Genetic determinants of daytime napping and effects on cardiometabolic health

Hassan S. Dashti1,2,3,14, Iyas Daghlas1,2,14, Jacqueline M. Lane1,2,3, Yunru Huang4, Miriam S. Udler1,2,5,6, Heming Wang2,7, Hanna M. Ollila1,2,8,9, Samuel E. Jones8,10, Jaegil Kim11, Andrew R. Wood10, 23andMe Research Team*, Michael N. Weedon10, Stella Aslibekyan4, Marta Garaulet7,12,13,15✉ & Richa Saxena1,2,3,15✉

Daytime napping is a common, heritable behavior, but its genetic basis and causal relationship with cardiometabolic health remain unclear. Here, we perform a genome-wide association study of self-reported daytime napping in the UK Biobank (n = 452,633) and identify 123 loci of which 61 replicate in the 23andMe research cohort (n = 541,333). Findings include missense variants in established drug targets for sleep disorders (HCRTR1, HCRTR2), genes with roles in arousal (TRPC6, PNOC), and genes suggesting an obesity-hypersomnolence pathway (PNOC, PATJ). Association signals are concordant with accelerometer-measured daytime inactivity duration and 33 loci colocalize with loci for other sleep phenotypes. Cluster analysis identifies three distinct clusters of nap-promoting mechanisms with heterogeneous associations with cardiometabolic outcomes. Mendelian randomization shows potential causal links between more frequent daytime napping and higher blood pressure and waist circumference.
Naps are short daytime sleep episodes that are evolutionarily conserved across diverse diurnal species ranging from flies\(^1\) to polyphasic mammals\(^2\). In human adults, daytime napping is highly prevalent in Mediterranean cultures and is also common in non-Mediterranean countries including the United States\(^3\). In modern society, napping is encouraged in sleep-deprived populations, such as night shift workers\(^4\) and airline pilots\(^5\), to acutely improve performance and alertness. Although an acute benefit of napping on increased arousal in the setting of sleep deprivation is well-established\(^6\), the long-term effects of habitual napping on chronic disease risk remain controversial. Indeed, cross-sectional studies have provided conflicting evidence on the effects of habitual napping on cognition, blood pressure, obesity, metabolic traits, and mortality\(^7\). As napping behavior may be confounded by inadequate nighttime sleep or underlying poor health\(^8\), causal inference from these observational studies is limited.

Genetic variation constitutes an important contributor to inter-individual differences in napping preference. A twin study estimated heritability of self-reported napping and objective daytime sleep duration to be 65% and 61%, respectively, demonstrating heritability similar or even higher than heritability found for other traits such as nighttime sleep duration and timing\(^8\). Indeed, up to seven genetic loci for daytime napping have been discovered in genome-wide association study (GWAS) of self-reported napping or related accelerometer-derived sleep measures\(^9\)-\(^11\). Discovery of additional genetic loci may reveal biological pathways regulating sleep, elucidate genetic links with other sleep and metabolic traits, and clarify the potential causal effects of habitual napping on cardiometabolic disease.

In this work, we leverage the full UK Biobank dataset of European ancestry, including related individuals (\(n = 452,633\)), and an independent replication sample from 23andMe research participants of European ancestry (\(n = 541,333\)), to define the genetic architecture of daytime napping and to assess links with other sleep and cardiometabolic traits. We identify 123 loci of which 61 replicate in the 23andMe research cohort, including variants in established drug targets for sleep disorders (\(HCRTR1, HCRTR2\)), genes with roles in arousal (\(TRPC6, PNOC\)), and genes suggesting an obesity-hypersomnolence pathway (\(PNOC, PATJ\)). Cluster analysis identifies three distinct clusters of nap-promoting mechanisms and Mendelian randomization shows potential causal links between more frequent daytime napping and higher blood pressure and waist circumference.

Results

Among UK Biobank participants of European ancestry (\(n = 452,633\)), 38.2% and 5.3% of participants reported sometimes and always napping, respectively (Supplementary Table 1). Participants reporting always napping were more likely to be older males, report longer 24 h sleep duration and more frequent daytime sleepiness, have higher body-mass index (BMI), waist circumference, systolic and diastolic blood pressures, have diagnosed sleep apnea, have a higher Townsend deprivation index (i.e., greater degree of socio-economic deprivation), and report being current smokers, unemployed or retired, and shift workers (all \(P < 0.001\); Supplementary Table 1).

Discovery, validation, and replication of 123 genetic loci for daytime napping in UK Biobank and 23andMe. We conducted GWAS using 13,304,133 high-quality imputed genetic variants across 452,633 participants. We identified 123 distinct loci, with \((P < 5 \times 10^{-8}\); Fig. 1A, Supplementary Data 1, Supplementary Fig. 1a) genome-wide SNP-based heritability estimated at 11.9% (standard error = 0.1%). The 123 loci explained 1.1% of the variance in daytime napping. The LD score regression intercept was 1.04 and therefore did not indicate uncontrolled confounding. Effect estimates were largely consistent in GWAS restricted to 338,764 participants self-reporting excellent or good overall health (Supplementary Table 1, Supplementary Data 1). As higher BMI is associated with more frequent napping\(^10\),\(^11\), we conducted a GWAS adjusting for BMI alone or BMI and BMI × BMI and found that 110 of the 123 loci retained genome-wide significance (Supplementary Data 1). Accounting for sleep apnea in GWAS models excluding participants with diagnosed sleep apnea (\(n = 5553\) excluded) or adjusting by a modified STOP-BANG risk score\(^12\) did not influence findings (Supplementary Data 1). Finally, when adjusting for daytime sleepiness, we observed modest attenuation of effect estimates, with 60 of the 123 loci retaining genome-wide significance (Supplementary Data 1).

We found no evidence of sexual dimorphism in the autosomal genetic determinants of daytime napping behavior\(^12\) as indicated by the lack of statistical heterogeneity by sex at any of the lead loci (all \(P > 0.005\) (Supplementary Data 1) and a genome-wide genetic correlation \((r_g)\) of male and female stratified GWAS of 0.94 (standard error = 0.03). We conducted association analyses on the X chromosome to further examine whether common variants on the X chromosome contribute to sex differences in daytime napping and identified five additional loci for daytime napping (Supplementary Table 2). Only one of these variants (rs66217175) had significantly different effect estimates in males and females (\(P = 0.006\)), and no additional GWAS signals were identified on the X chromosome in sex-stratified analysis.

Five of seven loci for daytime napping reported in earlier GWAS in a subset of unrelated UK Biobank participants of European ancestry (\(n = 386,577\))\(^18\) retained genome-wide significance in our analyses (Supplementary Table 3). However, none of the suggestive loci reported in GWAS of accelerometer-derived phenotypes related to napping behavior in the UK Biobank (\(n = 85,670\))\(^19\) and LIFE Adult Study (\(n = 956\))\(^17\) showed evidence of association in the current analysis.

