Genome-wide association analysis of self-reported daytime sleepiness identifies 42 loci that suggest biological subtypes

Heming Wang et al.

Excessive daytime sleepiness (EDS) affects 10–20% of the population and is associated with substantial functional deficits. Here, we identify 42 loci for self-reported daytime sleepiness in GWAS of 452,071 individuals from the UK Biobank, with enrichment for genes expressed in brain tissues and in neuronal transmission pathways. We confirm the aggregate effect of a genetic risk score of 42 SNPs on daytime sleepiness in independent Scandinavian cohorts and on other sleep disorders (restless legs syndrome, insomnia) and sleep traits (duration, chronotype, accelerometer-derived sleep efficiency and daytime naps or inactivity). However, individual daytime sleepiness signals vary in their associations with objective short vs long sleep, and with markers of sleep continuity. The 42 sleepiness variants primarily cluster into two predominant composite biological subtypes - sleep propensity and sleep fragmentation. Shared genetic links are also seen with obesity, coronary heart disease, psychiatric diseases, cognitive traits and reproductive ageing.

Correspondence and requests for materials should be addressed to R.S. (email: rsaxena@partners.org). A full list of authors and their affiliations appears at the end of the paper.
Excessive daytime sleepiness (EDS) is a chief symptom of chronic insufficient sleep as well as of several primary sleep disorders, such as sleep apnea, narcolepsy, and circadian rhythm disorders. Several disease processes and medications also associate with prevalent and incident EDS. EDS is estimated to contribute to risk for motor vehicle crashes, work-related accidents, and loss of productivity, highlighting its public health importance. The clinical impact of EDS extends to a negative impact on cognition, behavior, and quality of life. Therefore, sleep interventions often identify reduction in EDS as a chief goal. EDS is also associated with an increased risk for cardio-metabolic disorders, psychiatric problems, and mortality through pathways that may be causal, bi-directional, or reflect pleiotropic effects.

While EDS occurs in a variety of settings associated with insufficient sleep, there is large inter-individual variability in levels of EDS that is not fully explained by sleep duration, sleep quality, or chronic disease. Experimental studies have shown that there is also individual vulnerability to EDS following sleep restriction. The heritability of daytime sleepiness is estimated to be between 0.37 and 0.48 in twin studies and 0.17 in family studies, and between 0.084 and 0.17 in GWAS, suggesting that genetic factors contribute to variation in sleepiness. Despite multiple candidate gene studies and GWAS, few genome-wide significant genetic variants have been reported, likely reflecting the heterogeneous and multifactorial etiology of the phenotype and low statistical power.

Here, we extend our GWAS of self-reported daytime sleepiness to the full UK Biobank dataset and identify multiple genetic variants grouping into different biological subtypes that associate with sleepiness. Bioinformatics analyses further highlight relevant biological processes and reveal shared genetic background with other diseases.

Results

Sample characteristics. In the UK Biobank, 452,071 participants of European genetic ancestry self-reported the frequency of daytime sleepiness using the question: “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), with the answer categories “never” (N = 347,285), “sometimes” (N = 92,794), “often” (N = 11,963), or “all of the time” (N = 29). The severity of daytime sleepiness increased with older age, female sex, higher body mass index (BMI), various behavioral, social and environmental factors, and chronic diseases (Supplementary Table 1). Self-reported daytime sleepiness was positively but weakly correlated with self-reported insomnia symptoms, morning chronotype, ICD-10, or self-report of physician diagnosed sleep apnea and self-reported shorter and longer sleep duration, consistent with earlier reports or known clinical correlates (Spearman correlation <0.2; Supplementary Table 1 and Supplementary Fig. 1). Self-reported daytime sleepiness was also weakly correlated with shorter sleep duration, lower sleep efficiency (indicating more time awake during the sleep period), and longer daytime inactivity duration estimated using a 7-day accelerometer in a subset (N = 85,388) of UK Biobank participants (Methods; Supplementary Table 2).

GWAS, sensitivity, and replication analyses. We performed a GWAS of self-reported daytime sleepiness treating the four categories as a continuous variable using a linear mixed regression model adjusted for age, sex, genotyping array, ten principal components (PCs) of ancestry and genetic relatedness matrix, and identified 37 genome-wide significant loci (P < 5 x 10^-8) (Fig. 1, Supplementary Fig. 2, Supplementary Table 3). The most significant association was observed within the gene KSR2, a gene associated with multiple physiological pathways relevant to sleep and metabolism (see Discussion). Additional novel loci were identified within or near genes with known actions on sleep–wake control regulation or that are associated with sleep disorders (e.g. PCL1 (ref. 26), GABRA2 (ref. 27), BTBD9 (ref. 28), HTR7 (ref. 29), RA1 (ref. 30)), metabolic traits (e.g. GCKR3, SLC39A8 (ref. 32)), and psychiatric traits (e.g. AGAPI (ref. 33), CACNA1C). Regional association plots of genome-wide significant loci are shown in Supplementary Fig. 3. We identified 37 association signals driven by common lead variants with minor allele frequency 0.08–0.49. The previously identified rare variant signals for daytime sleepiness at AROPH1N (MAF = 0.002), ROBO1 (MAF = 0.003), and TMEM132B (MAF = 0.004) in the first release of the UK Biobank (N = 111,975) were not significantly associated with sleepiness in this study (P = 0.006–0.03; Supplementary Table 4). The lack of consistency across these analyses may relate to initial false-positive signals at rare variants (MAF = 0.001–0.005) and/or by selection bias in the initial sample in which heavy smokers had been oversampled. However, two genome-wide significant loci, HCRTR2 and PATJ, overlapped with those identified for a composite sleep trait in the interim release sample and a suggestive sleepiness signal at CEPE1 was replicated. No association was seen with single-nucleotide polymorphisms (SNPs) reported in smaller independent GWAS of EDS, hypersomnia, or narcolepsy (Supplementary Table 4).

Previous longitudinal research indicated obesity and weight gain are associated with incidence of daytime sleepiness; therefore, we performed an additional GWAS adjusting for BMI to identify loci that may operate in obesity-independent pathways. This analysis identified five additional loci (Supplementary Figs. 4 and 5; Supplementary Table 3). Effect estimates at the 37 loci identified in the primary model were largely unchanged.

Sensitivity analyses on autosomes additionally adjusted for potential confounders (including depression, socio-economic status, alcohol intake frequency, smoking status, caffeine intake, employment status, marital status, neurodegenerative disorders, and psychiatric problems) and stratified by obesity and sleep duration did not substantially alter effect estimates of the identified signals (Supplementary Data 1; Supplementary Tables 5 and 6). Secondary GWAS (N = 255,426), excluding shiftworkers and individuals with chronic health or psychiatric illnesses, additionally identified significant variants in SEMA3D and revealed marginally significant interactions with health status at PATJ, ZNF326/BARHL2, ECE2, ASAP1, and CYP1A1/CYP1A2 (interaction P < 0.05; Supplementary Fig. 6, Supplementary Table 7). Conditional analyses at each locus identified no secondary signals. Sex-stratified analyses on autosomes additionally identified CWC27 and DIAFPH3 in women but not in men; however, significant gene by sex interactions were not observed (Supplementary Fig. 7, Supplementary Table 8).

