Biological and clinical insights from genetics of insomnia symptoms

Jacqueline M. Lane1,3,4,10, Samuel E. Jones4,40, Hassan S. Dashti1,3, Andrew R. Wood4, Krishna G. Aragam1,5,6, Vincent T. van Hees7, Linn B. Strand8, Bendik S. Winsvold9, Heming Wang3,11,12, Jack Bowden13,14, Yanwei Song1,3,15, Krunal Patel1,15, Simon G. Anderson24,25, Max A. Little9,20, Annemarie I. Luik31,22, Andrew S. Loudon18, Shaun Purcell23, Rebecca C. Richmond24,25, Frank A. J. L. Scheer3,11,12, Barbara Schormair26, Jessica Tyrrell4, John W. Winkelman27, Juliane Winkelmann26,28,29,30, HUNT All In Sleep31, Kristian Hveem9, Chen Zhao26, Jonas B. Nielsen9, Cristen J. Willer32,33,34, Susan Redline35, Kai Spiegelhalder36, Simon D. Kyle21, David W. Ray18,37, John-Anker Zwart8,10, Ben Brumpton8,13,38, Timothy M. Frayling4, Deborah A. Lawlor10,13,14,41, Martin K. Rutter18,39,41, Michael N. Weedon4,41 and Richa Saxena1,2,3,41*

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 identified 57 loci for self-reported insomnia symptoms in the UK Biobank (n = 453,379) and confirmed their effects on self-reported insomnia symptoms in the HUNT Study (n = 14,923 cases and 47,610 controls), physician-diagnosed insomnia in the Partners Biobank (n = 2,217 cases and 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 glands. Evidence of shared genetic factors was found between frequent insomnia symptoms and restless legs syndrome, aging, and cardiometabolic, behavioral, psychiatric, and reproductive traits. Evidence was 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 population1. As much as one-third of the population experiences transient insomnia symptoms at any given time1. Longitudinal studies suggest that insomnia increases the risk of developing anxiety disorders, alcohol abuse, major depression, and cardiometabolic disease1. However, little is known about the underlying pathophysiologic mechanisms. Cognitive behavioral therapies are the recommended first-line treatment, but many barriers to treatment exist4,5. Common drug treatments target synaptic neurotransmission (via GABAergic pathways), hypothalamic neuropeptides (via hypocretin/orexin), cortical arousal (via histamine receptors), or the melanin system, but these drugs have variable effectiveness, may be habit forming, and have side effects6,7. Therefore, new personalized therapeutic strategies must be developed. Model-organism studies have identified genes involved in a variety of sleep processes8–11. Family-based heritability estimates suggest that insomnia has a genetic component (22–25%)1. Recent GWAS have reported four loci for insomnia symptoms15,16, but insights into the underlying biological pathways and causal genetic links with disease are limited.

UK Biobank participants of European ancestry (n = 453,379) self-reported insomnia symptoms in response 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”), and the prevalence was relatively higher in women (32% versus 24% in men) and in older participants, shift workers, and individuals with shorter self-reported sleep duration (Supplementary Table 1).

We performed two parallel GWAS: (i) frequent insomnia symptoms (“never/rarely” versus “usually”; n = 129,270 cases and 108,357 controls) and (ii) any insomnia symptoms (“never/rarely” versus “sometimes”/“usually”; n = 345,022 cases and 108,357 controls), adjusting for age, sex, ten principal components of ancestry, and genotyping array 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 were identified in analysis of frequent insomnia symptoms only, and 9 were identified 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, because 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 MEDI5 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], odds ratio (OR) and 95% confidence interval (CI) = 1.19 (1.15–1.23), P = 1.5 × 10−21), and nominal replication...
Fig. 3). We found 13 additional loci (eight in women and five in men), although there were no genome-wide-significant sex interactions. The effects in women were not modified by menopausal status (Supplementary Table 5).