We tested for independent replication of lead loci using data from 23andMe, Inc., a personal genetics company, where 541,333 research participants of European ancestry also provided data on the frequency of daytime napping (43.0% sometimes and 7.6% always napping; Supplementary Table 4). We replicated 61 of 109 tested loci (\(P < 4.6 \times 10^{-4}\), of which 18 of the 61 loci were genome-wide significant (i.e., \(P < 5.0 \times 10^{-8}\)). All 109 tested loci showed consistent direction of effect with the effect estimated in the UK Biobank (\(P_{\text{nominal}} = 3.21 \times 10^{-8}\) (Fig. 1B, Supplementary Data 2). In fixed-effects inverse-variance weighted meta-analysis of UK Biobank and 23andMe (total \(n = 993,966\)), 94 of the 109 lead variants remained genome-wide significant (Fig. 1C, Supplementary Data 2).

Given inherent limitations of self-reported data, we aimed to partly validate the specificity of our associations with an objective measure corresponding to daytime napping behavior. We thus compared effect estimates of the 123 loci with effect estimates for accelerometer-derived daytime inactivity duration\(^19\) from 7-day wrist accelerometry obtained in 85,499 participants of European ancestry in the UK Biobank >2 years after baseline assessment. Estimates of 90 variants were directionally concordant (\(P_{\text{nominal}} = 2.74 \times 10^{-7}\) and variants at \(ASCL4\) and \(SNAP91\) were strongly associated with longer duration of daytime inactivity (\(P_{\text{adj}} < 0.05\) (Supplementary Data 3)). We further quantified the impact of daytime napping on daytime inactivity duration using a polygenic score comprised of lead variants at all 123 loci. A category increase in frequency of daytime napping was associated with 18.9 min (95% confidence interval = 13.6, 24.2; \(P = 4.21 \times 10^{-12}\)) longer duration of daytime inactivity, but had no effect on other accelerometer-derived sleep duration, timing, or quality phenotypes (Supplementary Table 5).
Nap genetic variants share causal variants with other sleep phenotypes and lie near known genes that regulate arousal. Several nap-napping-associated variants had pleiotropic associations with other self-reported sleep traits and accelerometer-derived sleep measures (Supplementary Data 3, 4). This genetic overlap between nap-napping and other sleep traits was further supported by cross-trait LD score regression, where we observed the strongest evidence for a shared genetic basis with daytime sleepiness (\( r_g = 0.70, P = 7.94 \times 10^{-373} \)) and long sleep duration (\( r_g = 0.42, P = 1.94 \times 10^{-64} \)), and weaker correlations with other sleep duration, timing and quality phenotypes (Supplementary Table 6). In concordance with the null associations with other self-reported sleep traits (daytime sleepiness, sleep duration, insomnia, snoring, chronotype, and ease of awakening), we performed multi-trait colocalization analyses of nap-napping loci across six self-reported sleep traits, we performed multi-trait colocalization analyses of nap-napping loci across six self-reported sleep traits, and identified 33 shared signals (of which 25 corresponded to a genome-wide significant nap-napping locus) (Supplementary Data 5). These analyses prioritized putatively causal SNPs genes at several loci which may form hypotheses for experimental follow-up.

Several genetic variants for nap-napping were located in or near genes with known effects on sleep–wake regulation. Thus, to gain insights into putative causal variants driving nap-napping and sleep–wake biology, we integrated results from functional annotation, fine-mapping, multi-trait, and eQTL colocalization analyses (for each colocalization analysis we report a posterior probability for a shared causal variant in the association signal) (Supplementary Data 5, Supplementary Tables 7–9). Functional annotation of all variants identified an enrichment of variants in intronic (46.2%) and intergenic (31.5%) regions, suggesting that non-coding gene regulatory mechanisms may underlie nap-napping as they do for many other complex traits.

In order to identify association signals with evidence for shared causal variants with other sleep traits, we performed multi-trait colocalization analyses of nap-napping loci across six self-reported sleep traits day-time nap-napping, morning preference and ease of awakening, posterior probability of colocalization (pp) as 0.98, (ii) in a cytoplasmic domain of HCRTR1 [I408V; rs2271933, \( r^2 = 0.98 \) with lead nap-napping variant; A allele frequency = 0.21; associated with more frequent nap-napping], and (iii) a cytoplasmic domain of TRPC6 [P158; rs3802829, \( r^2 = 0.98 \) with lead rs11224896 variant; G allele frequency = 0.89; associated with more nap-napping and longer sleep duration; pp = 0.80], which encodes a subunit for transient receptor channels that maintains hypocretin/orexin neurons in a depolarized state. Although an intrinsic lead variant in HCRTR2 was previously reported in GWAS of daytime sleepiness, \( P \) value for association with daytime nap-napping 4.60 \( \times \) 10\(^{-18} \), \( r^2 = 0.29 \) with lead nap-napping variant rs2653349), the traits in the colocalization cluster excluded the daytime sleepiness phenotype, suggesting that the
observed napping signal is driven by a distinct causal variant in HCRTR2 (Supplementary Table 7). To further explore the independence of these signals, we used GCTA COJO to perform conditional analysis adjusting the regional napping associations for the lead napping signal in HCRTR2. We found substantial attenuation in the association with napping for the lead daytime sleepiness variant in HCRTR2 (rs3122170; $P$ value from $4.60 \times 10^{-18}$ to $4.56 \times 10^{-3}$ after conditioning). This further suggests that the identified signal for daytime napping is distinct from the previously reported daytime sleepiness signal in the HCRTR2 region.

Second, colocalization analyses revealed variants in PNOC and PATJ with effects on napping, daytime sleepiness, and BMI, suggesting a potential obesity-hypersomnolence pathway. An intronic candidate causal variant in PNOC (rs351776; C allele frequency = 0.55) associated with more frequent napping, more daytime sleepiness, and higher BMI. PNOC encodes a preproprotein that is proteolytically processed to generate the nociceptin neuropeptide, which opposes the effects of hypocretin to reduce arousal and spontaneous activity in zebrafish. The colocalization of daytime napping with BMI at this locus is consistent with known pleiotropic effects of PNOC in feeding behavior (pp = 0.84; Supplementary Table 8). The known missense variant in PATJ (rs12140153; G1543V; G allele frequency 0.90) has a stronger association with daytime napping than any previously studied sleep phenotypes (Supplementary Data 4), and is likely a shared causal variant with daytime sleepiness, chronotype, and with BMI (pp = 0.81 and pp = 0.99; Supplementary Tables 7 and 8).

