Supplementary Information

**Title: Genome-wide Analysis of Insomnia (N=1,331,010) Identifies Novel Loci and Functional Pathways**

Philip R. Jansen, Kyoko Watanabe, Sven Stringer, Nathan Skene, Julien Bryois, Anke R. Hammerschlag, Christiaan A. de Leeuw, Jeroen Benjamins, Ana B. Muñoz-Manchado, Mats Nagel, Jeanne E. Savage, Henning Tiemeier, Tonya White, The 23andMe Research Team, Joyce Y. Tung, David A. Hinds, Vladimir Vacic, Patrick F. Sullivan, Sophie van der Sluis, Tinca J.C. Polderman, August B. Smit, Jens Hjerling-Leffler, Eus JW Van Someren, Danielle Posthuma

**This PDF file includes:**

1. **Supplementary Methods**
   1. Sample description UK Biobank
   2. Sample description 23andMe
   3. Insomnia phenotype validation external sample
2. **Supplementary Discussion**

2.1. Sex-specific association results for insomnia

2.2. GWAS meta-analysis results for insomnia

2.3. Implicated genes for insomnia

2.4. Gene-set association results for insomnia

2.5. Results sleep-related phenotypes

2.6 Mendelian Randomization

1. **Supplementary Figures (1 to 2)**
2. **Supplementary Tables (1 to 28)**

**1. Supplementary Methods**

**1.1 Sample description UK Biobank**

1.1.1 UK Biobank study

The current study used data from the UK Biobank Study1 (UKB; [www.ukbiobank.ac.uk](http://www.ukbiobank.ac.uk)). The UKB is a large population-based cohort that includes over 500,000 participants and aims to improve insights into a wide variety of health-related determinants and outcomes across the UK. All participants provided written informed consent; the UKB received ethical approval from the National Research Ethics Service Committee North West-Haydock (reference 11/NW/0382), and all study procedures were in accordance with the World Medical Association for medical research. Between 2006 and 2010, approximately 9.2 million invitations to participate in the study were sent to all people aged 40-69 years who were registered with the National Health Service (NHS) and were living within 25 miles from one of the 22 study research centers. In total, 503,325 participants were recruited in the study. Besides phenotypic information from registries, extensive self-reported baseline data were collected by questionnaire, in addition to anthropometric assessments, DNA collection and magnetic resonance imaging of body and brain. We used self-report sleep questionnaire data collected during the first UKB research visit between 2006 and 2010, which covers seven sleep-related items: insomnia, morningness, sleep duration, ease of getting up in the morning, daytime napping, daytime sleepiness, and snoring.

1.1.2. UK Biobank phenotypes

Seven sleep-related measures were collected in the UK Biobank Study via a touchscreen device during the first research visit (2006-2010). All questions had additional answering options of ‘I do not know’ (except daytime napping) and ‘prefer not to answer’. These answers were set to missing in all following analyses.

*Insomnia*

Insomnia complaints were assessed by asking: “Do you have trouble falling asleep at night or do you wake up in the middle of the night?” Participants were instructed to answer this question in relation to the previous four weeks when in doubt. The participants were able to choose one of the following four answers: “never/rarely”, “sometimes”, “usually”, or “prefer not to answer”. Insomnia cases were defined as participants who answered this question with “usually”, while participants answering “never/rarely” or “sometimes” were defined as controls. The insomnia phenotype was available in 386,533 individuals of European descent after quality control. The prevalence of insomnia was 28.3% in this sample (N cases/controls = 109,402/277,131). In addition to the full sample analysis, we analyzed females and males separately (N males/females = 208,716/177,817). In females, the prevalence of insomnia was higher than in males (32.0% vs 23.9% respectively).

*Morningness*

Morningness (i.e. being a morning person rather than an evening person) was assessed by asking: “Do you consider yourself to be?”, followed by the answering categories “Definitely a ‘morning’ person”, “More a ‘morning’ than ‘evening’ person”, “More an ‘evening’ than a ‘morning’ person”, “Definitely an ‘evening’ person”, and “Do not know”, and was analyzed on a continuous scale. The morningness scale was inverted to harmonize the direction of effects with the phenotype definition in the 23andMe sample. After quality control, morningness data were available in 345,552 unrelated individuals of European descent l. Of these, 91,820 (26.6%) identified as definitely being a morning person, 124,151 (35.9%) as more a morning than an evening person, 98,313 (28.5%) as more an evening person than a morning person, and 31,268 (9.0%) as definitely being an evening person.