Self-reporting of insomnia symptoms has limitations, including recall bias and lack of granularity19. Therefore, we sought replication of association signals in the HUNT population study with self-reported insomnia symptoms16 (n = 14,923 cases and 47,610 controls; Supplementary Table 6) and the Partners Biobank with physician-diagnosed insomnia (n = 2,217 cases and 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 × 10⁻⁷) and the Partners Biobank (OR (95% CI) = 1.017 (1.007–1.027) per allele, P = 8.88 × 10⁻⁶) (Table 1 and Supplementary Table 8). A meta-analysis of the UK Biobank, HUNT, and Partners Biobank studies showed consistency across all three cohorts (Supplementary Table 5). Next, to investigate the effects of genetic variants on objective sleep patterns, we tested the 57 lead variants for association with eight activity-monitor sleep measures in a subset of the UK Biobank participants of European ancestry who had undergone up to 7 d of wrist-worn accelerometry

was observed for previously reported CYLC1 (P = 9.0 × 10⁻⁷). The TMEM132E and SCFD2 signals showed a concordant direction of effect in both UK Biobank subsamples but were not significant, a finding perhaps reflecting selection bias in the interim-release sample17. 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 use of hypnotic, antianxiolytic, or psychiatric medications, and/or having selected chronic diseases or psychiatric illnesses (excluding n = 76,470 participants) revealed strong pairwise genetic correlation to the primary GWAS (r = 1) and did not identify additional association signals (Supplementary Figs. 1–3). Thus, biological processes underlying the pathophysiology of insomnia symptoms may be common between the general population and those with comorbidities, in accordance with the recent clinical reclassification of primary and secondary insomnia diagnoses into an insomnia disorder16.

The prevalence of insomnia symptoms varies by sex; therefore, we performed secondary sex-stratified GWAS (Supplementary Table 5). As described previously15,16, the genetic architecture for frequent insomnia symptoms differed by sex, and we observed a genetic correlation between the stratified GWAS of r = 0.807 (Supplementary Fig. 3). We found 13 additional loci (eight in women and five in men), although there were no genome-wide-significant sex interactions. The effects in women were not modified by menopausal status (Supplementary Table 5).
Table 1 | Risk scores of genetic variants for self-reported insomnia symptoms (57 SNPs) are associated with self-reported insomnia symptoms in the HUNT Study, physician-diagnosed insomnia in the Partners Biobank, and activity-monitor-based measures of sleep fragmentation, timing, and duration from 7-d accelerometer in the UK Biobank

| Trait                                                      | Frequent insomnia symptoms | Frequent insomnia symptoms | OR      | 95% CI     | P          | OR      | 95% CI     | P          |
|------------------------------------------------------------|---------------------------|---------------------------|---------|------------|------------|---------|------------|------------|
| **HUNT Study** (n = 14,923 cases, 47,610 controls)         |                           |                           |         |            |            |         |            |            |
| Self-reported insomnia symptoms                            | 1.015                     | 1.01-1.02                 | 2.71 × 10^{-4} | 1.013     | 1.01-1.02 | 2.32 × 10^{-5} | |
| **Partners Biobank** (n = 2,217 cases, 14,240 controls)    |                           |                           |         |            |            |         |            |            |
| Physician-diagnosed insomnia                              | 1.017                     | 1.01-1.03                 | 8.88 × 10^{-4} | 1.012     | 1.00-1.02 | 0.015   | |

UK Biobank 7-d accelerometer (n = 84,745)

| Sleep-fragmentation measures:                              | β           | s.e.m. | P          | β           | s.e.m. | P          |
|------------------------------------------------------------|-------------|--------|------------|-------------|--------|------------|
| Sleep efficiency (%)                                       | -0.040      | 0.010  | 4.21 × 10^{-3} | -0.030      | 0.001  | 1.40 × 10^{-3} | |
| Number of sleep bouts (n)                                  | 0.005       | 0.002  | 0.030      | 0.004       | 0.003  | 0.073      | |

**Fig. 2 | Shared genetic architecture between frequent insomnia symptoms and behavioral and disease traits.** LD Score regression estimates of genetic correlation (\(r_p\)) of frequent insomnia symptoms, 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; \(r_p\) values are 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 cutoff of 0.0002). All genetic correlations in this report can be found in tabular form in Supplementary Table 19. IQ, intelligence quotient.

\(n = 84,745\); Supplementary Table 6). The lead MEIS1 risk variant was associated with a higher number of sleep episodes, thus 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 lower sleep efficiency (difference = −0.04 (0.01) percent 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).

