nature genetics

Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits

Jacqueline M Lane[1–3], Jingjing Liang[4], Irma Vlasac[1,3], Simon G Anderson[5,6], David A Bechtold[7], Jack Bowden[8,9], Richard Emsley[10], Shubhroz Gill[3], Max A Little[11,12], Annemarie I Luik[13], Andrew Loudon[7], Frank A J L Scheer[14], Shaun M Purcell[3,15,16], Simon D Kyle[13], Deborah A Lawlor[8,9], Xiaofeng Zhu[4], Susan Redline[17], David W Ray[7], Martin K Rutter[7,18,19] & Richa Saxena[1–3,14,19]

Chronic sleep disturbances, associated with cardiometabolic diseases, psychiatric disorders and all-cause mortality[1,2], affect 25–30% of adults worldwide[3]. Although environmental factors contribute substantially to self-reported habitual sleep duration and disruption, these traits are heritable[4–9] and identification of the genes involved should improve understanding of sleep, mechanisms linking sleep to disease and development of new therapies. We report single- and multiple-trait genome-wide association analyses of self-reported sleep duration, insomnia symptoms and excessive daytime sleepiness in the UK Biobank (n = 112,586). We discover loci associated with insomnia symptoms (near *MEIS1*, *TMEM132E*, *CYCL1* and *TGFBI* in females and *WDR27* in males), excessive daytime sleepiness (near *AR–OPHN1*) and a composite sleep trait (near *PATJ* (*INADL*) and *HCRTR2*) and replicate a locus associated with sleep duration (at *PAX8*). We also observe genetic correlation between longer sleep duration and schizophrenia risk ($r_g = 0.29$, $P = 1.90 \times 10^{-13}$) and between increased levels of excessive daytime sleepiness and increased measures for adiposity traits (body mass index (BMI): $r_g = 0.20$, $P = 3.12 \times 10^{-9}$; waist circumference: $r_g = 0.20$, $P = 2.12 \times 10^{-7}$).

Rather than being 'secondary', disordered sleep may have an important role in the etiology and maintenance of physical and mental health[1,2]. Heritability has been estimated at ~40% for sleep duration[4,6–8], 25–45% for insomnia[9] and 17% for excessive daytime sleepiness[9], but few genetic factors are known. A Mendelian mutation in *BHLHE41* (encoding p.Pro385Arg) associated with short sleep duration has been identified and was confirmed in mouse models[10]. Genome-wide association studies (GWAS) for sleep duration have been reported[11–14], but only an association at the *PAX8* locus reached genome-wide significance and was confirmed across ancestry groups[12]. There are several reported loci associated with restless legs syndrome (RLS) and narcolepsy, but no robust genetic loci have been associated with insomnia symptoms or excessive daytime sleepiness[15,16].

We and others have performed a GWAS for chronotype in the UK Biobank[17,18] and a 23andMe participant sample[19]. To identify genetic variants that contribute to self-reported sleep duration, insomnia symptoms and excessive daytime sleepiness and link them with other conditions, we performed GWAS using phenotype measures in UK Biobank participants of European ancestry. Variation in sleep duration, insomnia symptoms and excessive daytime sleepiness was significantly associated with age, sex, principal components of ancestry, genotyping array, depression, psychiatric medication use, self-reported sleep apnea and BMI (**Supplementary Table 1**), as previously reported[20–23]. Together, age, sex and principal components explained 0.4%, 3.0% and 1.3% of variation in sleep duration, insomnia symptoms and excessive daytime sleepiness, respectively. Strong and significant pairwise phenotypic correlation was seen between the

[1]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [2]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA. [3]Broad Institute, Cambridge, Massachusetts, USA. [4]Department of Epidemiology and Biostatistics, School of Medicine, Case Western Reserve University, Cleveland, Ohio, USA. [5]Division of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. [6]George Institute for Global Health, University of Oxford, Oxford Martin School, Oxford, UK. [7]Division of Endocrinology, Diabetes and Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. [8]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK. [9]School of Social and Community Medicine, University of Bristol, Bristol, UK. [10]Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. [11]Engineering and Applied Science, Aston University, Birmingham, UK. [12]Media Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [13]Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [14]Division of Sleep and Circadian Disorders, Brigham and Women's Hospital and Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts, USA. [15]Department of Psychiatry, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [16]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [17]Department of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA. [18]Manchester Diabetes Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK. [19]These authors contributed equally to this work. Correspondence should be addressed to R.S. (rsaxena@partners.org).

Received 28 March; accepted 21 November; published online 19 December 2016; doi:10.1038/ng.3749

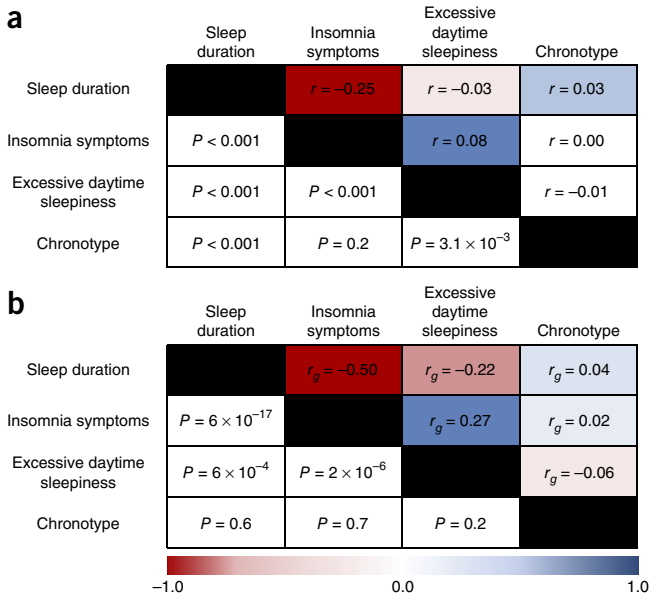

Figure 1 Sleep traits are phenotypically and genetically correlated. (**a**) Phenotypic correlation between the reported sleep traits, using Spearman correlation (r). (**b**) Genetic correlation (r_g) between the reported sleep traits, using LD score regression[67]. The color scale represents the strength of correlation. Chronotype ranges from extreme morning types to extreme evening types.

traits overall and within each sex, with limited correlation observed with chronotype (**Fig. 1** and **Supplementary Fig. 1**).