*Sleep duration*

Sleep duration was assessed by asking: “About how many hours sleep do you get in every 24 hours? (please include naps)". The answer could only contain integer values (round hours). Sleep duration was analyzed as a continuous outcome. Sleep duration was available in 384,317 unrelated individuals of European descent after quality control. The mean sleep duration was 7.10 (SD=1.30) hours per 24 hours.

*Ease of getting up in the morning*

Ease of getting up in the morning was assessed by asking: “On an average day, how easy do you find getting up in the morning?”. The possible answers included “not at all easy”, “not very easy”, “fairly easy” and “very easy”. Data were available 385,949 unrelated participants of European descent after quality control. Ease of getting up in the morning was analyzed as a continuous outcome, with four categories. In total, 14,686 (3.8%) participants answered finding it not at all easy to get up in the morning, 53,314 (13.8%) as not very easy, 192,668 (49.9%) as fairly easy and 125,281 (32.5%) as very easy.

*Daytime napping*

Daytime napping was assessed by asking: “Do you have a nap during the day?”. The three possible answers included categories “never/rarely”, “sometimes” and “usually”. Daytime napping was analyzed as a dichotomous trait (“never/rarely” and “sometimes”, vs. “usually”). Data were available in 386,577 unrelated participants of European descent after quality control. The prevalence of napping in this sample was 5.2%.

*Daytime sleepiness/Dozing*

Daytime sleepiness (dozing) was assessed by asking: “How likely are you to doze off or fall asleep during the daytime when you don't mean to? (e.g. when working, reading or driving)”. The possible answering options were: “never/rarely”, “sometimes”, “often” and “all of the time”. The answers were analyzed as a dichotomous trait (“never/rarely” and “sometimes” vs. “often” and “all the time”). Data were available in 385,333 unrelated participants of European descent after quality control. The prevalence of experiencing daytime sleepiness often or all the time in this sample was 2.6%.

*Snoring*

Snoring was assessed by asking: "Does your partner or a close relative or friend complain about your snoring?". Participants could answer with “yes” or “no”. Snoring data were available in 359,916 unrelated individuals of European descent after quality control. The prevalence of snoring in this sample was 37.3%.

1.1.3 UK Biobank genotype data

We used second-release genotype data that were released by UKB in July 2017. Genotype data collection and processing are described by the UKB in a previous overview paper2. Genotyping was performed on DNA extracted from blood samples. Genotyping was completed in 489,212 individuals on Affymetrix genotyping arrays with custom content: the UK BiLEVE Axiom array (n=50,520) and UKB Axiom array (n=438,692), covering 812,428 genetic markers. Of these, 488,377 individuals and 805,426 markers passed genotype quality controls. These genotypes were subsequently imputed to the Haplotype Reference Consortium3 reference panel (version 1.1), as well as a combined reference panel including the UK10K and 1000 Genomes project reference panels. In total, imputed genotype data was available for 487,422 individuals and 92,693,895 variants. For chromosome X, we used genotyped SNPs only, as this chromosome has not been imputed by UKB. Imputed variants were converted to hard-call genotypes at a certainty threshold of 0.9.

For this study, we selected unrelated individuals from European descent. To determine European descent, we projected 1000 Genomes project genetic principal components on the UK Biobank genotypes and assigned ancestry based on the closest Mahalanobis distance from the 1000 Genomes project population average. As recommended by UKB, we removed variants that were imputed from the UK10K reference panel due to technical errors in the imputation process.

Genome-wide association analysis (GWAS) on each of the seven sleep phenotypes in UKB was done in PLINK4, using logistic regression for dichotomous phenotypes (insomnia, napping, dozing and snoring), and linear regression for phenotypes analyzed as continuous outcomes (sleep duration, morningness, and ease of getting up). All association tests were adjusted for age, sex, genotype array, and 10 genetic principal components that were calculated on the subset of unrelated subjects of European descent, based on 145,432 independent SNPs (*r2*<0.1), MAF >0.01, INFO=1) using FlashPCA5. SNPs with a low imputation score (INFO score < 0.9) and low allele frequency (MAF<0.0001) were excluded, resulting in 10,848,137 SNPs. We used the genome-wide threshold for significance of *P*<5×10-8).

