OPEN ACCESS

University of BRISTOL

Peer reviewed version

License (if available):
Other

Link to published version (if available):
10.1038/s41588-019-0361-7

Link to publication record in Explore Bristol Research
PDF-document

University of Bristol - Explore Bristol Research
General rights

Biological and clinical insights from genetics of insomnia symptoms

Jacqueline M. Lane[1,2,3,40], Samuel E. Jones[4,40], Hassan S. Dashti[1,3], Andrew R. Wood[4], Krishna G. Aragam[1,3,5,6], Vincent T. van Hees[7], Linn B. Strand[8], Bendik S. Winsvold[8,9,10], Heming Wang[3,11,12], Jack Bowden[13,14], Yanwei Song[1,3,15], Krunal Patel[1,15], Simon G. Anderson[16,17], Robin N. Beaumont[4], David A. Bechtold[18], Brian E. Cade[3,11,12], Mary Haas[1,3], Sekar Kathiresan[1,3,5,6], Max A. Little[19,20], Annemarie I. Luik[21,22], Andrew S. Loudon[18], Shaun Purcell[23], Rebecca C. Richmond[24,25], Frank A. J. L. Scheer[3,11,12], Barbara Schormair[26], Jessica Tyrrell[4], John W. Winkelman[27], Juliane Winkelmann[26,28,29,30], HUNT All In Sleep[31], Kristian Hveem[8], Chen Zhao[26], Jonas B. Nielsen[9], Cristen J. Willer[32,33,34], Susan Redline[35], Kai Spiegelhalder[36], Simon D. Kyle[21], David W. Ray[18,37], John-Anker Zwart[8,10], Ben Brumpton[8,13,38], Timothy M. Frayling[4], Deborah A. Lawlor[13,14,41], Martin K. Rutter[18,39,41], Michael N. Weedon[4,41] and Richa Saxena[1,2,3,41]*

[1]Center for Genomic Medicine, 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]Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK.

[5]Cardiology Division, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA.

[6]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA.

[7]Netherlands eScience Center, Amsterdam, The Netherlands.

[8]K.G. Jebsen Centre for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Norway.

[9]FORMI and Department of Neurology, Oslo University Hospital, Oslo, Norway.

[10]Division of Clinical Neuroscience, Oslo University Hospital and University of Oslo, Oslo, Norway.

[11]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, Massachusetts, USA.

[12]Division of Sleep Medicine, Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA.

[13]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK.

[14]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK.

[15]College of Science, Northeastern University, Boston, Massachusetts, USA.

[16]Division of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK.

[17]The Farr Institute of Health Informatics Research, University College London, London, UK.

[18]Division of Endocrinology, Diabetes & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK.

[19]Department of Mathematics, Aston University, Birmingham, UK.

[20]Media Lab, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

[21]Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK.

[22]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, The Netherlands.

[23]Department of Psychiatry, Brigham & Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA.

[24]School of Social and Community Medicine, University of Bristol, Bristol, UK.

[25]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

[26]Institute of Neurogenomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.

[27]Departments of Psychiatry and Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA.

[28]Cluster for Systems Neurology (SyNergy), Munich, Germany.

[29]Institute of Human Genetics, Technische Universität München, Munich, Germany.

[30]Chair of Neurogenetics, Technische Universität München, Munich, Germany.

[31]A list of members and affiliations appears at the end of the paper.

[32]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan, USA.

[33]Department of Human Genetics, University of Michigan, Ann Arbor, Michigan, USA.

[34]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, USA.

[35]Departments of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts, USA.

[36]Clinic for Psychiatry and Psychotherapy, Medical Centre - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany.

[37]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, OX37LE/NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK .

[38]Department of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway.

[39]Manchester Diabetes Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK.

[40]These authors contributed equally to this work.

[41]These authors jointly directed this project.

*e-mail: rsaxena@partners.org

Insomnia is a common disorder linked with adverse long-term medical and psychiatric outcomes. The underlying pathophysiological processes and causal relationships of insomnia with disease are poorly understood. Here we identify 57 loci for self-reported insomnia symptoms in the UK Biobank (n = 453,379) and confirm their impact on self-reported insomnia symptoms in the HUNT study (n = 14,923 cases, 47,610 controls), physician-diagnosed insomnia in Partners Biobank (n = 2,217 cases, 14,240 controls), and accelerometer-derived measures of sleep efficiency and sleep duration in the UK Biobank (n = 83,726). Our results suggest enrichment of genes involved in ubiquitin-mediated proteolysis and of genes expressed in multiple brain regions, skeletal muscle, and adrenal gland. Evidence of shared genetic factors is found between frequent insomnia symptoms and restless legs syndrome, aging, cardio-metabolic, behavioral, psychiatric and reproductive traits. Evidence is found for a possible causal link between insomnia symptoms and coronary artery disease, depressive symptoms and subjective well-being.

Insomnia disorder, persistent difficulty in initiating or maintaining sleep, and corresponding daytime dysfunction, occurs in 10-20% of the population[1]. Up to one-third of the population experience transient insomnia symptoms at any given time[2]. Longitudinal studies suggest that insomnia increases the risk for developing anxiety disorders, alcohol abuse, major depression, and cardio-metabolic disease[3]. However, little is known about underlying pathophysiologic mechanisms. Cognitive-behavioral therapies are the recommended first-line treatment, but many barriers to treatment exist[4,5]. Common drug treatments target synaptic neurotransmission (via GABAergic pathways), hypothalamic neuropeptides (via hypocretin/orexin), cortical arousal (via histamine receptors), or the melatonin system, but these drugs have variable effectiveness, may be habit forming, and have side effects[6,7]. Therefore, it is necessary to develop new personalized therapeutic strategies. Model organism studies have identified genes involved in sleep processes[8–13]. Family-based heritability estimates suggest that insomnia has a genetic component (22%–25%)[14]. Recent GWAS reported four loci for insomnia symptoms[15,16], but insights into underlying biological pathways and causal genetic links with disease are limited.

UK Biobank participants of European ancestry (n = 453,379) self-reported insomnia symptoms to the question "do you have trouble falling asleep at night or do you wake up in the middle of the night?". In this sample, 29% of individuals self-reported frequent insomnia symptoms ("usually"), with a higher prevalence in women (32% vs. 24%) and in older participants, shift workers, and those with shorter self-reported sleep duration (Supplementary Table 1).