We performed GWAS analyses of sleep duration, insomnia symptoms and excessive daytime sleepiness using linear/logistic regression adjusting for age, sex, ten principal components and genotyping array. We identified 9 genome-wide significant ($P < 5 \times 10^{-8}$) and 14 suggestive ($P < 5 \times 10^{-7}$ to $P = 5 \times 10^{-8}$) loci (**Fig. 2**, **Table 1** and **Supplementary Figs. 2** and **3**). For sleep duration (n = 111,975), the strongest association was observed at the *PAX8* locus (rs62158211[T]: β (s.e.) = 2.34 (0.30) min/allele, $P = 4.7 \times 10^{-14}$, effect allele frequency (EAF) = 0.213; **Fig. 2a**), confirming a previously reported association (r^2 = 0.96, D' = 1 to lead SNP rs1823125 in 1000 Genomes Project CEU (European-ancestry) individuals)[12]. For insomnia symptoms (n = 32,155 cases and 26,973 controls), we observed significant associations within *MEIS1* (rs113851554[T]: odds ratio (OR) (95% confidence interval (CI)) = 1.26 (1.20–1.33), $P = 9.1 \times 10^{-19}$, EAF = 0.057; **Fig. 2b**), a homeobox gene implicated in motor neuron connectivity in *Drosophila melanogaster*[24,25], retinal and lens development in mouse[26] and substance P expression in the amygdala in humans[27]; near *TMEM132E* (rs145258459[C]: OR (95% CI) = 1.23 (1.13–1.35), $P = 2.1 \times 10^{-8}$, EAF = 0.983; **Fig. 2c**), which belongs to a gene family whose members have roles in brain development[28], panic/anxiety[29] and bipolar disorder[30], suggesting a link between insomnia symptoms and an underlying broader sensitivity to anxiety and stress; and near *CYCL1* (rs5922858[G]: OR (95% CI) = 1.12 (1.07–1.16), $P = 1.28 \times 10^{-8}$, EAF = 0.849; **Fig. 2d**), a locus previously associated ($P = 1 \times 10^{-6}$) with alcohol dependence comorbid with depressive symptoms[31]. Sex-stratified analyses identified an additional female-specific signal near *TGFBI* (rs3792900[C]: OR (95% CI) = 1.10 (1.07–1.14), $P = 2.16 \times 10^{-8}$, EAF = 0.470; **Table 1**, **Supplementary Fig. 3q,r** and **Supplementary Table 2**), which encodes an extracellular matrix protein responsible for human corneal dystrophy[32], and a male-specific signal near *WDR27* (rs13192566[G]: OR (95% CI) = 1.14 (1.09–1.20), $P = 3.2 \times 10^{-8}$, EAF = 0.860; **Table 1**, **Supplementary Figs. 3s,t** and **4**, and **Supplementary Table 2**), which encodes a scaffold protein. Independent associations at both loci are observed with type 1 diabetes, suggesting that these proteins may have a role in autoimmunity[33–35]. For excessive daytime sleepiness (n = 111,648), we identified a signal near the androgen receptor gene *AR* (rs73536079[T]: β (s.e.) = 0.634 (0.115) $P = 3.94 \times 10^{-8}$, EAF = 0.002; **Fig. 2e**), with no sex-specific effects. Secondary analyses after additional adjustment for depression or BMI identified a signal near *ROBO1* (depression adjustment n = 107,440; rs182765975[T]: β (s.e.) = 0.099 (0.018), $P = 3.33 \times 10^{-8}$, EAF = 0.003; **Table 1** and **Supplementary Fig. 3o**), which encodes a neuronal axon guidance receptor previously implicated in dyslexia[36], and a signal near another member of the TMEM132 family, *TMEM132B* (BMI adjustment n = 75,480; rs142261172[A]: β (s.e.) = 0.106(0.018), $P = 9.06 \times 10^{-9}$, EAF = 0.004; **Table 1** and **Supplementary Fig. 3p**). Conditional analyses did not identify independent association signals (**Supplementary Table 3**). Sensitivity analyses adjusting for factors influencing sleep traits, including self-reported sleep apnea, depression, psychiatric medication use, smoking, socioeconomic status, employment status, marital status and snoring, did not significantly alter results for primary association signals (**Supplementary Table 4**).

The leading associations overlap interesting candidate genes whose expression is enriched in hypocretin-expressing neurons in mice and zebrafish[37,38], that are differentially expressed in sleep-deprived rats[39] and/or that regulate sleep in *Drosophila*[40]. Credible set analyses[41] highlighted a number of potential causal variants at each locus (**Table 1**), and future experimental studies will be necessary to characterize the functions of these variants in regulating sleep traits. Bioinformatic annotations[42] offer an initial opportunity for *in silico* functional interpretation of the variants (**Supplementary Fig. 5** and **Supplementary Table 5**). For example, multiple variants for all three traits are predicted to disrupt binding of FOXP1, a neural transcriptional repressor implicated in intellectual disability, autism and language impairment[43]. Interestingly, the locus associated with sleep duration encompassing *PAX8* is adjacent to the only chromosomal fusion site that arose since the divergence of humans from other hominids ~5 million years ago[44,45], and the new genomic structure created by this unique evolutionary history may have a causal role in sleep duration. Pathway analysis[46] of significant and suggestive loci identified enrichment of genes associated with immune, neurodevelopmental, pituitary and communication disorders ($P < 0.01$) and genes enriched for binding sites for the stress-responsive transcription factor heat shock factor 1 (HSF1) and the endoplasmic reticulum stress- and unfolded-protein-responsive transcription factor HERPUD1 (**Supplementary Tables 6** and **7**).

Aside from the lead SNPs in the *PAX8* region and a variant in the *DRD2* region[47] for sleep duration, we found limited evidence of association for previously published candidate gene or GWAS signals ($P_{meta} < 5 \times 10^{-5}$; **Supplementary Table 8**) or for regions encompassing core clock genes (**Supplementary Fig. 6**). Our findings for the GWAS on sleep duration largely overlap with those of Jones *et al.*[18], despite differences in exclusion criteria and analytical approach. In particular, our study excluded shift workers (n = 6,557), sleep medication users (n = 1,184) and first- to third-degree relatives (n = 7,980), whereas the linear mixed-model analyses by Jones *et al.*[18] included these populations, leading to a larger sample size (n = 127,573). Likely because of this increased power, Jones *et al.*[18] identified two additional signals at *VRK2* that did not attain genome-wide significance in our study (rs1380703[A]: β (s.e.) = 1.50 (0.30) min/allele, $P = 8.43 \times 10^{-8}$ and rs17190618[T]: β (s.e.) = 1.60 (0.34) min/allele, $P = 3.80 \times 10^{-6}$).

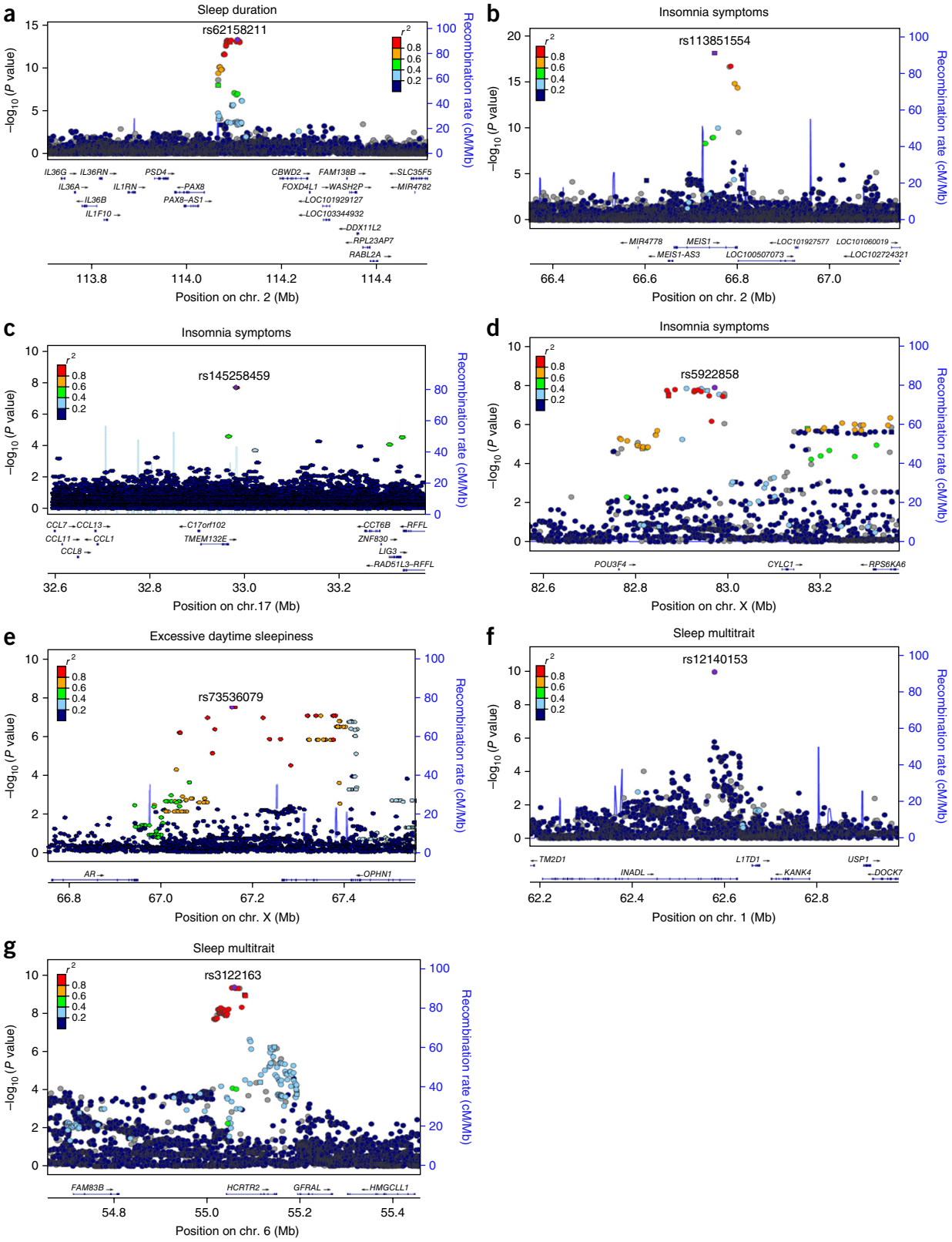

Figure 2 Regional association plots for genome-wide significant loci. (**a**) Sleep duration. (**b**–**d**) Insomnia symptoms. (**e**) Excessive daytime sleepiness. (**f**,**g**) Composite trait of sleep duration, insomnia symptoms, excessive daytime sleepiness and chronotype. Chromosomal position is indicated on the *x* axis, and the $-\log_{10}$ *P* value for each SNP (filled circles or squares) is indicated on the *y* axis, with the lead SNP shown in purple (the 400-kb window around each lead SNP is shown). The genes within each region are shown in the lower panel. The blue lines represent the recombination rate. Additional SNPs in each locus are colored according to linkage disequilibrium (r^2) with the lead SNP (estimated by LocusZoom on the basis of HapMap CEU haplotypes or within UK Biobank (**c**)). Squares represent directly genotyped SNPs, and circles represent imputed SNPs.