**1.2 Sample description 23andMe**

1.2.1. 23andMe study

Genome-wide association analysis results were obtained from the 23andMe, Inc. research participant database ([www.23andme.com](http://www.23andme.com))6,7, a personal genetics company that collects phenotypic data about a wide variety of common traits. Sleep outcomes were collected by 23andMe research participants through online surveys. Ethical and Independent Review Services, an external AAHRPP-accredited institutional review board, approved the study protocol, and participants provided informed consent. The phenotypes ‘insomnia’ and ‘morningness’ were used in the current study. Morningness was previously published as Hu et al. 20168.

1.2.2 23andMe phenotypes

GWAS summary statistics were obtained from 23andMe, Inc. for two sleep-related phenotypes: insomnia (total sample and males/females analyzed separately) and morningness.

*Insomnia*

Participants completed one or more questions concerning sleep. Insomnia cases affirmed at least one of the following questions:

"Have you ever been diagnosed with, or treated for: Insomnia?"; “Have you ever been diagnosed with, or treated for, any of the following conditions: Insomnia but not Narcolepsy, Sleep apnea or Restless leg syndrome”; "Has a doctor ever told you that you have any of these conditions: Insomnia (difficulty getting to sleep or staying asleep)?"; "Have you ever been diagnosed by a doctor with any of the following neurological conditions: Sleep disturbance"; "Do you routinely have trouble getting to sleep at night?"; "What sleep disorders have you been diagnosed with? Please select all that apply: Insomnia, trouble falling or staying asleep"; "Have you ever taken these medications? Prescription sleep aids"; "In the last 2 years, have you taken any of these medications? Prescription sleep aids".

Participants were classified as controls if they did not provide a positive or uncertain ("I don't know": "I am not sure") to any of the questions listed above, nor to any of the following questions: "Have you ever been diagnosed with, or treated for Insomnia, Narcolepsy, Sleep apnea, Restless leg syndrome?", "Have you ever been diagnosed with or treated for any of the following conditions? Post-traumatic stress disorder (PTSD); Autism; Asperger's; Sleep disorder", "Have you ever been diagnosed with or treated for a sleep disorder?".

Genome-wide analysis results in N=944,477 unrelated individuals of European descent and 15,648,586 SNPs were made available by 23andMe researchers. The prevalence of insomnia in this sample was 30.5% and was higher in females (178,919/501,270=35.7%) compared to males (109,638/443,207=24.7%). In addition to full-sample GWAS summary statistics, we obtained separate GWAS results in males (N=443,207) and females (N=501,270), allowing analysis of sex-specific effects by meta-analyzing with sex-specific GWAS results in UK Biobank.

*Morningness*

Morning type personality was assessed by a question that was repeated once: “Are you naturally a night person or a morning person?”. The first time, three possible answering options included: “night owl”, “early bird” and “neither”. The second time, five possible answers were given: “night person”, “morning person”, “neither”, “it depends”, and “I am not sure”. Morning type cases were defined as answering "Early bird" or "Morning person" at least once out of two times. while controls answered "Night owl" or "Night person". Individuals that consistently answered to be a morning person were defined as cases. They were contrasted with evening type participants who answered “Night owl” or “Night person” at least once out of two times. Individuals with discordant responses were removed from subsequent analysis. Genome-wide association results on 89,283 unrelated individuals and 8,158,278 SNPs were made available by 23andMe analysts for the purpose of the current project 8. Of these, 38,938 individuals were considered morning types (43.6%). This prevalence is higher than the general population, as individuals that were neither morning or evening type were not included in the analysis.