To gain insight into the probable causal variants underlying the 57 association signals, we performed fine-mapping in probabilistic identification of causal SNPs (PICS)\(^{21}\) and identified 38 variants with...
Fig. 3 | Causal relationships of insomnia symptoms. a, c, e. Associations between SNPs 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%CI for each polymorphism) are shown for three different MR association tests. b, d, f. Forest plots showing the estimates 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 genes are displayed to the right of the plots. Also shown for each SNP is the 95%CI (gray line segment) of the estimate and the IVW MR, MR-Egger, and weighted-median MR results in red. Sample sizes of each GWAS used in the MR analyses are as follows: frequent insomnia symptoms ($n_{\text{cases}} = 129,270; n_{\text{controls}} = 108,352$), CAD ($n_{\text{cases}} = 60,801; n_{\text{controls}} = 123,504$), subjective well-being ($n = 298,420$), and depressive symptoms ($n = 161,460$).
of 20 SNPs for RLS\textsuperscript{50} and found association with frequent insomnia symptoms (OR = 1.03 (1.02–1.04) per RLS risk allele, \(P = 2.57 \times 10^{-9}\)), driven partly by the \textit{MEIS1} region (GRS excluding the \textit{MEIS1} region OR = 1.02 (1.02–1.03) per RLS risk allele, \(P = 2.06 \times 10^{-11}\); 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), \(P = 1.70 \times 10^{-49}\)) driven partly by the \textit{MEIS1} signal (GRS excluding the \textit{MEIS1} region OR = 1.17 (1.13–1.21), \(P = 2.56 \times 10^{-20}\), Supplementary Table 18). We also observed a positive genetic correlation between insomnia and RLS (\(r_g = 0.291, P = 6.33 \times 10^{-12}\)). We performed BUHMBOX (breaking up heterogeneous mixture based on cross-locus correlations) analysis to distinguish between pleiotropy and heterogeneity. BUHMBOX\textsuperscript{10} analyses using all 20 RLS SNPs indicated heterogeneity (\(P = 4.09 \times 10^{-6}\)), probably as a result of undiagnosed RLS cases misclassified as insomnia. However, when we excluded the \textit{MEIS1} locus, we detected a possible shared genetic basis not explained by heterogeneity (\(P = 0.137\)).

To test the proportion of variance that frequent insomnia symptoms share with other traits on the basis of genetic overlap, we performed genetic correlation analyses with 233 traits with public GWAS summary statistics\textsuperscript{3,12,24,45}. We found strong positive genetic correlations (\(\rho \geq 2 \times 10^{-5}\)) between frequent insomnia symptoms and adiposity traits, coronary artery disease (CAD), number of children ever born, neuroticism, smoking behavior, and depressive symptoms and disorders. 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 in GWAS excluding subjects with chronic and psychiatric illnesses (Supplementary Table 19), thus 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 weighted (IVW) method\textsuperscript{46} (Fig. 3 and Supplementary Table 20), we found evidence of a causal association (\(P < 0.001\)) between frequent insomnia symptoms and prevalent CAD (OR 95\%CI = 2.15 (1.38–3.35)), diminished subjective well-being (difference in mean s.d. units of –0.29 (s.e. 0.06)), and greater depressive symptoms (difference in mean s.d. units of 0.42 (s.e. 0.08)); all associations had a consistent direction and similar magnitude of effect in sensitivity analyses, as compared with the results from MR-Egger\textsuperscript{47} and weighted-median methods\textsuperscript{46} (Fig. 3 and Supplementary Table 20). We found similar results by using effect estimates from GWAS, excluding individuals 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 indicated no evidence of reverse causality between CAD and insomnia symptoms. The one-sample MR causal association between insomnia status and CAD in the UK Biobank is consistent with the robust empirical evidence seen in prospective studies and meta-analyses\textsuperscript{47}.