Table 1 Genome-wide significant ($P < 5 \times 10^{-8}$) and suggestive ($P < 5 \times 10^{-7}$) loci associated with sleep duration, insomnia symptoms and excessive daytime sleepiness in subjects of European ancestry in the UK Biobank

SNP	Chr:position NCBI Build 37	Nearest gene(s)	Alleles (E/A)	EAF	INFO	β or OR	s.e. or CI	P	Most likely causal SNPs (probability)[a]
Sleep duration (n = 111,975)									
rs62158211	**2:114,106,139**	***PAX8***	**T/G**	**0.213**	0.99	**0.039**	**0.005**	**4.72×10^{-14}**	rs62158211 (0.16), rs62158213 (0.16), rs4618068 (0.16), rs1807282 (0.16), rs56093896 (0.16)
rs1380703	2:57,941,287	*VRK2, LOC647016, LOC100131953*	A/G	0.618	0.89	0.025	0.005	8.44×10^{-8}	rs1380703 (1)
rs10953765	7:114,291,435	*FOXP2*	G/A	0.447	0.98	0.022	0.004	2.96×10^{-7}	rs10953765 (0.27), rs1456031 (0.14)
rs146977851	10:56,570,954	*PCDH15*	C/T	0.971	0.97	0.065	0.013	3.53×10^{-7}	rs146977851 (0.85), rs75334053 (0.14)
rs61980273	14:94,218,949	*PRIMA1, UNC79*	A/G	0.039	1.00	0.058	0.011	1.30×10^{-7}	rs61980273 (1)
Insomnia symptoms (n up to 31,767 cases and 26,935 controls)									
rs576106307	1:18,007,282	*ARHGEF10L*	C/CT	0.934	0.89	1.07	1.10–1.04	2.66×10^{-7}	rs576106307 (1)
rs113851554	**2:66,750,564**	***MEIS1***	**T/G**	**0.057**	1.00	**1.26**	**1.20–1.33**	**9.11×10^{-19}**	rs113851554 (0.98)
rs376775068	8:145,604,659	*ADCK5*	G/C	0.934	0.67	1.11	1.16–1.06	6.81×10^{-8}	rs376775068 (1)
rs145258459	**17:32,986,155**	***TMEM132E***	**C/T**	**0.983**	0.69	**1.23**	**1.13–1.35**	**2.13×10^{-8}**	rs145258459 (1)
rs531814036	17:43219921	*ACBD4*	C/CT	0.419	0.91	1.06	1.03–1.08	2.92×10^{-7}	rs531814036 (1)
rs5922858	**X:82,971,008**	***CYCL1***	**G/T**	**0.849**	**0.99**	**1.12**	**1.07–1.16**	**1.28×10^{-8}**	rs5922858 (1)
rs13192566[b]	**6:169,961,635**	***WDR27***	**G/C**	**0.860**	**0.99**	**1.14**	**1.09–1.20**	**3.17×10^{-8}**	rs13192566 (0.50), rs13208844 (0.50)
rs3792900[c]	**5:135,393,754**	***TGFBI***	**C/T**	**0.470**	**0.99**	**1.1**	**1.07–1.14**	**2.16×10^{-8}**	rs3792900 (0.14), rs6894815 (0.07)
Excessive daytime sleepiness (n < 111,648)									
rs192315283	1:59,531,543	*HSD52*	C/T	0.010	0.76	0.126	0.025	3.55×10^{-7}	rs192315283 (1)
rs76645968	2:53,827,686	*ASB3*	G/C	0.977	0.99	0.073	0.014	1.79×10^{-7}	rs76645968 (0.26), rs12328289 (0.26)
rs920065	3:5,893,776	*MRPS35P1, MRPS36P1*	C/G	0.824	0.96	0.028	0.006	4.25×10^{-7}	rs920065 (0.49)
rs115320831	4:159,178,375	*TMEM144*	A/G	0.702	0.98	0.024	0.005	3.68×10^{-7}	rs115320831 (0.58)
rs35309287	5:146,775,386	*DPYSL3*	TA/T	0.970	0.94	0.067	0.013	1.25×10^{-7}	rs35309287 (0.45), rs34398961 (0.45)
rs189689339	6:82,375,372	*FAM46A*	T/C	0.003	0.67	0.226	0.044	2.13×10^{-7}	rs189689339 (1)
rs17507216	15:83,226,925	*CPEB1*	A/G	0.232	1.00	0.026	0.005	1.59×10^{-7}	rs17507216 (0.20), rs72751643 (0.11)
rs73536079	**X:67,154,206**	***AR, OPHN1***	**T/G**	**0.002**	**0.90**	**0.634**	**0.115**	**3.94×10^{-8}**	rs73536079 (1)
rs182765975[d]	**3:78,538,431**	***ROBO1***	**T/G**	**0.003**	**0.86**	**0.099**	**0.018**	**3.33×10^{-8}**	rs182765975 (0.33), rs191435135 (0.33), rs182979911 (0.33)
rs14226117[e]	**12:126,049,981**	***TMEM132B***	**A/G**	**0.004**	**0.92**	**0.106**	**0.018**	**9.06×10^{-9}**	rs142261172 (0.50), rs189248622 (0.50)

E, effect allele; A, alternative allele; chr, chromosome; OR, odds ratio; CI, confidence interval; INFO, imputation quality from IMPUTE2; EAF, effect allele frequency. Increasing β or odds ratio values correspond to longer sleep duration in hours, increased insomnia symptoms and increased daytime sleepiness. Analyses were adjusted for age, sex, genetic ancestry and genotyping array. Bold denotes genome-wide significant signals ($P < 5 \times 10^{-8}$).

[a]Using PICS. [b]Males only. [c]Females only. [d]Secondary analysis with additional adjustment for depression. [e]Secondary analysis with additional adjustment for BMI.

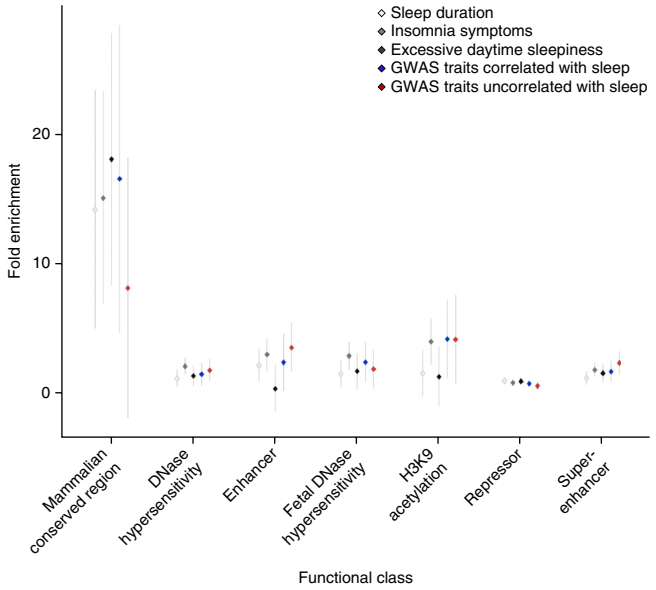

Figure 3 Partitioning of the genetic architecture of sleep duration, insomnia symptoms and excessive daytime sleepiness across functional annotation categories. For each trait, fold enrichment estimates for the main annotations from LD score regression[50] are indicated on the *y* axis across functional annotation classes on the *x* axis. Error bars represent the 95% confidence intervals around the estimates. We tested 25 functional annotations; annotations passing the multiple-testing significance threshold ($P < 0.005$) are shown. For context, the average enrichment across functional annotation categories is shown for nine traits with significant genetic correlation to at least one sleep trait (GWAS traits correlated with sleep; including GWAS for BMI, waist circumference, birth weight, depression, educational attainment, three glycemic traits in non-diabetics and schizophrenia) and for five traits with no significant genetic correlation to any sleep trait (GWAS traits uncorrelated with sleep; including GWAS for Alzheimer disease, type 2 diabetes, autism, rheumatoid arthritis and height). H3K9, histone H3 lysine 9.