Replication was attempted using self-reported daytime sleepiness indices (based on related questions) available in European whites from HUNT (N = 29,906; Supplementary Table 9) and Health 2000 studies (N = 4546; Supplementary Table 10) (Methods). Five individual signals, including KSR2, SUSD4, and CYP1A1/CYP1A2 were marginally significant (P < 0.05) and with consistent association direction in individual cohorts and/or meta-analysis (Supplementary Table 11). A genetic risk score (GRS) of 42 sleepiness loci weighted by the effect estimates from our primary daytime sleepiness GWAS was replicated in a meta-analysis of HUNT, and Health 2000 (Fisher’s P = 0.0003; Supplementary Table 11), and this remained significant after removing the three marginally associated loci from the meta-
ARTICLE

Fig. 1 Manhattan plot for genome-wide association analysis of self-reported daytime sleepiness. Dotted line indicates genome-wide significance. Genetic association signals are highlighted in green and annotated with the nearest genes.

analysis (Fisher’s \( P = 0.017 \)). Marginal association with a tiredness phenotype was observed for two signals in a meta-analysis of FINRISK\(^38\) (\( N = 20,344 \); Supplementary Table 12) and Finnish Twin\(^39\) (\( N = 5766 \); Supplementary Table 13); however, a combined GRS was not significant (Fisher’s \( P = 0.551 \); Supplementary Table 14).

We further validated our results through associations with subjective and objective measures of sleep patterns and disorders available in the UK Biobank. A GRS of 42 variants was associated with self-reported shorter sleep duration, morning chronotype, increased insomnia symptoms, increased frequency of daytime napping, and with accelerometer-assessed lower sleep efficiency and increased duration of daytime inactivity (Table 1).

Clustering of sleepiness loci suggest biological subtypes. Genetic variants may influence daytime sleepiness through different mechanisms. Therefore, to dissect this heterogeneity, we investigated associations of individual SNPs with other sleep traits (Supplementary Data 2).

Individual daytime sleepiness increasing alleles at \( PATJ \) and \( PLCL1 \) were also associated with morning chronotype; loci at metabolism regulatory genes \( KSR2 \), \( LOC644191/CRHR1 \), and \( SLC39A8 \) with self-reported sleep duration (\( KSR2 \) with increased sleep duration, \( LOC644191/CRHR1 \) with long sleep, and \( SLC39A8 \) with short sleep); \( LMOD1 \) and \( LOC644454/LOC730134 \) with both insomnia and short sleep duration; and at the orexin/hypocretin receptor \( HCRTR2 \) with both morning chronotype and short sleep duration, suggesting common genetic factors. Adjusting for sleep disturbance traits (ICD-10 code defined sleep apnea or narcolepsy, or self-reported sleep duration hours, frequent insomnia symptoms or chronotype) together attenuated effect estimates for several loci, suggesting that these genetic variants influence sleepiness through altered sleep patterns and sleep disorders; however, adjustment for any trait alone only minimally altered effect estimates at individual loci (Supplementary Table 1).

Using 7-day accelerometer-derived data available in a subset of the UK Biobank (\( N = 85,388 \)), we observed associations of several daytime sleepiness alleles with reduced sleep efficiency (e.g., \( SNX17 \)) whereas others were associated with increased sleep efficiency (e.g., \( PLCL1 \)), suggesting that genetic mechanisms may lead to sleepiness through effects on increased sleep fragmentation (i.e., low sleep efficiency) or increased sleep propensity (i.e., high sleep efficiency), respectively (Supplementary Data 2). Therefore, we performed hierarchical clustering analyses on risk alleles for sleepiness at 42 loci according to their association effect sizes (\( z \)-scores) with objective estimates of sleep efficiency, sleep duration, and number of sleep bouts, and self-reported frequent insomnia symptoms. An iterative approach based on silhouette coefficients was performed to remove cluster outliers (Methods; Supplementary Fig. 8). We interpreted sleepiness alleles showing patterns of association with higher sleep efficiency, longer sleep duration, fewer discrete sleep bouts and fewer insomnia symptoms as reflective of greater intrinsic sleep propensity, whereas sleepiness alleles associated with these sleep traits in a largely inverse manner were interpreted as reflective of disturbed sleep or a sleep fragmentation subtype resulting in less restorative sleep (Fig. 2). GRS of daytime sleepiness loci stratified by the two clusters support our interpretation, with sleep propensity loci showing robust associations with early circadian traits (e.g. morning chronotype \( P = 5.54 \times 10^{-4} \); Table 1).

Functional effects of loci. Sleepiness loci lie in genomic regions encompassing 164 genes (Supplementary Data 3), and 3 associations are in strong linkage disequilibrium with known GWAS associations for other traits, including blood cell count, high-density lipoprotein cholesterol, and caffeine metabolism. Genes at multiple loci have been implicated in Mendelian syndromes or in experimental studies in mouse or fly models. Eighteen locus harbor one or more genes with potential drug targets.

We performed fine mapping analyses for potential causal variants using PICS\(^40\) and identified 33 variants within 25 sleepiness loci with a causal probability larger than 0.2 (Supplementary Data 4). The majority of likely causal variants were intronic (65%) followed by non-coding transcript variants (8%) and nonsense-mediated decay transcript variants (7%) (Supplementary Fig. 9). Functional variants included a missense variant rs12140153 within \( PATJ \), a synonymous variant rs11078398 within RA11, regulatory variants rs10800796 in the promoter region of \( LMOD1 \), and rs239323 in a CTCF-binding site in the gene \( POM121L2 \). Using the Oxford Brain Imaging Genetics (BIG) server\(^41\), we further observed the pleiotropic locus at rs13135092 (\( SLC39A8 \)) associated with blood lipids, height, schizophrenia, and other traits\(^42\) to be significantly associated with bilateral putamen and striatum volume in the UK Biobank (\( P < 2.8 \times 10^{-7} \); \( N = 9,707 \); Supplementary Fig. 10). This could be of particular interest given the importance of these central brain centers in influencing motor and emotional behaviors, and emerging data implicating these centers in the integration of behavioral inputs that modulate arousal and sleep-wake states\(^43,44\).

Gene-based, pathway, and tissue enrichment analyses. Gene-based analyses using PASCAL\(^45\) identified 94 genes associated with self-reported daytime sleepiness (enrichment \( P < 2.29 \times 10^{-6} \); Supplementary Table 15), of which 61 overlapped with...
Table 1 Association of weighted genetic risk score (GRS) of all 42 daytime sleepiness loci, 10 sleep propensity loci, and 27 sleep fragmentation loci with (a) multiple self-reported sleep traits and (b) 7-day accelerometry-derived sleep, circadian, and activity traits in the UK Biobank