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

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at [https://doi.org/10.1038/s41588-019-0361-7](https://doi.org/10.1038/s41588-019-0361-7).
References

1. Morin, C. M. & Benca, R. Chronic insomnia. Lancet 379, 1129–1141 (2012).
2. Morin, C. M. et al. Insomnia disorder. Nat. Rev. Dis. Primers 1, 15026 (2016).
3. Hoevenaar-Blom, M. P., Spijkerman, A. M. W., Kromhout, D., van den Berg, J. F. & Verswuijs, 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. 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, 350–359 (2016).
6. Sateia, M. J., Buysse, D. J., Krystal, A. D., Neubauer, D. N. & Head, 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. 13, 307–349 (2017).
7. Zhang, F. & Lupski, J. R. Non-coding genetic variants in human disease. Nat. Genet. 48, 803–810 (2016).
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. 3, a027730 (2017).
11. Pimentel, D. et al. Operation of a homeostatic sleep switch. Nature 536, 333–337 (2016).
12. Fungaro, 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. & Gehrmann, P. R. Genetic pathways to insomnia. Brain Sci. 6, E64 (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. 5, 35–46 (2017).
18. Benjamins, J. S. et al. Insomnia heterogeneity: characteristics to consider for data-driven multivariate subtyping. Sleep Med. Rev. 36, 71–81 (2017).
19. Vahia, V. I. 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. 5, 35–46 (2017).
20. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web application for large-scale data annotation. Nucleic Acids Res. 43, W954–W960 (2015).
21. Han, B. et al. A method to decipher pleiotropy by detecting underlying heterogeneity driven by hidden subgroups applied to autoimmune and neuropsychiatric diseases. 1611.01506, arXiv (2016).
22. Friedman, A. A. H., Mandilaras, K., Missirlis, F. & Sanyal, S. An emerging role for Cullin 3-mediated ubiquitination in sleep and circadian rhythm: insights from Drosophila. Fly (Austin) 7, 39–43 (2013).
23. Alali, R. C. et al. Sleep is not just for the brain: transcriptional responses to sleep in peripheral tissues. BMC Genomics 14, 413 (2013).
24. Javaheri, S. & Redline, S. Insomnia and risk of cardiovascular disease. Arterioscler. Thromb. Vasc. Biol. 35, 214–219 (2015).
25. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. Nat. Genet. 47, 1236–1241 (2015).
26. Burgess, S., Butterworth, A. & Thompson, S. G. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet. Epidemiol. 37, 658–665 (2013).
27. 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).
28. Javarishi, S. & Redline, S. Insomnia and risk of cardiovascular disease. Chest 152, 435–444 (2017).

Acknowledgements

This research was conducted by using the UK Biobank Resource (UK Biobank applications 6818 and 9072). We would like to thank the participants and researchers from the UK Biobank who contributed or collected data. This work was supported by NIH grants R01DK107859 (R.S.), R21HL121728 (R.S.), F33DK102323 (J.M.L.), R01HL113338 (J.M.L., S.R., and R.S.), R01DK106296 (R.S. and F.S.), NHLBI R35 351731 (J.M.L. and S.R.), R01DK105072 (R.S. and F.S.), T32HL007567 (J.M.L.), K01HL136884 (J.M.L.), and HG003054 (H.W.). The MGH Research Scholar Fund (R.S.), The University of Manchester (Research Infrastructure Fund), the Wellcome Trust (salary support for D.W.R. and A.L.S.), UK Medical Research Council MC_UU_12013/5 (D.A.L.), UK Medical Research Council MC_UU_00106/1 (D.A.L.), and UK National Institute of Health Research NF-SE-1011-10196 (D.A.L.) AR.W. and T.M.F. are supported by a European Research Council grant (7Z-24S 50371: GLUCOSEGENES FP7-IDEAS-ERC). S.E.I. is funded by the Medical Research Council (MR/M005070/1). M.N.W. is supported by a Wellcome Trust Institutional Strategic Support Award (W079875/MF). The following groups provided summary statistics to LDHub and MR-base: ADIPOGen (Adiponectin Genetics Consortium), CAD (Coronary Artery Disease Genetics Consortium), CARDIOGRAM (Coronary Artery Disease Genome-wide Replication and Meta-analysis), CKDGen (Chronic Kidney Disease Genetics Consortium), dbGaP (Database of Genotypes and Phenotypes), DIABET (Diabetes Genetics Replication and Meta-analysis), ENIGMA (Enhancing Neuro Imaging Genetics through Meta Analysis), EAGLE (Early Genetics & Lifecourse Epidemiology (Emerging Genotypes, excluding 23andMe)), EGG (Early Growth Genetics Consortium), GABRIEL (a multidisciplinary study to identify the genetic and environmental causes of asthma in the European community), GCAN (Genetic Consortium for Anorexia Nervosa), GEFOS (Genetic Factors for Osteoporosis Consortium), GIANT (Genetic Investigation of Anthropometric Traits), GIS (Genetics of Iron Status Consortium), GLGC (Global Lipids Genetics Consortium), GPC (Genetics of Personality Consortium), GUCG (Global Catecholamine and Gout Consortium), Haemogen (Haematological and Platelet Traits Genetics Consortium), HMrG (Heart Rate Trait Consortium), IBDDGC (International Inflammatory Bowel Disease Genetics Consortium), ILCCO (International Lung Cancer Consortium), IMSGC (International Multiple Sclerosis Genetics Consortium), MAGIC (Meta-analyses of Glucose and Insulin-related Traits Consortium), MESA (Multi-ethnic Study of Pulmonary and Cardiovascular Diseases), Project MinE consortium, ReproGen (Reproductive Genetics Consortium), SSGAC (Social Science Genetics Association Consortium), TAG (Tobacco and Genetics Consortium), TRICL.
HUNT All In Sleep
Amy E. Martensena,42, Anne H. Skogholtc,7, Ben Brumpton7,8,9, Bendik S. Winsvold7,38, Børge Sivertsen43, John-Anker Zwart7,38, Jonas B. Nielsen, Kristian B. Nilsend, Kristian Hveemf, Lars Fritsche32, Linda M. Pedersen42, Linn B. Stranfg, Maiken E. Gabrielsn7, Marianne B. Johnsen48, Marie U. Lie42, Morten Engstrøm49, Trond Sand49 and Wei Zhou33