Trait heritability calculated as the proportion of trait variance due to additive genetic factors measured here (observed-scale SNP heritability, h^2 (s.e.)) was 10.3% (0.006%) for sleep duration, 20.6% (0.011%) for insomnia symptoms and 8.4% (0.006%) for daytime sleepiness (BOLT-REML variance-components analysis[48]). LD score regression analysis[49] confirmed no residual population stratification (intercept (s.e.): sleep duration, 1.012 (0.008); insomnia symptoms, 1.003 (0.008); excessive daytime sleepiness, 1.005 (0.007)). Tests for enrichment of heritability by functional class using an LD score regression approach[50] identified excess heritability across active transcriptional regions for insomnia symptoms and across genomic regions conserved in mammals for all three sleep traits. Consistent with these findings, heritability enrichment in conserved regions was seen for traits demonstrating significant genetic correlation with sleep (**Fig. 3** and **Supplementary Table 9**).

Sleep duration, insomnia symptoms, excessive daytime sleepiness and chronotype are significantly correlated, both at the phenotype and genetic levels (**Fig. 1**), with greater pairwise correlations in males than in females (**Supplementary Fig. 1**). Thus, to find loci common to sleep traits, we performed a multiple-trait GWAS[51]. We identified two new association signals near *HCRTR2* and *PATJ* (*INADL*) and found that *PAX8* and *MEIS1* associations influenced multiple sleep traits (**Fig. 2**, **Table 2** and **Supplementary Fig. 7**). *HCRTR2* encodes hypocretin receptor 2, the primary receptor of two receptors for wake-promoting orexin neuropeptides[52] involved in narcolepsy and regulation of sleep. Notably, the minor allele at rs3122163 (C) showed subthreshold association with shorter sleep duration and morningness chronotype, suggesting gain of function, but no association with insomnia symptoms. Assessment of objective sleep measures and functional and physiological follow-up should yield important insights into orexin receptor signaling, a pathway important for the pharmacological treatment of narcolepsy[53] and insomnia[54]. *PATJ* encodes a membrane protein involved in the formation of tight junctions and is implicated in photoreception in mice and *Drosophila*[55,56]. The INADL protein is reported to interact with HTR_{2A} (ref. 57), a serotonin receptor with a known role in sleep regulation[58,59].

Our strongest association for insomnia symptoms fell within *MEIS1*, a locus previously associated with RLS in GWAS[60]. Our lead SNP, rs113851554, and the correlated 3′ UTR variant rs11693221 (pairwise $r^2 = 0.69$, $D' = 0.90$ in 1000 Genomes Project European (EUR) individuals) represent the strongest known genetic risk factor for RLS and were identified in *MEIS1* sequencing studies[61,62] following up the original RLS GWAS signal (rs2300478)[60,63]. Conditional analysis suggests that only one underlying signal, detected by the lead SNP rs113851554 in our GWAS, explains the association of all three SNPs with insomnia symptoms (**Supplementary Fig. 8** and **Supplementary Table 10**). To further investigate the extent of overlap between RLS and insomnia symptoms, we tested a weighted genetic risk score (GRS) for RLS[64,65] and found that it was also associated with insomnia symptoms with concordant direction of allelic effects (OR (95% CI) = 1.06 (1.05–1.07) per RLS risk allele, $P = 1.17 \times 10^{-21}$; **Supplementary Table 11**). Weighting of the RLS GWAS alleles by SNP effect on periodic limb movements (PLMs) did not substantially alter the overall results (**Supplementary Table 11**). Interestingly, recent data indicating that thalamic glutamatergic activity is increased in RLS provide evidence of an underlying propensity for hyperarousal in RLS[66], which is also a core feature of insomnia. Future analyses of pairwise bidirectional causal effects for all three traits will be necessary to determine whether shared genetic associations correspond to causality, partial mediation or pleiotropy.

Strong epidemiological associations of sleep duration, insomnia symptoms and daytime sleepiness have been observed with disease traits, but the extent to which the underlying genetics are shared is unknown. Therefore, we tested for genome-wide genetic correlation between findings from our sleep GWAS and those from publicly available GWAS for 20 phenotypes spanning a range of cognitive, neuropsychiatric, anthropometric, cardiometabolic and autoimmune traits using LD score regression[67] (**Fig. 4** and **Supplementary Table 12**).

Genetic correlations demonstrated a strong biological link between longer sleep duration and risk of schizophrenia ($r_g = 0.29$, $P = 1 \times 10^{-13}$), as suggested by previous reports[18,47,68]. Furthermore, a schizophrenia GRS (96 variants) was associated with longer sleep duration (β (s.e.) = 1.44 (0.36) min/allele, $P = 2.56 \times 10^{-4}$) (interquartile range = 2.3 h), although a variety of sleep behaviors are seen in patients with schizophrenia[69–71]. Significant genetic correlation between low birth weight and longer sleep duration ($r_g = -0.27$, $P = 1 \times 10^{-4}$) may reflect shared links between genetically determined insulin secretion or action pathways underlying fetal growth[72,73] and long sleep duration. In support of this hypothesis, significant genetic correlation was observed by Jones *et al.*[18] between oversleeping and both fasting insulin and risk of type 2 diabetes in UK Biobank. Genetic correlation between sleep duration and Crohn's disease risk ($r_g = 0.18$, $P = 1 \times 10^{-3}$) is also consistent with epidemiological observations[74].

Significant genetic correlation was also found between increased insomnia symptoms and major depression, adverse glycemic traits, increased adiposity and fewer years of education and between excessive daytime sleepiness and increased adiposity (all $P < 1 \times 10^{-3}$), further

Table 2 Genome-wide significant ($P < 5 \times 10^{-8}$) loci associated with a multiple-phenotype model of sleep duration, insomnia symptoms, excessive daytime sleepiness and categorical chronotype in subjects of European ancestry in UK Biobank

