nature genetics

Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits

Anke R Hammerschlag[1], Sven Stringer[1], Christiaan A de Leeuw[1], Suzanne Sniekers[1], Erdogan Taskesen[1,2], Kyoko Watanabe[1], Tessa F Blanken[3–5], Kim Dekker[3], Bart H W te Lindert[3], Rick Wassing[3], Ingileif Jonsdottir[6,7], Gudmar Thorleifsson[6], Hreinn Stefansson[6], Thorarinn Gislason[7,8], Klaus Berger[9], Barbara Schormair[10,11], Juergen Wellmann[9], Juliane Winkelmann[10–13], Kari Stefansson[6,7], Konrad Oexle[10,15], Eus J W Van Someren[3–5,15] & Danielle Posthuma[1,14,15]

Persistent insomnia is among the most frequent complaints in general practice. To identify genetic factors for insomnia complaints, we performed a genome-wide association study (GWAS) and a genome-wide gene-based association study (GWGAS) in 113,006 individuals. We identify three loci and seven genes associated with insomnia complaints, with the associations for one locus and five genes supported by joint analysis with an independent sample (n = 7,565). Our top association (*MEIS1*, $P < 5 \times 10^{-8}$) has previously been implicated in restless legs syndrome (RLS). Additional analyses favor the hypothesis that *MEIS1* exhibits pleiotropy for insomnia and RLS and show that the observed association with insomnia complaints cannot be explained only by the presence of an RLS subgroup within the cases. Sex-specific analyses suggest that there are different genetic architectures between the sexes in addition to shared genetic factors. We show substantial positive genetic correlation of insomnia complaints with internalizing personality traits and metabolic traits and negative correlation with subjective well-being and educational attainment. These findings provide new insight into the genetic architecture of insomnia.

Insomnia disorder is the second most prevalent mental disorder[1], with prevalence estimates ranging from 10% (adults) to 22% (the elderly)[2,3]. This disorder is characterized by lasting problems falling asleep or by waking up in the night or early morning, with subjective repercussions for daytime functioning. It is the primary risk factor for depression[4] and contributes to risks of cardiovascular disease[5–9], type 2 diabetes[10] and obesity[11]. Heritability estimates of 38% (males) and 59% (females)[12] suggest a substantial role for genetic factors in insomnia disorder. In contrast to the neurobiological mechanisms presumed to be involved in most mental disorders, it has been suggested that insomnia merely involves reversible maladaptive learning of sleep-related cognitions and behaviors[13,14]. Indeed, interventions that address these activities are at least partly effective in about two-thirds of cases[15], but they ameliorate complaints by only about 50% (ref. 16), often resulting in a persistent course for the disorder[17]. Family and twin studies suggest the involvement of genetic factors in the etiology of insomnia[18–21]. However, only a few linkage and association studies of insomnia-related phenotypes have been conducted, which have mainly been underpowered (n < 5,000), and recent larger studies have used nonvalidated proxy measures for identification of individuals with insomnia disorder[22–24]. For example, the habitual duration and timing of sleep[25,26] do not reliably discriminate between cases of insomnia disorder and controls[27], and genes found to be related to these phenotypes may therefore not be indicative of the biological mechanisms of insomnia. The most recent GWAS for sleep disturbance traits reported several new loci[24]. However, this study did not provide information on the predictive accuracy of the included traits for insomnia disorder, making it difficult to interpret the findings in terms of clinical relevance. In addition, the top finding in this study (*MEIS1*) is a known risk factor for RLS, and it is still unclear

[1]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [2]Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam, the Netherlands. [3]Department of Sleep and Cognition, Netherlands Institute for Neuroscience (an institute of the Royal Netherlands Academy of Arts and Sciences), Amsterdam, the Netherlands. [4]Department of Integrative Neurophysiology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. [5]Department of Psychiatry, Vrije Universiteit Medical Center, Amsterdam, the Netherlands. [6]deCODE Genetics, Amgen, Inc., Reykjavík, Iceland. [7]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [8]Department of Respiratory Medicine and Sleep, Landspitali, National University Hospital of Iceland, Reykjavik, Iceland. [9]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [10]Institute of Neurogenomics, Helmholtz Zentrum München, Munich, Germany. [11]Institute of Human Genetics, Technische Universität München, Munich, Germany. [12]Neurologische Klinik und Poliklinik, Klinikum Rechts der Isar der Technischen Universität München, Munich, Germany. [13]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [14]Department of Clinical Genetics, Amsterdam Neuroscience, Vrije Universiteit Medical Center, Amsterdam, the Netherlands. [15]These authors contributed equally to this work. Correspondence should be addressed to D.P. (d.posthuma@vu.nl).

Received 29 October 2016; accepted 3 May 2017; published online 12 June 2017; doi:10.1038/ng.3888

whether the shared associations in this gene reflect causality, partial mediation or pleiotropy[24].

Here we report a GWAS using the UK Biobank sample[28] (see URLs) including 113,006 individuals (mean age = 56.92 years, s.d. = 7.94 years) to identify genetic risk factors related to insomnia complaints (Online Methods and **Supplementary Table 1**). The discriminative value of insomnia complaints for identifying insomnia disorder cases versus controls was validated in an independent cohort (n = 1,918 individuals). In addition, we report extensively on the possible mechanisms of action for the shared genetic signal in *MEIS1* on insomnia complaints and RLS.

RESULTS

Insomnia complaints are predictive of insomnia disorder

Insomnia disorder was assessed using a single question asking individuals whether they have trouble falling asleep at night or wake up in the middle of the night (**Supplementary Note** and **Supplementary Fig. 1**). Individuals who answered "usually" were scored as cases, and individuals reporting "never/rarely" or "sometimes" were scored as controls. We note that this operationalization differs from that in ref. 24, in which the same question in the UK Biobank sample was used but cases were defined as scoring "usually" and controls were defined as scoring "never/rarely" (**Supplementary Note**). We validated the predictive utility of this question for insomnia disorder in an independent sample of 1,918 participants (845 insomniacs and 1,073 controls) from the Netherlands Sleep Registry[29] (NSR; see URLs; **Supplementary Note**). The equivalent of the UK Biobank question (using our response category cutoff) in the NSR had a sensitivity of 98% and a specificity of 96% in discriminating questionnaire-defined insomnia disorder cases from unaffected controls (χ^2 = 1356.45, P < 0.0001) and a sensitivity of 0.96 and a specificity of 0.97 in discriminating insomnia disorder cases from cases exclusively having RLS (χ^2 = 639.06, P < 0.0001; **Supplementary Note** and **Supplementary Table 2**). It did not discriminate well between RLS cases and controls, with a sensitivity of only 0.43 and a specificity of 0.74 (χ^2 = 1.28, P = 0.26; **Supplementary Figs. 1** and **2**, and **Supplementary Table 2**), nor between cases and controls for 19 disease categories possibly related to insomnia (**Supplementary Note** and **Supplementary Fig. 3**). Further strong support for the validity of the UK Biobank insomnia phenotype was provided by an accuracy of 91% in the classification of clinical insomnia disorder in NSR participants, as diagnosed by a structured interview (**Supplementary Note**). Moreover, the sleep and mood characteristics of individuals who reported difficulties falling and staying asleep closely resembled the corresponding profile for NSR participants with insomnia disorder but not the profile for those with RLS (**Supplementary Note** and **Supplementary Fig. 4**). These findings show that the UK Biobank insomnia phenotype is predictive of insomnia disorder with little confounding from comorbidity. We refer to this classification as 'insomnia complaints' hereafter.

Genes implicated by GWAS and functional mechanisms

The prevalence of insomnia complaints was 29% in the UK Biobank sample (**Supplementary Table 1** and further descriptives in **Supplementary Table 3**), in keeping with previous estimates for people of advanced age in the UK[30] and elsewhere[31,32]. Females had a higher prevalence (33%) than males (24%). The sex odds ratio (OR) of 1.37 matches the previously published meta-analysis-derived estimate of 1.41 (ref. 2). GWAS was performed on all individuals of European descent, and standard quality control procedures included correction for population stratification and filtering based on minor allele frequency (MAF) and imputation quality (Online Methods). The GWAS included 12,444,916 SNPs. The pooled-sex analysis identified two genome-wide significant loci (**Fig. 1a** and **Table 1**), implicating two genes in insomnia disorder: *MEIS1* on chromosome 2 and *SCFD2* on chromosome 4 (**Supplementary Fig. 5a**,**b**). Sex-specific analyses implicated *MEIS1* in females as well, including the same significant SNPs found in the pooled-sex analysis, and an additional locus encompassing *WDR27* was identified for males on chromosome 6 (**Fig. 1b**,**c**, **Table 1** and **Supplementary Fig. 5c**,**d**). Both *MEIS1* and *WDR27* were identified by the recent GWAS for sleep disturbance traits[24], which also used UK Biobank data but considered a slightly different insomnia phenotype (**Supplementary Note** and **Supplementary Figs. 6** and **7**). Our most significant SNPs in both genes were the same as in this study, with similar association signals (*MEIS1*: rs113851554, $P = 9.11 \times 10^{-19}$; *WDR27*: rs13192566, $P = 3.17 \times 10^{-8}$).