| Trait | All 42 sleepiness loci | | | 10 sleep propensity loci | | | 27 sleep fragmentation loci | |
|-------|------------------------|---|---|------------------------|---|---|------------------------|---|
| | Trait Beta (SE)/OR [95% CI] per GRS effect | P value | Trait Beta (SE)/OR [95% CI] per GRS effect | P value | Trait Beta (SE)/OR [95% CI] per GRS effect | P value |
| **UK Biobank self-reported sleep traits** | | | | | | | |
| Sleep duration (hours), n = 446,118 | −0.164 (0.053) | 0.002* | 0.642 (0.106) | 1.24 × 10⁻⁹* | −0.586 (0.067) | 3.3 × 10⁻¹⁸* |
| Long sleep duration (>8 hours), n = 34,184 cases and 305,742 controls | 1.048 [1.014,1.084] | 0.006 | 1.137 [1.064,1.216] | 1.46 × 10⁻⁴* | 0.991 [0.951,1.034] | 0.689 |
| Short sleep duration (<7 hours), n = 106,192 cases and 305,742 controls | 1.154 [1.105,1.205] | 1.44 × 10⁻¹⁰* | 0.887 [0.812,0.968] | 0.007 | 1.306 [1.235,1.382] | 7.352 × 10⁻²²* |
| Frequent insomnia symptomsa (usually), n = 75,508 cases and 64,403 controls | 1.206 [1.132,1.284] | 6.47 × 10⁻⁹* | 0.765 [0.674,0.868] | 3.424 × 10⁻⁹* | 1.491 [1.376,1.617] | 2.5 × 10⁻²²* |
| Morning chronotype (4-level continuous variable), n = 697,828 | 0.364 (0.062) | 5.47 × 10⁻⁹* | 0.433 (0.125) | 5.54 × 10⁻⁴* | 0.241 (0.08) | 0.003* |
| Day naps (frequency), n = 450,918 | 0.974 (0.028) | 1.85 × 10⁻²⁶* | 1.198 (0.057) | 3.23 × 10⁻⁹⁹* | 0.813 (0.036) | 2.91 × 10⁻¹¹²* |
| **UK Biobank 7-day accelerometry, n = 85,388** | | | | | | | |
| Sleep durationa (minutes) | −0.150 (0.099) | 0.116 | 1.152 (0.19) | 1.34 × 10⁻⁹* | −0.776 (0.121) | 1.37 × 10⁻¹⁰* |
| Sleep duration variability (standard deviation; minutes) | 0.069 (0.068) | 0.305 | 0.134 (0.134) | 0.317 | 0.006 (0.086) | 0.85 |
| Sleep midpoint (minutes) | −0.151 (0.059) | 0.01 | −0.285 (0.117) | 0.035 | −0.041 (0.074) | 0.579 |
| Midpoint of 5-h daily period of minimum activity (L5 timing; minutes) | −0.206 (0.12) | 0.084 | −0.651 (0.237) | 6.13 × 10⁻³* | 0.18 (0.151) | 0.235 |
| Midpoint of 10-h daily period of maximum activity (M10 timing; minutes) | −0.262 (0.136) | 0.054 | −0.295 (0.27) | 0.274 | −0.118 (0.172) | 0.493 |
| Sleep efficiencya (%) sleep in sleep period | −0.028 (0.008) | 6.43 × 10⁻⁴* | 0.065 (0.017) | 7.74 × 10⁻⁵* | −0.067 (0.011) | 1.53 × 10⁻¹⁰* |
| Number of sleep boutsa (n) | −0.105 (0.405) | 0.796 | −4.231 (0.804) | 1.43 × 10⁻⁷* | 1.722 (0.512) | 7.67 × 10⁻⁴* |
| Daytime inactivity duration (minutes) | 0.398 (0.076) | 1.43 × 10⁻⁷* | 0.749 (0.15) | 6.24 × 10⁻⁷* | 0.155 (0.096) | 0.106 |

*P values significant after correction for 14 traits

aSleep traits used to cluster sleep propensity and sleep fragmentation biological subtypes
Daytime sleepiness risk alleles associate predominantly with sleep propensity or sleep fragmentation phenotypes. Each cell shows effect sizes (z-scores) of associations between sleepiness risk alleles (positively associated with self-reported daytime sleepiness) and sleep traits (accelerometry-derived sleep efficiency, sleep duration, number of sleep bouts, and self-reported insomnia symptoms). Blue color indicates positive z-scores and red color indicates negative z-scores. Sleep propensity alleles were defined as more likely associated with higher sleep efficiency, longer sleep duration, fewer sleep bouts, and fewer insomnia symptoms. Sleep fragmentation alleles were defined as more likely associated with lower sleep efficiency, shorter sleep duration, more sleep bouts, and more insomnia symptoms.

Genetic correlation and Mendelian randomization. Consistent with daytime sleepiness being a symptom of several sleep disorders, GRSs of genome-wide significant SNPs for restless legs syndrome\(^48\) \((P = 0.0002)\), insomnia\(^49\) \((P = 4 \times 10^{-7})\), and coffee consumption\(^50\) \((P = 1.87 \times 10^{-12})\) (often used as a sleepiness “counter-measure”) were significantly associated with self-reported daytime sleepiness phenotype (Table 2). Although EDS is a key symptom of narcolepsy, the GRS of narcolepsy\(^51\) was not associated with self-reported daytime sleepiness \((P = 0.126)\), suggesting narcolepsy loci did not explain sleepiness variation in this sample. We could not examine the genetic overlap of sleep apnea loci and sleepiness because few significant loci for sleep...
To control for bias due to pleiotropy, we performed two-sample summary-level Mendelian Randomization (MR) analyses using independent genetic variants from published GWAS summary statistics in LDSC. All genetic correlations in this report can be found in tabular form in Supplementary Data 6.

To evaluate the causal relationship between sleep disorders or other disease traits and daytime sleepiness, we performed two-sample summary-level Mendelian Randomization (MR) analyses using independent genetic variants from published GWAS summary statistics in LDSC. After adjusting for multiple comparisons, significant positive genetic correlations were observed for daytime sleepiness with obesity traits, coronary heart disease, and psychiatric traits \((P < 0.0001)\) (Supplementary Data 6). The genetic correlations with obesity traits, coronary heart disease, and psychiatric traits were also observed (IVW approach (Methods) 34 to control for bias due to pleiotropy, and observed an effect that was consistent with our main IVW analyses but less precisely estimated (wider confidence intervals) because this method is statistically relatively inefficient (MR-Egger \(\beta = 0.025; 95\% \ CI \ [-0.005, 0.055]; P = 0.103; \) Fig. 4 and Supplementary Table 21). An additional suggestive causal association of type 2 diabetes with increased daytime sleepiness was also observed (IVW \(\beta = 0.005; 95\% \ CI \ [0.001, 0.009]; P = 0.014)\) with evidence of heterogeneity (Cochran’s \(Q = 88.38; P = 0.005\), but broadly consistent results when using Radial MR-Egger again showed a consistent effect direction (MR-Egger \(\beta = 0.002; 95\% \ CI \ [-0.006, 0.01]; P = 0.637; \) Supplementary Table 21). Reverse MR did not identify any strong evidence for daytime sleepiness having a causal effect on any of the outcomes we examined (Supplementary Table 22).

Fig. 3 Top significant genetic correlations \(r_g\) between self-reported daytime sleepiness and published summary statistics of independent traits using genome-wide summary statistics using LD score regression (LDSC). Blue color indicates positive genetic correlation and red color indicates negative genetic correlation. Larger colored squares correspond to more significant \(P\) values, and asterisks indicate significant \(P < 2.2 \times 10^{-8}\) genetic correlations after adjusting for multiple comparisons of 224 available traits. All genetic correlations in this report can be found in tabular form in Supplementary Data 6.

To investigate the genetic correlation between sleepiness and other common disorders, we tested the proportion of genetic variation of self-reported daytime sleepiness shared with 233 other traits with published GWAS summary statistics in LDSC 352. After adjusting for multiple comparisons, significant positive genetic correlations were observed for daytime sleepiness with obesity traits, coronary heart disease, and psychiatric traits \((P < 0.0001)\) (Supplementary Data 6). The genetic correlations with coronary artery disease and psychiatric traits persisted after adjusting for BMI (Fig. 3). Consistently suggestive negative genetic correlations for daytime sleepiness with subjective wellbeing and reproductive traits (age at menarche and age at first birth) were also observed \((P < 0.005)\).