Third, colocalization analyses refined genetic effects previously described at the KSR2 locus implicated in ERK/EGFR signaling, a pathway with an established causal role in sleep regulation in C. elegans, Drosophila, and zebrafish. This included an intronic variant in KSR2 (rs1846644; T allele frequency = 0.60; pp = 0.91), that is associated with more frequent napping, longer sleep duration, and increased daytime sleepiness.

Fourth, several genetic variants were prioritized at or near genes (a) coding for proteins constituting or interacting with potassium channels (rs7715432, rs10875606, KCTD16), (b) involved in glutamate transmission (rs60920123, GRIN2A), rs2284015 (CAGN2), and (c) previously associated with periodic leg movements and restless legs syndrome with periodic leg movements (rs4236060; BTBD9).

Fifth, we found evidence of association for variants in PRRC2C, one of three orthologs of the Drosophila nocice gene. Nocice targets clock neurons to synchronize molecular and behavioral rhythms to temperature cycles and influences siesta sleep in flies. We observed no gene-by-season (a proxy for ambient temperature) statistical interaction at this and any other loci (Supplementary Data 1).

We performed colocalization analyses using gene expression data from the frontal cortex in the GTEx data release v7 (n = 129), the brain tissue predominantly enriched for daytime napping signals. Daytime napping variants at FADS1 associated with increased expression of FADS1 (p = 0.89) and at ECE2 associated with increased expression of ECE2 (p = 0.99) (Fig. 2A, B; Supplementary Table 9). Another lead variant is near FNDC5 (rs2786547), a gene coding for irisin, a muscle-derived hormone with putative effects on expression of sleep-regulating neuropeptides. We found strong evidence for colocalization of the daytime napping signal with gene expression of FNDC5 in skeletal muscle in the GTEx data release v7 (n = 706, pp = 0.93), with higher gene expression relating to less frequent napping (Supplementary Fig. 2). This suggests a role for FNDC5 in a sleep-regulating mechanism outside of the central nervous system.

Finally, multi-variant clustering suggested the possibility of at least three distinct pathways influencing daytime napping. Bayesian nonnegative matrix factorization (bNMF) clustering for 123 variants with 17 self-reported and accelerometer-derived sleep traits identified 3 clusters (63% of 1000 iterations) and these same 3 clusters were also present in an additional 34% of iterations with 4 clusters (Table 1, Supplementary Data 6, Supplementary Fig. 3a) reflecting (a) sleep propensity (cluster 1; 6 contributing loci with CRHR1, SKOR2, KSR2, ASCL4, RERE, and ECE2); (b) disrupted sleep (cluster 2; 5 contributing loci with SHHIA4, ADO, NRXN3, FNDC5, and GSI-259H13.13 as lead); and (c) early sleep timing (cluster 3; 9 contributing loci with HCRTR2, ALG10, ALG10B, PATJ, BTBD9, MTNR1B, AGAP1, RP11-6N13.1, and ZBTB5 as lead loci, notably not at known core clock genes). A fourth possible cluster, obstructive sleep apnea, was observed in 34% of 1000 iterations (Supplementary Fig. 3b). Results were corroborated with findings from an alternative unsupervised hierarchical clustering method (Supplementary Fig. 3c), with clusters 1 and 2 partly overlapping with previously observed clusters for daytime sleepiness.

Genes at association signals are enriched in brain and GABAergic neurons, and in neural development and opioid signaling pathways. In order to identify genes, neuronal subtypes and annotated pathways relevant to daytime napping, we first mapped the genes near association signals and then tested for their over-representation relative to all genes in experimental genome-wide datasets. Gene-based associations for 21,761 genes mapped with Pascal are listed in Supplementary Data 7; 324 genes showed association after Bonferroni correction. The identified signals were enriched for genes predominantly expressed in brain tissues, including the frontal cortex ($P = 1.18 \times 10^{-7}$) and nucleus accumbens ($P = 1.26 \times 10^{-7}$) (Fig. 2A, B; Supplementary Table 10). Single-cell enrichment analyses in FUMA using human brain datasets (listed in Fig. 3B) showed consistent enrichment in GABAergic neurons across several brain tissues including the prefrontal cortex and midbrain. In addition, pathway enrichment analysis using MAGMA and Pascal indicated enrichment of genes involved in regulation of transmission across chemical synapses, neuronal system, and opioid signaling.

The genetic contributors to daytime napping are shared with cardiometabolic diseases. To gain insights into shared heritability of daytime napping with other disease and behavior traits, we performed cross-trait LD score regression using publicly available GWAS data for 257 traits. Modest positive correlations were observed between daytime napping and several anthropometric and cardiometabolic diseases and traits including BMI, triglycerides, and type 2 diabetes (Fig. 4A, Supplementary Data 10), of which correlations with triglycerides remained significant in the GWAS model adjusting for BMI. To further characterize shared genetic links between daytime napping and diseases in a disease-enriched and independent health system-based clinical cohort, we conducted a phenome-wide association study (PheWAS) in the Mass General Brigham Biobank (n = 23,561 participants of European ancestry with genetic data). We generated a daytime napping genome-wide polygenic score (GPS) and tested associations with 951 ICD-code based disease categories. PheWAS showed 3 Bonferroni-significant associations (18 FDR-significant), including positive associations with essential hypertension (GPS q10 vs q1 odds ratio [95% confidence interval]: 1.30 [1.13, 1.51]), obesity (GPS q10 vs q1: 1.38 [1.18, 1.62]), and chronic nonalcoholic liver disease (GPS q10 vs q1: 1.51 [1.18, 1.92]), which encompasses diagnosis codes for chronic non-specific or nonalcoholic liver disease (Fig. 4B, C, Supplementary Data 11). We also observed associations of a polygenic
score of the 123 napping variants, and polygenic sub-scores for each of the 3 clusters with cardiometabolic traits from large-scale public GWAS (Table 1, Supplementary Table 11). Cluster-specific polygenic score associations varied across outcomes, and included associations of cluster 1 with higher blood pressure, and clusters 2 and 3 with adiposity traits (Table 1).