SNP	Chr:position NCBI Build 37	Nearest gene	Alleles (E/A)	EAF	INFO	Multitrait *P*	Trait	*β* or OR	s.e. or CI	*P*	Causal SNPs (probability)
rs12140153	1:62,352,479	*PATJ*	T/G	0.099	0.93	1.06×10^{-10}	Sleep duration	–0.009	0.007	0.22	rs12140153 (1)
							Insomnia symptoms	1.039	0.99–1.08	0.05	
							Excessive daytime sleepiness	–0.036	0.007	6.60×10^{-7}	
							Chronotype	0.036	0.008	2.59×10^{-6}	
rs76681500	1:77,247,749	*AK5*	A/G	0.159	0.99	1.03×10^{-9}	Sleep duration	–0.002	0.006	0.79	rs76681500 (0.57)
							Insomnia symptoms	0.98	0.95–1.011	0.27	
							Excessive daytime sleepiness	–0.008	0.006	0.15	
							Chronotype	–0.043	0.006	1.50×10^{-12}	
rs694383	1:180,834,827	*RGS16*	C/G	0.030	1.00	2.72×10^{-11}	Sleep duration	0.018	0.012	0.14	rs694383 (0.22), rs509476 (0.22), rs1144566 (0.22), rs12743617 (0.22)
							Insomnia symptoms	0.98	0.91–1.048	0.75	
							Excessive daytime sleepiness	–0.009	0.012	0.47	
							Chronotype	0.099	0.013	2.61×10^{-14}	
rs113851554	2:66,523,432	*MEIS1*	T/G	0.056	1.00	3.97×10^{-16}	Sleep duration	0.001	0.009	0.95	rs113851554 (0.96)
							Insomnia symptoms	1.264	1.20–1.329	9.11×10^{-19}	
							Excessive daytime sleepiness	–0.002	0.009	0.85	
							Chronotype	0.033	0.01	5.64×10^{-4}	
rs62158211	2:113,822,609	*PAX8*	T/G	0.214	0.99	8.18×10^{-13}	Sleep duration	0.039	0.005	4.72×10^{-14}	rs62158211 (0.15), rs62158213 (0.15), rs4618068 (0.15), rs1807282 (0.15), rs56093896 (0.15)
							Insomnia symptoms	0.943	0.91–0.969	1.31×10^{-5}	
							Excessive daytime sleepiness	0.005	0.005	0.37	
							Chronotype	0.014	0.005	7.93×10^{-3}	
rs3122163	6:55,164,327	*HCRTR2*	T/C	0.768	0.99	4.18×10^{-10}	Sleep duration	0.019	0.005	9.97×10^{-5}	rs3122163 (0.083), rs34694541 (0.083), rs3122170 (0.083)
							Insomnia symptoms	0.984	0.95–1.011	0.52	
							Excessive daytime sleepiness	–0.023	0.005	5.51×10^{-6}	
							Chronotype	0.021	0.005	8.68×10^{-5}	

E, effect allele; A, alternative allele; chr, chromosome; OR, odds ratio; CI, confidence interval; INFO, imputation quality from IMPUTE2; EAF, effect allele frequency. Increasing *β* or odds ratio values indicate longer sleep duration, increased insomnia symptoms, increased daytime sleepiness and later chronotype.

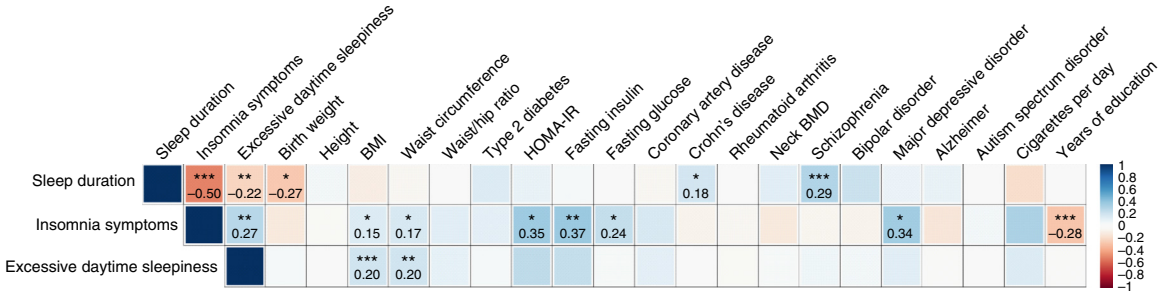

Figure 4 Genetic architectures shared between sleep duration, insomnia symptoms or excessive daytime sleepiness and 20 behavioral and disease traits. LD score regression[67] estimates of genetic correlation (r_g) were obtained by comparing our GWAS findings for sleep duration, insomnia symptoms and excessive daytime sleepiness with summary statistics from 20 publicly available GWAS for psychiatric and metabolic disorders, immune diseases and other traits of natural variation. Blue, positive genetic correlation; red, negative genetic correlation; r_g values are displayed for significant correlations. Larger squares correspond to more significant P values. Genetic correlations that were significantly different from zero after Bonferroni correction are marked (*$P < 1 \times 10^{-3}$, **$P < 1 \times 10^{-5}$, ***$P < 1 \times 10^{-7}$); the P-value cutoff after Bonferroni correction was 0.0025. All genetic correlations in this report can be found in tabular form in **Supplementary Table 12**. BMI, body mass index; BMD, bone mineral density, HOMA-IR, homeostatic model assessment of insulin resistance.

highlighting biological overlap of sleep traits with metabolic traits, psychiatric traits and educational attainment[17]. In support of this overlap, studies have shown that experimentally suppressing slow-wave sleep leads to decreased insulin sensitivity and impaired glucose tolerance[75,76]. Notably, a GRS for fasting insulin was not significantly associated with insomnia symptoms (7 SNPs, OR (95% CI) = 1.01 (0.99–1.03), $P = 0.51$). Finally, consistent with a well-established but poorly understood link between excessive daytime sleepiness and obesity[77,78], a GRS for BMI was associated with excessive daytime sleepiness (95 SNPs, β (s.e.) = 0.002 (0.0004) sleepiness category/allele, $P = 1.67 \times 10^{-4}$) but not with insomnia symptoms (OR (95% CI) = 1.00 (0.998–1.003), $P = 0.73$).

Moving forward, replication and systematic testing of genetic correlations in larger samples will be needed. Notably, genetic correlation testing between insomnia and RLS should be examined but was not possible here because RLS consortium GWAS results were not available. Additionally, identifying causal relationships between genetically correlated traits may be difficult, and findings using Mendelian randomization approaches will need cautious interpretation given potential selection biases in UK Biobank[79–81].

In summary, in a GWAS of sleep traits, we identified new genetic loci that point to previously unstudied variants that might modulate the hypocretin–orexin system and retinal development and influence genes expressed in the cerebral cortex. Furthermore, genome-wide analysis suggests that sleep traits have underlying genetic pathways in common with neuropsychiatric and metabolic diseases. This work should advance understanding of the molecular processes underlying sleep disturbances and open new avenues of treatment for sleep disorders and related disorders.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This research has been conducted using the UK Biobank Resource under application number 6818. We would like to thank the participants and researchers from the UK Biobank who contributed or collected data. This work was supported by US NIH grants R01DK107859 (R.S.), R21HL121728 (R.S.), F32DK102323 (J.M.L.), R01HL113338 (J.M.L., S.R. and R.S.), R01DK102696 (R.S. and F.A.J.L.S.), R01DK105072 (R.S. and F.A.J.L.S.), T32HL007567 (J.L.) and HG003054 (X.Z.), the University of Manchester (Research Infrastructure Fund), the Wellcome Trust (salary support for D.W.R. and A.L.) and UK Medical Research Council MC_UU_12013/5 (D.A.L.). Data on glycemic traits have been contributed by MAGIC investigators and were downloaded from http://www.magicinvestigators.org/. Data on coronary artery disease and myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and were downloaded from http://www.cardiogramplusc4d.org/. We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses.

AUTHOR CONTRIBUTIONS

J.M.L., M.K.R. and R.S. designed the study. J.M.L., J.L., I.V. and R.S. performed genetic analyses. J.M.L. and R.S. wrote the manuscript, and all co-authors helped interpret data and 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.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Fernandez-Mendoza, J. & Vgontzas, A.N. Insomnia and its impact on physical and mental health. *Curr. Psychiatry Rep.* **15**, 418 (2013).
2. Luyster, F.S., Strollo, P.J. Jr., Zee, P.C. & Walsh, J.K. Sleep: a health imperative. *Sleep* **35**, 727–734 (2012).
3. Stranges, S., Tigbe, W., Gómez-Olivé, F.X., Thorogood, M. & Kandala, N.B. Sleep problems: an emerging global epidemic? Findings from the INDEPTH WHO-SAGE study among more than 40,000 older adults from 8 countries across Africa and Asia. *Sleep* **35**, 1173–1181 (2012).
4. de Castro, J.M. The influence of heredity on self-reported sleep patterns in free-living humans. *Physiol. Behav.* **76**, 479–486 (2002).
5. Evans, D.S. *et al.* Habitual sleep/wake patterns in the Old Order Amish: heritability and association with non-genetic factors. *Sleep* **34**, 661–669 (2011).
6. Heath, A.C., Eaves, L.J., Kirk, K.M. & Martin, N.G. Effects of lifestyle, personality, symptoms of anxiety and depression, and genetic predisposition on subjective sleep disturbance and sleep pattern. *Twin Res.* **1**, 176–188 (1998).
7. Heath, A.C., Kendler, K.S., Eaves, L.J. & Martin, N.G. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. *Sleep* **13**, 318–335 (1990).
8. Partinen, M., Kaprio, J., Koskenvuo, M., Putkonen, P. & Langinvainio, H. Genetic and environmental determination of human sleep. *Sleep* **6**, 179–185 (1983).
9. Wing, Y.K. *et al.* Familial aggregation and heritability of insomnia in a community-based study. *Sleep Med.* **13**, 985–990 (2012).
10. He, Y. *et al.* The transcriptional repressor DEC2 regulates sleep length in mammals. *Science* **325**, 866–870 (2009).
11. Gottlieb, D.J., O'Connor, G.T. & Wilk, J.B. Genome-wide association of sleep and circadian phenotypes. *BMC Med. Genet.* **8** (Suppl. 1), S9 (2007).
12. Gottlieb, D.J. *et al.* Novel loci associated with usual sleep duration: the CHARGE Consortium Genome-Wide Association Study. *Mol. Psychiatry* **20**, 1232–1239 (2015).
13. 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).
14. Allebrandt, K.V. *et al.* A K_{ATP} channel gene effect on sleep duration: from genome-wide association studies to function in *Drosophila*. *Mol. Psychiatry* **18**, 122–132 (2013).