To evaluate the causal relationship between sleep disorders or other disease traits and daytime sleepiness, we performed two-sample summary-level Mendelian Randomization (MR) analyses using independent genetic variants from published GWAS summary statistics from GWAS of BMI, type 2 diabetes, coronary heart disease, and psychiatric traits \((P < 0.0004), \) which was significant after accounting for multiple comparisons \((P < 0.003; \) Supplementary Table 20). However, there was evidence of variant heterogeneity potentially due to horizontal pleiotropy (Cochran’s \(Q = 677.17; P = 1.09 \times 10^{-37}; \) Supplementary Table 21). Therefore, we performed sensitivity analysis using the Radial MR-Egger approach (Methods) 34 to control for bias due to pleiotropy, and observed an effect that was consistent with our main IVW analyses but less precisely estimated (wider confidence intervals) because this method is statistically relatively inefficient (MR-Egger \(\beta = 0.025; 95\% \ CI \ [-0.005, 0.055]; P = 0.103; \) Fig. 4 and Supplementary Table 21). An additional suggestive causal association of type 2 diabetes with increased daytime sleepiness was also observed (IVW \(\beta = 0.005; 95\% \ CI \ [0.001, 0.009]; P = 0.014)\) with evidence of heterogeneity (Cochran’s \(Q = 88.38; P = 0.005\), but broadly consistent results when using Radial MR-Egger again showed a consistent effect direction (MR-Egger \(\beta = 0.002; 95\% \ CI \ [-0.006, 0.01]; P = 0.637; \) Supplementary Table 21). Reverse MR did not identify any strong evidence for daytime sleepiness having a causal effect on any of the outcomes we examined (Supplementary Table 22).

**Discussion**

This study expands our knowledge of the genetic architecture of daytime sleepiness. Despite the modest SNP-heritability \(\hat{h}^2 = 6.9\%\), consistent with previous reports 17 18, we identified 42 genome-wide significant loci \((P < 5 \times 10^{-8})\) given boosted statistical power with 452,071 samples. The association effects were largely unchanged adjusting for BMI, depression, socio-economic status, alcohol, smoking, caffeine, employment, neurodegenerative disorders, sleep disturbance traits individually, and upon exclusion of shiftworkers, sleep/psychiatric medication users. We did not evaluate the effect of restless legs syndrome and periodic apnea have been reported in the literature and there was limited sleep apnea information in this cohort.

**Table 2 Association between weighted genetic risk scores (GRS) of significant SNPs \(P < 5 \times 10^{-8}\) for other sleep behavioral traits and sleep disorders with self-reported daytime sleepiness phenotype in UK Biobank**

| Trait | N | nSNP | Beta (SE) per GRS effect | \(P\) value |
|-------|---|------|--------------------------|------------|
| Frequent insomnia symptoms 49 | 237,620 | 57 | 0.0007 (0.0001) | 4.00 \times 10^{-7} |
| Sleep duration (hours) 59 | 446,118 | 78 | -0.0002 (0.0001) | 0.193 |
| Short sleep 59 | 106,192 | 27 | 0.0009 (0.0002) | 1.25 \times 10^{-4} |
| Long sleep 59 | 34,184 | 8 | 0.0009 (0.0005) | 0.068 |
| Day naps | 450,918 | 125 | 0.4078 (0.011) | 6.61 \times 10^{-28} |
| Morning chronotype 67 | 697,828 | 348 | 0.00004 (0.0001) | 0.524 |
| Restless legs syndrome 38 | 110,851 | 20 | 0.0009 (0.0002) | 2.21 \times 10^{-4} |
| Narcolepsy 51 | 25,857 | 8 | 0.0007 (0.0004) | 0.126 |
| Coffee consumption (cups) 50 | 91,462 | 8 | 0.033 (0.005) | 1.87 \times 10^{-12} |

*\(P\) values significant after correction for nine traits. Increasing beta reflects increasing frequency of daytime sleepiness per increase in one risk allele for other sleep trait.
limb movement disorder because information on these disorders were not collected in the UKB.

An aggregate effect of a genetic risk score of 42 loci was confirmed in independent Scandinavian cohorts with different self-reported daytime sleepiness. Despite the challenges of individual loci replication with insufficient power (5–57% in replication cohorts; Supplementary Table 11), variable questionnaires in different scales across different cohorts and the multifactorial etiology of sleepiness, we observed nominal replication at five loci including our most significant association observed at KSR2, a gene regulating multiple signaling pathways (e.g. the ERK/MEK signaling pathway), affecting energy balance, cellular fatty acid, and glucose oxidation that is implicated in obesity, insulin resistance, and heart rate during sleep in previous studies in humans and mice24,25. While the GRS association was highly significant including the three loci with nominal significance in the meta-analysis, the effect estimates removing the three loci remained at 67% of the original effect in HUNT and 86% of the original effect in Health 2000, suggesting that additional individual sleepiness loci contribute to the combined effect of the GRS. However, replication in additional, well-powered cohorts will be important.

The validation of our results was further supported by associations between sleepiness GRS with self-reported shorter sleep duration, morning chronotype, increased insomnia risk, increased frequency of daytime napping, and with accelerometer-derived lower sleep efficiency and increased duration of daytime inactivity. These associations also suggest sleepiness loci impact other sleep parameters such as sleep latency, sleep efficiency, and sleep timing. The sleepiness GRS was not associated with 7-day accelerometer-derived continuous sleep duration, largely reflecting the heterogeneity of self-reported sleepiness (sleep propensity vs sleep fragmentation).

Our results were strengthened by previous GWAS associations with related traits (e.g. metabolic and psychiatric traits), model organism evidence for sleep phenotypes (HCRTR2, SEMA7A, and RAI1), and tissue and pathway enrichment analyses. Genes under association peaks were enriched in multiple brain tissues, including brain regions implicated in sleep–wake and arousal disorders53 as well as centers responsive to sleep deprivation and pathways involved with the central nervous system, neurotransmitters, and metabolic processes. Enrichment of partitioning heritability were observed in variants in central nervous system and highly conserved regions shared by human and other 28 mammals56, suggesting strong conservation of sleep regulation throughout evolution.

We also investigated the heterogeneity of daytime sleepiness loci for the first time by performing clustering analysis according to individual SNP associations with four major sleep parameters: 7-day accelerometer-derived sleep efficiency, sleep duration, and number of sleep bouts, and self-reported frequent insomnia symptoms. We discovered risk sleepiness variants at 10 loci (e.g. PLCL1 and KSR2) and their GRS that associated with sleep propensity traits (higher sleep efficiency, longer sleep duration, fewer discrete sleep bouts, and fewer insomnia symptoms); whereas sleepiness variants at 27 loci (e.g. LMOD1, HCRTR2, and GABRA2, known to play a central role in sleep/wake control and narcolepsy57) and their GRS were more likely to contribute to sleep fragmentation (lower sleep efficiency, shorter sleep duration, more sleep bouts and more insomnia symptoms). Sleep propensity GRS revealed significant associations with early chronotype, reflective of circadian influences on sleep drive. Genes at sleep propensity loci also showed enriched expression in brain tissues whereas no tissues were enriched in sleep fragmentation loci, suggesting that the mechanisms associated with sleep fragmentation may be more complex, reflective of multifactorial influences. Future experimental and statistically robust clustering analysis that include other sleep and related traits are needed to validate and distinguish the biological subtypes of daytime sleepiness58.