**Mendelian randomization suggests a causal effect of more frequent daytime napping on increased blood pressure and waist circumference.** To explore whether daytime napping may causally increase cardiometabolic disease risk, we performed two-sample Mendelian randomization (MR) analyses using the 123 loci as genetic proxies for daytime napping (Supplementary Table 11). We observed a potentially causal effect of more frequent daytime napping on higher diastolic blood pressure (DBP; 0.25 standard deviation (SD) unit increase per category increase in daytime napping, 95% CI [0.15, 0.34], \( P = 2.99 \times 10^{-7} \), systolic blood pressure (SBP; 0.18 SD units, [0.09, 0.27], \( P = 5.15 \times 10^{-5} \), and waist circumference (0.28 SD units, [0.11, 0.45], \( P = 1.3 \times 10^{-3} \), all of which surpassed multiple testing correction (Fig. 5A, B). In sensitivity analysis, we found a consistent effect, although attenuated in magnitude for the outcome of DBP, of genetically proxied more frequent daytime napping on higher blood pressure when using variant association statistics from 23andMe as the exposure, and blood pressure in the ICB-UKB meta-analysis\(^{48}\) as the outcome (DBP: 0.08 SD units, [0.003, 1.18], \( P = 0.04 \); SBP: 0.21 SD units, [−0.02, 0.43], \( P = 0.07 \)). As the MR effects may be explained by pleiotropic effects of the napping variants on pathways independent of napping, we performed five sensitivity analyses and found consistent evidence of effect (Supplementary Data 12, Supplementary Table 12; Supplementary Fig. 4). Given prior evidence for a causal effect of higher BMI on daytime sleepiness\(^{24}\), we tested the hypothesis that adiposity traits (waist circumference, waist-to-hip ratio adjusted for BMI (WHRadjBMI), and BMI) influenced daytime napping frequency. Genetically proxied WHRadjBMI was nominally associated with a modest increase in daytime napping frequency (inverse-variance weighted: 0.03 category increase in daytime napping per SD increase in WHRadjBMI, [0.01, 0.05], \( P = 0.01 \)) (Fig. 5B, Supplementary Data 12).

**Leveraging HCRTR1 and HCRTR2 genetic associations to predict the cardiovascular safety profile of dual orexin antagonists.** Given our observation that the hypocretin pathway contributed to variation in daytime napping behavior (variants in HCRTR1 and HCRTR2), and recent reports suggesting that mammalian orexin signaling has cardioprotective effects\(^{49}\), we examined whether these variants may serve as instruments to predict the cardiovascular safety of orexin receptors as drug targets. This has clinical relevance, as dual orexin receptor antagonists (DORAs) are currently used as sleep medications, and orexin receptor agonists are currently in development for.
Table 1 Cluster-specific daytime napping polygenic scores associations with self-reported and accelerometer-derived sleep traits and other cardiometabolic traits.

| Trait, units (sleep trait is a defining feature of cluster #) | Cluster 1: Sleep propensity N loci = 6 | Cluster 2: Disrupted sleep N loci = 5 | Cluster 3: Early morning awakening N loci = 9 |
|-------------------------------------------------------------|----------------------------------------|----------------------------------------|---------------------------------------------|
| Sleep duration, minutes (1) | Beta | SE | P Value | Beta | SE | P Value | Beta | SE | P Value |
| 0.72 | 0.07 | 1.8×10^{-23} | -0.60 | 0.11 | 1.05×10^{-07} | 0.14 | 0.08 | 6.7×10^{-02} |
| Short sleep duration, log-odds (1) | -0.12 | 0.03 | 1.5×10^{-04} | 0.34 | 0.05 | 1.54×10^{-12} | 0.07 | 0.03 | 1.98×10^{-02} |
| Long sleep duration, log-odds (1) | 0.19 | 0.02 | 9.0×10^{-17} | 0.06 | 0.04 | 9.0×10^{-02} | 0.12 | 0.02 | 6.0×10^{-03} |
| Ease of awakening, more ease (3) | -0.18 | 0.05 | 3.3×10^{-04} | -0.60 | 0.08 | 3.08×10^{-14} | 0.61 | 0.05 | 4.12×10^{-03} |
| Snoring, log-odds | -0.02 | 0.03 | 5.4×10^{-01} | 0.14 | 0.05 | 5.14×10^{-03} | 0.16 | 0.03 | 6.4×10^{-03} |
| Daytime sleepiness, more sleepiness (1, 2) | 0.10 | 0.03 | 6.3×10^{-07} | 0.48 | 0.05 | 1.76×10^{-20} | 0.62 | 0.03 | 1.53×10^{-05} |
| Insomnia, log-odds (2) | -0.01 | 0.04 | 7.7×10^{-01} | 0.45 | 0.07 | 4.89×10^{-11} | 0.11 | 0.05 | 1.27×10^{-02} |
| Sleep duration, minutes (1, 2) | -0.12 | 0.03 | 1.6×10^{-04} | 0.34 | 0.05 | 1.54×10^{-12} | 0.07 | 0.03 | 1.98×10^{-02} |
| 0.04 | 0.10 | 6.9×10^{-01} | 0.53 | 0.15 | 4.19×10^{-04} | 0.53 | 0.10 | 1.00×10^{-07} |

Table 1 Cluster-specific daytime napping polygenic scores associations with self-reported and accelerometer-derived sleep traits and other cardiometabolic traits.

- Narcolepsy.
- To test for such potential on-target cardiovascular side effects, we used missense variants in HCRTR1 (A allele of rs2271933) and HCRTR2 (A allele of rs2653349), both associated with more frequent daytime napping and daytime sleepiness43,44, as proxies for pharmacologic inhibition of these proteins, and tested for associations with cardiovascular phenotypes in large GWAS (Supplemental Table 13). This analysis revealed no associations of the variants with cardiovascular outcomes, but showed opposing effects on systolic blood pressure at HCRTR1 (-0.10 mmHg, 95% CI [−0.17, −0.04], P = 1.00×10^{-4}) and HCRTR2 (0.14 mmHg, 95% CI [0.07, 0.21], P = 1.00×10^{-04}, Fig. 6). We further performed a hypothesis-free scan across 1402 ICD-code defined phenotypes in the UK Biobank30 and found no variant-disease associations (Supplementary Fig. 5; Supplementary Data 13). The present human genetic evidence therefore does not support a net excess adverse cardiovascular risk from on-target inhibition of HCRTR1 and HCRTR2, but suggests potential opposing effects on blood pressure regulation by the two receptors.

Discussion

We comprehensively investigated the genomic influences of daytime napping using the largest discovery and replication sample sizes to date. We identified 123 independent loci in the UK Biobank with strong evidence of replication in 23andMe, an independent study with different demographic characteristics. Variant effects were largely independent of BMI and sleep apnea, and the associations retained significance when GWAS was restricted to healthier participants, a strong determinant of 5-year mortality in the UK Biobank51, suggesting that signals were not driven by poor health. In addition, despite higher prevalence of daytime napping among men compared to women52, we identified only one sex-specific signal on the X chromosome, suggesting sex differences may be attributed to environmental factors or possibly rare genetic variants. Our results advance the understanding of the biology of daytime napping, refine the understanding of pleiotropy and causality in the relationship of napping with sleep and cardiometabolic traits, and inform pharmacologic investigations of orexin antagonism.