We performed two parallel GWAS: (i) frequent insomnia symptoms ("never/rarely" vs. "usually" insomnia symptoms, n = 129,270 cases and 108,357 controls); (ii) any insomnia symptoms ("never/rarely" vs. "sometimes"/"usually" insomnia symptoms, n = 345,022 cases and 108,357 controls), adjusting for age and sex using 14,661,600 genotyped and imputed genetic variants across the autosomes and genotyped variants on the X chromosome. We identified 57 association signals explaining 1% of the variance (Fig. 1, Supplementary Table 2, and Supplementary Figs. 1 and 2). Of these, 20 loci were identified in both analyses, 28 loci in analysis of frequent insomnia symptoms only, and 9 in analysis of any insomnia symptoms only (Supplementary Table 2). Conditional analyses identified no secondary association signals. The 57 associations were independent of insomnia risk factors, as sensitivity analyses adjusting for body mass index (BMI), lifestyle, caffeine consumption, and depression or recent stress did not notably alter the magnitude or direction of effect estimates (Supplementary Table 3). The *MEIS1* association signal identified in the interim release of the UK Biobank was confirmed in the remainder of the UK Biobank sample (n = 75,508 cases of frequent insomnia symptoms and 64,403 controls; rs113851554 T, OR (95% CI) = 1.19 (1.15-1.23), P = 1.5 x 10^{-21}), and nominal replication was seen for previously reported *CYCL1* (P = 9.0 x 10^{-3}). The *TMEM132E* and *SCFD2* signals showed concordant direction of effect in both UK Biobank sub-samples, but were not significant, perhaps reflecting selection bias in the interim release sample[17]. No other findings from previous candidate gene association studies or smaller GWAS were confirmed (Supplementary Table 4).

Secondary GWAS excluding current shift workers or individuals reporting hypnotic, anti-anxiolytic or psychiatric medication usage, and/or with selected chronic diseases or psychiatric illnesses (excluding n = 76,470 participants) revealed strong pair-wise genetic correlation to the primary GWAS ($r_g \sim 1$) and did not identify additional association signals (Supplementary Figs. 1-3). Thus, biological processes underlying pathophysiology of insomnia symptoms may be common between the general population and those with co-morbidities, in accordance with the recent clinical reclassification of primary and secondary insomnia diagnoses into an insomnia disorder[18].

The prevalence of insomnia symptoms varies by sex; therefore, we performed secondary sex stratified GWAS (Supplementary Table 5). As described previously[15,16], the genetic architecture for frequent insomnia symptoms differed by sex, with a genetic correlation between the stratified GWAS of r_g = 0.807 (Supplementary Fig. 3). We found 13 additional loci (8 in women and 5 in men), although there were no genome-wide significant sex interactions. Effects in women were not modified by menopausal status.

Self-report of insomnia symptoms has limitations, including recall bias and lack of granularity[19]. Therefore, we sought replication of association signals in the HUNT population study with self-reported insomnia symptoms (n = 14,923 cases, 47,610 controls; Supplementary Table 6)[20] and the Partners Biobank with physician-diagnosed insomnia (n = 2,217 cases, 14,240 controls). Replication was observed for the *MEIS1* variant in both cohorts, and a genetic risk score (GRS) using our 57 variants and weighted by the effect estimates from GWAS of frequent insomnia symptoms (provided in Supplementary Table 2) was also associated with insomnia symptoms in HUNT (OR (95%CI) = 1.015 (1.01-1.02) per allele, P = 2.71 x 10^{-11}) and the Partners Biobank (OR (95%CI) 1.017 (1.007-1.027) per allele, P = 8.88 x 10^{-4}) (Table 1 and Supplementary Table 8). A meta-analysis of the UK Biobank, HUNT, and Partners Biobank studies shows consistency across all three cohorts (Supplementary Table 2). Next, to investigate impact of genetic variants on objective sleep patterns, we tested the 57 lead variants for association with 8

activity-monitor sleep measures in a subset of the UK Biobank participants of European ancestry who had undergone up to 7 days of wrist-worn accelerometry (n = 84,745; Supplementary Table 6). The lead *MEIS1* risk variant was associated with a higher number of sleep episodes, indicating an interrupted sleep pattern, lower sleep efficiency, shorter sleep duration and later sleep timing (P < 0.0008; Supplementary Table 9). The GRS was associated with reduced sleep efficiency (difference = -0.04 (0.01) % per allele; $P = 4 \times 10^{-14}$), shorter sleep duration (difference = -0.25 (0.035) minutes per allele; $P = 8 \times 10^{-13}$), and greater day-to-day variability in sleep duration (difference = 0.077 (0.025) minutes per allele; P = 0.0017) but not with the number of sleep episodes or diurnal inactivity duration (Table 1 and Supplementary Table 9).

In order to gain insight into the probable causal variants underlying the 57 association signals, we performed fine-mapping in PICS[21] and identified 38 variants with a causal probability of 0.20 or greater (Supplementary Table 10 and Supplementary Fig. 4)[22]. This list includes missense variants in NAD kinase *NADK* (N262K) and *MDGA1* (L61P), as well as rs324017, a SNP within the gene encoding transcriptional repressor *NAB1* (EGR-1 binding protein) that is predicted to disrupt a binding site for EGR1 (Supplementary Table 10)*,* a transcription factor involved in response to stress[23] and synaptic plasticity during REM sleep[23].

The 57 insomnia symptoms loci lie in genomic regions encompassing up to 236 genes, and a summary of annotations for signals or genes within each locus is shown in Supplementary Table 11. Association signals at 14 loci overlapped with NHGRI GWAS signals (r^2 > 0.7 in 1KG CEU) for one or more complex traits, with the insomnia symptoms risk allele associated with a greater risk of restless legs syndrome, schizophrenia, Tourette's syndrome, or obsessive-compulsive disorder, and associated with higher systolic blood pressure, greater carotid plaque burden, lower age at menarche and influencing adiposity traits, height, and educational attainment. Notably, experimental studies in mouse or fly models have implicated five genes at three associated loci in sleep regulation (*DVL1, LRP1, NR1H3, PRKAR2A*, and *SEMA3F)*. Genes within 16 loci are known drug targets.