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

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. E.A.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 GlaxoSmithKline, Novo Nordisk, Roche, and Merck Sharp & Dohme (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 GlaxoSmithKline.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41588-019-0361-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to R.S.

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Methods
UK Biobank population and study design. Study participants were from the UK Biobank study, as described in detail elsewhere. 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 40–69 years of age and living <25 miles from a study center were invited to participate between 2006 and 2010. In total, 503,325 participants (5%) were recruited from more than 9.2 million mailed invitations. Self-reported baseline data were collected by questionnaire, and anthropometric assessments were performed. For the current analysis, individuals of nonwhite ancestry were excluded (n = 48,665) to avoid confounding effects. A subset analysis was also performed by excluding UK Biobank subjects from the interim release and their relatives (exclusion n = 190,216).

Insomnia and covariate measures. The study subjects self-reported insomnia symptoms, depression, anxiety, and engagement in touch-screen questionnaire at baseline assessment. Height and weight were measured at baseline. To assess insomnia symptoms, we asked subjects, “Do you have trouble falling asleep at night, or do you wake up in the middle of the night?” and asked them to select from responses of “never/rarely,” “sometimes,” “usually,” and “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 another in which participants were dichotomized into controls (“never/rarely”) and frequent insomnia symptoms (“usually”), with those reporting “sometimes” excluded. Additional covariates used in sensitivity analyses included reported and proxy area deprivation index, alcohol intake, smoking, depression, stress, BMI, and sex. Haplotype-based tests were performed using BOLT-REML with a hard-call genotype threshold of 0.1, a SNP imputation-quality threshold of 0.80, and a MAF threshold of 0.001. We performed sex-specific GWAS in PLINK 1.9 with logistic regression stratified for age and sex. Analysis of the European-ancestry subset (n = 643,844), which was identified on the basis of principal components of ancestry, SNPs were tested for both primary and secondary effects. Hardy–Weinberg equilibrium tests and linkage disequilibrium were performed using Haploviz (722 SNPs per batch), sex effects (45 SNPs per batch), and discordance across control replicates (622 on the UK BILEVE Axiom array and 632 on the UK Biobank Axiom array) (P < 10^{-12} or < 0.05% 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 passed QC in at least one batch (>99% of the array content). Population structure was captured by principal component analysis on the samples, by using a subset of high-quality (missingness <1.5%), high-frequency SNPs (2.5%) = 100,000 SNPs, and the subsample of white British descent was identified. We further clustered subjects into four ancestry clusters by 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 with the current analysis, by using only those SNPs imputed to the HaploType Reference Consortium reference panel. Autosomal SNPs were prephased with SHAPEIT3 (ref. 49) and imputed with IMPUTE4. In total ~96 million SNPs were imputed. Risk-scores were generated by estimating kinship coefficients and generating pairs of samples, by using only markers weakly informative of ancestral background. In total, there were 107,162 related pairs comprising 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) with BOLT-LMM linear mixed models and an additive genetic model adjusted for age, sex, ten principal components, 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 × 10^{-8} for each GWAS. To determine SNP effects on self-reported insomnia symptoms, we also performed genetic association analysis in unrelated subjects of white British ancestry (n = 337,545) with PLINK logistic regression and an additive genetic model adjusted for age, sex, ten principal components, and genotyping array. We used a hard-call genotype threshold of 0.1, a SNP imputation-quality threshold of 0.80, and a MAF threshold of 0.001. Genetic association analysis for the X chromosome was performed with the same model, with the addition of sex and ancestry as covariates. The results were consistent with the analysis of autosomal SNPs and the X chromosome, with the addition of sex as a fixed effect. We computed Q2 values for the first 5,000 SNPs, which were significant at the genome-wide level. We performed a genome-wide association analysis of 40% or more (that is, an OR of 1.04 or 0.96, assuming a MAF of 0.1, P = 5 × 10^{-8}). We performed sex-specific GWAS in PLINK 1.9 with 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, a SNP imputation-quality threshold of 0.80, and a MAF threshold of 0.001. SNP sex interactions (13 tests) and SNP sex interactions 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 with BOLT-REML, to leverage the power of whole genome data together with other covariates (MAF ≥ 0.001). Additional independent risk loci were identified by using the approximate conditional and joint association method implemented in GCTA (GCTA-COJO)

