurological impairments with similarities to Smith-Magenis syndrome. *Genome Med.* **11**, 12 (2019).
38. Zhou, J. et al. Disruption of MeCP2-TCF20 complex underlies distinct neurodevelopmental disorders. *Proc. Natl Acad. Sci. USA* **119** (2022).
39. Hoshina, N. et al. Protocadherin 17 regulates presynaptic assembly in topographic corticobasal ganglia circuits. *Neuron* **78**, 839–854 (2013).
40. Sharma, V. & Hiller, M. Loss of enzymes in the bile acid synthesis pathway explains differences in bile composition among mammals. *Genome Biol. Evol.* **10**, 3211–3217 (2018).
41. Xing, C. et al. Roles of bile acids signaling in neuromodulation under physiological and pathological conditions. *Cell Biosci.* **13**, 1–18 (2023).
42. Zhang, M. H. et al. Chronic intermittent hypoxia impaired collagen synthesis in mouse genioglossus via ROS accumulation: A transcriptomic analysis. *Respir. Physiol. Neurobiol.* **308**, 103980 (2023).
43. Lamparter, D., Marbach, D., Ruedi, R., Kutalik, Z. & Bergmann, S. Fast and rigorous computation of gene and pathway scores from SNP-based summary statistics. *PLoS Comput. Biol.* **12**, e1004714 (2016).
44. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
45. Zhou, J. J. et al. Integrating multiple correlated phenotypes for genetic association analysis by maximizing heritability. *Hum. Hered.* **79**, 93–104 (2015).
46. Ukita, Y., Okumura, M. & Chihara, T. Ubiquitin proteasome system in circadian rhythm and sleep homeostasis: lessons from Drosophila. *Genes Cells* **27**, 381–391 (2022).
47. Ben Simon, E., Rossi, A., Harvey, A. G. & Walker, M. P. Overanxious and underslept. *Nat. Hum. Behav.* **4**, 100–110 (2020).
48. Staner, L. Sleep and anxiety disorders. *Dialogues in Clinical Neuroscience* **5**, 249–258 (2003).
49. Huang, T. et al. A population-based study of the bidirectional association between obstructive sleep apnea and type 2 diabetes in three prospective US cohorts. *Diab. Care* **41**, 2111–2119 (2018).
50. Andrade da Silva Dantas, D. et al. The extracellular matrix of the lateral pharyngeal wall in obstructive sleep apnea. *Sleep* **35**, 483–490 (2012).
51. Huang, T. et al. C-reactive protein and risk of OSA in four US cohorts. *Chest* **159**, 2439–2448 (2021).
52. Huang, T. et al. Genetic predisposition to elevated C-reactive protein and risk of obstructive sleep apnea. *Am. J. Respir. Crit. Care Med.* **209**, 329–331 (2024).
53. Kajiwara, Y. et al. GJA1 (connexin43) is a key regulator of Alzheimer's disease pathogenesis. *Acta Neuropathol. Commun.* **6**, 1–20 (2018).
54. Sollis, E. et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res.* **51**, D977–D985 (2023).
55. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
56. Zhang, L.-B. et al. Neuronal activity in the cerebellum during the sleep-wakefulness transition in mice. *Neurosci. Bull.* **36**, 919–931 (2020).
57. Cook, J. P., Mahajan, A. & Morris, A. P. Fine-scale population structure in the UK Biobank: implications for genome-wide association studies. *Hum. Mol. Genet.* **29**, 2803–2811 (2020).
58. Ried, J. S. et al. A principal component meta-analysis on multiple anthropometric traits identifies novel loci for body shape. *Nat. Commun.* **7**, 13357 (2016).
59. 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).
60. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
61. Jones, S. E. et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat. Commun.* **10**, 1585 (2019).
62. van Hees, V. T. et al. Estimating sleep parameters using an accelerometer without sleep diary. *Sci. Rep.* **8**, 12975 (2018).
63. Windred, D. P. et al. Objective assessment of sleep regularity in 60 000 UK Biobank participants using an open-source package. *Sleep* **44**, zsab254 (2021).
64. Dashti, H. S. et al. Genetic determinants of daytime napping and effects on cardiometabolic health. *Nat. Commun.* **12**, 900 (2021).
65. Strausz, S. et al. Genetic analysis of obstructive sleep apnoea discovers a strong association with cardiometabolic health. *Eur. Respir. J.* **57**, 2003091 (2021).
66. Didriksen, M. et al. Large genome-wide association study identifies three novel risk variants for restless legs syndrome. *Commun. Biol.* **3**, 703 (2020).
67. Bi, W. et al. Efficient mixed model approach for large-scale genome-wide association studies of ordinal categorical phenotypes. *Am. J. Hum. Genet.* **108**, 825–839 (2021).
68. Foley, C. N. et al. A fast and efficient colocalization algorithm for identifying shared genetic risk factors across multiple traits. *Nat. Commun.* **12**, 764 (2021).
69. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *elife* **7**, e34408 (2018).
70. Davies, N. M., Holmes, M. V. & Davey Smith, G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *BMJ* **362**, k601 (2018).
71. Bowden, J. et al. Improving the visualization, interpretation and analysis of two-sample summary data Mendelian randomization via the Radial plot and Radial regression. *Int. J. Epidemiol.* **47**, 2100 (2018).
72. Sorlie, P. D. et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann. Epidemiol.* **20**, 629–641 (2010).
73. Lavange, L. M. et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann. Epidemiol.* **20**, 642–649 (2010).
74. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
75. Dashti, H. S. et al. Polygenic risk score for obesity and the quality, quantity, and timing of workplace food purchases: A secondary analysis from the ChooseWell 365 randomized trial. *PLoS Med.* **17**, e1003219 (2020).
76. Cann, H. M. et al. A human genome diversity cell line panel. *Science* **296**, 261–262 (2002).
77. Wang, C., Zhan, X., Liang, L., Abecasis, G. R. & Lin, X. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *Am. J. Hum. Genet.* **96**, 926–937 (2015).
78. Denny, J. C., Bastarache, L. & Roden, D. M. Phenome-wide association studies as a tool to advance precision medicine. *Annu. Rev. Genomics Hum. Genet.* **17**, 353–373 (2016).