Two lead SNPs were associated with one or more of 3,144 human brain structure and function traits assessed in the UK Biobank ($P < 2.8 \times 10^{-7}$, n = 9,707; Oxford Brain Imaging Genetics Server)[24], including rs1544637 (within a transcript *LOC642659*) with several large white matter tracts and with cingulate gyrus morphometry, which has previously been connected to insomnia[25], and rs62158170 (near *PAX8*) with resting-state fMRI networks (Supplementary Fig. 5).

Gene-based tests[26] identified 135 associated genes ($P \leq 2.29 \times 10^{-6}$; Supplementary Table 12). These genes are enriched for expression in brain regions[9] (Supplementary Table 13), including the cerebellum ($P = 1.3 \times 10^{-6}$), frontal cortex ($P = 1.3 \times 10^{-5}$), anterior cingulate cortex ($P = 1.7 \times 10^{-5}$), hypothalamus ($P = 2.2 \times 10^{-5}$), basal ganglia ($P = 7.0 \times 10^{-4}$), amygdala ($P = 3.4 \times 10^{-4}$), and hippocampus ($P = 8.4 \times 10^{-4}$), in line with previous research linking these brain regions to insomnia[25,27] and mouse single-cell studies linking insomnia GWAS hits to specific populations of brain cells[28]. Integration of gene expression data with GWAS using transcriptome-wide association analyses[29] identified 24 genes where insomnia-SNPs influence gene expression in one or more of 14 tissue types tested, including eight brain regions, muscles, peripheral nerves, whole blood, pituitary, thyroid, and adrenal gland tissue (Supplementary Table 16).

SNP-based heritability of frequent insomnia symptoms was estimated at $h^2 = 16.7\%$[30]. Partitioning of heritability across tissue types[31,32] confirmed enrichment in the central nervous

system, adrenal/pancreas tissue, and skeletal muscle ($P < 10^{-5}$) (Supplementary Table 17). Partitioning across functional class implicated activation and repression of enhancers (Supplementary Table 17). Pathway and ontology analyses[26,33,34] reveal a significant role for ubiquitin mediated proteolysis ($P_{bonf} = 0.04$; Supplementary Table 14), similar to findings in recent GWAS of RLS[35], and suggestive roles for phototransduction and muscle tissue development, across several databases (Supplementary Tables 14 and 15). Evidence from model organisms supports the link between Cullin-3 mediated ubiquitination and sleep/circadian rhythms[36–39]. Furthermore, the restless legs syndrome (RLS) gene *BTBD9* has been implicated as a substrate adaptor for E3 ubiquitin ligases[40]. We find no evidence for enrichment of neurotransmission receptor or biosynthesis genes known to be involved in the regulation of sleep (e.g. GABA, glutamate, adenosine).

We investigated the genetic link between frequent insomnia symptoms and other behavioral and/or disease states. Based on previous links between RLS and insomnia symptoms[15,16], we tested a GRS of 20 SNPs for RLS[35] and found association with frequent insomnia symptoms (OR = 1.03 (1.02-1.04) per RLS risk allele, $P = 2.57 \times 10^{-57}$), driven partly by the *MEIS1* region (GRS excluding *MEIS1* region OR = 1.02 (1.02-1.03) per RLS risk allele, $P = 2.06 \times 10^{-31}$)(Supplementary Table 18 and Supplementary Fig. 6). We also tested a GRS of our 57 insomnia SNPs in RLS and found an association with RLS (OR = 1.39 (1.34-1.44), $P = 1.70 \times 10^{-80}$) driven partly by the *MEIS1* signal (GRS excluding *MEIS1* region OR = 1.17 (1.13-1.21), $P = 2.56 \times 10^{-20}$) (Supplementary Table 18). We also observe a positive genetic correlation between insomnia and RLS ($r_g = 0.291$, $P = 6.33 \times 10^{-12}$). We performed BUHMBOX analysis to distinguish between pleiotropy and heterogeneity. BUHMBOX[41] analyses using all 20 RLS SNPs points towards heterogeneity ($P = 4.09 \times 10^{-6}$), most likely due to undiagnosed RLS cases misclassified as insomnia. However, when we exclude the *MEIS1* locus, we detect a possible shared genetic basis not explained by heterogeneity ($P = 0.137$).

To test the proportion of variance frequent insomnia symptoms shares with other traits based on genetic overlap, we performed genetic correlation analyses with 233 traits with public GWAS summary statistics[31,32,42,43]. We found strong positive genetic correlations ($P \le 2 \times 10^{-3}$) between frequent insomnia symptoms and adiposity traits, coronary artery disease, number of children ever born, neuroticism, smoking behavior, and depressive symptoms and disorder. Strong negative genetic correlations were observed with self-reported sleep duration, subjective well-being, cognitive measures, proxy longevity measures, and the remaining reproductive traits (Fig. 2 and Supplementary Table 19). These genetic links persisted with GWAS excluding subjects with chronic and psychiatric illnesses (Supplementary Table 19), indicating a relationship not driven by the presence of concomitant conditions.

To test for causal links between frequent insomnia symptoms and seven clusters of genetically correlated traits, we performed Mendelian randomization (MR) analyses mostly within a two-sample summary data framework. Using the inverse variance (IVW) method[44] (Fig. 3 and Supplementary Table 20), we found evidence of a causal association ($P < 0.001$) between frequent insomnia symptoms and prevalent coronary artery disease (CAD; OR (95%CI) = 2.15 (1.38-3.35)) reduced subjective well-being (difference in mean SD units of -0.29 (s.e. 0.06)), greater depressive symptoms (difference in mean SD units of 0.42 (s.e. 0.08)); all associations had a consistent direction and similar magnitude of effect in sensitivity analyses using MR-Egger[45] and weighted median methods (Fig. 3 and Supplementary Table 20)[46]. We found similar results using effect estimates from GWAS excluding those with preexisting conditions. We validated the causal

association between insomnia and CAD with one-sample MR in the UK Biobank ($n = 23{,}980$ cases and 361,706 controls, OR (95%CI) = 2.95 (2.18–3.99), $P = 2.3 \times 10^{-12}$) (Supplementary Table 20 and Supplementary Fig. 7). Bidirectional MR identified no evidence of reverse causality. The one-sample MR causal association between insomnia status and CAD in UK Biobank is consistent with the robust empirical evidence seen in prospective studies and meta-analyses[3,47].