The identified variants highlight a central role for arousal-regulating neuropeptide signaling pathways in daytime napping propensity. Most prominent among these pathways was the well-established hypocretin arousal pathway53 (including missense variants in HCRTR1, HCRTR2, and TRPC6). It is thus possible that orexin receptor agonism, a therapeutic strategy currently under investigation for narcolepsy, may have roles in the treatment of patients with more mild deficits in the arousal/wake drive system54. Additional pathways with known roles in sleep-wake biology in model organisms55 include neuronal excitability driven by variation in the function of potassium channels and glutamate signaling, EEG signaling pathway, and opioid signaling.
Expression of genes under association peaks was most enriched in the frontal cortex, similar to observations for daytime inactivity duration\textsuperscript{19}, and other brain regions prominently implicated in sleep duration, timing, and quality traits\textsuperscript{23,25,26}. Cross-trait clustering of the identified loci suggest at least three underlying physiologic mechanisms, including (1) propensity for longer sleep, (2) consequence of poor and disrupted sleep, and (3) napping concomitant with early sleep timing, potentially reflecting loss of function in arousal pathways. Notably, genetic links between daytime napping and sleep disorders, e.g., sleep apnea or restless legs syndrome, may be partially undetected by our study because of incomplete ascertainment of these disorders in the UK Biobank and the lack of available summary statistics in public repositories and databases. We found that the genetic architecture of daytime napping is shared with cardiometabolic diseases and traits, consistent with previous epidemiologic associations of more frequent daytime napping with increased cardiometabolic risk\textsuperscript{7–13,56}. At the locus level, we observed colocalization of the daytime napping loci with daytime sleepiness, snoring, chronotype, and BMI loci at \textit{PNOC} and \textit{PATJ}, suggesting an obesity-hypersomnolence pathway\textsuperscript{57}. Furthermore, colocalization of \textit{FADS1} gene expression in the frontal cortex with the daytime napping signal suggests uncharacterized pleiotropic effects of lipid metabolism on sleep. Positive genome-wide genetic correlations were observed with multiple anthropometric, glycemic, and cardiometabolic traits, of which several correlations were attenuated after accounting for BMI. In a large health system-based clinical cohort, phenome-wide association analyses using a daytime napping genome-wide polygenic score further supported associations with obesity and hypertension, in addition to other cardiometabolic diseases. Although daytime napping shares biological determinants with other sleep traits, most prominently daytime sleepiness\textsuperscript{24}, there were several genetic findings unique to daytime napping. There were 26/123 loci unique to daytime napping, with several other loci exhibiting stronger relationships with daytime napping relative to other traits (e.g., \textit{KSR2} locus). The SNP-based heritability of daytime napping (11.9%) was almost double that previously reported for daytime sleepiness (6.9%)\textsuperscript{24}, and daytime napping variants were modestly attenuated in GWAS models accounting for daytime sleepiness. Although prior analyses related higher BMI to more frequent daytime sleepiness\textsuperscript{24}, we observed no such relationship with frequency of daytime napping. Taken together, these data suggest that daytime napping and daytime sleepiness should be considered related, but distinct features of the impaired arousal continuum.

A key clinical question is whether habitual daytime napping has causal effects on cardiometabolic health. Findings from our Mendelian randomization analyses suggest potentially deleterious effects of daytime napping frequency on cardiometabolic health, with effects on increased blood pressure and waist circumference. A causal effect of more frequent napping with higher blood pressure is consistent with earlier epidemiologic findings between self-reported and actigraphy-measured daytime napping and hypertension\textsuperscript{58–60}. Mechanisms driving this relationship are unknown but may include detrimental effects of napping on nighttime sleep quality, or chronic effects related to transient

**Fig. 3** Tissue expression, single-cell, and pathway-based enrichment analyses for daytime napping. **A** MAGMA tissue expression analysis using gene expression per tissue based on GTEx RNA-seq data for 53 specific tissue types. Significant tissues ($P < 9.43 \times 10^{-4}$) are shown in red. **B** Significant single-cell types from single-cell enrichment analyses using human brain datasets in FUMA. **C** Top pathways determined from analysis using MAGMA gene sets and Pascal (gene-set enrichment analysis using 1077 pathways from KEGG, REACTOME, BIOCARTA). Significant pathways are shown in red ($P_{adj} < 0.05$). All pathway and tissue expression analyses in this figure can be found in tabular form in Supplementary Table 10, Supplementary Data 8, 9.
obtained by comparing genome-wide association estimates for daytime napping (without and with BMI adjustment) with summary statistics estimates from 257 publicly available genome-wide association studies. Blue indicates positive genetic correlation and red indicates negative genetic correlation; \( r_g \) values are displayed for significant correlations. Larger colored squares correspond to more significant \( P \) values. Asterisk denotes significant false discovery rate (FDR) corrected \( P \) values. Full genetic correlations for all 257 traits can be found in Supplementary Data 10. B Manhattan plot of phenome-wide association findings for daytime napping genome-wide polygenic score in Mass General Brigham Biobank \((n = 23,561)\). The \( x \)-axis is color-coded phecodes organized by broad disease categories and the \( y \)-axis is \( P \) value of association \((-\log_{10} P)\). The horizontal red line depicts phenome-wide significance using Bonferroni correction for all tested diseases \((951 \text{ diseases})\), and the horizontal blue line depicts phenome-wide significance using FDR correction. Upward arrows denote positive associations \((OR > 1)\), and downward arrows denote inverse associations \((OR < 1)\). Full results for all 951 diseases can be found in Supplementary Data 11. C Cross-sectional association between quartile 10 and quartile 1 \((\text{reference group})\) of daytime napping genome-wide polygenic score and essential hypertension, obesity, and chronic nonalcoholic liver disease in the Mass General Brigham Biobank \((n = 23,561)\). Error bars represent the 95% confidence intervals for association.

Fig. 4 Genome-wide genetic architecture of daytime napping correlations and associations with diseases and traits. A Shared genetic architecture between daytime napping and cardiometabolic diseases and traits. Linkage disequilibrium (LD) score regression estimates of genetic correlation \((r_g)\) were obtained by comparing genome-wide association estimates for daytime napping (without and with BMI adjustment) with summary statistics estimates from 257 publicly available genome-wide association studies. Blue indicates positive genetic correlation and red indicates negative genetic correlation; \( r_g \) values are displayed for significant correlations. Larger colored squares correspond to more significant \( P \) values. Asterisk denotes significant false discovery rate (FDR) corrected \( P \) values. Full genetic correlations for all 257 traits can be found in Supplementary Data 10. B Manhattan plot of phenome-wide association findings for daytime napping genome-wide polygenic score in Mass General Brigham Biobank \((n = 23,561)\). The \( x \)-axis is color-coded phecodes organized by broad disease categories and the \( y \)-axis is \( P \) value of association \((-\log_{10} P)\). The horizontal red line depicts phenome-wide significance using Bonferroni correction for all tested diseases \((951 \text{ diseases})\), and the horizontal blue line depicts phenome-wide significance using FDR correction. Upward arrows denote positive associations \((OR > 1)\), and downward arrows denote inverse associations \((OR < 1)\). Full results for all 951 diseases can be found in Supplementary Data 11. C Cross-sectional association between quartile 10 and quartile 1 \((\text{reference group})\) of daytime napping genome-wide polygenic score and essential hypertension, obesity, and chronic nonalcoholic liver disease in the Mass General Brigham Biobank \((n = 23,561)\). Error bars represent the 95% confidence intervals for association.

evening blood pressure surges following daytime napping\(^{61,62}\). Similarly, mechanisms underlying the link between daytime napping and body fat distribution are poorly understood\(^{63}\).