15. Gehrman, P.R., Keenan, B.T., Byrne, E.M. & Pack, A.I. Genetics of sleep disorders. *Psychiatr. Clin. North Am.* **38**, 667–681 (2015).
16. Sehgal, A. & Mignot, E. Genetics of sleep and sleep disorders. *Cell* **146**, 194–207 (2011).
17. Lane, J.M. *et al.* Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nat. Commun.* **7**, 10889 (2016).
18. Jones, S.E. *et al.* Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS Genet.* **12**, e1006125 (2016).
19. 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).
20. Pemberton, R. & Fuller Tyszkiewicz, M.D. Factors contributing to depressive mood states in everyday life: a systematic review. *J. Affect. Disord.* **200**, 103–110 (2016).
21. Foral, P., Knezevich, J., Dewan, N. & Malesker, M. Medication-induced sleep disturbances. *Consult Pharm.* **26**, 414–425 (2011).
22. Rosenberg, R.P. Clinical assessment of excessive daytime sleepiness in the diagnosis of sleep disorders. *J. Clin. Psychiatry* **76**, e1602 (2015).
23. Gonnissen, H.K. *et al.* Sleep duration, sleep quality and body weight: parallel developments. *Physiol. Behav.* **121**, 112–116 (2013).
24. Kurant, E. *et al.* Dorsotonals/homothorax, the *Drosophila* homologue of meis1, interacts with extradenticle in patterning of the embryonic PNS. *Development* **125**, 1037–1048 (1998).
25. Casares, F. & Mann, R.S. Control of antennal versus leg development in *Drosophila*. *Nature* **392**, 723–726 (1998).
26. Hisa, T. *et al.* Hematopoietic, angiogenic and eye defects in *Meis1* mutant animals. *EMBO J.* **23**, 450–459 (2004).
27. Davidson, S., Miller, K.A., Dowell, A., Gildea, A. & Mackenzie, A. A remote and highly conserved enhancer supports amygdala specific expression of the gene encoding the anxiogenic neuropeptide substance-P. *Mol. Psychiatry* **11**323, 410–421 (2006).
28. Oh-hashi, K., Naruse, Y., Amaya, F., Shimosato, G. & Tanaka, M. Cloning and characterization of a novel GRP78-binding protein in the rat brain. *J. Biol. Chem.* **278**, 10531–10537 (2003).
29. Erhardt, A. *et al.* Replication and meta-analysis of *TMEM132D* gene variants in panic disorder. *Transl. Psychiatry* **2**, e156 (2012).
30. Sklar, P. *et al.* Whole-genome association study of bipolar disorder. *Mol. Psychiatry* **13**, 558–569 (2008).
31. Edwards, A.C. *et al.* Genome-wide association study of comorbid depressive syndrome and alcohol dependence. *Psychiatr. Genet.* **22**, 31–41 (2012).
32. Han, K.E. *et al.* Pathogenesis and treatments of *TGFBI* corneal dystrophies. *Prog. Retin. Eye Res.* **50**, 67–88 (2016).
33. Bradfield, J.P. *et al.* A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet.* **7**, e1002293 (2011).
34. Patry, M. *et al.* βig-h3 represses T-cell activation in type 1 diabetes. *Diabetes* **64**, 4212–4219 (2015).
35. Han, B. *et al. TGFBI* (βIG-H3) is a diabetes-risk gene based on mouse and human genetic studies. *Hum. Mol. Genet.* **23**, 4597–4611 (2014).
36. Poelmans, G., Buitelaar, J.K., Pauls, D.L. & Franke, B. A theoretical molecular network for dyslexia: integrating available genetic findings. *Mol. Psychiatry* **16**, 365–382 (2011).
37. Dalal, J. *et al.* Translational profiling of hypocretin neurons identifies candidate molecules for sleep regulation. *Genes Dev.* **27**, 565–578 (2013).
38. Yelin-Bekerman, L. *et al.* Hypocretin neuron–specific transcriptome profiling identifies the sleep modulator Kcnh4a. *eLife* **4**, e08638 (2015).
39. Mackiewicz, M. *et al.* Macromolecule biosynthesis: a key function of sleep. *Physiol. Genomics* **31**, 441–457 (2007).
40. Takahama, K. *et al.* Pan-neuronal knockdown of the c-Jun N-terminal kinase (JNK) results in a reduction in sleep and longevity in *Drosophila*. *Biochem. Biophys. Res. Commun.* **417**, 807–811 (2012).
41. Farh, K.K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2015).
42. Ward, L.D. & Kellis, M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* **44** D1, D877–D881 (2016).
43. Hamdan, F.F. *et al. De novo* mutations in *FOXP1* in cases with intellectual disability, autism, and language impairment. *Am. J. Hum. Genet.* **87**, 671–678 (2010).
44. Fan, Y., Newman, T., Linardopoulou, E. & Trask, B.J. Gene content and function of the ancestral chromosome fusion site in human chromosome 2q13–2q14.1 and paralogous regions. *Genome Res.* **12**, 1663–1672 (2002).
45. Fan, Y., Linardopoulou, E., Friedman, C., Williams, E. & Trask, B.J. Genomic structure and evolution of the ancestral chromosome fusion site in 2q13–2q14.1 and paralogous regions on other human chromosomes. *Genome Res.* **12**, 1651–1662 (2002).
46. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
47. Cade, B.E. *et al.* Common variants in *DRD2* are associated with sleep duration: the CARe consortium. *Hum. Mol. Genet.* **25**, 167–179 (2016).
48. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
49. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
50. Finucane, H.K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
51. Zhu, X. *et al.* Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. *Am. J. Hum. Genet.* **96**, 21–36 (2015).
52. Mignot, E. Sleep, sleep disorders and hypocretin (orexin). *Sleep Med.* **5** (Suppl. 1), S2–S8 (2004).
53. Thompson, M.D., Xhaard, H., Sakurai, T., Rainero, I. & Kukkonen, J.P. OX_1 and OX_2 orexin/hypocretin receptor pharmacogenetics. *Front. Neurosci.* **8**, 57 (2014).
54. Herring, W.J. *et al.* Suvorexant in patients with insomnia: results from two 3-month randomized controlled clinical trials. *Biol. Psychiatry* **79**, 136–148 (2016).
55. Shieh, B.H. & Niemeyer, B. A novel protein encoded by the *InaD* gene regulates recovery of visual transduction in *Drosophila*. *Neuron* **14**, 201–210 (1995).
56. Peirson, S.N. *et al.* Microarray analysis and functional genomics identify novel components of melanopsin signaling. *Curr. Biol.* **17**, 1363–1372 (2007).
57. Bécamel, C. *et al.* The serotonin 5-HT_{2A} and 5-HT_{2C} receptors interact with specific sets of PDZ proteins. *J. Biol. Chem.* **279**, 20257–20266 (2004).
58. Sharpley, A.L., Elliott, J.M., Attenburrow, M.J. & Cowen, P.J. Slow wave sleep in humans: role of 5-HT_{2A} and 5-HT_{2C} receptors. *Neuropharmacology* **33**, 467–471 (1994).
59. Rosenberg, R. *et al.* APD125, a selective serotonin 5-HT_{2A} receptor inverse agonist, significantly improves sleep maintenance in primary insomnia. *Sleep* **31**, 1663–1671 (2008).
60. Winkelmann, J. *et al.* Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat. Genet.* **39**, 1000–1006 (2007).
61. Xiong, L. *et al. MEIS1* intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. *Hum. Mol. Genet.* **18**, 1065–1074 (2009).
62. Schulte, E.C. *et al.* Targeted resequencing and systematic *in vivo* functional testing identifies rare variants in *MEIS1* as significant contributors to restless legs syndrome. *Am. J. Hum. Genet.* **95**, 85–95 (2014).
63. Spieler, D. *et al.* Restless legs syndrome–associated intronic common variant in *Meis1* alters enhancer function in the developing telencephalon. *Genome Res.* **24**, 592–603 (2014).
64. Moore, H. IV *et al.* Periodic leg movements during sleep are associated with polymorphisms in *BTBD9*, *TOX3*/*BC034767*, *MEIS1*, *MAP2K5*/*SKOR1*, and *PTPRD*. *Sleep* **37**, 1535–1542 (2014).
65. Winkelmann, J. *et al.* Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet.* **7**, e1002171 (2011).
66. Allen, R.P., Barker, P.B., Horská, A. & Earley, C.J. Thalamic glutamate/glutamine in restless legs syndrome: increased and related to disturbed sleep. *Neurology* **80**, 2028–2034 (2013).
67. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
68. Byrne, E.M., Gehrman, P.R., Trzaskowski, M., Tiemeier, H. & Pack, A.I. Genetic correlation analysis suggests association between increased self-reported sleep duration in adults and schizophrenia and type 2 diabetes. *Sleep* **39**, 1853–1857 (2016).
69. Wulff, K., Dijk, D.J., Middleton, B., Foster, R.G. & Joyce, E.M. Sleep and circadian rhythm disruption in schizophrenia. *Br. J. Psychiatry* **200**, 308–316 (2012).
70. Poulin, J. *et al.* Sleep habits in middle-aged, non-hospitalized men and women with schizophrenia: a comparison with healthy controls. *Psychiatry Res.* **179**, 274–278 (2010).
71. Chouinard, S., Poulin, J., Stip, E. & Godbout, R. Sleep in untreated patients with schizophrenia: a meta-analysis. *Schizophr. Bull.* **30**, 957–967 (2004).
72. Hattersley, A.T. & Tooke, J.E. The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* **353**, 1789–1792 (1999).
73. Horikoshi, M. *et al.* New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. *Nat. Genet.* **45**, 76–82 (2013).
74. Ananthakrishnan, A.N. *et al.* Sleep duration affects risk for ulcerative colitis: a prospective cohort study. *Clin. Gastroenterol. Hepatol.* **12**, 1879–1886 (2014).
75. Tasali, E., Leproult, R., Ehrmann, D.A. & Van Cauter, E. Slow-wave sleep and the risk of type 2 diabetes in humans. *Proc. Natl. Acad. Sci. USA* **105**, 1044–1049 (2008).
76. Nedeltcheva, A.V. & Scheer, F.A. Metabolic effects of sleep disruption, links to obesity and diabetes. *Curr. Opin. Endocrinol. Diabetes Obes.* **21**, 293–298 (2014).
77. Vgontzas, A.N. *et al.* Obesity without sleep apnea is associated with daytime sleepiness. *Arch. Intern. Med.* **158**, 1333–1337 (1998).
78. Bixler, 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).
79. Swanson, J.M. The UK Biobank and selection bias. *Lancet* **380**, 110 (2012).
80. Collins, R. What makes UK Biobank special? *Lancet* **379**, 1173–1174 (2012).
81. Sudlow, C. *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).