This study provides a comprehensive description of the genetic architecture of frequent or persistent insomnia symptoms, pointing to putative causal variants and candidate genes, pathways and tissues for functional studies. Further, we define physiological correlates for insomnia symptoms and meaningful clinical links, including a causal link with CAD.

HUNT All-In Sleep

Amy E. Martinsen[42], Anne H. Skogholt[7], Ben Brumpton[7,8,9], Bendik S. Winsvold[7,31,38], Børge Sivertsen[43], Cristen J. Willer[32,33,34], Daniela Bragantini[44], Håvard Kallestad[45], Imre Janszky[46], Ismail C. Guzey[45], John-Anker Zwart[7,38], Jonas B. Nielsen[31], Kristian B. Nilsen[47], Kristian Hveem[7], Lars Fritsche[32], Linda M. Pedersen[42], Linn B. Strand[7], Maiken E. Gabrielsen[7], Marianne B. Johnsen[48], Marie U. Lie[42], Morten Engstrøm[49], Trond Sand[49], and Wei Zhou[33]

[42]FORMI, Oslo University Hospital, Oslo, Norway.
[43]Department of Health Promotion, Norwegian Institute of Public Health, Bergen, Norway.
[44]Institute for Mental Health, Faculty of Medicine, Norwegian Univeristy of Science and Technology, Trondheim, Norway.
[45]Department of Research and Development, Division of Psychiatry, St. Olavs University Hospital, Trondheim, Norway.
[46]Department of Public Health and Nursing, Norwegian University of Science and Technology, NTNU, Trondheim, Norway.
[47]Oslo University Hospital, Department of Neurology, Oslo, Norway.
[48]Institute of Clinical Medicine, University of Oslo, Oslo, Norway.
[49]Department of Neurology and Clinical Neurophysiology, St. Olavs Hospital, Trondheim, Norway.

Acknowledgements

Author Contributions

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

Competing Interests

J.W.W. is a consultant for Merck and Flex Pharma. He receives royalties from UpToDate. He has received speaker fees and travel support from Otsuka. He has received research grants from UCB Pharma, NeuroMetrix, NIMH, the RLS Foundation, and Luitpold Pharma. F.A.J.L.S. has received lecture fees from Bayer HealthCare, Sentara HealthCare, Philips, Vanda Pharmaceuticals, and Pfizer. D.A.L. has received funding from Medtronic and Roche Diagnostics for biomarker research unrelated to this study. M.K.R. has acted as a consultant for GSK, Novo Nordisk, Roche and MSD, and also participated in advisory board meetings on their behalf. M.K.R. has received lecture fees from MSD and grant support from Novo Nordisk, MSD and GSK.

1. Morin, C. M. & Benca, R. Chronic insomnia. *Lancet Lond. Engl.* **379**, 1129–1141 (2012).
2. Morin, C. M. *et al.* Insomnia disorder. *Nat. Rev. Dis. Primer* **1**, 15026 (2015).
3. Hoevenaar-Blom, M. P., Spijkerman, A. M. W., Kromhout, D., van den Berg, J. F. & Verschuren, W. M. M. Sleep duration and sleep quality in relation to 12-year cardiovascular disease incidence: the MORGEN study. *Sleep* **34**, 1487–1492 (2011).
4. Riemann, D. *et al.* European guideline for the diagnosis and treatment of insomnia. *J. Sleep Res.* **26**, 675–700 (2017).
5. Qaseem, A., Barry, M. J., Kansagara, D. & for the Clinical Guidelines Committee of the American College of Physicians. Nonpharmacologic versus pharmacologic treatment of adult patients with major depressive disorder: A clinical practice guideline from the American College of Physicians. *Ann. Intern. Med.* **164**, 350 (2016).
6. Sateia, M. J., Buysse, D. J., Krystal, A. D., Neubauer, D. N. & Heald, J. L. Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: An American Academy of Sleep Medicine clinical practice guideline. *J. Clin. Sleep Med. JCSM Off. Publ. Am. Acad. Sleep Med.* **13**, 307–349 (2017).
7. Kapil, V., Green, J. L., Le Lait, C., Wood, D. M. & Dargan, P. I. Misuse of benzodiazepines and Z-drugs in the UK. *Br. J. Psychiatry J. Ment. Sci.* **205**, 407–408 (2014).
8. Naylor, E. *et al.* The Circadian clock mutation alters sleep homeostasis in the mouse. *J. Neurosci.* **20**, 8138–8143 (2000).
9. Cirelli, C. The genetic and molecular regulation of sleep: from fruit flies to humans. *Nat. Rev. Neurosci.* **10**, 549–560 (2009).
10. Allada, R., Cirelli, C. & Sehgal, A. Molecular mechanisms of sleep homeostasis in flies and mammals. *Cold Spring Harb. Perspect. Biol.* **9**, a027730 (2017).
11. Pimentel, D. *et al.* Operation of a homeostatic sleep switch. *Nature* **536**, 333–337 (2016).
12. Funato, H. *et al.* Forward-genetics analysis of sleep in randomly mutagenized mice. *Nature* **539**, 378–383 (2016).
13. Chung, S. *et al.* Identification of preoptic sleep neurons using retrograde labelling and gene profiling. *Nature* **545**, 477–481 (2017).
14. Lind, M. J. & Gehrman, P. R. Genetic pathways to insomnia. *Brain Sci.* **6**, (2016).
15. Lane, J. M. *et al.* Genome-wide association analyses of sleep disturbance traits identify new loci and highlight shared genetics with neuropsychiatric and metabolic traits. *Nat. Genet.* **49**, 274–281 (2017).
16. Hammerschlag, A. R. *et al.* Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. *Nat. Genet.* **49**, 1584–1592 (2017).
17. Wain, L. V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir. Med.* **3**, 769–781 (2015).
18. Vahia, V. N. Diagnostic and statistical manual of mental disorders 5: A quick glance. *Indian J. Psychiatry* **55**, 220–223 (2013).
19. Benjamins, J. S. *et al.* Insomnia heterogeneity: Characteristics to consider for data-driven multivariate subtyping. *Sleep Med. Rev.* **36**, 71–81 (2017).
20. Krokstad, S. *et al.* Cohort profile: the HUNT Study, Norway. *Int. J. Epidemiol.* **42**, 968–977 (2013).
21. Farh, K. K.-H. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2014).

