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

Anke R Hammerschlag1, Sven Stringer1, Christiaan A de Leeuw1, Suzanne Sniekers1, Erdoghan Taskesen1,2, Kyoko Watanabe1, Tessa F Blanken3–5, Kim Dekker3, Bart H W te Lindert3, Rick Wassing3, Ingileif Jonsdottir6,7,8, Gudmar Thorleifsson6, Hreinn Stefansson6, Thorarinn Gislason7,8, Klaus Berger9, Barbara Schormair10,11, Juergen Wellmann9, Juliane Winkelmann10–13, Kari Stefansson6,7, Konrad Oexle10,15, Eus J W Van Someren3–5,15 & Danielle Posthuma1,14,15

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 × 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, with prevalence estimates ranging from 10% (adults) to 22% (the elderly). 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 and contributes to risks of cardiovascular disease, type 2 diabetes and obesity. Heritability estimates of 38% (males) and 59% (females) 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. Indeed, interventions that address these activities are at least partly effective in about two-thirds of cases, but they ameliorate complaints by only about 50% (ref. 16), often resulting in a persistent course for the disorder.

Family and twin studies suggest the involvement of genetic factors in the etiology of insomnia. 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. For example, the habitual duration and timing of sleep do not reliably discriminate between cases of insomnia disorder and controls. 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.

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 pleiotropy24).

Here we report a GWAS using the UK Biobank sample28 (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 Registry29 (NSR; see URLs; Supplementary Note). The equivalent of the UK Biobank question (using our response category cutoff) in the NSR had a sensitivity of 0.96 and a specificity of 0.97 in discriminating insomnia disorder cases from unaffected controls (\( \chi^2 = 1356.45, P < 0.0001 \)) 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 traits24, 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 \( P = 3.65 \times 10^{-13} \); Coorl20, 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 regression33 (LDSC) and 0.11 (s.e.m. = 0.0093) by BOLT-REML34 (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 (\( \lambda_{\text{male}} = 1.11; \lambda_{\text{female}} = 1.06; \lambda_{\text{sex}} = 1.05 \); 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 GWAS

A GWAS, as implemented in MAGMA35 (Online Methods), on all individuals identified three genes associated with insomnia complaints: MEIS1 (also implicated by the GWAS), DCBLD1 and MED27. Sex-specific GWAS identified two additional genes (HHHEX and RHCG) for males and two additional genes (IP07 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 development36. MEIS1 shows the highest expression levels in female internal reproductive organs, but it is also expressed in many other
tissues, including the brain. 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 based threshold of association in the meta-analysis. Six of the seven genes detected in the GWAS were significant at the genome-wide gene-based threshold of P = 5 × 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). The effects of 11 of the 12 SNPs from the full GWAS (both sexes combined) and all 5 SNPs from the female GWAS showed a stronger association signal for insomnia complaints in the meta-analysis. Six of the seven genes detected in the GWAS were significant at the genome-wide gene-based threshold of P = 2.72 × 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 GWAS 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 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 and gene expression 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.

| SNP          | Annotation  | Chr. | Position (bp) | EA   | Non-EA | INFO | EAF | OR    | 95% CI       | P      |
|--------------|-------------|------|---------------|------|--------|------|-----|-------|--------------|--------|
| rs792143917  | MEIS1 intronic | 2    | 66,728,627    | GT   | G      | 0.942| 0.105| 1.09  | 1.05–1.12    | 1.21 × 10⁻⁸ |
| rs214449535  | MEIS1 intronic | 2    | 66,745,864    | A    | G      | 0.964| 0.095| 1.10  | 1.06–1.13    | 1.79 × 10⁻³ |
| rs214449535  | MEIS1 intronic | 2    | 66,747,480    | A    | G      | 0.976| 0.094| 1.10  | 1.06–1.13    | 1.69 × 10⁻⁷ |
| rs11679120   | MEIS1 intronic | 2    | 66,750,564    | T    | G      | 1    | 0.056| 1.19  | 1.14–1.24    | 2.14 × 10⁻¹⁸ |
| rs182588061  | MEIS1 intronic | 2    | 66,757,709    | T    | G      | 0.852| 0.020| 1.21  | 1.13–1.28    | 7.00 × 10⁻¹⁷ |
| rs139775539  | MEIS1 intronic | 2    | 66,782,432    | A    | AC     | 0.961| 0.048| 1.19  | 1.14–1.24    | 2.09 × 10⁻⁴  |
| rs549771308  | MEIS1 intronic | 2    | 66,799,986    | T    | C      | 0.943| 0.048| 1.17  | 1.12–1.22    | 2.88 × 10⁻¹⁰ |
| rs71554396   | Intergenic   | 6    | 169,841,072   | GT   | G      | 0.946| 0.137| 1.95  | 0.93–0.98    | 1.43 × 10⁻³  |
| rs113851554  | Intergenic   | 6    | 169,961,603   | G    | A      | 1    | 0.147| 1.96  | 0.93–0.98    | 1.50 × 10⁻³  |
| rs13192566   | Intergenic   | 6    | 169,961,635   | C    | G      | 0.999| 0.146| 1.95  | 0.93–0.98    | 1.89 × 10⁻¹⁰ |