We extended our analysis to compare the genetic architecture between daytime sleepiness and other common disorders, and observed significant genetic correlations with obesity, coronary heart disease, and psychiatric traits. The genetic correlations of sleepiness with coronary artery disease and psychiatric traits persisted after adjusting for BMI, perhaps partially reflecting shared neurologic or neuroendocrine factors, such as those that underlay insomnia and short sleep with cardiac and psychiatric traits49,59. Using MR analysis, we identified potential causal association of higher BMI with increased daytime sleepiness, consistent with prospective epidemiological studies, which likely reflect metabolic and/or circadian dysfunction in obese people5. Suggestive causal association of type 2 diabetes and daytime sleepiness were also identified, which may reflect a high prevalence of sleep disturbances in diabetes (e.g., sleep apnea) or systemic inflammation. Reverse MR did not identify any strong
evidence for sleepiness having a causal effect on any of the outcomes we examined, implying that sleepiness is often a "symptom" of other disorders. However, replications with large samples are required to confirm the causal relationships. Future systematic MR analyses with other common disorders may be of particular interest.

This study has several strengths. It is the largest GWAS of self-reported daytime sleepiness with better power than previous studies. We identified 42 loci and confirmed the aggregated association effect in independent cohorts. Moreover, using a wide range of data on individual sleep traits—both self-reported and objectively measured—we showed that individual daytime sleepiness variants associate with unique patterns of sleep and circadian traits that largely cluster into two biological subtypes, intrinsic sleep propensity and sleep fragmentation. These findings were extended with knowledge from published databases of tissue-based expression, pathway annotations, and GWAS summary statistics for other traits.

This study also has several limitations. Primary analyses used self-reported daytime sleepiness expressed as a continuous variable derived from a 4-point scale. Future work should evaluate the psychometric properties of this question and compare it to other frequently used measures of daytime sleepiness, such as the Epworth Sleepiness Scale (ESS) or Maintenance of Wakefulness Test. It is likely that there was some loss of power due to use of a single measure of self-reported sleepiness resulting in random misclassification. However, the large sample for which questionnaire data were available provided results that were able to be further studied in a smaller sample of 7-day accelerometer-derived sleep data (which has been shown to agree well with polysomnography). Future work using objective measurements of sleepiness, such as from vigilance tests, may provide further insights into the genetics of sleepiness-related traits. The statistical power of this study may also be limited by the heterogeneity of daytime sleepiness which could be addressed by adjusting or restricting analyses using the available covariate data. Only individuals of European ancestries aged 40–69 years old in the UK were included, which limits generalizability to other populations and age groups, especially considering that sleep patterns change with age.

In summary, we conducted an extensive series of analysis from a large-scale GWAS and identified the heterogeneous genetic architecture of daytime sleepiness. Multiple genetic loci were identified, including genes expressed in brain areas implicated in sleep–wake control and genes influencing metabolism. Shared genetic factors were identified for daytime sleepiness and other sleep disorders, with evidence that sleepiness variants clustered with two predominant phenotypes—sleep propensity and sleep fragmentation—with the former showing stronger evidence for enrichment in central nervous system tissues, suggesting two unique mechanistic pathways. Genetic variants for daytime sleepiness also overlapped those for other diseases and lifestyle traits, with evidence that higher BMI and possibly diabetes are causally associated with increased daytime sleepiness. This work will advance understanding of biological mechanisms relating to sleepiness and underlying sleep and circadian regulation, and open new avenues for future study.

Methods

Population and study design. The discovery analysis was conducted on participants of European ancestry from the UK Biobank study. 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 and 2010, including self-reported health questionnaire and anthropometric assessments were used in this analysis. Participants taking any self-reported sleep medication (Supplementary Note 1) were excluded. The UK Biobank study was approved by the National Health Service National Research Ethics Service (ref. 11/NW0382), and all participants provided written informed consent to participate in the UK Biobank study. In total, 452,071 individuals of European ancestry were studied with available phenotypes and genotyping passing quality control, as described below.

EDS and covariate measurements. Self-reported daytime sleepiness was ascertained in the UK Biobank using the question "How likely are you to dose off or fall asleep during the daytime when you don’t mean to? (e.g. when working, reading or driving)" with the response options of "Never/rarely", "sometimes", "often", "all of the time", "do not know", and "prefer not to answer". Participants reporting "do not know" and "prefer not to answer" were set to missing. Other responses were coded continuously as 1 to 4 corresponding to the severity of daytime sleepiness. The primary covariates used were self-reported age and sex, and BMI calculated as weight/height². Covariates used in the sensitivity analyses include potential founders (depression, social economic status, alcohol intake frequency, smoking status, caffeine intake, employment status, marital status, neurodegenerative disorders, and use of psychiatric medications) and indices of sleep disorders and sleep traits (daytime napping, sleep apnea, narcolepsy, sleep duration, insomnia, and chronotype). Depression was recorded as a binary variable (yes/no) corresponding to the question "Ever depressed for a whole week?". Social economic status was measured using the Townsend Deprivation Index based on the 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", "former", "never" or "part-time student". Neurodegenerative disorder cases (N = 517) were identified from a validated International Classification of Disease (ICD)-10-based diagnosis generated by the UK Biobank for up to 7 days derived from accelerometer data using this algorithm has been shown to provide robust phenotypes and genotyping passing quality control, as described below.

Activity-monitor-derived measures of sleep. Raw accelerometer data (.cwa) were collected using open source Axivity AX3 wrist-worn triaxial accelerometers (https://github.com/digitalinteraction/openmovement) in 103,711 individuals from the UK Biobank for up to 7 days. These data were converted and aggregated to daily scores using Openomc (https://github.com/digitalinteraction/openmovement/tree/master/Software/AX3/openomc)⁵⁰,⁶¹. Time windows of sleep (SPT-window) and activity levels were extracted for each 24-h period using a heuristic algorithm using the R package GGIR (https://cran.r-project.org/web/packages/GGIR/GGIR.pdf)⁵²,⁵³, typically 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 and inactivity bouts with less than 60 min gaps were categorized as 'transition'. 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. The sleep measurements derived from accelerometer data using this algorithm has been shown to provide reliable estimates for sleep onset time, waking time, SPT-window duration, and sleep duration within the SPT-window⁵⁰,⁵¹. We applied exclusion criteria based on accelerometer data quality including (1) zero or missing in "data problem indicator" (Field 90002); (2) 0 in "good wear time" (Field 90015); (3) 0 in "good calibration" (Field 90016); (4) 0 in "calibrated on own data" (Field 90017); (5) "data recording errors" (Field 90182) >788 (Q + 1.5 x IQR); and (6) recorded recording periods (Field 90180). Accelerometer data from 85,388 participants of European ancestry passed quality control and were analyzed in this study.

The distributions of accelerometer data are described in Supplementary Table 2. The details of each measurement are as follows. L5 and M10 were the least-active 5-h window and most-active 10-h window for each day estimated from a moving average of a contiguous 5/10-h window. The L5 timing was defined as the number of hours elapsed from the previous midnight whereas M10 was defined as the...
number of hours elapsed from the previous midday. Sleep midpoint was the midpoint between the start and end of the SPT-window. LS, M10, and sleep midpoint were calculated as the circadian characteristics of an individual. EPIS episodes within the SPT-window were defined as periods of the z-axis angle change less than 5° for at least 5 min. Sleep duration in an SPT-window was calculated as the sum of all sleep episodes. The mean and standard deviation of sleep duration across all SPT-windows were investigated in this study. Sleep efficiency was calculated as sleep duration divided the total SPT-window duration in an SPT-window. Snoring duration was calculated as the total duration of estimated bouts of inactivity that fell outside of the SPT-window in 24 h, which included both inactivity and naps.