22.   Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102-110 (2015).

23.   Duclot, F. & Kabbaj, M. The role of early growth response 1 (EGR1) in brain plasticity and neuropsychiatric disorders. *Front. Behav. Neurosci.* **11**, (2017).

24.   Elliott, L. *et al.* Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature* **562**, 210-216 (2018).

25.   Winkelman, J. W. *et al.* Increased Rostral Anterior Cingulate Cortex Volume in Chronic Primary Insomnia. *Sleep* **36**, 991–998 (2013).

26.   Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput. Biol.* **12**, (2016).

27.   Vetrivelan, R., Qiu, M.-H., Chang, C. & Lu, J. Role of Basal Ganglia in Sleep–Wake Regulation: Neural Circuitry and Clinical Significance. *Front. Neuroanat.* **4**, (2010).

28.   Jansen, P. R. *et al.* Genome-wide analysis of insomnia (N=1,331,010) identifies novel loci and functional pathways. *bioRxiv* 214973 (2018). doi:10.1101/214973

29.   Gusev, A. *et al.* Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

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

31.   Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* (2015). doi:10.1038/ng.3404

32.   Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

33.   Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

34.   Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

35.   Schormair, B. *et al.* Identification of novel risk loci for restless legs syndrome in genome-wide association studies in individuals of European ancestry: a meta-analysis. *Lancet Neurol.* **16**, 898–907 (2017).

36.   Stavropoulos, N. & Young, M. W. Insomniac and Cullin-3 regulate sleep and wakefulness in Drosophila. *Neuron* **72**, 964–976 (2011).

37.   Livingston, W. S. *et al.* Improved sleep in military personnel is associated with changes in the expression of inflammatory genes and improvement in depression symptoms. *Front. Psychiatry* **6**, (2015).

38.   Freeman, A. A. H., Mandilaras, K., Missirlis, F. & Sanyal, S. An emerging role for Cullin-3 mediated ubiquitination in sleep and circadian rhythm. *Fly (Austin)* **7**, 39–43 (2013).

39.   Anafi, R. C. *et al.* Sleep is not just for the brain: transcriptional responses to sleep in peripheral tissues. *BMC Genomics* **14**, 362 (2013).

40.   Trotti, L. M. Restless Legs Syndrome and Sleep-Related Movement Disorders. *Contin. Minneap. Minn* **23**, 1005–1016 (2017).

41.   Han, B. *et al.* A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. *Nat. Genet.* **48**, 803–810 (2016).

42.   Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and

genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).

43.   Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* (2015). doi:10.1038/ng.3406

44.   Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665 (2013).

45.   Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

46.   Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian Randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).

47.   Javaheri, S. & Redline, S. Insomnia and risk of cardiovascular disease. *Chest* **152**, 435–444 (2017).

Table 1 | A risk score of genetic variants for self-reported insomnia symptoms (57 SNPs) associates with self-reported insomnia symptoms
insomnia in the Partners Biobank, and activity-monitor based measures of sleep fragmentation, timing and duration from 7 day acceleromet

	Trait	Frequent insomnia symptoms Weighted GRS			Fr Wei
		OR	95% CI	*P*	OR [95% CI]
HUNT Study (*n* = 14,923 cases, 47,610 controls)	**Self-reported insomnia symptoms**	**1.015**	**1.01-1.02**	**2.71x10^{-11}**	**1.0130**
Partners Biobank (*n* = 2,217 cases, 14,240 controls)	Physician diagnosed insomnia	1.017	1.01-1.03	8.88x10^{-4}	1.0120
		Beta	SE	*P*	Beta (SE)
UK Biobank 7-day accelerometry (*n* = 84,745)	*Sleep fragmentation measures*				
	Sleep efficiency %	**-0.0400**	**0.010**	**4.21x10^{-14}**	**-0.0300**
	Number of sleep bouts (n)	0.0053	0.002	0.030	0.0044
	Sleep duration measures				
	Sleep duration (minutes)	**-0.2460**	**0.036**	**8.39x10^{-13}**	**-0.2210**
	Sleep duration standard deviation (minutes)	**0.0780**	**0.024**	**1.72x10^{-3}**	**0.0750**
	Sleep timing measures Midpoint of 5h daily period of minimum activity (L5 timing) (minutes)	0.0960	0.042	0.025	0.0580
	Midpoint of 10 daily period of maximum activity (M10 timing) (minutes)	0.0420	0.048	0.429	0.0300
	Diurnal inactivity duration (minutes)	0.0060	0.030	0.901	0.0080
	Sleep midpoint (minutes)	-0.0124	0.024	0.975	-0.0070

Tested using a risk score weighted by effect of SNP on self-reported frequent insomnia symptoms in UK Biobank sample (Supplementary Table 2) in t
SE, standard error; OR, odds ratio per risk allele; Beta, effect per risk allele; CI, 95% confidence interval.

Figure 1 | Manhattan plots for genome-wide association analyses of insomnia. a, Frequent insomnia symptoms (n_{cases} = 129,270, $n_{controls}$ = 108,352). **b**, Any insomnia symptoms (n_{cases} = 345,022, $n_{controls}$ = 108,352). Dotted line is genome-wide significant (P = 5 x 10^{-8}) results of linear mixed models adjusted for age, sex, principal components of ancestry, and genotyping array. SNP-based heritability estimates were calculated using BOLT-REML variance components analysis. Chromosomes are annotated with the nearest gene to each association signal.