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 × 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. SNP association results in the deCODE sample and the meta-analysis are reported in Supplementary Table 11 and 13. Chr., chromosome. EA, effect allele frequency. EAF and allele frequency. OR, odds ratio; CI, confidence interval.

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 on our top SNP, rs113851554, on the RLS-associated SNPs (Online Methods) showed that this SNP has an effect on insomnia complaints (OR = 2.90 × 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 BUMHBOX, 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 BUMHBOX 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, we investigated the RLS associations. 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 on our top SNP, rs113851554, on the RLS-associated SNPs (Online Methods) showed that this SNP has an effect on insomnia complaints (OR = 2.90 × 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 BUMHBOX, 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 BUMHBOX 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, we investigated the RLS associations. 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 on our top SNP, rs113851554, on the RLS-associated SNPs (Online Methods) showed that this SNP has an effect on insomnia complaints (OR = 2.90 × 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 BUMHBOX, 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 BUMHBOX 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, we investigated the RLS associations. 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 on our top SNP, rs113851554, on the RLS-associated SNPs (Online Methods) showed that this SNP has an effect on insomnia complaints (OR = 2.90 × 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).

Third, we performed a genetic risk score analysis (Online Methods) to interpret the BUMHBOX 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 BUMHBOX (Supplementary Table 17): (i) five RLS loci excluding the MEIS1 locus (P = 7.28 × 10⁻³); (ii) six RLS loci including the MEIS1 association from the RLS GWAS (P = 6.23 × 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 × 10⁻¹³). As in the BUMHBOX analysis, including the top signal for insomnia complaints strongly increased the association. The results of the BUMHBOX 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.
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, explaining why an effect of MEIS1 on insomnia was not strongly observed in the COR sample. 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 GWAS40 was insufficient to obtain a reliable estimate of genetic correlation). Seventy-eight percent of the independent top SNPs (P = 1 x 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 duration24,26, chronotype25,26 and excessive daytime sleepiness24. 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 dozing or sleeping (r_g = 0.51, P = 3.25 x 10^−4) and napping during the day (r_g = 0.42, P = 3.95 x 10^−6) and a negative genetic correlation with sleep duration (r_g = 0.47, P = 1.97 x 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

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

| Gene    | Entrez ID | Chr. | Start position (bp) | Stop position (bp) | n SNPs | P | P_corrected | Male n SNPs | P | P_corrected | Female n SNPs | P | P_corrected |
|---------|-----------|------|---------------------|--------------------|--------|---|-------------|-------------|---|-------------|---------------|---|-------------|
| DCDL1   | 285761    | 6    | 117,801,803         | 117,892,021        | 384    | 4.54 x 10^−7 | 0.0083 | 385 | 0.0856     | 1              | 386 | 7.12 x 10^−4 | 0.131          |
| MEIS1   | 4211      | 2    | 66,660,257          | 66,800,891         | 426    | 4.60 x 10^−7 | 0.0121 | 424 | 7.56 x 10^−3 | 1              | 428 | 3.58 x 10^−4 | 1              |
| MED27   | 9442      | 9    | 134,734,497         | 134,957,274        | 608    | 7.81 x 10^−7 | 0.0143 | 608 | 9.56 x 10^−3 | 1              | 603 | 5.41 x 10^−4 | 1              |
| HHX5    | 3087      | 10   | 94,447,681          | 94,456,408         | 19     | 2.49 x 10^−3 | 1    | 19 | 1.71 x 10^−6 | 0.031          | 19  | 0.5832      | 1              |
| RHCG    | 51458     | 15   | 90,013,638          | 90,041,799         | 35     | 1.15 x 10^−4 | 1    | 34 | 2.19 x 10^−6 | 0.040          | 35  | 0.7115      | 1              |
| IPT7    | 10527     | 11   | 9,404,169           | 9,470,674          | 241    | 7.80 x 10^−4 | 1    | 234 | 0.7839     | 1              | 245 | 1.67 x 10^−6 | 0.031          |
| TSNARE1 | 203062    | 8    | 143,292,441         | 143,486,543        | 1,083  | 0.029503    | 1    | 1,074 | 0.9498  | 1              | 1,079 | 1.76 x 10^−6 | 0.032          |