## Acknowledgements

This research has been conducted using the UK Biobank Resource under application 6818. We would like to thank the participants and researchers from the UK Biobank who contributed or collected data. This work was supported by the National Institute of Health (NIH) grants R01HL153814 (to H.W.), R01HL146751 (to R.S.), R35HL135818 (to S.R.), and NIHR Manchester Biomedical Research Centre. S.D.K. is supported by the National Institute of Health and Care Research (NIHR) Oxford Health

Biomedical Research Centre, NIHR Efficacy and Mechanisms Evaluation Programme (Ref: 131789), NIHR Programme Grants for Applied Research (Ref: 203667) and the Wellcome Trust (226784/Z/22/Z and 227093/Z/23/Z). D.A.L. is supported by the UK Medical Research Council (MC\_UU\_00032/05), British Heart Foundation (CH/F/20/90003 and AA/18/1/34219) and Diabetes UK (17/0005700). J.I.R. is supported by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The Hispanic Community Health Study/Study of Latinos was supported by contracts from the National Heart Lung and Blood Institute to the University of North Carolina (N01-HC65233), University of Miami (N01-HC65234), Albert Einstein College of Medicine (N01-HC65235), the University of Illinois at Chicago (HHSN268201300003I), Northwestern University (N01-HC65236), and San Diego State University (N01-HC65237). The following Institutes/Centers/Offices contributed to the HCHS/SOL through a transfer of funds to the National Heart Lung and Blood Institute: National Center on Minority Health and Health Disparities, the National Institute of Deafness and Other Communications Disorders, the National Institute of Dental and Craniofacial Research, the National Institute of Diabetes and Digestive and Kidney Diseases, the National Institute of Neurological Disorders and Stroke, and the Office of Dietary Supplements.

### Author contributions

M.O.G., T.F., V.P., and H.W. conducted the primary analyses. J.M.L., B.S., M.M., B.E.C., T.S., H.S.D., and R.S. contributed additional analysis scripts and expertise. M.O.G., T.F., V.P., P.N., and H.W. prepared the tables and figures. M.O.G. wrote the initial draft with input from H.W. The study concept and analysis plan were developed by H.W., M.O.G., J.M.L., J.C., V.P., H.S.D., T.S., S.R., and R.S. All remaining authors (S.M.P., X.Z., R.N., A.J.K.P., S.D.K., K.S., M.N.W., D.A.L., J.I.R., K.D.T., C.R.I., and M.K.R.) provided conceptual guidance, scientific critique, and substantive feedback on the manuscript. All authors discussed the results and their implications and revised and commented on the manuscript at multiple stages.

### Competing interests

The authors declare the following competing interests: A.J.K.P. has received research funding from Versalux and Delos, and he is co-founder and co-director of Circadian Health Innovations PTY LTD. S.R. received consulting

fees from Jazz Pharma, Eli Lilly, and ApniMed Inc. Martin Rutter has received consulting fees from Eli Lilly. All other authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-07514-0>.

**Correspondence** and requests for materials should be addressed to Heming Wang.

**Peer review information** *Communications Biology* thanks Thomas Hansen and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: H  l  ne Choquet and Mengtan Xing. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

   The Author(s) 2025

<sup>1</sup>Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA. <sup>2</sup>Department of Neurology and Medicine, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA. <sup>3</sup>Broad Institute, Cambridge, MA, USA. <sup>4</sup>Instituto de Psicolog  a Cl  nica, Facultad de Psicolog  a, Universidad de la Rep  blica, Montevideo, Uruguay. <sup>5</sup>MRC Unit for Lifelong Health & Ageing, Institute of Cardiovascular Science, University College London, London, United Kingdom. <sup>6</sup>Center for Genomic Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. <sup>7</sup>Division of Cardiology, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. <sup>8</sup>Department of Informatics and Health Data Science, Miller School of Medicine, University of Miami, Miami, FL, USA. <sup>9</sup>Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, USA. <sup>10</sup>Department of Population and Quantitative Health Sciences, Case Western Reserve University School of Medicine, Cleveland, OH, USA. <sup>11</sup>Department of Internal Medicine, Section of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands. <sup>12</sup>School of Psychological Sciences, Turner Institute for Brain and Mental Health, Monash University, Melbourne, VIC, Australia. <sup>13</sup>Sir Jules Thorn Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. <sup>14</sup>Department of Psychiatry and Psychotherapy, Medical Centre - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. <sup>15</sup>Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK. <sup>16</sup>MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK. <sup>17</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. <sup>18</sup>The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA. <sup>19</sup>Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA. <sup>20</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. <sup>21</sup>Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA. <sup>22</sup>Division of Endocrinology, Diabetes & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. <sup>23</sup>Diabetes, Endocrinology and Metabolism Centre, Manchester University NHS Foundation Trust, NIHR Manchester Biomedical Research Centre, Manchester Academic Health Science Centre, Manchester, UK.    e-mail: [hwang@bwh.harvard.edu](mailto:hwang@bwh.harvard.edu)