EDITORIAL SUMMARY

AOP: Richa Saxena and colleagues report genome-wide association analyses of sleep disturbance traits in the UK Biobank cohort. They discover loci associated with insomnia symptoms and excessive daytime sleepiness and identify genetic correlations with several neuropsychiatric and metabolic traits.

ONLINE METHODS

Population and study design. Study participants were from the UK Biobank study, described in detail elsewhere[80–82]. In brief, the UK Biobank is a prospective study of >500,000 people living in the UK. All people in the National Health Service registry who were aged 40–69 years and living <25 miles from a study center were invited to participate from 2006–2010. In total, 503,325 participants were recruited from over 9.2 million mailed invitations. Self-reported baseline data were collected by questionnaire, and anthropometric assessments were performed. For the current analysis, individuals of non-white ethnicity were excluded to avoid confounding effects. All participants provided informed consent to the UK Biobank.

Sleep quality, sleep quantity and covariate measures. Study subjects self-reported sleep duration, insomnia symptoms, excessive daytime sleepiness, depression, medication use, age, sex, height and weight on a touchscreen questionnaire. For sleep duration, subjects were asked, "About how many hours sleep do you get in every 24 hours? (please include naps)," with responses in hour increments. To assess insomnia symptoms, subjects were asked, "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" and "prefer not to answer." To assess daytime sleepiness, subjects were asked, "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 responses "never/rarely," "sometimes," "often," "all the time," "don't know" and "prefer not to answer." Approximately 500,000 subjects answered these questions, but only the 120,286 unrelated individuals with genetic data and European ancestry were considered for analysis. Subjects with self-reported shift work (n = 6,557) or sleep medication use (n = 1,184) were excluded. Subjects who responded "do not know" or "prefer not to answer" were set to missing. Sleep duration and excessive daytime sleepiness were untransformed and treated as continuous variables, with daytime sleepiness coded as 1–4. The insomnia symptoms trait was dichotomized into controls ("never/rarely") and cases ("usually"). Covariates used in sensitivity analyses included self-reported sleep apnea, BMI, depression, psychiatric medication use, socioeconomic, smoking, employment and marital status, and snoring, and secondary GWAS for sleepiness included adjustment for BMI or depression. Sleep apnea cases were defined on the basis of ICD-10 diagnosis code (391 cases). BMI at baseline visit was calculated from entries of height and weight (n = 75,540 with available data). Depression was reported in answer to the question "How often did you feel down, depressed or hopeless mood in last 2 weeks?" (cases (n = 4,242) answered "more than half the days" or "nearly every day"). Medication use was self-reported as part of the initial UK Biobank interview. Our list of psychiatric medications for sensitivity analysis included the four most widely used: fluoxetine (Prozac), citalopram (Cipranol), paroxetine (Seroxat) and sertraline (Lustral). Our list of sleep medications included the 21 most widely used sleep medications in the UK Biobank (oxazepam, meprobamate, medazepam, bromazepam, lorazepam, clobazam, chlormezanone, temazepam, nitrazepam, lormetazepam, diazepam, zopiclone, triclofos, methyprylone, prazepam, triazolam, ketazolam, dichloralphenazone, clomethiazole, zaleplon and butobarbital). Smoking status was self-reported as past and current smoking behavior, and individuals were classified into those with "current" or "past" smoking or "never" smokers. Socioeconomic status was represented by the Townsend deprivation index, on the basis of national census data obtained immediately preceding participation in UK Biobank. Employment status was self-reported (cases were retired; controls were currently employed). Marital status was derived from self-reported household occupancy and relatedness data. Snoring was reported in answer to the question "Does your partner or a close relative or friend complain about your snoring?"