**Genotyping and quality control.** 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). In all, 488,377 samples and 805,426 genotyped markers passed standard QC criteria and were available in the full data release. SNPs were with MAF > 0.001 and genome-wide significance for EnrichR ARTICLE. Genome-wide summary statistics of sleep duration, insomnia, chronotype, long sleep duration, short sleep duration, and 7-day actometry using BOLT-LMM were available in public database [56, 58]. We performed hierarchical cluster analyses using the pairwise Euclidean distances between 42 loci: $d_i = \sqrt{\sum_{j=1}^{42} (x_{ij} - x_{kj})^2}$, where $x_{ij}$ is the $i$th SNP at locus $j$, $x_{kj}$ is the $k$th SNP at locus $j$, and $x_{ik}$ is the $i$th SNP at locus $k$. The detailed description of genotyping, QC, and imputation are available elsewhere [54]. We further performed K-means clustering using the PCs of all ~100,000 high-quality genotyped SNPs (missingness < 1.5% and MAF > 2.5%) and identified 453,964 participants of European ancestry.

**Genome-wide association analysis.** We performed a genome-wide association analysis (GWAS) of self-reported daytime sleepiness as a continuous variable derived from 453,964 participants of European ancestry in the UK Biobank. A linear mixed regression model was applied adjusting for age, sex, genotyping array, and 10 PCs. Genome-wide summary statistics of sleep duration, insomnia, chronotype, long sleep duration, short sleep duration, and 7-day actometry were available in public database [56, 58]. We performed hierarchical cluster analysis using the pairwise Euclidean distances between 42 loci: $d_i = \sqrt{\sum_{j=1}^{42} (x_{ij} - x_{kj})^2}$, where $x_{ij}$ is the $i$th SNP at locus $j$, $x_{kj}$ is the $k$th SNP at locus $j$, and $x_{ik}$ is the $i$th SNP at locus $k$. The detailed description of genotyping, QC, and imputation are available elsewhere [54]. We further performed K-means clustering using the PCs of all ~100,000 high-quality genotyped SNPs (missingness < 1.5% and MAF > 2.5%) and identified 453,964 participants of European ancestry.

**Gene, pathway, and tissue enrichment analyses.** We further examined the genes within genome-wide significant loci using gene-based pathway and tissue enrichment analysis [46, 47]. Gene-based pathway enrichment was performed using PASCAL, which used a combined association $P$ value from the fixed-effect analysis on singlenucleotide polymorphisms in a gene [45]. Pathway and ontology enrichment analyses were performed using FUMA and EnrichR [46]. Tissue enrichment analysis was performed using MAGMA [46] in FUMA, which controlled for gene size. Pathway and tissue enrichment analyses were also performed on genes within loci belonging to sleep propensity and sleep fragmentation clusters separately.

We constructed a weighted GRMs comprising the 42 significant sleepiness loci and tested for associations with other self-reported sleep traits (sleep duration, long sleep duration, short sleep duration, insomnia, chronotype, and day naps), and 7-day actometry traits in the UK Biobank. Weighted GS analyses were performed by summing the products or risk allele count multiplied by the effect estimate reported in the primary GWAS of self-reported daytime sleepiness using R package gds (https://cran.r-project.org/web/packages/gds/gds.pdf). We also tested the GRMs of reported loci for insomnia, sleep duration, short sleep, long sleep, day naps, chronotype, restless legs syndrome (RLS), narcolepsy, and coffee consumption associated with self-reported daytime sleepiness using the same approach. The SNPs selected for each trait include 57 genome-wide significant loci for frequent insomnia [30, 38, 39] and 8 loci for sleep duration, long sleep, and short sleep, respectively [40, 41]. 348 loci for chronotype [42] for day napping; 20 genome-wide significant loci for RLS [43]; and 8 non-HLA suggestive significant loci (P < 0.4) in a narcolepsy case-control study of European Americans [44, 45].

**Genetic correlation analysis.** Genetic correlation analysis using LD Score regression was performed on genome-wide SNPs mapped to the HapMap3 reference panel between daytime sleepiness (with and without adjustment for BMI) and 233 published GWAS available in LHD [25]. The significance level was determined as 0.05 after correcting for multiple comparisons. Pathway enrichment among daytime sleepiness, frequent insomnia, sleep duration, long sleep duration, short sleep duration, and chronotype were performed locally using LDSC. We also partitioned heritability across 8 cell-type regions and 25 functional annotation categories available in LDSC [26]. Enrichment of the partitioning heritability was calculated in each region with and without extension (±500 bp).

**MR analyses.** To investigate the causal relationship between daytime sleepiness and other traits, we performed two-sample MR using MBbase package in R [50]. 1IVW approach, assuming no horizontal pleiotropy effect, was implemented as the primary analysis in this study. BMI, type 2 diabetes, coronary heart disease, psychiatric, reproductive traits, and other sleep and circadian traits (narcolepsy, insomnia, sleep duration, and chronotype) were tested as exposures for daytime sleepiness. Independent genome-wide significant SNPs extracted from publicly available summary statistics of exposures of interest (Supplementary Table 20) were tested as instruments for their effect on daytime sleepiness. The significance level was determined as IVW P < 0.003 after accounting for multiple comparisons. We identified a putative causal association of higher BMI with increased sleepiness risk (IVW β = 0.0189; 95% CI [0.008, 0.029]; P = 0.0004; Supplementary Table 21). The mean F statistic was 32.7, indicating the instruments were sufficiently strong. However, Cochran’s Q statistic was calculated to be 677.16 (P = 1.09 × 10−37), indicating substantial heterogeneity about the IVW slope. This is an indicator of potential horizontal pleiotropy that violates the traditional IV assumptions. Therefore, we applied MR-Egger regression approach, assuming no horizontal pleiotropy effect, to investigate the association between sleepiness risk alleles with BMI and other sleep traits in the UK Biobank (including self-reported sleep duration, insomnia, chronotype, long sleep duration [≥8 h], short sleep duration [<7 h], snoring, obstructive sleep apnea defined by ICD-10 code [G37.3], hypersomnia [defined as sleepiness plus long sleep duration without any chronic or psychiatric diseases], and 7-day actometry data). Linear or logistic regression analyses were performed adjusting for age, sex, genotyping array, and 10 PCs. Genome-wide summary statistics of sleep duration, insomnia, chronotype, long sleep duration, short sleep duration, and 7-day actometry using BOLT-LMM were available in public database [56, 58]. We performed hierarchical cluster analyses using the pairwise Euclidean distances between 42 loci: $d_i = \sqrt{\sum_{j=1}^{42} (x_{ij} - x_{kj})^2}$, where $x_{ij}$ is the $i$th SNP at locus $j$, $x_{kj}$ is the $k$th SNP at locus $j$, and $x_{ik}$ is the $i$th SNP at locus $k$. The detailed description of genotyping, QC, and imputation are available elsewhere [54]. We further performed K-means clustering using the PCs of all ~100,000 high-quality genotyped SNPs (missingness < 1.5% and MAF > 2.5%) and identified 453,964 participants of European ancestry.
Radial MR-Egger (β = 0.025; 95% CI [−0.005, 0.055]; P value = 0.103). Rucker's Q statistic for Radial MR-Egger is 676.58, indicating that the IVW and Radial MR-Egger fit the a priori equally well (Supplementary Table 21). We then investigated the suggestive putative causal association of type 2 diabetes with increased sleepiness risk (IVW mean F = 29; β = 0.005; 95% CI [0.001, 0.009]; P value = 0.014; Supplementary Table 21). Given variants heterogeneity evidence (Cochran’s Q = 88.38; P < 0.005), we performed sensitivity analysis using Radial MR-Egger again (Supplementary Table 21) and observed consistency (95% CI [−0.005, 0.055]; P = 0.637; Supplementary Table 21). Reverse MR between daytime sleepiness and other outcome were conducted using genome-wide significant sleepiness SNPs as instruments, and did not identify any causal association (IVW P > 0.05; Supplementary Table 22).