Possible functional mechanisms for the identified SNPs and the SNPs in high linkage disequilibrium (LD) with them ($r^2 > 0.6$) are reported in the **Supplementary Note** and **Supplementary Table 4**. Most of the genome-wide significant SNPs were intronic and unlikely to be deleterious or were part of a regulatory element. However, the SNPs in the locus on chromosome 6 were associated with increased expression of two neighboring genes in blood cells (*PHF10*, lowest $P = 3.65 \times 10^{-13}$; *C6orf120*, lowest $P = 3.81 \times 10^{-13}$). One SNP (rs113851554) in the *MEIS1* locus showed evidence ($P = 1.08 \times 10^{-6}$, false discovery rate (FDR) < 0.05) of acting as a *cis* methylation quantitative trait locus (meQTL). Credible set analysis (Online Methods) of the SNPs in the *MEIS1* locus identified two variants (rs113851554 and rs182588061) within the 99% confidence set of associated variants that are plausibly the causal variants (**Supplementary Table 5**). When incorporating functional annotation, rs113851554 accounted for the full posterior probability of association, suggesting that the SNP in the *MEIS1* locus most strongly associated with insomnia disorder is also the most likely causal SNP.

SNP heritability

SNP-based heritability was estimated to be 0.09 (s.e.m. = 0.0082) by LD score regression[33] (LDSC) and 0.11 (s.e.m. = 0.0093) by BOLT-REML[34] (BR). The sex difference was small, with estimates of 0.12 (s.e.m. = 0.018; LDSC) and 0.11 (s.e.m. = 0.02; BR) in males versus 0.08 (s.e.m. = 0.014; LDSC) and 0.09 (s.e.m. = 0.02; BR) in females. The quantile–quantile plots for all SNPs exhibited only mild inflation (λ_{all} = 1.11; λ_{males} = 1.06; $\lambda_{females}$ = 1.05; **Supplementary Fig. 8**), as is expected for a polygenic trait using the current sample size. The intercepts estimated by LD score regression of 1.00, 0.99 and 1.00 for the sex-combined, male-only and female-only analyses, respectively, suggest that this mild inflation is unlikely to be due to population stratification.

Genes implicated by GWGAS

A GWGAS, as implemented in MAGMA[35] (Online Methods), on all individuals identified three genes associated with insomnia complaints: *MEIS1* (also implicated by the GWAS), *DCBLD1* and *MED27*. Sex-specific GWGAS identified two additional genes (*HHEX* and *RHCG*) for males and two additional genes (*IPO7* and *TSNARE1*) for females (**Fig. 1d–f**, **Table 2** and **Supplementary Table 6**). Some of these genes have previously been associated with other phenotypes, such as diabetes and schizophrenia (**Supplementary Table 7** and **Supplementary Note**). The most strongly associated gene, *MEIS1*, encodes a homeobox protein that acts as a transcriptional regulator and activator, and it is thought to be important for normal development[36]. *MEIS1* shows the highest expression levels in female internal reproductive organs, but it is also expressed in many other

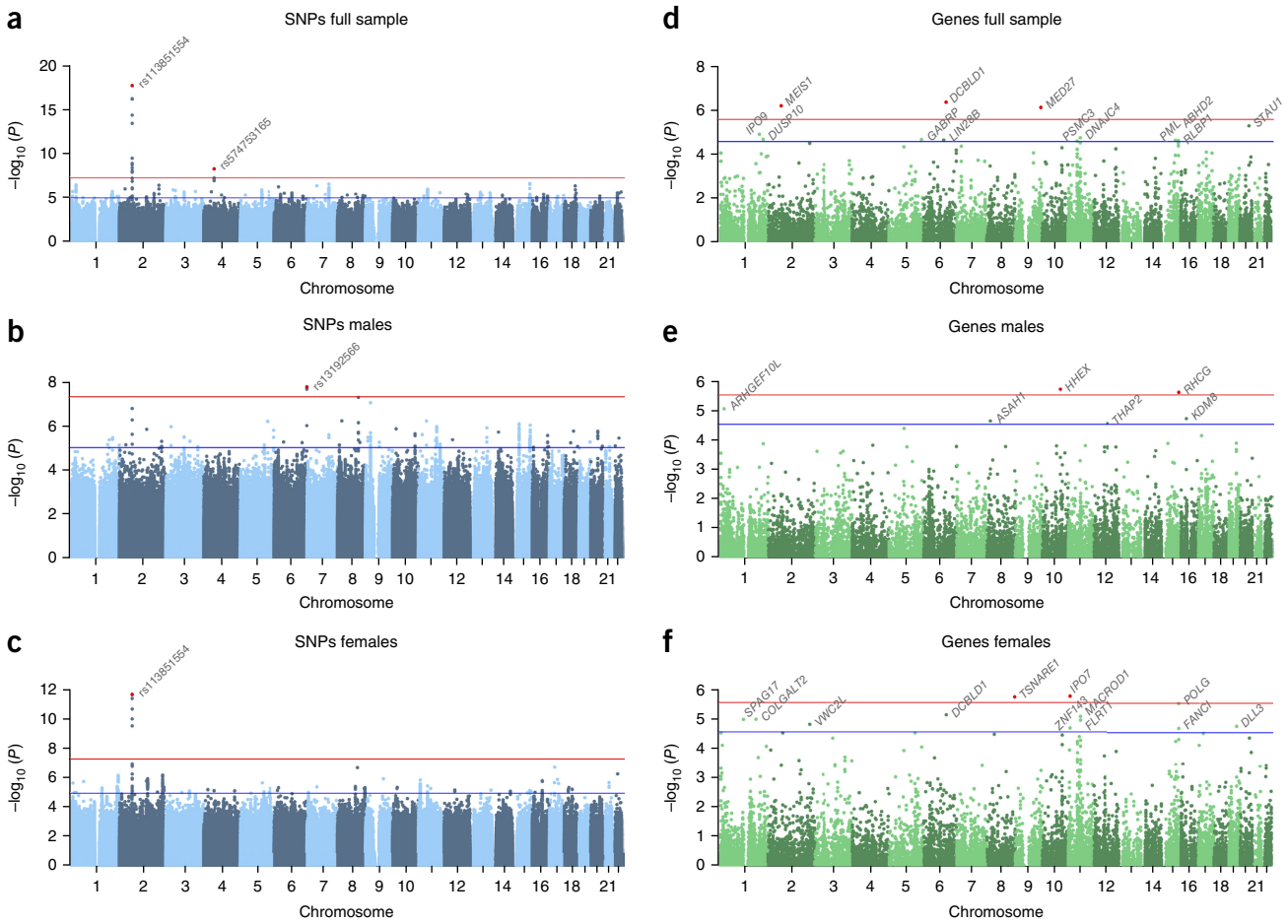

Figure 1 Manhattan plots showing SNP and gene associations with insomnia complaints. Association results are shown for the frequency of experiencing trouble falling asleep or waking up in the middle of the night in 113,006 individuals of European descent in the UK Biobank study, of whom those experiencing these complaints usually (n = 32,384 cases) were contrasted with those experiencing these complaints never, rarely or only sometimes (n = 80,622 controls). (**a**–**c**) Manhattan plots of the GWAS performed in SNPTEST including all individuals (**a**), males only (n = 12,863 cases and 40,776 controls) (**b**) and females only (n = 19,521 cases and 39,846 controls) (**c**). Negative $\log_{10}$-transformed P values for each SNP (y axis) are plotted by chromosomal position (x axis). The red and blue lines represent the thresholds for genome-wide significant association ($P = 5 \times 10^{-8}$) and suggestive association ($P = 1 \times 10^{-5}$), respectively. Each dot represents an individual SNP. Red dots represent top SNPs. (**d**–**f**) Manhattan plots of the gene analysis performed in MAGMA including all individuals (**d**), males only (**e**) and females only (**f**). Each dot represents a gene, and the red and blue lines represent the thresholds for gene-wide significant association ($P = 2.72 \times 10^{-6}$) and suggestive association ($P = 2.72 \times 10^{-5}$), respectively.

tissues, including the brain[37]. *HHEX* and *MED27* are also involved in the regulation of transcription. *TSNARE1* and *SCFD2* (implicated in the GWAS) play a role in exocytosis (**Supplementary Note**, **Supplementary Tables 8** and **9**, and **Supplementary Figs. 9**–**12**).

Joint analysis with an independent sample

To examine the robustness of the three loci and seven genes that reached genome-wide significance in the primary analyses, we tested their association with a well-defined insomnia phenotype in a sample from deCODE comprising n = 7,565 individuals (Online Methods, **Supplementary Note** and **Supplementary Tables 10**–**12**) and performed meta-analysis of these results together with the UK Biobank association results while adhering to the GWAS significance threshold of $P = 5 \times 10^{-8}$ (Online Methods and **Supplementary Tables 13** and **14**). The probability of replicating significant SNPs in the deCODE sample was low, owing to the difference in sample size (**Supplementary Note** and **Supplementary Table 15**), whereas meta-analysis takes into account the standard errors (sample size) for the observed effects and allows for evaluation of whether discovery P values increase (suggesting no replication) or decrease (supporting a similar effect in the added sample)[38]. The effects of 11 of the 12 SNPs from the full GWAS (both sexes combined) and all 5 SNPs from the female GWAS were sign concordant, whereas those for the 3 SNPs from the male GWAS were not. Only SNPs in the *MEIS1* locus (six SNPs in the full GWAS and two SNPs in the female GWAS) showed a stronger association signal for insomnia complaints in the meta-analysis. Six of the seven genes detected in the GWGAS were significant at the genome-wide gene-based threshold of $P = 2.72 \times 10^{-6}$ in the meta-analysis. The signal for *MEIS1*, as well as for the four genes found to be associated in the sex-specific analyses, was stronger than in the initial GWGAS and remained below the genome-wide gene-based threshold of association in the meta-analysis.