Figure 2 | Shared genetic architecture between frequent insomnia symptoms and behavioral and disease traits. LD-score regression estimates of genetic correlation (r_G) of frequent insomnia symptoms are compared with the summary statistics from 224 publicly available genome-wide association studies of psychiatric and metabolic disorders, immune diseases, and other traits of natural variation. Blue, positive genetic correlation; red, negative genetic correlation, r_G values displayed for significant correlations. Larger squares correspond to more significant P-values. Genetic correlations that are significantly different from zero after Bonferroni correction are shown on the plot (after Bonferroni correction, P-value cut-off is 0.0002). All genetic correlations in this report can be found in tabular form in Supplementary Table 19. IQ, intelligence quotient.

Figure 3 | Causal relationships of insomnia symptoms. a-f, Association between single nucleotide polymorphisms associated with frequent insomnia symptoms and CAD (**a**), subjective well-being (**c**), and depressive symptoms (**e**). Per allele associations with risk plotted against per allele associations with frequent insomnia symptom risk (vertical and horizontal black lines around points show 95% confidence interval for each polymorphism) results are shown for three different MR association tests. Forest plots show the estimate of the effect of genetically increased insomnia risk on CAD (**b**), depressive symptoms (**d**), and subjective well-being (**f**) as assessed for each SNP. Nearest gene(s) is displayed to the right of plot. Also shown for each SNP is the 95% confidence interval (gray line segment) of the estimate and the Inverse Variance MR, MR-Egger, and Weighted Median MR results in red. Sample size of each GWAS used in the MR analyses is as follows: frequent insomnia symptoms (n_{cases} = 129,270, $n_{controls}$ = 108,352), CAD (n_{cases} = 60,801, $n_{controls}$ = 123,504), subjective well-being (n = 298,420), and depressive symptoms (n = 161,460).

Methods

UK Biobank population and study design. Study participants were from the UK Biobank study, described in detail elsewhere[48]. In brief, the UK Biobank is a prospective study of >500,000 people living in the United Kingdom. All people in the National Health Service registry aged 40-69 and living <25 miles from a study center were invited to participate between 2006-2010. In total, 503,325 participants (5%) 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 (n = 48,667) were excluded to avoid confounding effects. A subset analysis was also performed excluding UK Biobank subjects from the interim release and their relatives (exclusion n = 190,216).

Insomnia and covariate measures. Study subjects self-reported insomnia symptoms, depression, medication use, age, and sex on a touch-screen questionnaire at baseline assessment. Height and weight were measured at baseline. 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", "prefer not to answer". Subjects who responded "Prefer not to answer" (n = 637) were set to missing. We undertook two GWAS, one in which insomnia symptoms were dichotomized into controls ("never/rarely") and cases with any symptoms ("sometimes" and "usually"), and the second in which participants were dichotomized into controls ("never/rarely") or frequent insomnia symptoms ("usually"), with those reporting "sometimes" excluded. Additional covariates used in sensitivity analyses included BMI, self-reported and prevalent sleep apnea diagnosis, area deprivation index, alcohol intake, snoring, nap behavior, smoking status, menopause status, weekly physical activity, tea and coffee intake, depression, and extreme stress. BMI was calculated from measured height and weight at baseline visit. Prevalent sleep apnea cases were defined based on primary or secondary ICD10 diagnosis code (G47.3; n = 391 cases) at the time of baseline assessment. Social deprivation for participant area of residence at the time of recruitment was represented by the Townsend deprivation index, based on national census data immediately preceding participation in the UK Biobank. The Townsend deprivation index was log transformed for use in the analyses. Alcohol intake was self-reported in response to the question "About how often do you drink alcohol?" with answers ranging from "daily or almost daily" to "never". Snoring was reported in answer to the question "Does your partner or a close relative or friend complain about your snoring?". Nap behavior was self-reported in response to the question "Do you have a nap during the day" with answers "never/rarely", "sometimes", or "usually". Smoking status was self-reported as past smoking behavior and current smoking behavior, and classified into "current", "past", or "never" smoked. Menopause status was reported in answer to the question "Have you had your menopause (periods stopped)?" with answers "yes", "no", "not sure – had a hysterectomy", "not sure – other reason". Weekly physical activity in total metabolic equivalents per week, total MET-h/week, was calculated using self-reported estimates of type and duration of physical activity. Tea and coffee intake was estimated using the questions "How many cups of tea do you drink each day? (Include black and green tea)" and "How many cups of coffee do you drink each day? (Include decaffeinated coffee)," respectively, with responses in cups/day. Employment status was self-reported in response to the following question "Which of the following describes your current situation?" with responses "in paid employment or self-employed", "retired", "looking after home

and/or family", "unable to work because of sickness or disability", "unemployed", doing unpair or voluntary work", "full or part-time student", "none of the above", "prefer not to answer". Marital status was derived from self-reported household occupancy and relatedness data as follows: married/partner was derived from those reporting husband/wife/partner in household with >1 person reported to live in household. Depression was reported in answer to the question "How often did you feel down, depressed or hopeless mood in last 2 weeks?" (cases, $n = 4,242$ based on answers "more than half the days", or "nearly every day"). Extreme stress was assessed with the question "In the last two years have you experiences any of the following (you can select more than one answer" with responses "serious illness, injury or assault to yourself", "serious illness, injury or assault of a close relative", "death of a close relative", "death of a spouse or partner", "marital separation/divorce", "financial difficulties", "none of the above". A secondary GWAS further excluded shift workers, sleep and psychiatric medication users, and subjects with chronic and psychiatric illness ($n = 76,470$). Cases of psychiatric disorder was determined using any of the following definitions (derived from Howard et al.): (i) ICD-10 codes for major depressive disorder (F32, F33), bipolar disorder (F30, F31), schizophrenia (F20-F29), autism (F84.0, F84.3, F84.5), intellectual disability (F70.0, F70.1, F70.9), anxiety disorder (F40 - F43), multiple personality disorder (F44.8), or mood disorder (F30 - F39); (ii) Self-reported antidepressant, antipsychotic, or anxiolytic use at nurse-led interview; (iii) Self-reported depression, major depressive disorder, bipolar disorder, or schizophrenia at nurse-led interview at baseline visit; (iv) "Broad depression": responded yes to the question "Have you ever seen a general practitioner (GP) for nerves, anxiety, tension or depression?" and yes to either, "have you ever had a time when you were feeling depressed or down for at least a whole week," or, "Have you ever had a time when you were uninterested in things or unable to enjoy the things you used to for at least a whole week?" lasting for more than 1 week; (v) Questionnaire-assessed bipolar disorder (Smith algorithm) responded yes to the question, "Have you ever had a period of time lasting at least two days when you were feeling so good, "high", excited or "hyper" that other people thought you were not your normal self or you were so "hyper" that you got into trouble?" or "Have you ever had a period of time lasting at least two days when you were so irritable that you found yourself shouting at people or starting fights or arguments?" for a duration of at least one week and with at least 3 manic/hyper symptoms[49]. A list of sleep medications can be found in the Supplementary Note.