1.2.3 23andMe genotype data

Genotypes were derived from saliva samples by the Laboratory Corporation of America. Genotyping was performed on either two variants of Illumina HumanHap550+ BeadChip, the Illumina OmniExpress+ BeadChip or a custom Illumina array that aimed to improve coverage of lower-frequency coding variation. These genotyping platforms contained 586,916, 584,942, 960,838, and 620,444 SNPs, respectively. Individual samples that did not reach a 98.5% call rate were reanalyzed. Research participants whose samples repeatedly did not reach the call rate threshold were contacted again by 23andMe to provide an additional saliva sample. Genotypes (including chromosome X) were phased with Finch, 23andMe’s in-house variation on Beagle9 and subsequently imputed against the 1000 Genomes Project Phase 1 all-ancestry reference haplotypes (September 2013 release)10. SNPs with Hardy-Weinberg equilibrium *P-*value <1×10-20, call rate <95%, or large allele frequency discrepancies in comparison to European 1000 Genomes Project reference data were excluded. Imputation was run using Minimac211 on chunked chromosome segments of no more than 10,000 genotypes SNPs, with 200 SNPs overlap between segments. Association testing for each SNP was done using logistic regression, while correcting for age, sex and the top five principal components. GWAS was run by 23andMe, and the resulting summary statistics were provided for meta-analysis. Variants were additionally filtered based on allele frequency (minor allele count (MAC) >100).

**1.3 Insomnia phenotype validation external sample**

We previously showed an excellent sensitivity (0.98) and specificity (0.96) of the UK Biobank insomnia phenotype to differentiate cases with diagnosed Insomnia Disorder from controls among participants of the Netherlands Sleep Registry 12. Using this same independent sample (n=1,918, 72% female, m=50 (SD=15) years of age, n=845 cases and n=1,073 controls), the 23andMe insomnia phenotype had a lower but still acceptable sensitivity (0.84) and specificity (0.80).

**2. Supplementary Discussion  
2.1 Sex-specific association results for insomnia**

Epidemiological studies have clearly demonstrated population differences in risk of insomnia between males and females, as well as a genetic correlation significantly different from 1 suggesting sex-specific effects12. To investigate sex-specific genetic effects, we conducted sex-specific meta-analyses of insomnia in males (N=621,024) and females (N=709,986) (**Extended Data Fig. 2, 3**). SNP heritability from LD score regression was slightly higher in females than in males (*h2SNP*=0.08, SE=0.003 vs *h2SNP*=0.07 SE=0.003, respectively) and this sex difference in SNP heritability was consistent across the two samples (**Extended Data Table 1**). The genetic correlation between males and females was *rg*=0.86 (SE=0.05) in the UKB sample and *rg*=0.93 (SE=0.02) in the 23andMe sample.

The observed genetic correlation of combined results in males and females was high (*rg*=0.92, SE=0.02)**,** indicating strong overlap of genetic effects. However, it was also significantly different from 1 (one-sided Wald test, *P*=2.54×10-6) suggesting a small role for sex-specific genetic risk factors, confirming previous reports12.

The sex-specific GWAS resulted in 87 lead SNPs in 83 risk loci for females and 24 lead SNPs in 23 risk loci for males. Of these, ten risk loci were genome-wide significant in both males and females (**Supplementary Table 1**). GWGAS identified 239 GWS genes for females and 60 for males, while the three gene-mapping strategies in FUMA led to 320 genes for females and 41 for males (**Supplementary Table 2**).

**2.2 GWAS meta-analysis results for insomnia**

*2.2.1 GWAS results for insomnia per cohort*

For the two cohorts included in the meta-analysis (UKB and 23andMe), an individual GWAS was run after cohort-specific quality control of the data was performed. Individual GWAS results were inspected for inflation possibly due to insufficiently corrected population stratification, by calculating the lambda inflation factor and the LD score intercept using LD score regression13. A lambda > 1 suggests an inflation of the genetic effects which can be due to both spurious and genuine effects. It is likely to increase with sample size and degree of polygenicity of the trait14, as the distribution of effect sizes begins to differ substantially from a null distribution when more variants have true associations. An LD Score intercept > 1 suggests that there is spurious association, and an intercept < 1.10 is generally considered to suggest that the signal is mostly due to genuine association effects. For the UKB cohort the lambda inflation factor was 1.307, and the LD Score intercept was 1.014 (SE=0.008), while for the 23andMe cohort these were 1.699 and 1.066 (SE=0.011) respectively. The genetic correlation between the two samples was 0.69 (SE=0.02). For the combined meta-analysis, the lambda inflation factor was 1.808, the LD Score intercept 1.075 (SE=0.011), and the ratio between LD Score intercept and mean χ2 was 0.06. These results suggest the genetic signal is mostly due to polygenicity and unlikely to be driven by population stratification. Similar results were observed for the sex-specific meta-analyses of the UKB and 23andMe cohorts (**Extended Data Table 1**).