Although results from the MR Egger sensitivity analysis of waist circumference on daytime napping were inconsistent with findings from our primary MR analysis, the genetic overlap we demonstrated with BMI indicates that the Egger analysis may be biased by violation of the “instrument strength independent of direct effect (INSIDE)” assumption\(^{64}\). Polygenic scores of each napping subtype showed heterogeneous associations with cardiometabolic outcomes across clusters, including associations with higher blood pressure for cluster 1, and other adiposity traits for clusters 2 and 3. Exploring causal relationships with biologically distinct subtypes of daytime napping will be important to understand the beneficial or detrimental role of different aspects of napping biology with disease outcomes.

We leveraged coding variation in HCRTR1 and HCRTR2 to predict the cardiovascular consequences of long-term pharmacologic modulation of orexin receptors. We found no net effect of these genetic proxies on cardiovascular outcomes, nor on any ICD-code defined disease outcomes in a PheWAS. These results predict that pharmacologic agonism or antagonism of orexin receptors therapies is unlikely to increase the risk of cardiovascular disease. An association of HCRTR1 and HCRTR2 with blood pressure was observed, however, the direction of effect differed for the two variants. This suggests a neutral net blood pressure effect of dual orexin receptor antagonism, and more broadly suggests pleiotropic effects of these proteins on blood pressure regulation. However, it is possible that these genetic variants do not proxy peripheral effects of HCRTR1 and HCRTR2 inhibition (e.g., bone marrow)\(^{49}\). The application of PheWAS to study on-target side effects of sleep medications sets the stage for future use of these genetic proxies to understand the health consequences of orexin receptor modulation.

Our analyses are limited by the crude assessment of daytime napping frequency via questionnaire with no information on duration or timing. Our effort to partly validate the specificity of our discovered loci from self-report with an objectively determined daytime napping behavior from accelerometer was likely limited as a result of phenotypic differences between self-report and accelerometer (self-report was based on daytime napping frequency whereas accelerometer measures were based on daytime inactivity duration in the absence of sleep diaries; Pearson correlation \( r^2 = 0.17 \), relatively smaller sample size in the accelerometer subsample \((n = 85,670)\), or lapsed time between measurements as the accelerometer was worn between 2.8 and
Fig. 6 Cardiovascular risk factor and disease associations of missense variants in HCCTR1 (rs2271933) and HCCTR2 (rs2653349), which encode targets of Suvorexant, an FDA-approved sleep medication with an unknown cardiovascular safety profile. Sample size either reflects the total number of subjects (for continuous outcomes) or the number of cases and controls (for binary traits) that were included in each of the genome-wide association studies. All associations are oriented to the napping-increasing allele of the variants. Additional details regarding the included studies are provided in Supplementary Table 13 and Supplementary Data 13. Black boxes show Mendelian randomization effect estimates and surrounding lines display 95% confidence intervals. All $P$ values are two-sided. * significant at Bonferroni-corrected alpha threshold and robust in sensitivity analyses. BMI body-mass index, CAD coronary artery disease, CI confidence interval, DBP diastolic blood pressure, HOMA homeostatic model assessment of insulin resistance, HOMA2 homeostasis model assessment of $\beta$-cell function, LDL low-density lipoprotein, HDL high-density lipoprotein, OR odds ratio, WMH white matter hyperintensities, WHR waist-to-hip ratio.

**Table**: Sample size either reflects the total number of subjects (for continuous outcomes) or the number of cases and controls (for binary traits) that were included in each of the genome-wide association studies. All associations are oriented to the napping-increasing allele of the variants.
9.7 years after study baseline. Replication of most loci and the specific association with daytime activity duration, but not other accelerometer measures, however, support our findings. The low participation rate of the UK Biobank at 5.5% may have introduced selection bias. However, consistency of the genetic signals between the UK Biobank and 23andMe, an independent study with different demographic, and various findings with the Mass General Brigham Biobank, an independent clinical cohort, supports the generalizability of our findings. In addition, the identification of variants in pathways with known relevance to sleep (e.g., HCRTR1 and HCRTR2) suggests that the GWAS is capturing true biological signal. Nonetheless, continued evaluation in other demographics, including age-groups and ancestries, is necessary. It remains possible that rare and structural variation have an important contribution to the genetic architecture of daytime napping, however, these data were not tested in the present analysis. In addition, our analysis was limited in scope to cardiometabolic health, and future studies should evaluate the impact of daytime napping on other health outcomes including mental health. Finally, despite consistency in Mendelian randomization estimates, these analyses require strong, unverifiable assumptions for the determination of causality and therefore require confirmation in randomized controlled trials of sleep interventions. Further dissection of the heterogeneity of daytime napping is necessary to determine which types of daytime napping behavior are most detrimental to cardiometabolic health. In addition, future analyses investigating sex heterogeneity in daytime napping frequency is warranted. In summary, our genetic analyses contribute important insight into the biology and cardiometabolic consequence of habitual daytime napping in adults.

Methods
UK Biobank. The UK Biobank is a large population-based study established to facilitate detailed investigations of the genetic and lifestyle determinants of a wide range of phenotypes. Data from >500,000 participants living in the United Kingdom who were aged 40–69 and living <25 miles from a study center participated in the study between 2006 and 2010. Extensive phenotypic data were self-reported upon baseline assessment by participants using touchscreen tests and questionnaires and at nurse-led interviews. 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. The current study was conducted under UK Biobank application 6818.