Gene-based P values are reported for all genes significant after Bonferroni correction (α = 2.72 x 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 study and the meta-analysis are reported in Supplementary Tables 12 and 14. GWAS, genome-wide gene-association study; chr., chromosome; P_corrected, P value corrected for multiple testing.

*Supplemented on GRC37, including a window of 2.1 kb. †Genome-wide significant in the combined analysis of UK Biobank and deCODE. ‡Stronger association signal in the combined analysis of UK Biobank and deCODE.
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 males4, 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 found47. 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 evolving specific genetic risk factors of insomnia. Whether X-chromosome loci and sex-specific imprinting48,49 play a role still needs to be determined. Our finding is in line with sex differences across most sleep variables50, including subjective sleep complaints51, prevalence2 and heritability12 of insomnia, and physiological signatures of sleep, both in the general population52 and within the population that has insomnia53.

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 states49. This finding is especially interesting given that stronger maternal transmission21, hypermetabolism35 and instability of these sleep states56 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 diseases5–9, obesity11 and psychiatric disorders like depression4. Using whole-genome LD score regression33, 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 (\( r_g = 0.53, P = 1.03 \times 10^{-11} \)), neuroticism (\( r_g = 0.44, P = 1.20 \times 10^{-22} \)) and major depressive disorder (\( r_g = 0.41, P = 6.50 \times 10^{-9} \)). 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
These findings provide starting points for subsequent functional analyses to unravel the molecular neurobiological mechanisms underlying vulnerability to insomnia disorder.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, 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) 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 Genotyping Center of Clinical Genetics and clinical data collection of 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 Leg), and a consortium formed by Boehringer 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.L.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., S. Snekers, 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 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.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.
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. Sleep 29, 12–13 (2006).
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. Buysse, D.J. et al. EEG spectral analysis in primary insomnia: NREM period effects and sex differences. Sleep 31, 1673–1682 (2008).
53. 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).
54. Bonnet, M.H. & Arand, D.L. 24-hour metabolic rate in insomniacs and matched normal sleepers. Sleep 18, 581–588 (1995).
55. Feige, B. et al. The microstructure of sleep in primary insomnia: an overview and extension. Int. J. Psychophysiol. 89, 171–180 (2013).
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, 113,066 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%) detect-
ition of SNPs with small effect sizes explaining only 0.1% of the variance58,
indicating sufficient statistical power to detect SNPs associated with insomnia
in our sample.
Association tests were performed in SNPTEST59 (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 GWAS in MAGMA50 (see URLs). We
annotated all SNPs in our association analysis to these genes, resulting in
18,353 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 GWAS 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 GWAS 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.
Conditional analyses. We performed two types of conditional analysis for our SNPs associated with insomnia complaints: (i) conditioning of our top SNP rs113851554 in MEFS1 on three SNPs (rs6710341, rs12469063 and rs2300478) representing the association signals detected in RLS GWAS39,40 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 SNPTEST29 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 PLINK62 (--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 (BUHMBOX25; 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 RLS40 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 PLINK62 (--indep-pairwise 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 (1.05, 0.05, 1 × 10−1, 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).

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 × 10−1, 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 METAL64 (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 MAGMA30 (see URLs), which uses Stouffer’s weighted z-transform method.

HotNet2 analysis. We applied the HotNet2 algorithm34 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 −log10 (P value) was defined as the input gene score. HotNet2 was performed based on protein–protein interactions reported by iRefIndex45. 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 ≤ 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 regression22 (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 NSR29, 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 χ2 tests (dichotomous phenotypes).

Data availability. Summary statistics from our insomnia GWAS are available for download at http://ctg.cncc.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. Vischer, 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).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
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).