*2.2.2 Genomic risk loci for insomnia*

The meta-analysis resulted in 250 independent lead SNPs (*r2* <0.1) in 204 independent genomic risk loci (loci < 250 KB apart were merged into one) (**Supplementary Table 3**). The borders of the genomic risk loci were defined by taking all independent GWS SNPs at *r2*<0.6 (N=536) and then identifying all SNPs that were in LD with one of the independent GWS SNPs (N=22,068), and were only based on SNPs that were available in both 23andMe and the UKB samples. Annotated SNPs are all SNPs that are located in a risk locus and that are in LD with one of the independent GWS SNPs (Supplementary **Table 8**).

Regional association plots were visually inspected for each GWS locus (**Supplementary Fig. 1**). We observed 2 loci (#s 13 and 91) for which the patterns of association appeared suspicious. Specifically, these regions had only one (locus 13) or 3 (locus 91) SNPs whose association was GWS, in contrast to a pattern of broad enrichment for SNPs in a local region that is seen in the majority of credibly associated loci due to the correlation of test statistics between SNPs in LD. We investigated these loci further and did not find obvious signs that would indicate genotyping errors or other statistical artefacts, such as poor imputation quality, inconsistent association statistics between the two cohorts, major allele frequency differences from a reference panel, or patterns of regional LD that differed from a known reference panel. The lead SNPs in these loci had low MAF (0.004 for locus 13 and 0.01 for locus 91), which may explain the absence of LD proxies. These 2 loci appeared to be regions for which there were simply few other SNPs in strong LD with the lead SNP, and so there is a lack of information available to infer whether these represent credible association signals. None of the SNPs in these regions have been previously reported to have an association with insomnia or other sleep-related phenotypes. Compared to the other risk loci, we have less confidence that these loci represent true genetic associations until they are replicated in future research, and so we report the total number of genomic loci for insomnia as 202 (204**−**2) and the number of lead SNPs as 248 (250**−**2). We do, however, include these two tentative loci in all tables for reference and future replication, with a notation of any results that stem from their inclusion. All 248 lead SNPs showed concordant directions of effect in UK Biobank and 23andMe (**Extended Data Fig. 4b)**.

There were no genes mapped via positional, eQTL or chromatin interaction mapping based on SNPs in these two loci, and there were no genes GWS in the MAGMA gene-based analysis that were located in these two loci.

The 202 identified genomic risk loci were spread throughout the genome, with one or more risk loci found on every autosome, and 2 risk loci on the X-chromosome. Of these, 200 were novel, which we defined as loci for which no SNPs within the positional bounds of the genomic risk locus had been previously linked to “insomnia complaints”, or “insomnia” in the NHGRI-EBI catalog or in our manual lookup of recently published GWAS for these phenotypes (**Supplementary Table 5**).

*2.2.3 SNPs only available in 23andMe*

Identification of lead SNPs and genomic risk loci was based only on SNPs that were available in both the UKB and 23andMe sample (N=8,806,247 SNPs). An additional 3,195,826 SNPs were available in 23andMe but not in UKB and 2,056,212 SNPs available in UKB and not in 23andMe. The reason that several of these SNPs were not available in both samples was three-fold: 1. the current UKB v2 release included badly imputed UK10K SNPs - we deleted these from the analysis in UKB following UKB recommendations; 2. the current UKB v2 release does not contain X-chromosome imputed SNPs, while 23andMe does; 3. a small proportion of SNPs may not have passed our QC in UKB while they did pass in 23andMe or vice versa.

There were 1,280 SNPs that were only available in 23andMe and that reached genome-wide significance. Of these, 1096 were inside risk loci, and were thus supported by surrounding SNPs that were available in both datasets and that reached GWS. This was the case for all 23andMe-imputed SNPs on chromosome X. There were also 184 SNPs available only in the 23andme sample and that were GWS but were outside of these risk loci and were not supported by surrounding SNPs available in both samples. These 184 GWS SNPs were thus excluded from the definition of lead SNPs and genomic risk loci, as well as from functional annotation and all follow-up analyses (see **Supplementary** T**able 4** for more details on these 184 SNPs).