Daytime napping, covariates, and other self-reported and objectively measured sleep traits. At baseline assessment, all study participants reported their daytime napping habit. For daytime napping analysis (n = 501,646), participants were asked whether you had a nap during the day? with responses Never/rarely, Sometimes, Usually, Prefer not to answer. Responses were treated as a continuous variable in the GWAS. Prefer not to answer responses were set to missing. Participants further self-reported age, gender, sleep duration, chronotype, insomnia symptoms, sleep apnea, smoking, and overall health. Weight, height, and waist circumference were measured and body-mass index (BMI) was calculated as weight (kg)/height² (m²). Systolic and diastolic blood pressure were measured at baseline and the average of two automatically-readings was used. Socio-economic status was represented by the Townsend deprivation index based on national census data immediately preceding participation. Assessment season was determined from self-report of assessment visit date and categorized as 1 for winter [January–March], 2 for spring [April–June], 3 for summer [July–September], and 4 for fall [October–December], as previously conducted. Participants rated their overall health in response to the question, In general how would you rate your overall health?, with responses excellent, good, fair, poor, and do not know. Cases of sleep apnea were determined from self-report during nurse-led interviews or health records using International Classification of Diseases (ICD)-10 codes for sleep apnea (G47.3). For each participant, a modified STOP-BANG risk scale was calculated to assess the risk of sleep apnea in the UK Biobank to account for undiagnosed sleep apnea, as was calculated. The modified STOP-BANG risk scale for sleep apnea is missing the question, Has anyone observed you stop breathing during sleep? and replacing neck circumference with waist circumference dichotomized to the threshold for metabolic syndrome. Insomnia symptoms were ascertained from self-report to the question, Do you have trouble falling asleep at night or do you wake up in the middle of the night? with responses never/rarely, sometimes, usually, prefer not to answer. Participants who responded usually or never as insomnia cases, and remaining participants were set as controls. Smoking status (never, former, current) was further self-reported. Missing covariates were imputed by using sex-specific median values for continuous variables (i.e., BMI and Townsend index).

A subset of 103,711 participants from the UK Biobank wore actigraphy devices (Axivity AX3) for up to 7 days, ~2.8–9.7 years after their study baseline visits. Details on quality control and data processing have been described previously. The following sleep measures were derived by processing raw actimeter data: daytime activity, sleep duration, sleep efficiency, and estimated bouts of inactivity that fell outside of the sleep period time window, sleep mid-point, midpoint of the least-active 5 h of the day (L5 timing), and midpoint of the most-active 10 h of the day (M10 timing). Specifically, daytime inactivity duration was estimated by the total daily duration of estimated bouts of inactivity that fell outside of the sleep period time window. These inactivity bouts are any inactivity lasting >30 min. Inactivity bouts that are <60 min are combined to form inactivity blocks. This measure captures very inactive states such as napping and wakeful rest but not inactivity such as sitting and reading or watching television, which are associated with a low but detectable level of movement.

Genome-wide association study for daytime napping in UK Biobank. Genotyping was performed by the UK Biobank, and genotyping, quality control, and imputation procedures are described in detail previously. In brief, blood, saliva, and urine were collected from participants, and DNA was extracted from the buffy coat samples. Participant DNA was genotyped on two arrays, UK BiLEVE and UK Biobank Axiom with >95% common content and genotypes for ~800,000 autosomal SNPs were imputed to two reference panels. Genotypes were called using Affymetrix Power Tools software. Sample and SNPs for quality control were selected from a set of 489,212 samples across 812,428 unique markers. Sample and SNP missingness (QCs) was controlled using 605,876 high-quality markers. Samples were removed for high missingness or heterozygosity (968 samples) and sex chromosome abnormalities (652 samples). Genotypes for 488,377 samples passed sample QCs (~99.9% of total samples). Marker-based QC measures were tested in the European ancestry subset (n = 463,844), which was identified based on ancestry informative markers. At baseline, all study participants reported their daytime napping data (never/rarely, sometimes, Usually, Prefer not to answer). Responses were treated as a continuous variable in the GWAS. In total ~96 million SNPs were imputed. Related individuals were clustered subjects into four ancestry clusters using K-means clustering on ancestry, genotyping array and genetic correlation matrix with a maximum per SNP missingness >2.5% (~100,000 SNPs) and identified values were consistent with those estimated for other highly polygenic traits such as sitting and reading or watching television, which are associated with a low but detectable level of movement.
Within a 1 Mb window. Annotation of the lead variants, including predicted sequence consequence, was obtained from the FUMA output. We determined the PICs used to assign a variant to be the causal variant at the peak signal.

For the 123 lead variants, we tested for gene-by-season interaction in PLINK among unrelated participants of white British ancestry (n = 337,409) using linear regression and an additive genetic model. Interaction analyses were adjusted for age, sex, 10 principal components of ancestry, genotyping array, and season to determine SNP interaction with season on daytime sleeping. In addition, for each lead variant, corresponding summary statistics for other self-reported and accelerometer-derived sleep measures were obtained from the Sleep Disorder Knowledge Portal (http://sleepdisordergenetics.org/). As earlier UK Biobank GWASs were restricted to the HRC-imputed variants, if the lead signal was unavailable, a proxy SNP (r² > 0.8) was used instead.

**23andMe, Inc. replication.** 23andMe, Inc. is a personal genetics company. DNA extraction and genotyping were performed on saliva samples by National Genetics Institute, a CLIA licensed clinical laboratory and a subsidiary of Laboratory Corporation of America. Samples were genotyped on one of five genotyping platforms. Samples that failed to reach 98.5% call rate were re-analyzed. A single unified imputation of colocalization was created by combining the May 2015 release of the 100 Genomes Phase 3 haplotypes with the UK10K imputation reference panel. For each chromosome, Minimac was used to impute the reference panels against each other, reporting the best-guess genotype at each site. Ancestry was determined through an analysis of local ancestry. A principal component analysis was performed separately for each ancestry group (n = 65,000 haplotypes with imputed variants present in all five genotyping platforms. In addition, a maximal set of unrelated individuals was chosen for each analysis using a segmental identity-by-descent estimation algorithm. All individuals included in the analyses provided informed consent and answered surveys online according to human subject protocol, which was reviewed and approved by the Ethical Independent Review Services, a private institutional review board (http://www.eandireview.com).

For the present daytime napping replication, we restricted analyses to 541,333 participants of European ancestry with survey responses to a question on frequency of daytime napping. Participants were asked, How many days per week do you take naps during the daytime? (0, <1 day/week or rarely) with responses on a continuous scale. Responses in days per week were scaled to never/rarely if 0 or 1 (n = 267,271), sometimes if 2 to 5 (n = 232,868), and usually if 6 or 7 (n = 41,194) to more closely resemble the UK Biobank categories. Replication for the 123 daytime napping loci or proxy for lead SNP (r² > 0.80) were generated through linear regression (using an additive model) of the phenotype. Adjustment for ancestry, sex, age, sex by age, the first four principal components, and a categorical variable representing genotyping platform. Furthermore, meta-analysis of UK Biobank and 23andMe associations for the daytime napping loci was performed using METAL by weighting effect-size estimates using the inverse of the corresponding squared standard errors (version released 25 March 2011).