The role of *MEIS1* in insomnia complaints and RLS

The gene most strongly associated with insomnia complaints was *MEIS1*. Winkelmann and colleagues[39,40] previously reported an association of multiple SNPs in *MEIS1* with RLS, and two of our top SNPs were previously associated with clinically diagnosed RLS in a sequencing[41] and gene expression[42] study of *MEIS1*. RLS is a prevalent disorder characterized by the urge to move the legs, a symptom

Table 1 Three genome-wide significant loci associated with insomnia complaints in a full GWAS including 113,006 individuals and sex-specific GWAS

								Full			Male			Female		
rsID	Annotation	Chr.	Position (bp)[a]	EA	Non-EA	INFO	EAF	OR	95% CI	*P*	OR	95% CI	*P*	OR	95% CI	*P*
rs375216017	*MEIS1* intronic	2	66,728,627	GT	G	0.942	0.105	1.09	1.05–1.12	**1.21×10^{-8}**	1.09	1.04–1.14	2.27×10^{-8}	1.09	1.04–1.13	1.60×10^{-5}
rs62144051	*MEIS1* intronic	2	66,730,783	G	A	0.964	0.094	1.10	1.06–1.13	**3.81×10^{-9}**[c]	1.09	1.04–1.14	9.21×10^{-4}	1.10	1.06–1.15	1.01×10^{-6}
rs62144053	*MEIS1* intronic	2	66,745,864	A	G	0.964	0.095	1.10	1.06–1.13	**1.20×10^{-9}**[c]	1.08	1.03–1.13	1.79×10^{-3}	1.11	1.07–1.16	1.18×10^{-7}
rs62144054	*MEIS1* intronic	2	66,747,480	A	G	0.976	0.094	1.10	1.06–1.13	**1.68×10^{-9}**[c]	1.08	1.03–1.13	2.45×10^{-3}	1.11	1.07–1.16	1.09×10^{-7}
rs113851554	*MEIS1* intronic	2	66,750,564	T	G	1	0.056	1.19	1.14–1.24	**2.14×10^{-18}**[c]	1.18	1.11–1.25	1.69×10^{-7}	1.20	1.14–1.26	**2.25×10^{-12}**[c]
rs182588061	*MEIS1* intronic	2	66,757,709	T	G	0.852	0.020	1.21	1.13–1.28	**3.18×10^{-10}**[b]	1.21	1.10–1.33	7.08×10^{-5}	1.21	1.11–1.32	1.13×10^{-6}
rs139775539	*MEIS1* intronic	2	66,782,432	A	AC	0.961	0.048	1.19	1.14–1.24	**7.00×10^{-17}**[b]	1.17	1.09–1.24	2.49×10^{-6}	1.21	1.14–1.28	**4.08×10^{-12}**[b]
rs11679120	*MEIS1* intronic	2	66,785,180	A	G	0.960	0.047	1.19	1.14–1.24	**6.18×10^{-17}**[b]	1.18	1.11–1.26	5.54×10^{-7}	1.20	1.14–1.27	**2.09×10^{-11}**[b]
rs115087496	*MEIS1* intronic	2	66,793,725	C	G	0.950	0.047	1.18	1.13–1.23	**4.43×10^{-15}**[b]	1.16	1.09–1.24	7.00×10^{-6}	1.19	1.13–1.26	**9.91×10^{-11}**[b]
rs549771308	*MEIS1* intronic	2	66,795,237	C	CT	0.832	0.120	1.08	1.05–1.11	**9.51×10^{-9}**[c]	1.06	1.02–1.11	3.14×10^{-3}	1.09	1.05–1.13	5.17×10^{-7}
rs11693221	*MEIS1* downstream	2	66,799,986	T	C	0.943	0.048	1.17	1.12–1.22	**3.79×10^{-14}**[c]	1.15	1.08–1.23	1.92×10^{-5}	1.19	1.12–1.25	**2.88×10^{-10}**[c]
rs574753165	*SCFD2* intronic	4	53,977,261	G	A	0.930	0.005	1.40	1.24–1.57	**4.98×10^{-9}**[b]	1.38	1.15–1.65	1.85×10^{-4}	1.42	1.21–1.66	6.84×10^{-6}
rs71554396	Intergenic	6	169,841,072	GT	G	0.946	0.137	1.95	0.93–0.98	1.43×10^{-3}	1.89	0.85–0.93	**1.95×10^{-8}**	1.01	0.98–1.05	5.17×10^{-1}
rs13208844	*WDR27* intronic	6	169,961,603	G	A	1	0.147	1.96	0.93–0.98	1.50×10^{-3}	1.89	0.85–0.93	**2.25×10^{-8}**[b]	1.01	0.98–1.05	5.21×10^{-1}
rs13192566	*WDR27* intronic	6	169,961,635	C	G	0.999	0.146	1.95	0.93–0.98	1.26×10^{-3}	1.89	0.85–0.93	**1.80×10^{-8}**[b]	1.01	0.98–1.05	5.41×10^{-1}

Summary of the three significant loci present in the full GWAS and/or the male (n = 53,639) and female (n = 59,367) GWAS. All SNPs with $P < 5 \times 10^{-8}$ in one of the analyses are reported for all three analyses. SNP P values and odds ratios were calculated for each GWAS with an additive genetic model using logistic regression adjusted for age, sex in the full GWAS, genotyping array and principal components. P values for loci significantly associated with insomnia ($P < 5 \times 10^{-8}$) are shown in bold. SNP association results in the deCODE sample and the meta-analysis are reported in **Supplementary Tables 11** and **13**. Chr., chromosome; EA, effect allele; EAF, effect allele frequency; OR, odds ratio; CI, confidence interval.

[a]Reported on GRCh37. [b]Genome-wide significant in in the combined analysis of UK Biobank and deCODE. [c]Stronger association signal in the combined analysis of UK Biobank and deCODE.

that has been suggested to involve deficiencies in the dopaminergic system, whereas arousals during sleep, a symptom present in some individuals with RLS, is related to the balance between glutamate and GABA[43]. The latter activity has also been implicated in hyperarousal in insomnia[44]. RLS and insomnia have some form of agitation or restlessness in common, which is expressed primarily in the cognitive domain in insomnia and in the sensorimotor domain in RLS. Given the possible phenotypic overlap between insomnia and RLS and the reported genetic associations in *MEIS1* for RLS, we investigated the mechanisms of action that could explain the shared signal in *MEIS1* for insomnia complaints and RLS.

First, we investigated whether the observed associations in *MEIS1* with insomnia complaints and RLS were independent. At least two signals in *MEIS1* are associated with RLS: one including common SNPs reported in refs. 39,40 and a second including low-frequency SNPs reported by refs. 41,42. The RLS-associated SNPs in *MEIS1* reported by Winkelmann and colleagues were not genome-wide significant in our insomnia complaints GWAS and were in low LD with our associated SNPs (**Supplementary Table 16**). Conditioning our top SNP, rs113851554, on the RLS-associated SNPs (Online Methods) showed that this SNP has an effect on insomnia complaints ($P = 2.90 \times 10^{-13}$) independent of those of the RLS-associated SNPs in *MEIS1*. We did not condition on the two SNPs from the second RLS signal in *MEIS1* (our top SNP, rs113851554 (ref. 42), and rs11693221 (ref. 41)), as these SNPs were part of our top association findings for insomnia complaints (**Table 1**).

Second, we applied BUHMBOX[45], which provided information on the likelihood that a subgroup of individuals with genetic characteristics of RLS existed within the group of individuals with insomnia complaints that could explain our association results (Online Methods). After establishing sufficient power (0.82 reported by the BUHMBOX power calculator, calculated on the basis of sample size, the effect sizes of RLS-associated SNPs and RLS prevalence; Online Methods) to detect heterogeneity when defining RLS genetic structure by the six RLS-associated loci reported in the RLS GWAS[40] (**Supplementary Table 17**), we found no evidence that an RLS subgroup by itself was driving the reported associations, both when excluding the *MEIS1* locus ($P = 0.33$) and when including it ($P = 0.36$). However, after adding our top SNP in *MEIS1* for insomnia complaints, which was also previously associated with RLS[42] ($P = 4.80 \times 10^{-12}$), we did find excessive positive correlation between the RLS-associated loci and a subgroup of the individuals reporting insomnia ($P = 0.029$), although the correlation was driven by this single SNP. Note that, because of a lack of power (due to the small RLS sample size and smaller effect sizes of the insomnia-associated loci), we could not test the reverse hypothesis that an insomnia subgroup drives the RLS associations.

Third, we performed a genetic risk score analysis (Online Methods) to interpret the BUHMBOX results. This yielded a significant association between insomnia complaints and the RLS-associated loci when including the same RLS-associated loci that were tested in BUHMBOX (**Supplementary Table 17**): (i) five RLS loci excluding the *MEIS1* locus ($P = 7.28 \times 10^{-3}$); (ii) six RLS loci including the *MEIS1* association from the RLS GWAS ($P = 6.23 \times 10^{-4}$); and (iii) six RLS loci including the *MEIS1* RLS-associated SNP that was also the top hit for insomnia complaints ($P = 5.30 \times 10^{-13}$). As in the BUHMBOX analysis, including the top signal for insomnia complaints strongly increased the association. The results of the BUHMBOX and genetic risk score analyses together are compatible with pleiotropy, but phenotypic overlap between RLS and insomnia complaints might contribute to the association found in the *MEIS1* locus.