A description of coronary artery disease and activity monitor derived measures of sleep in the UK Biobank can be found in the Supplementary Note.

Genotyping and quality control. Phenotype data is available for 502,631 subjects in the UK Biobank. Genotyping was performed by the UK Biobank, and genotyping, quality control, and imputation procedures are described in detail here[50]. In brief, blood, saliva, and urine was collected from participants, and DNA was extracted from the buffy coat samples. Participant DNA was genotyped on two arrays, UK BiLEVE and UKB 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 QC was conducted using 605,876 high quality autosomal markers. Samples were removed for high missingness or heterozygosity (968 samples) and sex chromosome abnormalities (652 samples). Genotypes for 488,377 samples passed sample QC (~99.9% of total samples). Marker based QC measures were

tested in the European ancestry subset (n = 463,844), which was identified based on principal components of ancestry. SNPs were tested for batch effects (197 SNPs/batch), plate effects (284 SNPs/batch), Hardy-Weinberg equilibrium (572 SNPs/batch), sex effects (45 SNPs/batch), array effects (5417 SNPs), and discordance across control replicates (622 on UK BiLEVE Axiom array and 632 UK Biobank Axiom array) ($P < 10^{-12}$ or <95% for all tests). For each batch (106 batches total), markers that failed at least one test were set to missing. Before imputation, 805,426 SNPs pass QC 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 SNPs (>2.5%) (~100,000 SNPs) and identified the sub-sample of white British descent. We further clustered subjects into four ancestry clusters using K-means clustering on the principal components, identifying 453,964 subjects of European ancestry. Imputation of autosomal SNPs was performed to UK10K haplotype, 1000 Genomes Phase 3, and Haplotype Reference Consortium (HRC) with the current analysis using only those SNPs imputed to the HRC reference panel. Autosomal SNPs were pre-phased using SHAPEIT3[51] and imputed using IMPUTE4. In total ~96 million SNPs were imputed. Related individuals were identified by estimating kinship coefficients for all pairs of samples, using only markers weakly informative of ancestral background. In total, there are 107,162 related pairs comprised of 147,731 individuals related to at least one other subject in the UK Biobank.

Genome-wide association analysis. Genetic association analysis across the autosomes was performed in related subjects of European ancestry (n = 453,964) using BOLT-LMM[30] linear mixed models and an additive genetic model adjusted for age, sex, 10 PCs, genotyping array and genetic correlation matrix with a maximum per SNP missingness of 10% and per sample missingness of 40%. We used a genome-wide significance threshold of 5 x 10^{-8} for each GWAS. Genetic association analysis was also performed in unrelated subjects of white British ancestry (n = 337,545) using PLINK[52] logistic regression and an additive genetic model adjusted for age, sex, 10 PCs and genotyping array to determine SNP effects on self-reported insomnia symptoms. We used a hard-call genotype threshold of 0.1, SNP imputation quality threshold of 0.80, and a MAF threshold of 0.001. Genetic association analysis for the X chromosome was performed using the genotyped markers on the X chromosome with the additional –sex flag in PLINK. We are 80% powered to detect a relative difference of 4% or more (i.e. an OR of 1.04 or 0.96 assuming a MAF 0.1, P = 5 x 10^{-8}). Sex-specific GWAS were performed in PLINK 1.9 using logistic regression stratified by sex adjusting for age, 10 principal components of ancestry, and genotyping array. We used a hard-call genotype threshold of 0.1, SNP imputation quality threshold of 0.80, and a MAF threshold of 0.001. SNP x sex interactions (13 tests) and SNP x menopause interaction in females (13 tests) were tested for genome-wide significant signals with the significance threshold defined by Bonferroni correction. SNP-based trait heritability was calculated as the proportion of trait variance due to additive genetic factors measured in this study using BOLT-REML[30], to leverage the power of raw genotype data together with low frequency variants (MAF ≥ 0.001). Additional independent risk loci were identified using the approximate conditional and joint association method implemented in GCTA (GCTA-COJO)[53]. Fixed effects meta-analysis was performed using METAL[54] with the standard error scheme.

Sensitivity analyses on top signals. Follow-up analyses on genome-wide significant loci in the primary analyses included covariate sensitivity analysis individually adjusting for sleep apnea, coffee/tea intake, physical activity, severe stress, depression, psychiatric medication use, socio-economic status, smoking, employment and marital status, and snoring, or BMI in addition to baseline adjustments for age, sex, 10 PCs and genotyping array. Sensitivity and sex-specific analyses were conducted only in unrelated subjects of white British ancestry.

Gene, pathway and tissue-enrichment analyses. Gene-based analysis was performed using PASCAL[26]. Tissue enrichment analysis was conducted using FUMA[55]. Enrichment for pathways and ontologies was performed in EnrichR[33,34] using the human genome as the reference set and a minimum number of 2 genes per category. A genetic risk score for RLS was tested using the weighted genetic risk score calculated by summing the products of the RLS risk allele count for 20 genome-wide significant SNPs multiplied by the scaled RLS effect reported by Schormair et al.[35] using the summary statistics from our frequent insomnia symptom GWAS using the GTX package in R[56]. Integrative transcriptome-wide association analyses with GWAS were performed using the FUSION TWAS package[29] with weights generated from gene expression in 8 brain regions and 6 tissues from the GTEX consortium (v6) , and SNPs common to the 1000 Genomes LD reference panel and our frequent insomnia symptoms GWAS summary statistics. Tissues for TWAS testing were selected from the FUMA tissue enrichment analyses and here we present significant results that survive Bonferroni correction for the number of genes tested per tissue and for all 14 tissues. In the results table, we show the number of individuals per tissue type used to generate expression weights, the total number of genes expressed per tissue, gene symbol for significant gene, rsID for the best GWAS and eQTL SNPs in the reference panel, and the TWAS *P* value. GWAS and eQTL SNPs analyzed include only those SNPs present in the 1000 Genomes LD reference panel used by the FUSION TWAS software; therefore, the best GWAS SNP listed in this table may be in strong linkage disequilibrium with, but not identical to, the lead GWAS SNP for hits in Supplementary Table 2.