*2.2.4 Polygenic score validation*

To evaluate the replicability and predictive utility of our SNP-based results, we estimated the variation in insomnia that could be explained by the current GWAS meta-analysis by calculating a polygenic score (PGS) in three hold-out samples (N=3,000) randomly drawn from the UK Biobank study. The three samples were excluded from the meta-analysis, which was then re-run so the results could be used to calculate PGS in each independent hold-out sample. PGS were calculated using LDpred (modelling LD)15 and PRSice16 (*P*-value thresholding + clumping).The PGS explained at most 2.6% of the variance in each sample (*P*=1.20×10-13) when filtering the included SNPs by meta-analysis *P*-value thresholds of *P*<0.01 (**Supplementary Table 6**). Individuals in the highest decile of genetic risk had an odds ratio of approximately 3.2 compared to the lowest risk group (**Extended Data Fig. 5**).

*2.2.5 Description and functional annotation of genomic risk loci*

FUMA annotated all 22,068 SNPs that were in LD (*r2*>=0.6) with one of the independent significant SNPs and allowed further inspection of specific coding variants. Functional annotations included ANNOVAR17, which identifies the SNP’s genic position (e.g. intron, exon, intergenic), Combined Annotation Dependent Depletion (CADD) score18, which predicts how deleterious the effect of a SNP is on protein structure/function, RegulomeDB (RDB) scores19, which predict likelihood of regulatory functionality, and chromatin states from the Roadmap ChromHMM model20, which predict transcription/regulatory effects from chromatin states at the SNP locus.

The most likely genetic variants to have a substantial impact on a phenotype are those that have a direct consequence on a protein, specifically variants located in coding exons that have a nonsynonymous change that alters the protein’s amino acid sequence (ExNS SNPs). Of the annotated 22,068 SNPs in GWS loci, we identified 71 ExNS located in 58 unique genes (**Supplementary Table 9**). Eleven genes included 2 ExNS: *ADO*, *ASCC3, ASXL3, CCDC71, DPP3*, *FAM120AOS, FES, LAMB2, MICB, MST1R, SYCE1L*. One gene (*WDR90*) included four ExNS. Twenty-eight ExNS had CADD scores above 12.37 (the threshold suggested to be deleterious18), and eight had a RegulomeDB (RDB) score < 2, suggesting they were likely affecting binding sites.

From the full GWAS results, the strongest signal was on chromosome 2 (risk locus 20), with the most strongly associated lead SNP rs113851554 (*P*=1.57×10-51). This was also the most strongly associated SNP in previous reports12,21. This SNP is located in an intron of *MEIS1*. There were eight other independent lead SNPs in the same locus, indicative of multiple strong signals co-localized in this region of the genome (**Supplementary Fig. 1** and **Supplementary Table 3**). There were no exonic SNPs in LD with the lead SNPs in this locus, but there were 932 SNPs with CADD scores above 12.37, suggesting that these SNPs are likely to have deleterious effects18.

The second most associated locus (locus 155, **Supplementary Table 3**) on chromosome 13 included 8 lead SNPs. The top lead SNP was rs9527083 (*P*=1.61×10-32), which is intergenic. Of the other 7 lead SNPs, rs8181889 (*P*=8.90×10-10) is perhaps the most interesting as it is located in an intronic region of *OLFM4*, and has a CADD score of 15.36, indicating it is likely to have a deleterious effect (**Supplementary Table 10**).

**2.3 Implicated genes for insomnia**

*2.3.1 Functional mapping of genes in FUMA*

We used the three gene-mapping strategies implemented in FUMA22 to select genes of interest based on GWS SNPs in the risk loci. For positional mapping, we mapped SNPs in the risk loci and in LD with the independent GWS SNPs to genes using a window of 10 Kb, resulting in 394 mapped genes. eQTL mapping implicated 596 genes, of which 320 were located outside of the genic boundaries of the genomic risk loci, and chromatin interaction mapping implicated 133 genes, of which 30 were located outside of the genomic risk loci. Of the resulting set of 651 unique genes mapped by FUMA, 152 had a probability of loss-of-function intolerance (pLI) score >0.90, indicating that these were extremely sensitive to mutations within the gene that result in truncation or loss of function of the protein product (**Supplementary Table 11**)**.**