Table 2 Genome-wide significant genes associated with insomnia complaints in a GWGAS including 113,006 individuals and sex-specific GWGAS

Gene	Entrez ID	Chr.	Start position (bp)[a]	Stop position (bp)[a]	Full			Male			Female		
					n SNPs	*P*	$P_{corrected}$	*n* SNPs	*P*	$P_{corrected}$	*n* SNPs	*P*	$P_{corrected}$
DCBLD1	285761	6	117,801,803	117,892,021	384	4.54×10^{-7}	**0.0083**	385	0.0856	1	386	7.12×10^{-4}	0.131
MEIS1	4211	2	66,660,257	66,800,891	426	6.60×10^{-7}	**0.0121**[c]	424	7.56×10^{-3}	1	428	3.58×10^{-4}	1
MED27	9442	9	134,734,497	134,957,274	608	7.81×10^{-7}	**0.0143**[b]	608	9.56×10^{-3}	1	603	5.41×10^{-4}	1
HHEX	3087	10	94,447,681	94,456,408	19	2.49×10^{-3}	1	19	1.71×10^{-6}	**0.031**[c]	19	0.5832	1
RHCG	51458	15	90,013,638	90,041,799	35	1.15×10^{-4}	1	34	2.19×10^{-6}	**0.040**[c]	35	0.7115	1
IPO7	10527	11	9,404,169	9,470,674	241	7.80×10^{-4}	1	234	0.7839	1	245	1.67×10^{-6}	**0.031**[c]
TSNARE1	203062	8	143,292,441	143,486,543	1,083	0.029503	1	1,074	0.9498	1	1,079	1.76×10^{-6}	**0.032**[c]

Gene-based *P* values are reported for all genes significant after Bonferroni correction ($\alpha = 2.72 \times 10^{-6}$) in at least one of the three analyses (full, males, females). *P* values for loci significantly associated with insomnia are shown in bold. Gene association results in the deCODE sample and the meta-analysis are reported in **Supplementary Tables 12** and **14**. GWGAS, genome-wide gene-association study; chr., chromosome; $P_{corrected}$, *P* value corrected for multiple testing.

[a]Reported on GRCh37, including a window of 2.1 kb. [b]Genome-wide significant in the combined analysis of UK Biobank and deCODE. [c]Stronger association signal in the combined analysis of UK Biobank and deCODE.

Fourth, we investigated possible confounding between RLS and insomnia in the *MEIS1* locus using conditional phenotypic analysis in data from the Course of Restless Legs Syndrome (COR) study (individuals with RLS; included in ref. 40) and the Dortmund Health Study (DHS; individuals without RLS); each study contains information on insomnia complaints (Online Methods and **Supplementary Note**). The combined 'COR + DHS' sample included 1,985 individuals with quality-controlled genotypes (**Supplementary Table 18**). We note that this sample has strong ascertainment biases due to the COR sample (53% of the combined sample) consisting only of individuals who are relatively old (65 years, on average) and, as members of RLS support groups, tend to have severe RLS. The resulting biases are (i) oversampling of insomnia evoked by severe RLS, as all individuals with insomnia complaints in the COR study necessarily have RLS, and (ii) over-representation of RLS comorbidity with insomnia, as the prevalence of insomnia increases with age and people seeking help for RLS may be more likely to have insomnia as a comorbidity. Taking these biases into consideration, *MEIS1* was found to be strongly associated with insomnia complaints, supporting our initial finding, and was associated with RLS as expected (**Supplementary Tables 19** and **20**). Conditioning the signals for insomnia complaints on those for RLS and vice versa reduced the association signals, indicating that phenotypic overlap contributes to the associations for both phenotypes. However, this reduction in signal, which was complete when conditioning insomnia complaints on RLS but incomplete when conditioning RLS on insomnia complaints, cannot exclude the possibility of pleiotropy at *MEIS1* because of the strong ascertainment biases. Up to 83% of all insomnia cases in the combined COR + DHS group may belong to the subgroup in which insomnia complaints are evoked by severe RLS or RLS is a comorbidity (**Supplementary Table 18**), explaining why an effect of *MEIS1* on insomnia was not visible after conditioning on RLS.

Fifth, we predicted association *P* values for insomnia complaints in the UK Biobank sample under the assumption that RLS alone drives the association. To this end, we calculated the expected proportions of individuals with RLS in the UK Biobank insomnia cases and controls, using data on age-specific RLS prevalence[46], the sensitivity and specificity of the UK Biobank question for RLS, as derived from the NSR and DHS sample, and the reported rs113851554 effect size in RLS cases and controls[42] (**Supplementary Note**). A χ^2 test yielded a predicted association *P* value of 2×10^{-4} (95% confidence interval (CI) = 0.056 to 1.1×10^{-11}, based on sampling variances; **Supplementary Note**). This predicted *P* value under the assumption that the complete signal is driven by an RLS subgroup was much weaker than the actual *P* value we observed (2.14×10^{-18}), and the latter was outside the 95% CI of the predicted *P* value. This finding thus supports the notion that the effect of *MEIS1* on insomnia complaints can, at most, be explained only in part by contamination of the UK Biobank insomnia cases with RLS cases.

Finally, we conducted tests for sign concordance and enrichment of low *P* values (Online Methods) on the summary statistics for insomnia complaints and RLS (unfortunately, the sample size for the RLS GWAS[40] was insufficient to obtain a reliable estimate of genetic correlation). Seventy-eight percent of the independent top SNPs ($P = 1 \times 10^{-4}$) for insomnia complaints had sign-concordant effects in RLS, whereas 83% of the top SNPs for RLS had sign-discordant effects in insomnia complaints (**Supplementary Table 21**). The top signals from the two studies showed little overlap (**Supplementary Tables 22** and **23**, and **Supplementary Fig. 13**). This suggests that, besides pleiotropy at some loci, there are genetic factors specific to each of the two disorders.

Taken together, the above results suggest that phenotypic overlap between RLS and insomnia can drive some, but not all, of the association of *MEIS1* with insomnia. This confounding effect of RLS on insomnia complaints association likely also occurs in the opposite direction (with insomnia complaints confounding RLS association). Hence, we conclude that *MEIS1* is likely to have pleiotropic effects on both RLS and insomnia.

Overlap with sleep-related phenotypes

Data on multiple sleep-related phenotypes are available in the UK Biobank, and multiple loci have been found to be associated with sleep duration[24,26], chronotype[25,26] and excessive daytime sleepiness[24]. We performed GWAS on six additional sleep phenotypes in the UK Biobank (**Supplementary Fig. 14** and **Supplementary Table 24**) and investigated genetic and phenotypic correlations with insomnia complaints (**Supplementary Note**). Phenotypically, individuals reporting insomnia complaints have shorter sleep duration, more trouble getting up and more periods of unintentional dozing, but they do not show a systematically different chronotype. Furthermore, insomnia complaints showed a significant positive genetic correlation with daytime dozing or sleeping ($r_g = 0.51$, $P = 3.25 \times 10^{-4}$) and napping during the day ($r_g = 0.42$, $P = 3.95 \times 10^{-6}$) and a negative genetic correlation with sleep duration ($r_g = 0.47$, $P = 1.97 \times 10^{-16}$; **Supplementary Table 25**). The loci found to be associated with insomnia complaints showed no significant association with the six additional sleep phenotypes (**Supplementary Table 26**). In addition, we investigated possible confounding by other psychiatric, metabolic and socioeconomic

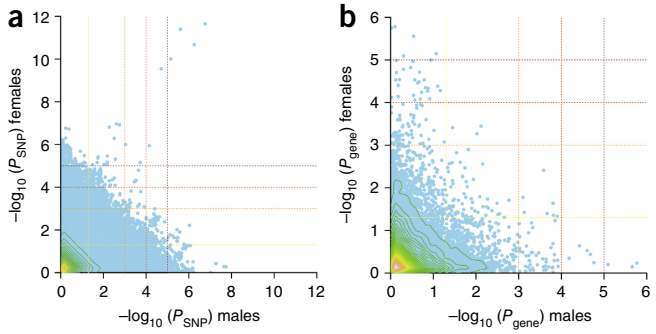

Figure 2 Comparison of association results for insomnia complaints in males and females. (**a**,**b**) SNP (**a**) and gene (**b**) associations with insomnia complaints, calculated in SNPTEST and MAGMA, respectively, in males plotted against those in females. Contour lines indicate the density of the data in each region. The lines are colored from green to yellow, indicating increasing data density. Dotted lines indicate the *P*-value thresholds used in the test for enrichment of low *P* values: from yellow to red, $P = 0.05$, $P = 1 \times 10^{-3}$, $P = 1 \times 10^{-4}$ and $P = 1 \times 10^{-5}$ (note that all SNPs present in both GWAS are displayed, although the enrichment tests were performed on pruned data).

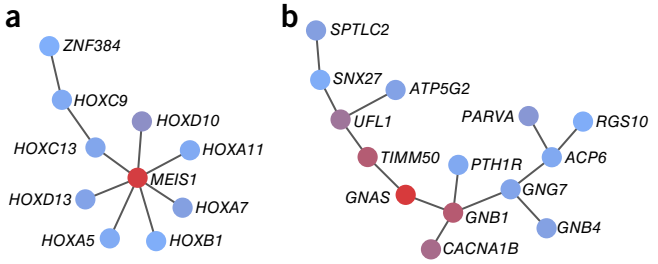

Figure 3 Protein–protein interaction subnetworks identified by the heat diffusion algorithm HotNet2. The genes most strongly related to insomnia ($P < 0.1$) in the full and sex-specific GWGAS were used as input to investigate the enrichment of protein–protein interaction networks. Each node (protein) is assigned a score on the basis of the gene *P* value from the GWGAS. The scores, denoted as 'heat' in HotNet2, diffuse across the edges of the network. We report subnetworks that include genes with a wide range of heat scores: red nodes send and receive a substantial amount of heat, whereas blue nodes do not. (**a**) Subnetwork identified for females including *MEIS1*. (**b**) Subnetwork identified for females including *GNAS*. Other identified subnetworks for females and males are shown in **Supplementary Figure 18**.

traits (Online Methods and **Supplementary Note**). Adjustment of the associations of the significant SNPs for insomnia complaints by these traits did not show evidence that they have confounding effects (**Supplementary Table 27**).