Fixed-effects meta-analysis was performed in METAL with the standard-error scheme.
Sensitivity analyses on top signals. We performed follow-up analyses on genome-wide-significant loci in the primary analyses, including covariate sensitivity analysis individually adjusting for sleep apnea, coffee/tea intake, physical activity, severe stress, depression, use of psychiatric medication, socioeconomic status, smoking, employment status, marital status, snoring, or BMI in addition to baseline adjustments for age, sex, ten principal components, 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 in PASCAL⁴⁶. Tissue enrichment analysis was conducted in FUMA⁴⁷. Enrichment for pathways and ontologies was performed in EnrichR ³¹,²⁴ by using the human genome as the reference set and a minimum number of two genes per category. A genetic risk score for RLS was tested by 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, by using the summary statistics from our frequent-insomnia-symptom GWAS and the GTX package in R ⁴⁸. Integrative transcriptome-wide association analyses with GWAS were performed with the FUSION TWAS package²⁹, and weights were generated from gene expression in eight brain regions and six tissues from the GTEX consortium (v6), and SNPs common to the 1000 Genomes LD reference panel and our frequent-insomnia-symptom GWAS summary statistics. Tissues for TWAS testing were selected from the FUMA tissue enrichment analyses, and we presented significant results after Bonferroni correction for the number of genes tested per tissue and for all 14 tissues. The results table shows the number of individuals per tissue type used to generate expression weights, the total number of genes expressed per tissue, the gene symbol for the significant gene, the rsID for the best GWAS and eQTL SNPs in the reference panel, and the TWAS P value. GWAS and eQTL SNPs analyzed included only those SNPs present in the 1000 Genomes LD reference panel used by 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 were performed with BUMHBOX²⁹, which tests for the presence of heterogeneity between two traits. BUMHBOX analysis was performed in the insomnia GWAS, by using all 20 RLS SNPs and weights reported by Schormair et al.³⁵ and in the RLS GWAS, by 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) ³²,³³ by using LDHub was conducted on all UK Biobank SNPs also found in HapMap3, including publicly available data from 224 published genome-wide association studies, with a significance threshold of P = 6.002 after Bonferroni correction for all tests performed. LDSC estimates the genetic correlation between two traits from summary statistics (ranging from -1 to 1) on the basis of the incorporation of effects of all SNPs in LD with that SNP into the GWAS effect-size estimate for each SNP; SNPs with high LD had higher chi-square statistics than SNPs with low LD, and a similar relationship was observed when single study test statistics were replaced with the products 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 by using the eight precomputed cell-type regions and 25 precomputed functional annotations available through LDSC, which were curated from large-scale robust datasets³¹. Enrichment both in the functional regions and in an expanded region (+ 500 bp) around each functional class was calculated to prevent the estimates from being biased upward by enrichment in nearby regions. The multiple testing threshold for the partitioning of heritability was determined by using 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 with MR-Base³⁶, with the IVW approach as our main analysis method³⁷, and with MR-Egger ³⁸ and weighted-median estimation³⁸ as sensitivity analyses. MR results may be biased by horizontal pleiotropy, that is, when the genetic variants that are robustly related to the exposure of interest (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 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 (that is, 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, our causal inference is strengthened if all three methods are broadly consistent. 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 publicly available summary data in the MR-Base platform. The results are therefore a measure of ‘any insomnia’, and sample 1 is UK Biobank (our GWAS), whereas sample 2 is a number of different GWAS consortia covering the outcomes that 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 did not find one of the 57 SNPs in the outcome database, we substituted for a proxy where possible; LD proxies were defined by using the 1000 Genomes European sample with r² > 0.8. The number of SNPs used in each MR analysis varied 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/) by using imputed genotype dosages and a logistic regression model adjusting for age at first visit, sex, genotyping array, and the first ten principal components of ancestry. A total of 23,980 CAD cases were compared to 388,326 referents. For MR, a fixed-effects IVW meta-analysis was performed of the SNP-specific association estimates with CAD, by aligning each insomnia SNP allele/allele coefficient to increased risk of insomnia. Sensitivity analyses were performed by 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, QC, imputation, and analyses are described in the Supplementary Note.