**Colocalization.** To identify genomic regions which harbor causal variants that influence multiple sleep traits, we performed multi-trait colocalization analyses using the Hypothesis Prioritization Colocalization (HyPrColo) package. This package performs multi-trait colocalization using a computationally efficient algorithm that facilitates colocalization of large numbers of traits. To identify clusters of colocalized traits, we implemented the branch and bound divisive clustering algorithm that facilitates colocalization of large numbers of traits. To identify clusters of colocalized traits, we implemented the branch and bound divisive clustering algorithm.

The outputs from the algorithm include: (i) colocalized traits, (ii) the posterior correction probability for all tests performed (v) the proportion of the posterior probability of colocalization explained by the genetic variant (interpreted as a multi-trait fine-mapping probability). We report loci with posterior probability (pp) for colocalization above this cutoff corresponds to a false discovery rate of 5%. We performed two additional colocalization analyses. Using summary statistics from a meta-analysis (n ~ 700,000) of UK Biobank and the GIANT consortium, we performed genome-wide colocalization analyses using LDSC. Genetic correlation between two traits from summary statistics (ranging from ~ 1 to 1) using the fact that the GWAS effect-size estimate for each SNP incorporates effects of all SNPs with LD with that SNP, SNPs with high LD have higher statistics than SNPs with low LD, and a similar relationship is observed when single study test statistics are replaced with the product of z-scores from two studies of traits with some correlation. Significance was considered at the Bonferroni correction for all tests performed (P < 0.05/257 tests). In addition to publicly available summary statistics from LDHub, we also used publicly available summary statistics from earlier UK Biobank GWASs for self-reported and accelerometer-derived sleep traits from the Sleep Disorder Knowledge Portal (http://sleepdisordergenetics.org/). We performed genome-wide genomewide gene expression and correlation analyses using LDSC. Finally, we calculated genetic correlations between the sex-specific napping GWAS to determine the similarity in male and female genetic architecture.

**Phenome-wide association study in the Mass General Brigham Biobank.** The Phenome-wide association study (PheWAS) in the Mass General Brigham Biobank (formerly Partners Biobank) is a hospital-based cohort study from the Mass General Brigham healthcare network in Boston, MA with electronic health record (EHR) and genetic data. Recruitment for the Mass General Brigham Biobank launched in 2010 and is active at participating clinics at Brigham and Women’s Hospital, Massachusetts General Hospital, Spaulding...
Rehabilitation Hospital, Faulkner Hospital, McLean Hospital, Newton-Wellesley Hospital, and North Short Medical Center. All patients provided consent upon enrollment. Data from this protocol was approved by the Institutional Review Board. To date (07/2019), a total of 104,965 subjects were consented.

Genomic data for 30,683 participants were generated with the Illumina Multi-Ethnic Genotyping Array. The genotyping data were harmonized, and quality controlled with a three-step protocol, including two stages of genetic variant removal and an intermediate stage of sample exclusion. The example criteria for variants were: (1) missing call rate ≥0.05, (2) MAF < 0.001, and (3) deviation from Hardy–Weinberg equilibrium (P < 10^{-6}). The exclusion criteria for samples were: (1) sex discordanse between the reported and genetically predicted sex, (2) missing or infrequent sample ≥0.02, (3) subject relatedness (pairs with estimated identity-by-descent ≥0.125, from which we removed the individual with the highest proportion of missingness), and (4) population structure using the first four principal components. Phasing was performed with SHAPEIT290 and then imputations were performed with the Haplotype Reference Consortium Panel38 using the Michigan Implementation Server35. Written consent was provided by all study participants. Approval for analysis of Biobank data was obtained by Mass General Brigham IRB, protocol #2018P002276.

Participant ancestry was determined using TRAC32 and the Human Genome Diversity Project (HGDP)38 as a reference panel. Principal component analysis outliers were determined by using data from the largest available GWAS for ancestry33, and subsequently excluded from analysis. To correct for population stratification, we computed principal components using TRAC32 in the subset with genetically European ancestry. Furthermore, sample relatedness was determined using PLINK75, and subsequently one-sample related trio was identified and removed.

In aggregate, participants had a total of 7,422,726 ICD-9 and ICD-10 diagnostic codes corresponding to 784,878 instances of phecodes with at least 2 distinct values within the distribution of the study population, according to the first four principal components. Phasing was performed with SHAPEIT290 and then imputations were performed with the Haplotype Reference Consortium Panel38 using the Michigan Implementation Server35. Written consent was provided by all study participants. Approval for analysis of Biobank data was obtained by Mass General Brigham IRB, protocol #2018P002276.

Participant ancestry was determined using TRAC32 and the Human Genome Diversity Project (HGDP)38 as a reference panel. Principal component analysis outliers were determined by using data from the largest available GWAS for ancestry33, and subsequently excluded from analysis. To correct for population stratification, we computed principal components using TRAC32 in the subset with genetically European ancestry. Furthermore, sample relatedness was determined using PLINK75, and subsequently one-sample related trio was identified and removed.

In aggregate, participants had a total of 7,422,726 ICD-9 and ICD-10 diagnostic codes corresponding to 784,878 instances of phecodes with at least 2 distinct diagnostic codes. The most prevalent codes were 401.1 (essential hypertension: n = 11,397) and 745 (pain in joint: n = 10,333 cases). A total of 951 distinct phecodes from participants (n = 30,683) as determined from EHR. Same-day duplicated diagnoses were performed to adjust the napping proxies for their associations with insomnia and sleep duration35. We considered consistent effects across multiple methods to strengthen causal evidence.

(references)
23andMe Research Team
Michelle Agee4, Adam Auton4, Robert K. Bell4, Katarzyna Bryc4, Sarah K. Clark4, Sarah L. Elson4,
Kipper Fletez-Brant4, Pierre Fontanillas4, Nicholas A. Furlotte4, Pooja M. Gandhi4, Karl Heilbron4, Barry Hicks4,
David A. Hinds4, Karen E. Huber4, Ethan M. Jewett4, Yunxuan Jiang4, Aaron Kleinman4, Keng-Han Lin4,
Nadia K. Litterman4, Marie K. Luff4, Jennifer C. McCreight4, Matthew H. McIntyre4, Kimberly F. McManus4,
Joanna L. Mountain4, Sahar V. Mozaffari4, Priyanka Nandakumar4, Elizabeth S. Noblin4, Carrie A. M. Northover4,
Jared O’Connell4, Aaron A. Petrakovitz4, Steven J. Pitts4, G. David Poznik4, J. Fah Sathirapongsasuti4,
Anjali J. Shastri4, Janie F. Shelton4, Suyash Shringarpure4, Chao Tian4, Joyce Y. Tung4, Robert J. Tunney4,
Vladimir Vacic4, Xin Wang4 & Amir S. Zare4