Genotyping, quality control and imputation. Of the ~500,000 subjects with phenotype data in the UK Biobank, ~153,000 are currently genotyped. Genotyping was performed by the UK Biobank, and genotyping, quality control and imputation procedures are described in detail at the UK Biobank website (http://biobank.ctsu.ox.ac.uk/). In brief, blood, saliva and urine were collected from participants, and DNA was extracted from buffy coat samples. Participant DNA was genotyped on two arrays, UK BiLEVE and UKB Axiom, with >95% common content. Genotypes were called using Affymetrix Power Tools software. Sample and SNP quality control were performed. Samples were removed for high rates of missingness or heterozygosity (480 samples), short runs of homozygosity (8 samples), relatedness (1,856 samples) and sex mismatches (191 samples). Genotypes for 152,736 samples passed sample quality control (~99.9% of total samples). SNPs were excluded if they did not pass quality control filters across all 33 genotyping batches. Batch effects were identified through frequency and Hardy–Weinberg equilibrium tests ($P < 1 \times 10^{-12}$). Before imputation, 806,466 SNPs passed quality control in at least one batch (>99% of the array content). Population structure was captured by principal-component analysis on the samples using a subset of high-quality (missingness <1.5%), high-frequency (>2.5%) SNPs (~100,000 SNPs) and identified the subsample of European descent. Imputation of autosomal SNPs was performed to a merged reference panel comprising the Phase 3 1000 Genomes Project and UK10K panels using IMPUTE2 (ref. 83). Data were prephased using SHAPEIT3 (ref. 84). In total, 73,355,677 SNPs, short indels and large structural variants were imputed. X-chromosome data were imputed separately, using Eagle 2.0 for prephasing with the --X chromosome flag (no reference panel) for the entire cohort[85] and IMPUTE2 with the Phase 3 1000 Genomes Project reference panel for imputation with the --chrX flag on 500-kb chunks in randomly assigned subsets of 30,000 individuals. Post-imputation quality control was performed as previously outlined (http://biobank.ctsu.ox.ac.uk/), and an imputation INFO score cutoff of 0.8 was applied. For GWAS, we further excluded SNPs with minor allele frequency (MAF) <0.001, per-SNP missingness >10% and per-sample missingness >40%. In total, up to 112,586 samples of European descent with high-quality genotyping and complete phenotype/covariate data were used for these analyses.

Statistical analysis. Phenotypic correlation analysis was performed with the Spearman test in R using the Hmisc package. Genetic association analysis for autosomes was performed in SNPTEST[86,87] with the 'expected' method using an additive genetic model adjusted for age, sex, ten principal components and genotyping array. Genome-wide association analysis was performed separately for sleep duration, insomnia symptoms and excessive daytime sleepiness with a genome-wide significance threshold of 5×10^{-8} for each GWAS. We had 80% power to detect the following effects: sleep duration, $\beta = 0.045$ h (2.7 min); insomnia symptoms, OR = 1.07; excessive daytime sleepiness, $\beta = 0.021$ units (assuming MAF = 0.1, $P = 5 \times 10^{-7}$); we had 80% power to detect the following effects: sleep duration, $\beta = 0.048$ h (2.9 min); insomnia symptoms, OR = 1.08; excessive daytime sleepiness, $\beta = 0.023$ units (assuming MAF = 0.1, $P = 5 \times 10^{-8}$). X-chromosome analysis was performed in PLINK 1.9 (ref. 88) using linear/logistic regression with separate analysis of the pseudoautosomal regions using the split-chromosome flag, adjusting for sex, age, ten principal components and genotyping array. For the X-chromosome signal at rs73536079, we verified using principal-component analysis that all carriers of the minor allele fell within the major European-ancestry cluster. Follow-up analyses of genome-wide suggestive and significant loci in the primary analyses included covariate sensitivity analyses individually adjusting for sleep apnea, depression, psychiatric medication use, socioeconomic, smoking, employment and marital status, and snoring or BMI (on top of the baseline model adjusting for age, sex, ten principal components and genotyping array). Sensitivity analysis was conducted only in the subset of subjects with data for all secondary covariates (n = 75,477 for sleep duration, n = 39,812 for insomnia symptoms and n = 75,640 for excessive daytime sleepiness). Analysis of enrichment for disease-associated gene sets and transcription factors was performed in WebGestalt[46] using the human genome as the reference set, Benjamini–Hochberg adjustment for multiple testing and a minimum of two genes per category. Sex-specific GWAS were performed in PLINK 1.9 using linear/logistic regression stratified by sex adjusting for age, ten principal components of ancestry and genotyping array. We used a hard-call genotype threshold of 0.1 (calls with greater than 0.1 were treated as missing), a SNP imputation quality threshold of 0.80 and a MAF threshold of 0.001. Regional association plots were generated with LocusZoom using the hg19 Nov2014 EUR reference panel to determine background linkage disequilibrium[89].

Trait heritability was calculated as the proportion of trait variance due to additive genetic factors across the autosomes, measured in this study using BOLT-REML[48] to leverage the power of raw genotype data together with data for low-frequency variants (MAF ≥ 0.001). For multiple-trait genome-wide

association analysis, we applied the CPASSOC package developed by Zhu *et al.*[51] to combine association evidence for chronotype, sleep duration, insomnia symptoms and excessive daytime sleepiness. CPASSOC provides two statistics, SHom and SHet. SHom is similar to statistics generated by the fixed-effects meta-analysis method[90] but accounts for the correlation of summary statistics due to trait correlation. SHom uses the sample size for a trait as a weight instead of variance, so that it is possible to combine traits with different measurement scales. SHet is an extension of SHom, but power can be improved when the genetic effect sizes are different for different traits. The distribution of SHet values under the null hypothesis of no association was obtained through an estimated beta distribution. To calculate the SHom and Shet statistics, a correlation matrix is required to account for the correlation among traits or resulting from overlapping or related samples from different cohorts. In this study, we directly provide the correlation matrix calculated from the residuals of four sleep traits after adjusting for age, sex, principal components of ancestry and genotyping array. Post-GWAS genome-wide genetic correlation analysis by LD score regression (LDSC)[67] was conducted using all UK Biobank SNPs also found in HapMap 3 (ref. 89) and included publicly available data from 20 published GWAS, using a significance threshold of $P = 0.0026$ after Bonferroni correction for all 20 tests performed. As expected, the observed heritability estimates from LDSC[67] using summary statistics for HapMap 3 were lower (5.7% (0.0065%) for sleep duration, 13.3% (0.0123%) for insomnia symptoms and 5.3% (0.005%) for sleepiness) than those calculated by BOLT-REML[48] using primary data (10.3% (0.006%) for sleep duration, 20.6% (0.011%) for insomnia symptoms and 8.4% (0.006%) for sleepiness) because the HapMap 3 panel was restricted to variants with MAF >5%. LDSC estimates genetic correlation between two traits (ranging from −1 to 1) from summary statistics using the facts that the GWAS effect size estimate for each SNP incorporates the effects of all SNPs in linkage disequilibrium with that SNP, SNPs with high linkage disequilibrium have higher χ^2 statistics than SNPs with low linkage disequilibrium, and a similar relationship is observed when single-study test statistics are replaced with the product of the z scores from two studies of traits with some correlation[67]. Furthermore, genetic correlation is possible between case–control studies and studies of quantitative traits, as well as within these study types. We performed a weighted GRS analysis using risk scores for RLS, schizophrenia, BMI and fasting insulin. Risk score SNPs passed the genome-wide significance threshold ($P < 5 \times 10^{-8}$) in recent large-scale GWAS and were present in the UK Biobank (RLS, 7 SNPs[65], **Supplementary Table 11**; schizophrenia, 96 SNPs[91]; BMI, 95 SNPs[92]; fasting insulin, 7 SNPs[93]). Independent SNPs were identified, and β estimates were recorded for calculation of the weighted risk score. The GRS was calculated by summing the products of the risk allele count multiplied by the effect reported in the discovery GWAS. The additive genotype model was used for all SNPs. We performed partitioning of heritability using the 25 precomputed functional annotations available through LDSC, which were curated from large-scale robust data sets[50]. Enrichment both in functional regions and in an expanded region (+500 bp) around loci corresponding to each functional class was calculated to prevent the estimates from being biased upward by enrichment in nearby regions. The multiple-testing threshold was determined using conservative Bonferroni correction ($P = 0.05/25$ classes).

Data availability. Summary GWAS statistics will be made available at the UK Biobank website (http://biobank.ctsu.ox.ac.uk/).

82. Allen, N.E., Sudlow, C., Peakman, T. & Collins, R. UK Biobank. UK Biobank data: come and get it. *Sci. Transl. Med.* **6**, 224ed4 (2014).
83. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
84. O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nat. Genet.* **48**, 817–820 (2016).
85. Loh, P.R., Palamara, P.F. & Price, A.L. Fast and accurate long-range phasing in a UK Biobank cohort. *Nat. Genet.* **48**, 811–816 (2016).
86. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
87. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
88. Chang, C.C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
89. International HapMap Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
90. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
91. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
92. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
93. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).

doi:10.1038/ng.3749