Replication analyses. Replication analyses were conducted using self-reported day sleepiness or fatigue in Scandinavian individuals from four population-based studies, including Nord-Trøndelag Health Study (HUNT), Health 2000 Survey, FINRISK, and The Finnish Twin Cohort Study.

The HUNT is a large longitudinal population health study, investigating the county of Nord-Trøndelag, Norway since 1984 (ref. 36). Three surveys (HUNT1 [1984–1986], HUNT2 [1995–1997], and HUNT3 [2006–2008]) have been completed including more than 120,000 individuals. Daytime sleepiness phenotype was collected in HUNT3 by asking the question “How often in the last 3 months have you felt sleepy during the day?” with the choices “Never/seldom”, “Sometimes”, and “Several times”. Individuals with self-reported stroke, myocardial infarction, angina pectoris, diabetes mellitus, hypo- and hyperthyroidism, fibromyalgia, and arthritis were excluded from the replication analysis.

DNA samples were collected in 71,860 HUNT samples and genotyped on one of three Illumina arrays: HumanCoreExome12 v1.0, HumanCoreExome12 v1.1, HumanOmniExpress, HumanCoreExome). The data were imputed using the 1000 Genomes project phase 3 haplotypes and a custom haplotype set of 2000 whole genome sequenced HUNT participants. In total, 29,906 individuals with both phenotype and imputed genotype data were available for this analysis. Sample distributions are presented in Supplementary Table 9. A generalized linear mixed model association for EDS performance was performed with RYTESTS v2.0.9 adjusted for age, sex, and the genetic kinship matrix as a random effect controlling for sample relatedness and population structure.

A GRS of all sleepiness loci were also tested in the four cohorts. Meta-analyses of the sleepiness cohorts (HUNT and Health 2000) and the tiredness cohorts (FINRISK and Finnish Twin) were performed using Fisher’s method.

Data Availability
UK Biobank Sleep Traits GWAS summary statistics are available at the Sleep Disorder Knowledge Portal (SDKP) website (http://www.sleepdisordergenetics.org). All other data are contained within the article and its supplementary information or available upon request.

Received: 26 November 2018 Accepted: 27 June 2019

Published online: 13 August 2019

References
1. Cohen, D. A. et al. Uncovering residual effects of chronic sleep loss on human performance. Sci. Transl. Med. 2, 141ra3 (2010).
2. Ohayon, M. M. From wakefulness to excessive sleepiness: what we know and still need to know. Sleep Med. Rev. 12, 129–141 (2008).
3. Slater, G. & Steier, J. Excessive daytime sleepiness in sleep disorders. J. Thorac. Dis. 4, 608–618 (2012).
4. Bisler, E. O. et al. Excessive daytime sleepiness in a general population sample: the role of sleep apnea, age, obesity, diabetes, and depression. J. Clin. Endocrinol. Metab. 90, 4510–4515 (2005).
5. Fernandez-Mendoza, J. et al. Natural history of excessive daytime sleepiness: role of obesity, weight loss, depression, and sleep propensity. Sleep 38, 351–360 (2015).
6. Breidau, N., Roth, T., Rosenthal, L. & Andreski, P. Sleep disturbance and psychiatric disorders: a longitudinal epidemiological study of young adults. Biol. Psychiatry 39, 411–418 (1996).
7. Ferrie, J. E., Kumari, M., Salo, P., Singh-Manoux, A. & Kivimaki, M. Sleep epidemiology—a rapidly growing field. Int. J. Epidemiol. 40, 1431–1437 (2011).
8. Lloberes, P. et al. Self-reported sleepiness while driving as a risk factor for traffic accidents in patients with obstructive sleep apnoea syndrome and in non-apnoeic snorers. Respir. Med. 94, 971–976 (2000).
9. Ohayon, M. M. & Vecchierini, M. F. Daytime sleepiness and cognitive impairment in the elderly population. Arch. Intern. Med. 162, 201–208 (2002).
10. Gangwisch, J. E. et al. Daytime sleepiness and risk of coronary heart disease and stroke: results from the Nurses’ Health Study II. Sleep Med. 15, 782–788 (2014).
11. Van Dongen, H. P., Baynard, M. D., Maislin, G. & Dinges, D. F. Systematic interindividual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. Sleep 27, 423–433 (2004).
12. Van Dongen, H. P., Vitelaro, K. M. & Dinges, D. F. Individual differences in adult human sleep and wakefulness: Leitmotif for a research agenda. Sleep 28, 479–496 (2005).
13. Carmelli, D., Blizard, D. L., Swan, G. E. & Reed, T. A genetic analysis of the sleep-wakefulness interchanges in late adulthood. Arch. Intern. Med. 195, 789–794 (2005).
14. Desai, A. V., Cherkas, L. F., Spector, T. D. & Williams, A. J. Genetic influences in self-reported symptoms of obstructive sleep apnoea and restless legs: a twin study. Twin Res. 7, 589–595 (2004).
15. Watson, N. F., Goldberg, J., Arguelles, L. & Buchwald, D. Genetic and environmental influences on insomnia, daytime sleepiness, and obesity in twins. Sleep 29, 645–649 (2006).
16. Wing, Y. E. et al. Familial aggregation and heritability of insomnia in a community-based study. Sleep Med. 13, 985–990 (2012).
17. Gottlieb, D. J.,Connor, G. T. & Wilk, J. B. Genome-wide association of sleep and circadian phenotypes. BMC Med. Genet. 8(Suppl 1), S9 (2007).

18. Lane, J. M. et al. Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. Nat. Genet. 49, 274–281 (2017).

19. Pak, V. M. et al. Candidate gene analysis in the Sao Paulo Epidemiological Sleep Episodio (SEPIsOnO) shows an association of variant in PDE4D and sleepiness. Sleep Med. 16, 968–973 (2015).

20. Khor, S. S. et al. Genome-wide association study of HLA-DQB1*06:02 negative essential hyporesponsia. Peer J, e66 (2013).

21. Byrne, E. M. et al. A genome-wide association study of sleep habits and insomnia. Am. J. Med. Genet. B Neuropsychiatr. Genet. 162B, 439–451 (2013).

22. Sudlow, C. et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases and of old age. PLoS Med. 12, e1001775 (2015).

23. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat. Genet. 47, 284–290 (2015).

24. Pearce, L. R. et al. KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. Cell 155, 765–777 (2013).

25. Costanzo-Garvey, D. L. et al. KSR2 is an essential regulator of AMP kinase, energy expenditure, and insulin sensitivity. Cell Metab. 10, 366–378 (2009).