Sex-related differences in genetic associations

Females have a higher predisposition for insomnia than males[2], which might result from sex-related differences in genetic architecture. The genetic correlation between the sexes was estimated to be 0.79 (s.e.m. = 0.13), which is just significantly smaller than 1 (one-sided Wald test, $P = 0.045$). This estimate is comparable with, for example, that for waist circumference, for which between-sex genetic heterogeneity is expected, in contrast to estimates for height and BMI, where no heterogeneity is found[47]. In keeping with these overall differences, the significant SNP- and gene-based association results also differed between the sexes, except for those in *MEIS1* (**Fig. 2**). Adding sex as an interaction term to the GWAS on the full sample (Online Methods) did not result in genome-wide significant interactions (**Supplementary Fig. 15**), although this finding might also be due to low statistical power for interaction analyses. Tests of sign concordance and enrichment of low *P* values for sex-specific results showed little evidence for overlap in the top signals between the sexes (**Fig. 2**, Online Methods and **Supplementary Tables 28–30**), suggesting that sex has a role in evoking specific genetic risk factors of insomnia. Whether X-chromosome loci and sex-specific imprinting[48,49] play a role still needs to be determined. Our finding is in line with sex differences across most sleep variables[50], including subjective sleep complaints[51], prevalence[2] and heritability[12] of insomnia, and physiological signatures of sleep, both in the general population[52] and within the population that has insomnia[53].

Functional networks

We applied the heat diffusion algorithm HotNet2 (ref. 54) (Online Methods) to investigate protein–protein interaction networks enriched for the genes most strongly associated with insomnia complaints ($P < 0.1$) in the full and sex-specific GWGAS (**Supplementary Note**, **Supplementary Table 31** and **Supplementary Figs. 16** and **17**). For each input gene, HotNet2 processes a heat diffusion algorithm on the protein–protein interaction network to define a local neighborhood of 'influence', which is followed by a two-stage multiple-hypothesis test to identify recurrent subnetworks. As input for HotNet2, we selected genes with $P < 0.1$ from the GWGAS, thereby considering crosstalk across pathways and network topology. In total, we observed 12 subnetworks of genes for males ($P = 0.01$) and 9 subnetworks for females ($P = 0.02$; **Supplementary Fig. 18**). These subnetworks significantly overlapped with known pathways that were mostly involved in transcription (Online Methods and **Supplementary Table 32**). In females, one subnetwork involved *MEIS1* (**Fig. 3a**) together with multiple homeobox genes encoding a family of transcription factors important for development. Other subnetworks presented candidate genes at 'hotspots' that were not detected by GWGAS alone; among these, was *GNAS* in the largest of the subnetworks in females (**Fig. 3b**). *GNAS* is an imprinted gene that is expressed from the maternal chromosome. It has metabolic functions and modulates REM and NREM sleep states[49]. This finding is especially interesting given that stronger maternal transmission[21], hypermetabolism[55] and instability of these sleep states[56] are all characteristic of insomnia. Future studies will be needed to confirm the involvement of the identified subnetworks in insomnia.

Genetic overlap with other traits

Insomnia implies an increased risk for major health problems, notably in the domains of cardiovascular diseases[5–9], obesity[11] and psychiatric disorders like depression[4]. Using whole-genome LD score regression[33], we assessed the genetic correlation of insomnia with 29 traits from these domains and additional anthropometric and lifestyle traits (Online Methods). Significant genetic correlations (conservatively adjusted for multiple testing: $P < 1.72 \times 10^{-3}$ (= 0.05/29)) were observed between insomnia complaints and ten other traits (**Fig. 4**, left and **Supplementary Table 33**). Strong positive genetic correlations were observed with anxiety ($r_g = 0.59$, $P = 7.14 \times 10^{-5}$), depressive symptoms ($r_g = 0.53$, $P = 1.03 \times 10^{-17}$), neuroticism ($r_g = 0.44$, $P = 1.20 \times 10^{-25}$) and major depressive disorder ($r_g = 0.41$, $P = 6.50 \times 10^{-4}$). Other positive yet weaker genetic correlations were observed with metabolic traits, including type 2 diabetes, waist circumference, waist-to-hip ratio and body mass index. Strong but negative genetic correlations were observed with subjective well-being ($r_g = -0.44$, $P = 5.64 \times 10^{-11}$) and educational attainment ($r_g = -0.34$, $P = 1.81 \times 10^{-22}$). Of the 29 traits, 18 have also been assessed in the NSR, allowing investigation of the phenotypic differences between these phenotypes in insomnia cases versus controls. We found that the profile of magnitudes for the differences (*d*) between phenotypic groups

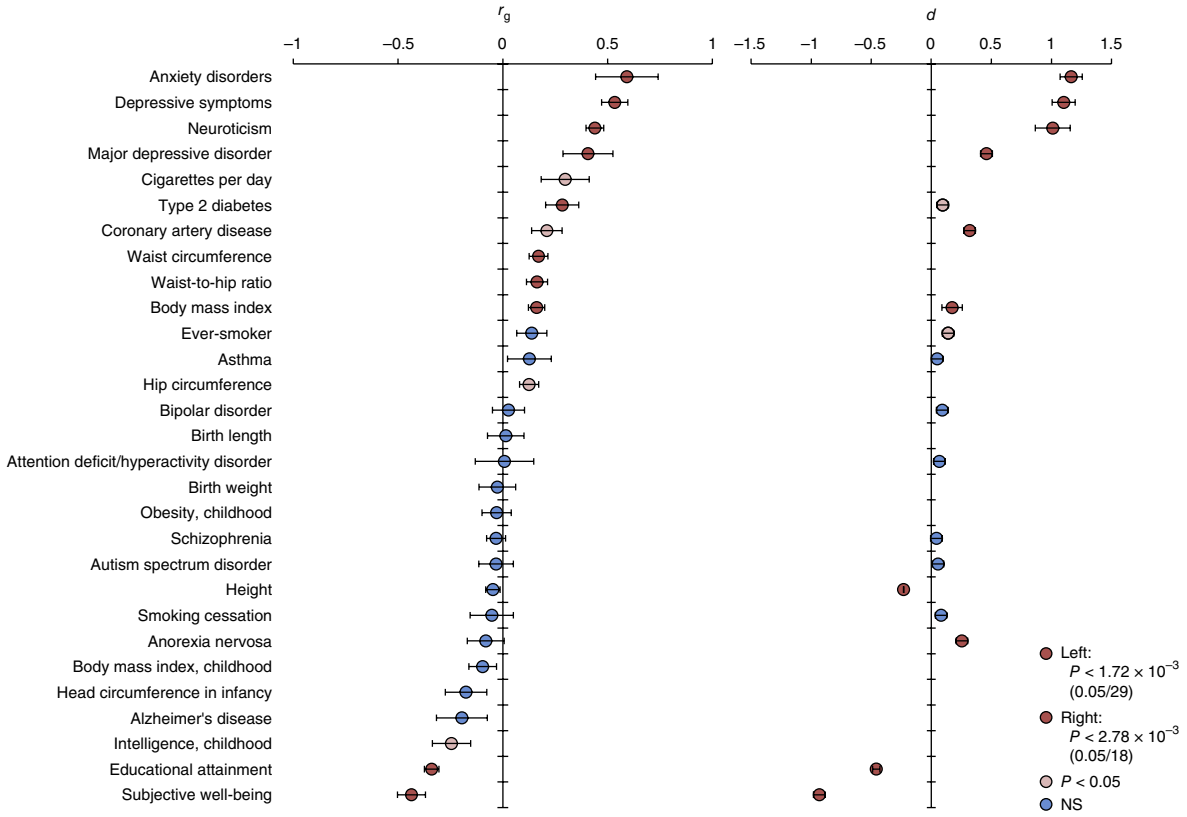

Figure 4 Genetic and phenotypic overlap between insomnia complaints and other traits and disorders. Left, genetic correlations (r_g) between the frequency of experiencing trouble falling asleep or waking up in the middle of the night and various other traits and diseases. LD score regression tested genome-wide SNP associations for these insomnia complaints against similar data for 29 other anthropometric and cardiovascular traits and neuropsychiatric outcomes (**Supplementary Table 33**). Error bars represent standard errors on these estimates. Red bars represent traits that showed a significant genetic correlation after correction for multiple testing ($P < 1.72 \times 10^{-3}$), pink bars represent traits that showed nominal association ($P < 0.05$) and blue bars represent traits that did not show a significant genetic association. Right, the genetic correlations profile was strikingly similar to phenotypic overlap of insomnia with the same subject characteristics assessed in an independent sample. Of the 29 disorders, traits and characteristics, 18 had been assessed in the NSR as well. Group differences between the 1,073 individuals without insomnia complaints and the 845 individuals likely to have insomnia disorders were evaluated using *t* tests (continuous phenotypes) or χ^2 tests (dichotomous phenotypes) (**Supplementary Table 34**). The profile of the magnitudes (*d*) of phenotypic group differences strongly resembled the genetic correlation profiles. NS, not significant.

(Online Methods) was strikingly similar to the profile for genetic correlations (rank correlation = 0.82; **Fig. 4**, right and **Supplementary Table 34**), providing further evidence for a link between the above-mentioned traits and insomnia.

DISCUSSION

We conducted a large-scale GWAS and GWGAS on insomnia complaints, using a measure that reliably discriminates individuals with insomnia disorders from unaffected controls. We identified five new genes (*MEIS1*, *HHEX*, *RHCG*, *IPO7* and *TSNARE1*) and one locus (encompassing *MEIS1*) associated with insomnia complaints whose association was supported by joint analysis with an independent sample. These findings partly (two genes) overlapped with those reported in a recently published GWAS[24] using a slightly different operationalization of the phenotype, which was less discriminative of clinical insomnia disorder cases versus controls. The gene most significantly associated with insomnia in our study was *MEIS1*, which was also found in ref. 24 and has previously been implicated in RLS. Our extensive analyses now show that both residual phenotypic overlap and pleiotropy are relevant in the involvement of *MEIS1* in insomnia as well as RLS.