Data availability

GWAS summary statistics are available at the Sleep Disorders Knowledge Portal data download page (http://sleepdisordergenetics.info/GenomeData/).

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. Howard, D. M. et al. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nat. Commun. 9, 1470 (2018).
50. 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).
51. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203–209 (2018).
52. O’Connell, J. et al. Haplotype estimation for biobank-scale data sets. Nat. Genet. 48, 817–820 (2016).
53. 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).
54. 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). SI–S3.
55. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 26, 2190–2191 (2010).
56. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. Nat. Commun. 8, 1826 (2017).
57. R Development Core Team. R: a Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, 2018).
58. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. eLife 7, e34408 (2018).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
- Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection
All data was collected by the UK Biobank, HUNT study, and Partners Biobank. This is a secondary use of data.

Data analysis
The following softwares were used: R 3.12, BOLT-LMM 2.3.2, PLINK 1.9, LocusZoom 0.4.8, omconvert, FUSION TWAS, Affymetrix Power Tools 1.16.1, MRbase 1.2.1, LDhub 1.9.0, LDSC 1.0.0, FUMA 1.3.3, MS Office2016, SHAPEIT3, Impute4, METAL 2011-03-25, GCTA 1.90.0, and Pascal. We used no custom code.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data
All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Summary GWAS statistics will be made available at the Sleep Disorders Genetic Portal (http://www.sleepdisordergenetics.org/informational/data)
Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences  ☐ Behavioural & social sciences  ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | We used all available subjects in the UK Biobank of European ancestry with complete phenotype and genotype information. European ancestry was determined by cluster analysis of the genetic principal components of ancestry. |
| Data exclusions | Excluded subjects of non-European ancestry. |
| Replication | We replicated the genetic associations in the HUNT study, Partners Biobank, and UK Biobank accelerometer study. |
| Randomization | This is not relevant to this study, as there are no groups. |
| Blinding | This is not relevant to this study, as there are no groups. |

Reporting for specific materials, systems and methods

| Materials & experimental systems | Methods |
| n/a Involved in the study | n/a Involved in the study |
| ☑ ☐ Unique biological materials | ☑ ☐ ChiP-seq |
| ☑ ☐ Antibodies | ☑ ☐ Flow cytometry |
| ☑ ☐ Eukaryotic cell lines | ☑ ☐ MRI-based neuroimaging |
| ☑ ☐ Palaeontology | |
| ☑ ☐ Animals and other organisms | |
| ☑ ☐ Human research participants | |

Human research participants

Policy information about studies involving human research participants

| Population characteristics | See supplementary table 1 for a detailed description of the research participants. Briefly, participants were 45.7% male, 56 years old with 47.5% self-reporting insomnia symptoms “sometimes” and 28.5% self-reporting insomnia symptoms “usually”. |
| Recruitment | Participants were recruited by the UK Biobank by mailers to 9 million people in the UK Medical system. The UK Biobank population is healthier than average with a lower mortality rate. |