26. Hu, Y. et al. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. Nat. Commun. 7, 10448 (2016).

27. Nutt, D. GABA receptors: subtypes, regional distribution, and function. J. Clin. Neuropsychol. 18, 95–111 (2006).

28. Freeman, A. et al. Sleep fragmentation and motor restlessness in a Drosophila model of Restless Legs Syndrome. Curr. Biol. 22, 1142–1148 (2012).

29. Monti, J. M. & Jantos, H. The role of serotonin 5-HT7 receptor in regulating the sleep-wake behavior and cortical activation. Sleep Med. 97, 894–895 (2015).

30. Chen H., et al. Multi-ethnic Meta-analysis Identification of novel risk loci for restless legs syndrome with habitual coffee consumption. J. Appl. Physiol. 120, 507–512 (2010).

31. Raimondo, A., Rees, M. G. & Gloyn, A. L. Glucokinase regulatory protein: a disorder of manganese transport and diabetes. J. Med. Genet. 48, 554–558 (2011).

32. Bhat, S. et al. CACNA1C (Cav1.2) in the pathophysiology of psychiatric disorders. Curr. Opin. Neurobiol. 24, 238–246 (2013).

33. Duan, J., Sanders, A. R. & Gejman, P. V. Genome-wide approaches to the analysis of complex traits. Annu. Rev. Genomics Hum. Genet. 14, 37–58 (2013).

34. Wain, L. V. et al. Novel insights into the genetics of schizophrenia. Nature 492, 66–72 (2012).

35. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional enrichment: a disorder of manganese transport and diabetes variants. Curr. Biol. 24, 238–246 (2013).

36. Borodulin K., et al. Cohort profile: the HUNT Study, Norway. Int. J. Epidemiol. 47, 1–7 (2018).

37. Tabin, C. J. et al. The role of genetic and epigenetic interactions in human disease. Nat. Genet. 46, 369–375 (2014).

38. Bhat, S. et al. CACNA1C (Cav1.2) in the pathophysiology of psychiatric disorders. Curr. Opin. Neurobiol. 24, 238–246 (2013).

39. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomics. Nature 562, 203–209 (2018).

40. Chang, C. C. et al. Selection PUNK: rising to the challenge of larger and richer datasets. GigaScience 4, 7 (2015).

41. Yang, J. et al. Conditional and joint multiple SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat. Genet. 44, 369–375 (2012).

42. 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).

43. Jones, S. E. et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. Nat. Commun. 10, 1585 (2019).

44. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional enrichment: a disorder of manganese transport and diabetes variants. Curr. Biol. 24, 238–246 (2013).

45. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 47, 1228–1235 (2015).

46. Bycroft, C. et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. Int. J. Epidemiol. 45, 1961–1974 (2016).
Author contributions

The study was designed by H.W., J.M.L., S.E.J., H.S.D., H.M.O., A.R.W., V.T.V.H., D.R., D.A.W., M.K.R., M.N.W., S.R., and R.S. H.W., J.M.L., S.E.J., H.S.D., H.O., A.R.W., V.V.H., B.B., B.W., K.Kantojärvi, T.P., Y.S., K.P., J.B., M.A.L., F.A.J.L.S., S.M.P., J.A.K., K.Kristiansson, J.R.N., D.A.L., M.K.R., M.N.W., S.R. and R.S. participated in acquisition, analysis and/or interpretation of data. H.W., J.M.L., H.S.D., H.O., K.Kantojärvi, T.P., S.R. and R.S. wrote the manuscript and all co-authors reviewed and edited the manuscript, before approving its submission. R.S. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-019-11456-7.

Competing interests: H.O. is a consultant for Jazz Pharmaceuticals, Medica Biochemica, and Roche Holding. K. Kristiansson is a Consultant for Negen Ltd. F.A.J.L.S. has received speaker fees from Bayer Healthcare, Sentara Healthcare, Philips, Kellogg Company, Vanda Pharmaceuticals, and Pfizer. M.K.R. has acted as a consultant for Novo Nordisk and Roche Diabetes Care, and also participated in advisory board meetings on their behalf. The remaining authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Heming Wang 1,2, Jacqueline M. Lane 2,3,4, Samuel E. Jones 5, Hassan S. Dashti 2,3,4, Hanna M. Olli 2,3,6,7, Andrew R. Wood 5, Vincent T. van Hees 8, Ben Brumpton 9,10,11, Bendik S. Winsvold 9,12, Katri Kantojärvi 3,13,14, Teemu Palviainen 15, Brian E. Cade 1,2, Tamar Sofer 1,2,12, Yanwei Song 3,16, Krunal Patel 3,16, Simon G. Anderson 17,18, David A. Bechtold 19, Jack Bowden 10,20, Richard Emsley 19, Simon D. Kyle 21, Max A. Little 22,23, Andrew S. Loudon 19, Frank A.J.L. Scheer 1,2, Shaun M. Purcell 1,2, Rebecca C. Richmond 10,24, Kai Spiegelhalter 25, Jessica Tyrrell 5, Xiaofeng Zhu 26, Christer Hublin 15,27, Jaakko A. Kaprio 15,27, Kati Kristiansson 13, Tiina Paunio 13,14, Kristian Hveem 9,10, Jonas B. Nielsen 27, Cristen J. Willer 28, John-Anker Zwart 12, Linn B. Strand 9, Timothy M. Frayling 5, David Ray 29, Deborah A. Lawlor 10,20,32, Martin K. Rutter 19,30,32, Michael N. Weedon 32, Susan Redline 1,2,31,32 & Richa Saxena 1,2,3,4,32

1Division of Sleep and Circadian Disorders, Brigham and Women’s Hospital and Harvard Medical School, Boston, MA, USA. 2Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. 3Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. 4Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. 5Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom. 6Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA, USA. 7Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. 8Netherlands eScience Center, Amsterdam, Netherlands. 9K.G. Jebsen Centre for Genetic Epidemiology, Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway. 10MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK. 11Department of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. 12Division of Clinical Neuroscience, Oslo University Hospital and University of Oslo, Oslo, Norway. 13Genomics and Biomarkers Unit, National Institute for Health and Welfare, Helsinki, Finland. 14Department of Psychiatry and SleepWell Research Program, Faculty of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. 15Institute for Molecular Medicine FIMM, HIIFLE, University of Helsinki, Helsinki, Finland. 16Northeastern University College of Science, Boston, MA, USA. 17Division of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK. 18The George Aileyne Chronic Disease Research Centre, Caribbean Institute for Health Research, University of the West Indies, Cave Hill, Barbados. 19Division of Endocrinology, Diabetes & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. 20Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. 21Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. 22Division of Mathematics, Aston University, Birmingham, UK. 23Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. 24School of Social and Community Medicine, University of Bristol, Bristol, UK. 25Clinic for Psychiatry and Psychotherapy, Medical Centre, University of Freiburg, Freiburg, Germany. 26Department of Population and Quantitative Health Sciences, Case Western Reserve University, Cleveland, OH, USA. 27Department of Public Health, University of Helsinki, Helsinki, Finland. 28Department of Internal Medicine, Division of Cardiology, University of Michigan, Ann Arbor, MI, USA. 29NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford OX39D, UK. 30Manchester Diabetes Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK. 31Department of Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. 32These authors jointly supervised this work: Deborah A. Lawlor, Martin K. Rutter, Michael N. Weedon, Susan Redline, Richa Saxena.