We also provide evidence of sex-specific genetic effects and show genetic overlap with several psychiatric and metabolic disorders. These findings provide starting points for subsequent functional analyses to unravel the molecular neurobiological mechanisms underlying vulnerability to insomnia disorder.

URLs. UK Biobank, http://www.ukbiobank.ac.uk/; Netherlands Sleep Registry, https://www.sleepregistry.nl/; UK Biobank genotyping and quality control, http://biobank.ctsu.ox.ac.uk/crystal/refer.cgi?id=155580; SNPTEST, https://mathgen.stats.ox.ac.uk/genetics_software/snptest/snptest.html; MAGMA, http://ctg.cncr.nl/software/magma; Sanger Imputation Service, https://imputation.sanger.ac.uk/; genotyping chip strand files, http://www.well.ox.ac.uk/~wrayner/strand/; dbSNP data, ftp://ftp.ncbi.nih.gov/snp/; PLINK, https://www.cog-genomics.org/plink2; BUHMBOX, http://software.broadinstitute.org/mpg/buhmbox/; METAL, http://genome.sph.umich.edu/wiki/METAL_Program; MSigDB, http://software.broadinstitute.org/gsea/msigdb/; LD score regression, https://github.com/bulik/ldsc.

METHODS

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

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

ACKNOWLEDGMENTS

This work was funded by the Netherlands Organization for Scientific Research (NWO Brain & Cognition 433-09-228, NWO VICI 453-14-005 and 453-07-001, 645-000-003) and by the European Research Council (ERC-ADG-2014-671084 INSOMNIA). The analyses were carried out on the Genetic Cluster Computer, which is financed by the Netherlands Scientific Organization (NWO; 480-05-003), by VU University (Amsterdam, the Netherlands) and by the Dutch Brain Foundation and is hosted by the Dutch National Computing and Networking Services (SurfSARA). This research has been conducted using the UK Biobank Resource under application number 16406. We thank the participants and researchers who contributed and collected the data. We thank the participants of the Netherlands Sleep Registry for providing extensive phenotypic data. We thank the participants who provided samples and data for the Icelandic study and our valued colleagues who contributed to data collection and the phenotypic characterization of clinical samples, genotyping and analysis of genome sequence data. We also thank the EU-RLS consortium and the Cooperative Research in the Region of Augsburg (KORA) study for providing the RLS summary statistics. KORA was initiated and is financed by the Helmholtz Zentrum München, which is funded by the German Federal Ministry of Education and Research and by the state of Bavaria. The collection of sociodemographic and clinical data in the DHS was supported by the German Migraine & Headache Society (DMKG) and by unrestricted grants of equal share from Almirall, AstraZeneca, Berlin Chemie, Boehringer, Boots Health Care, GlaxoSmithKline, Janssen Cilag, McNeil Pharma, MSD Sharp & Dohme, and Pfizer to the University of Münster. Blood collection in the DHS was done through funds from the Institute of Epidemiology and Social Medicine at the University of Münster. Genotyping for the Human Omni chip was supported by the German Ministry of Education and Research (BMBF; grant 01ER0816). Researchers interested in using DHS data are required to sign and follow the terms of a cooperation agreement that includes a number of clauses designed to ensure protection of privacy and compliance with relevant laws. The COR study was supported by unrestricted grants to the University of Münster from the German Restless Legs Patient Organisation (RLS Deutsche Restless Legs Vereinigung), the Swiss RLS Patient Association (Schweizerische Restless Legs Selbsthilfegruppe) and a consortium formed by Boeringer Ingelheim Pharma, Mundipharma Research, Neurobiotec, Roche Pharma, UCB (Germany + Switzerland) and Vifor Pharma. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Researchers interested in using COR data are required to sign and follow the terms of a cooperation agreement that includes a number of clauses designed to ensure protection of privacy and compliance with relevant laws. For further information on DHS and COR, contact K.B. (bergerk@uni-muenster.de). Acknowledgments for data contributed by other consortia that were used for secondary analyses are presented in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

D.P. and E.J.W.V.S conceived the study. A.R.H. and D.P. performed the analyses. T.F.B., K.D., B.H.W.t.L, R.W. and E.J.W.V.S. recruited participants from the NSR and collected and analyzed data for phenotypic validation. C.A.d.L., S. Sniekers, K.W. and E.T. performed secondary analyses. S. Stringer prepared the UK Biobank data for analyses and wrote a pipeline to facilitate efficient data processing. G.T. and I.J. performed the deCODE analyses. K.O. performed the COR and DHS analyses. H.S., T.G., K.B., B.S., J. Wellmann, J. Winkelmann, K.S., K.O. and E.J.W.V.S. contributed data analyzed in this study. A.R.H., K.O., E.J.W.V.S. and D.P. wrote the paper. All authors discussed the results and commented on the paper.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Wittchen, H.U. *et al.* The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679 (2011).
2. Zhang, B. & Wing, Y.-K. Sex differences in insomnia: a meta-analysis. *Sleep* **29**, 85–93 (2006).
3. Morin, C.M. *et al.* Insomnia disorder. *Nat. Rev. Dis. Primers* **1**, 15026 (2015).
4. Baglioni, C. *et al.* Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies. *J. Affect. Disord.* **135**, 10–19 (2011).
5. Palagini, L. *et al.* Sleep loss and hypertension: a systematic review. *Curr. Pharm. Des.* **19**, 2409–2419 (2013).
6. Mallon, L., Broman, J.E. & Hetta, J. Sleep complaints predict coronary artery disease mortality in males: a 12-year follow-up study of a middle-aged Swedish population. *J. Intern. Med.* **251**, 207–216 (2002).
7. Nilsson, P.M., Nilsson, J.Å., Hedblad, B. & Berglund, G. Sleep disturbance in association with elevated pulse rate for prediction of mortality—consequences of mental strain? *J. Intern. Med.* **250**, 521–529 (2001).
8. Schwartz, S.W. *et al.* Are sleep complaints an independent risk factor for myocardial infarction? *Ann. Epidemiol.* **8**, 384–392 (1998).
9. Clark, A., Lange, T., Hallqvist, J., Jennum, P. & Rod, N.H. Sleep impairment and prognosis of acute myocardial infarction: a prospective cohort study. *Sleep* **37**, 851–858 (2014).
10. Cappuccio, F.P., D'Elia, L., Strazzullo, P. & Miller, M.A. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. *Diabetes Care* **33**, 414–420 (2010).
11. Hargens, T.A., Kaleth, A.S., Edwards, E.S. & Butner, K.L. Association between sleep disorders, obesity, and exercise: a review. *Nat. Sci. Sleep* **5**, 27–35 (2013).
12. Lind, M.J., Aggen, S.H., Kirkpatrick, R.M., Kendler, K.S. & Amstadter, A.B. A longitudinal twin study of insomnia symptoms in adults. *Sleep* **38**, 1423–1430 (2015).
13. Bootzin, R.R. & Epstein, D.R. Understanding and treating insomnia. *Annu. Rev. Clin. Psychol.* **7**, 435–458 (2011).
14. Bettolo, A. L'insonnia e sua importanza clinica: le sue cause predisponenti e determinanti, i suoi caratteri nelle diverse malattie, la sua importanza prognostica ed il suo trattamento generale e speciale. *Studium* **21**, 54–65 (1931).
15. Morin, C.M. *et al.* Cognitive behavioral therapy, singly and combined with medication, for persistent insomnia: a randomized controlled trial. *J. Am. Med. Assoc.* **301**, 2005–2015 (2009).
16. Harvey, A.G. & Tang, N.K.Y. Cognitive behaviour therapy for primary insomnia: can we rest yet? *Sleep Med. Rev.* **7**, 237–262 (2003).
17. Morin, C.M. *et al.* The natural history of insomnia: a population-based 3-year longitudinal study. *Arch. Intern. Med.* **169**, 447–453 (2009).
18. Bastien, C.H. & Morin, C.M. Familial incidence of insomnia. *J. Sleep Res.* **9**, 49–54 (2000).
19. Dauvilliers, Y. *et al.* Family studies in insomnia. *J. Psychosom. Res.* **58**, 271–278 (2005).
20. Beaulieu-Bonneau, S., LeBlanc, M., Mérette, C., Dauvilliers, Y. & Morin, C.M. Family history of insomnia in a population-based sample. *Sleep* **30**, 1739–1745 (2007).
21. Wing, Y.K. *et al.* Familial aggregation and heritability of insomnia in a community-based study. *Sleep Med.* **13**, 985–990 (2012).
22. Byrne, E.M. *et al.* A genome-wide association study of sleep habits and insomnia. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **162B**, 439–451 (2013).
23. Amin, N. *et al.* Genetic variants in *RBFOX3* are associated with sleep latency. *Eur. J. Hum. Genet.* **24**, 1488–1495 (2016).
24. 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).
25. Lane, J.M. *et al.* Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nat. Commun.* **7**, 10889 (2016).
26. Jones, S.E. *et al.* Genome-wide association analyses in 128,266 individuals identifies new morningness and sleep duration loci. *PLoS Genet.* **12**, e1006125 (2016).
27. Vgontzas, A.N. *et al.* Persistent insomnia: the role of objective short sleep duration and mental health. *Sleep* **35**, 61–68 (2012).
28. 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).
29. Benjamins, J.S. *et al.* Insomnia heterogeneity: characteristics to consider for data-driven multivariate subtyping. *Sleep Med. Rev.* http://dx.doi.org/10.1016/j.smrv.2016.10.005 (2016).
30. Morphy, H., Dunn, K.M., Lewis, M., Boardman, H.F. & Croft, P.R. Epidemiology of insomnia: a longitudinal study in a UK population. *Sleep* **30**, 274–280 (2007).
31. Paparrigopoulos, T. *et al.* Insomnia and its correlates in a representative sample of the Greek population. *BMC Public Health* **10**, 531 (2010).
32. Cho, Y.W. *et al.* Epidemiology of insomnia in Korean adults: prevalence and associated factors. *J. Clin. Neurol.* **5**, 20–23 (2009).
33. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
34. Loh, P.-R. *et al.* Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet.* **47**, 1385–1392 (2015).
35. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
36. Cai, M. *et al.* Dual actions of Meis1 inhibit erythroid progenitor development and sustain general hematopoietic cell proliferation. *Blood* **120**, 335–346 (2012).
37. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
38. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).