Heterogeneity analysis. Analyses to distinguish pleiotropy and heterogeneity between insomnia and RLS was performed using Breaking Up Heterogeneous Mixture Based On Cross-locus correlations (BUHMBOX)[41]. BUHMBOX analyses tests the presence of heterogeneity between two traits. BUHMBOX was performed in the insomnia GWAS using all 20 RLS SNPs and weights reported by Schormair et al.[35] and in the RLS GWAS using all 57 SNPs reported in Supplementary Table 2. Additional sensitivity analyses were performed excluding SNPs in the *MEIS1* region.

Genetic correlation analyses. Post-GWAS genome-wide genetic correlation analysis of LD Score Regression (LDSC)[32,42,43] using LDHub was conducted using all UK Biobank SNPs also found in HapMap3 and included publicly available data from 224 published genome-wide association studies, with a significance threshold of $P = 0.002$ after Bonferroni correction for all tests performed. LDSC estimates 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 in LD with that SNP, SNPs with high LD have higher X^2 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. Furthermore, genetic correlation is possible between case/control studies and quantitative traits, as well as within these trait types. We performed partitioning of heritability using the 8 pre-computed cell-type regions, and 25 pre-computed functional annotations available through LDSC, which were curated from large-scale robust datasets[31]. Enrichment both in the functional regions and in an expanded region (+500 bp) around each functional class was calculated in order to prevent the estimates from being biased upward by enrichment in nearby regions. The multiple testing threshhold for the partitioning of heritability was determined using the conservative Bonferroni correction (P of 0.05/25 classes). Summary GWAS statistics will be made available at the UK Biobank web portal.

Mendelian randomization analyses. MR analysis was carried out using MR-Base[57], using the inverse variance weighted approach as our main analysis method[44], and MR-Egger[45] and weighted median estimation[46] as sensitivity analyses. MR results may be biased by horizontal pleiotropy, i.e. where the genetic variants that are robustly related to the exposure of interest (here frequent insomnia symptoms) independently influence levels of a causal risk factor for the outcome. IVW assumes that there is no horizontal pleiotropy. MR-Egger provides unbiased causal estimates even if all of the genetic instruments have horizontal pleiotropic effects, but it assumes that the association of genetic instruments with risk factor is not correlated with any pleiotropic genetic instrument associations with outcome. The weighted median approach is valid if less than 50% of the weight is pleiotropic (i.e. no single SNP that contributes 50% of the weight or a number of SNPs that together contribute 50% should be invalid because of horizontal pleiotropy. Given these different assumptions, if all three methods are broadly consistent, this strengthens our causal inference. For most of our MR analyses, we used two-sample MR, in which, for all 57 of the insomnia GWAS hits identified in this study, we looked for the per allele difference in odds (binary outcomes) or means (continuous) with outcomes from summary publicly available data in the MR-Base platform. Results are therefore a measure of 'any insomnia' and sample 1 is UK Biobank (our GWAS) and sample 2 a number of different GWAS consortia covering the outcomes we explored (Supplementary Table 17). For all four of the longevity outcomes and as follow up for CAD, we used one-sample MR with the SNP-outcome associations also obtained from UK Biobank. If we could not find one of the 57 SNPs in the outcome database we substituted for a proxy where possible, LD proxies are defined using 1000 Genomes European sample with $r^2 > 0.8$. The number of SNPs used in each MR analysis varies by outcome from 11 to 53 because of some SNPs (or proxies for them) not being located in the outcome GWAS (Table 1).

Primary association analyses of the 57 genome-wide significant insomnia SNPs with CAD were performed in Hail (https://github.com/hail-is/hail) using imputed genotype dosages and a logistic regression model adjusting for age at first visit, sex, genotyping array, and the first 10 principal components of ancestry. A total of 23,980 CAD cases were compared to 388,326 referents. For Mendelian randomization, a fixed-effects inverse-variance weighted meta-analysis was performed of the SNP-specific association estimates with CAD, aligning each insomnia SNP allele/beta coefficient to "increased risk of insomnia." Sensitivity analyses were performed excluding the variants with the strongest effect estimate and/or widest confidence intervals to account for SNP heterogeneity.

Replication cohort. Replication cohort sample ascertainment, phenotype definition, genotyping, quality control, imputation, and analyses are described in the Supplementary Note.

Reporting summary. An accompanying Life Sciences Reporting Summary is published along with this manuscript.

Data availability. GWAS summary statistics are available at the Sleep Disorders Knowledge Portal data download page (http://sleepdisordergenetics.org/informational/data).

Methods-only references

48.   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).
49.   Smith, D. J. *et al.* Prevalence and characteristics of probable Major Depression and Bipolar Disorder within UK Biobank: Cross-sectional study of 172,751 participants. *PLoS ONE* **8**, e75362 (2013).
50.   Bycroft, C *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018).
51.   O'Connell, J. *et al.* Haplotype estimation for biobank-scale data sets. *Nature Genet.* **48**, 817-820 (2016).
52.   Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575 (2007).
53.   Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
54.   Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinforma. Oxf. Engl.* **26**, 2190–2191 (2010).
55.   Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
56.   R Development Core Team. R: A language and environment for statistical computing. *R Found. Stat. Comput.* **Vienna**, ISBN 3-900051-07-0 (URL http://www.R-project.org/).
57.   Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).

Editorial summary:

Genome-wide association analyses identify 57 loci associated with insomnia symptoms and evidence of shared genetic architecture between insomnia and cardio-metabolic, behavioral, psychiatric and reproductive traits.