39. Winkelmann, J. *et al.* Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. *Nat. Genet.* **39**, 1000–1006 (2007).
40. Winkelmann, J. *et al.* Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet.* **7**, e1002171 (2011).
41. Schulte, E.C. *et al.* Targeted resequencing and systematic *in vivo* functional testing identifies rare variants in *MEIS1* as significant contributors to restless legs syndrome. *Am. J. Hum. Genet.* **95**, 85–95 (2014).
42. Xiong, L. *et al. MEIS1* intronic risk haplotype associated with restless legs syndrome affects its mRNA and protein expression levels. *Hum. Mol. Genet.* **18**, 1065–1074 (2009).
43. Allen, R.P., Barker, P.B., Horská, A. & Earley, C.J. Thalamic glutamate/glutamine in restless legs syndrome: increased and related to disturbed sleep. *Neurology* **80**, 2028–2034 (2013).
44. Spiegelhalder, K. *et al.* Magnetic resonance spectroscopy in patients with insomnia: a repeated measurement study. *PLoS One* **11**, e0156771 (2016).
45. 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).
46. Ohayon, M.M., O'Hara, R. & Vitiello, M.V. Epidemiology of restless legs syndrome: a synthesis of the literature. *Sleep Med. Rev.* **16**, 283–295 (2012).
47. Yang, J. *et al.* Genome-wide genetic homogeneity between sexes and populations for human height and body mass index. *Hum. Mol. Genet.* **24**, 7445–7449 (2015).
48. Tucci, V. Genomic imprinting: a new epigenetic perspective of sleep regulation. *PLoS Genet.* **12**, e1006004 (2016).
49. Lassi, G. *et al.* Loss of *Gnas* imprinting differentially affects REM/NREM sleep and cognition in mice. *PLoS Genet.* **8**, e1002706 (2012).
50. Dijk, D.-J. Sleep of aging women and men: back to basics. *Sleep* **29**, 12–13 (2006).
51. Ursin, R., Bjorvatn, B. & Holsten, F. Sleep duration, subjective sleep need, and sleep habits of 40- to 45-year-olds in the Hordaland Health Study. *Sleep* **28**, 1260–1269 (2005).
52. Redline, S. *et al.* The effects of age, sex, ethnicity, and sleep-disordered breathing on sleep architecture. *Arch. Intern. Med.* **164**, 406–418 (2004).
53. Buysse, D.J. *et al.* EEG spectral analysis in primary insomnia: NREM period effects and sex differences. *Sleep* **31**, 1673–1682 (2008).
54. Leiserson, M.D.M. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–114 (2015).
55. Bonnet, M.H. & Arand, D.L. 24-hour metabolic rate in insomniacs and matched normal sleepers. *Sleep* **18**, 581–588 (1995).
56. Feige, B. *et al.* The microstructure of sleep in primary insomnia: an overview and extension. *Int. J. Psychophysiol.* **89**, 171–180 (2013).

ONLINE METHODS

UK Biobank sample. We used data provided by the UK Biobank Study[28] (see URLs). UK Biobank is a major national health resource including >500,000 participants, with the aim of improving the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses—including cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders, depression and forms of dementia. All participants provided written informed consent; the UK Biobank received ethical approval from the National Research Ethics Service Committee North West–Haydock (reference 11/NW/0382), and all study procedures were performed in accordance with the World Medical Association Declaration of Helsinki ethical principles for medical research. The current study was conducted under UK Biobank application number 16406.

The study design of the UK Biobank has been described elsewhere[28,57]. Briefly, in 2006–2010, about 9.2 million invitation letters to participate in the study were sent to all people aged 40–69 years who were registered with the National Health Service and living up to ~25 miles from one of the 22 study assessment centers. A total of 503,325 participants were recruited into the study[28]. Apart from registry-based phenotypic information, extensive self-reported baseline data were collected by questionnaire, in addition to anthropometric assessments and DNA collection. For the present study, we focused on insomnia, which was measured as experiencing trouble falling asleep or waking up in the middle of the night (**Supplementary Note**,).

SNP analysis of the UK Biobank sample. We used imputed genetic data from UK Biobank (May 2015 release) including ~73 million genetic variants in 152,249 individuals. Details on the data are provided elsewhere (see URLs). In summary, the first ~50,000 samples were genotyped on the UK BiLEVE custom array, and the remaining ~100,000 samples were genotyped on the UK Biobank Axiom array. After standard quality control of the SNPs and samples, which was centrally performed by UK Biobank, the data set comprised 641,018 autosomal SNPs in 152,256 samples for phasing and imputation. Imputation was performed with a reference panel that included the UK10K haplotype panel and the 1000 Genomes Project Phase 3 reference panel.

For analyses in our study, we selected only individuals of European descent. After removal of related individuals and individuals with discordant sex, who withdrew consent and who had missing phenotype data, 113,006 individuals remained for analysis (**Supplementary Table 1**). This is the largest available GWAS sample for insomnia to date. Previous power analyses showed that a sample size of n = 40,000 individuals allows for high-power (>90%) detection of SNPs with small effect sizes explaining only 0.1% of the variance[58], indicating sufficient statistical power to detect SNPs associated with insomnia in our sample.

Association tests were performed in SNPTEST[59] (see URLs) using logistic regression with the covariates age, sex (for the full sample), genotyping array, the top five genetically determined principal components and additional principal components out of ten further ones that were associated with the phenotype (tested by logistic regression). SNPs with imputation quality of <0.8 (based on the total sample and only those of European ancestry) and a MAF of <0.001 were excluded after the association analysis, resulting in 12,444,916 SNPs, 12,428,592 SNPs and 12,432,937 SNPs for the full, male and female analyses, respectively.

Gene analysis. We used all 19,427 protein-coding genes from the NCBI 37.3 gene definitions as the basis for a GWGAS in MAGMA[35] (see URLs). We annotated all SNPs in our association analysis to these genes, resulting in 18,355 genes that were covered by at least one SNP. We included a window around each gene of 2 kb before the transcription start site and 1 kb after the transcription stop site. Gene association tests were performed taking into account the LD between SNPs. We applied a stringent Bonferroni correction to account for multiple testing.

A GWGAS can identify genes in which multiple genetic variants show mild effects that are not sufficiently strong to be detected by GWAS. On the other hand, although a GWAS analysis can indicate a significant locus encompassing a gene, it is possible for this gene not to be identified by GWGAS because a gene can harbor many more SNPs that do not show an association signal and the GWGAS takes all SNPs within the gene into account.

SNP analysis of the deCODE sample. The Icelandic GWAS data set used in the current study is based on whole-genome sequencing, chip genotyping and long-range phasing of Icelandic population samples[60]. In brief, we performed whole-genome sequencing of 15,220 Icelanders using Illumina technology to an average depth of at least 34×, resulting in the identification of ~94 million variants. Using imputation assisted by long-range haplotype phasing[61] and after removing variants with imputation information content of <0.8 as well as an imputed MAF of <0.01%, we successfully inferred the genotypes of 32,463,443 variants in 434,571 Icelanders, of whom 151,677 had been genotyped using the Illumina chip genotyping platform. The remaining 282,894 Icelanders are first- and second-degree relatives of the chip-typed individuals and were imputed through the aid of genealogical information. Of the 3,774 cases and 3,791 controls used in this study, 3,671 and 3,697 were directly genotyped, respectively.

Logistic regression was used to test for associations between variants and insomnia, assuming a multiplicative model treating disease status as the response and using expected genotype counts from imputation as covariates. For the Icelandic cohort, this was done using software developed at deCODE Genetics[60]. Testing was performed using the likelihood-ratio statistic, and population stratification was adjusted for by including county of birth as a covariate.

To account for inflation in test statistics due to cryptic relatedness and stratification within the case and control sample sets, we applied the method of LD score regression[33]. With a set of 1.1 million variants, we regressed the χ^2 statistics from our GWAS scan against LD score and used the intercept as a correction factor. The estimated correction factors were 1.059, 1.025 and 1.036 for the analyses including all individuals, males only and females only, respectively. All P values were adjusted using these correction factors.

Genotyping and association analyses of the COR and DHS samples. Genotyping was performed on 1,051 DHS participants using the Illumina HumanOmni Chip 2.5-4v1 and the GenomeStudio Genotype module. A total of 1,057 COR study participants were genotyped using the Affymetrix Axiom CEU array and the Axiom GT1 algorithm. Genome-wide imputation of autosomal SNPs according to Phase 3 of the 1000 Genomes Project was performed at the Sanger Imputation Service (see URLs). Quality control before imputation removed variants with a genotyping rate of <95%, a MAF of $<10 \times 10^{-3}$ or strong deviation from Hardy–Weinberg equilibrium ($P < 10 \times 10^{-8}$), and it excluded individuals with sex mismatch or a genotyping rate of <95%, finally selecting a maximum set of unrelated individuals ($\hat{\pi} > 0.16$). Variants were recoded before imputation according to human genome build 37 information on position, strand orientation and major alleles, as reported by genotyping chip strand files (see URLs) and dbSNP (see URLs). Recoded data sets were merged using PLINK software[62] (see URLs) and subjected to multidimensional scaling (MDS) on ten dimensions, which involved outlier detection (>4 s.d. above or below the population mean) and provided covariates besides age and sex for the association analyses. For the latter, the imputed genotypes were merged, yielding data on 1,985 individuals after quality control and outlier removal. For 1,772 of these individuals, the three-level insomnia severity score was available. In the case of individual SNPs, association analysis was applied as an additive model and linear or logistic regression, as implemented in PLINK. For the gene analysis, MAGMA was applied with the same phenotypes and covariates, using the same method in MAGMA as for the GWGAS.

Credible set analysis. For the *MEIS1* locus, we defined a credible set of SNPs that could plausibly be considered as causal using PAINTOR (probabilistic annotation integrator)[63]. PAINTOR uses a multivariate normal approximation to connect the LD structure of the SNPs in a locus to P values. Data on functional annotation are integrated through an empirical Bayes prior, resulting in a prior probability of a variant being causal that is governed by its score in the functional classes. The SNPs scoring high for certain functional annotations are upweighted, while the SNPs scoring low for a certain functional annotation are downweighted. We included the functional annotations reported in the **Supplementary Note** to prioritize the SNPs in the *MEIS1* locus, including deleteriousness (continuous score), regulatory function (continuous score), meQTL (yes/no) and mean chromatin state over different tissues (continuous score).

doi:10.1038/ng.3888

Conditional analyses. We performed two types of conditional analysis for our SNPs associated with insomnia complaints: (i) conditioning of our top SNP rs113851554 in *MEIS1* on three SNPs (rs6710341, rs12469063 and rs2300478) representing the association signals detected in RLS GWAS[39,40] and (ii) conditioning of all SNPs significantly associated with insomnia complaints on the signals for other traits and characteristics related to insomnia that were available in the UK Biobank study: waist-to-hip ratio, body mass index (BMI), Townsend deprivation index, years of education, depressive symptoms and neuroticism. Analyses were performed in SNPTEST[59] using the logistic regression model including covariates as described above. The additive effects of the SNPs in the first analysis were added to the regression model with the -condition on flag. The phenotypes in the second analysis we added to the other covariates with the -cov_names flag.

Genotype × sex interaction analysis. To investigate possible sex-related effects on insomnia complaints, we performed an association analysis adding sex as an interaction term in the original insomnia complaints GWAS using PLINK[62] (--linear interaction). We analyzed the same SNPs included in the main GWAS of this study and used the same covariates (sex, age, array and principal components).

BUHMBOX. We applied Breaking Up Heterogeneous Mixture Based On Cross-locus correlations (BUHMBOX[45]; see URLs) to test whether a heterogeneous subgroup showing genetic characteristics of RLS was present in our UK Biobank insomnia sample, which should otherwise be homogeneous (that is, to address whether the sharing of risk alleles by insomnia and RLS is driven by all individuals or a subset of individuals). We used the top SNPs from the six loci associated with RLS[40] along with their risk alleles and allele frequencies (**Supplementary Table 17**) to define the genetic architecture of RLS. We first ran the BUHMBOX power calculator for 1,000 simulated experiments including the UK Biobank sample size, the risk allele frequencies and odds ratios for the RLS SNPs (**Supplementary Table 17**), and the expected proportion of individuals with RLS in the insomnia complaints group (0.107). BUHMBOX tests whether the RLS risk alleles have higher allele frequencies only in a subset of insomnia cases (if pleiotropy exists, the RLS risk alleles are expected to have higher allele frequencies across the total sample of insomnia cases). If the RLS risk alleles are enriched in one subgroup of insomnia cases, the expected correlations between the numbers of risk alleles at the loci will consistently be positive. The pairwise correlations are combined in one statistic to test for excessive positive correlations.

Genetic risk score analysis. We used the top SNPs from the six loci associated with RLS (the same as for BUHMBOX) as input for the genetic risk score analysis. For each individual, we calculated the genetic risk score by summing the risk allele dosage (0, 1 or 2) multiplied by the effect size (log(OR)) for the six top SNPs. An association analysis was performed between this genetic risk score and the UK Biobank insomnia phenotype using logistic regression including the genetic principal components as covariates.

Sign concordance tests. As input for the sign concordance tests, we used independent SNPs that we defined by pruning the data with PLINK[62] (--indep-pairwise 1000 100 0.1; see URLs). For analysis with the RLS and insomnia complaints data, we first removed all SNPs with the allele combinations A/T and C/G to exclude strand ambiguity. In addition, all SNPs with non-matching alleles were removed. Sign concordance between two data sets was tested by a two-sided binominal test for a probability of 0.5, using SNPs selected for association with insomnia complaints (or with RLS) below six different *P*-value thresholds (1, 0.5, 0.05, 1×10^{-3}, 1×10^{-4} and 1×10^{-5}).

Tests for low *P* value enrichment. The pruned data used as input for the sign concordance tests were used for tests of low *P* value enrichment as well. Enrichment of low *P* values between two data sets was tested with a two-sided Fisher's exact test on the cross-tabs of the SNPs below and over four different *P*-value thresholds (0.05, 1×10^{-3}, 1×10^{-4} and 1×10^{-5}). In addition, because the RLS and insomnia complaints summary statistics were from samples that substantially differed in size (influencing *P* values), we performed the analysis for seven different ranked *P*-value thresholds as well (50, 100, 200, 400, 800, 1,600 and 3,200).

Meta-analysis. Meta-analysis of the SNPs in UK Biobank and deCODE was performed in METAL[64] (see URLs). The analysis was based on *P* values, taking sample size and direction of effect into account. Meta-analysis of the genes in UK Biobank and deCODE was performed in MAGMA[35] (see URLs), which uses Stouffer's weighted *z*-transform method.

HotNet2 analysis. We applied the HotNet2 algorithm[54] to identify networks of genes that are related to insomnia. HotNet2 is based on a heat diffusion model. The key advantage of HotNet2 as compared to conventional methods is the possibility to detect genes in connected subnetworks with associations to the phenotype stronger than expected by chance. Conventional gene enrichment or gene set analyses are limited by the rigid 'in or out' definition of a gene set, which does not allow for crosstalk between pathways that are represented by different gene sets. To depict an entire network topology, conventional enrichment tools therefore need to define a large number of gene sets, resulting in a loss of statistical power due to a high level of multiple testing.

As input for the HotNet2 analysis, we selected all genes from our GWGAS results with $P \leq 0.1$ (2,335, 2,101 and 2,077 genes for the full, female and male analyses, respectively). The $-\log_{10}$ (*P* value) was defined as the input gene score. HotNet2 was performed based on protein–protein interactions reported by iRefIndex[65]. For four δ thresholds (minimum edge weight) that were automatically chosen by HotNet2, the significance of n subnetworks at k (the minimum number of proteins in a subnetwork) was reported based on an influence matrix that was permuted 100 times.

Next, we performed an enrichment analysis of the identified subnetworks by calculating a *P* value for the fraction of genes that overlapped predefined pathways using the hypergeometric test. We selected all canonical pathways ($n = 1{,}330$) and Gene Ontology (GO) pathways ($n = 1{,}454$) from the molecular signature database (MSigDB v5.1 (ref. 66); see URLs). A pathway was considered statistically significant when the hypergeometric test showed $P \leq 0.05$ after correcting for multiple testing using the Benjamini–Hochberg method.

Genetic correlations. Genetic correlations (r_g) were calculated between (i) insomnia complaints and 6 other sleep-related phenotypes present in UK Biobank; (ii) insomnia complaints in males and females; and (iii) insomnia complaints and 29 other traits for which summary statistics from GWAS were publicly available (**Supplementary Table 33**), using LD score regression[33] (see URLs). We used precomputed LD scores that were provided by LD score regression, which were calculated using the European panel of the 1000 Genomes Project. No constraining of the intercept was applied. A conservative Bonferroni-corrected *P*-value threshold of 1.72×10^{-3} was used in the analysis of correlations with GWAS traits to define significant associations.

Phenotypic group differences between individuals with and without insomnia. For 18 disorders, traits and characteristics measured in the NSR[29], group differences between 1,073 individuals without insomnia complaints and 845 individuals likely to have insomnia disorder were evaluated using *t* tests (continuous phenotype) or χ^2 tests (dichotomous phenotypes).

Data availability. Summary statistics from our insomnia GWAS are available for download at http://ctg.cncr.nl/software/summary_statistics. The data generated in the secondary analyses of this study are included with this article in the supplementary tables. The genotype data analyzed during the current study were provided by the UK Biobank Study (see URLs), obtained under UK Biobank application number 16406. The genotype data from deCODE, DHS and COR were obtained through the principal investigators of those studies.

57. Allen, N.E., Sudlow, C., Peakman, T. & Collins, R. UK Biobank data: come and get it. *Sci. Transl. Med.* **6**, 224ed4 (2014).
58. Visscher, P.M. Sizing up human height variation. *Nat. Genet.* **40**, 489–490 (2008).
59. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
60. Gudbjartsson, D.F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat. Genet.* **47**, 435–444 (2015).

61. Kong, A. *et al.* Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat. Genet.* **40**, 1068–1075 (2008).
62. 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).
63. Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet.* **10**, e1004722 (2014).
64. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
65. Razick, S., Magklaras, G. & Donaldson, I.M. iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics* **9**, 405 (2008).
66. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).

doi:10.1038/ng.3888