*2.3.2 Gene-based associations in MAGMA*

We tested 18,185 protein-coding genes for association with insomnia using MAGMA’s gene-based test, which detected 517 (513 novel) genes significantly associated at the Bonferroni corrected (*P*<2.75×10-6) threshold. 222 genes were located outside the SNP-based genomic risk loci (**Supplementary Table 12**) The top gene *BTBD9* (*P*=8.51×10-23) on chromosome 6 in locus 81 was also mapped by positional and eQTL mapping in the left ventricle of the heart (**Supplementary Table 11**). This locus encodes a BTB/POZ domain-containing protein, which is known to be involved in protein-protein interactions. Alternatively spliced transcript variants have been described (description from [http://www.genecards.org](http://www.genecards.org/), provided by RefSeq, Feb 2016). This gene was positionally mapped by 160 SNPs that were either GWS or in LD (*r2*>0.6) with one of the independent significant SNPs, and interestingly the top lead SNP is located right next to the *BTBD9-AS1*, implicating antisense functionality. See **Supplementary Table 12** for a full list of all GWS significant genes.

**2.4 Gene-set association results for insomnia**

We conducted gene-set analyses on a total of 7,473 gene-sets: 7,246 sets derived from the MsigDB (*55*), 54 tissue specific gene-sets and 173 cell specific gene-sets. The threshold for significance of gene-sets was thus set at 0.05/7,473=6.7×10-6,and competitive testing was used throughout. Gene-set analysis using gene-sets from MsigDB version 6.0 in MAGMA resulted in 3 associated gene-sets: *GO:locomotory behavior (P*=8.95×10-7), *GO:behavior (P*=5.23×10-6)*, GO:axon part (P*=4.25×10-6)(**Supplementary Table 15**). Many genes that were included in each of these sets and thus conditional gene-set analyses were carried out to test whether the observed associations were independent. Conditioning on *GO:axon part* led to a slight decrease of the significance of *GO:behavior (*conditioned *P*=3.88×10-5)and *GO:locomotory behavior* (conditioned *P*=8.57×10-6).Conditioning on GO:*behavior* led to a slight decrease of the significance of *GO:axon part (*conditioned *P*=3.16×10-5)and a stronger decrease in *GO:locomotory behavior* (conditioned *P*=3.35×10-3). Conditioning on *GO:locomotory behavior* led to a slight decrease of the significance of *GO:axon (*conditioned *P*=4.13×10-5)and a strong decrease in *GO:behavior* (conditioned *P*=2.26×10-2). We thus conclude that *GO:axon part* and *GO:locomotory behavior* are two independently associated sets and that the effect of *GO:behavior* is explained by *GO:locomotory behavior.*

**2.5 Results sleep-related traits**

We conducted additional GWAS analyses for six sleep-related traits. Genetic and phenotypic correlations with insomnia are in **Supplementary Table 18**, and below we describe the main findings for each trait separately.

*2.5.1 Morningness*

The genome-wide meta-analysis on morningness included 434,835 subjects and 11,597,492 SNPs. The genetic correlation between the two samples included in the meta-analysis (UKB, N=345,552 and 23andMe, N=89,283) was estimated at 0.92 (SE=0.02). The individual GWAS’s showed some inflation in genetic signal (λ=1.603 for UKB and λ=1.253 for 23andMe) and mean χ2 statistic (1.815 and 1.302 respectively) (**Fig. 3a**). The LD Score regression (LDSC) intercept was 1.046; (SE=0.011) for UKB and 1.007 (SE=0.008) for 23andMe. The two cohorts were meta-analyzed in METAL. The quantile-quantile (Q-Q) plot of the meta-analyzed results also showed moderate inflation in λ (1.749) and mean χ2 statistic (2.073). The LDSC intercept (1.054; SE=0.012) was consistent with inflation due to true polygenicity and large sample size. The LDSC SNP-based heritability (*h2*SNP) of morningness was 0.186 (SE=0.006).

The morningness GWAS analysis identified 16,805 GWS SNPs (*P*<5×10-8), represented by 274 independent lead SNPs, which were mapped to 207 independent genomic loci (**Supplementary Table 17**). Of these, 24 were in common with risk loci for insomnia (morningness locus id - insomnia locus id: 12-6, 17-9, 31-19, 33-20, 43-31, 45-36, 49-42, 70-63, 71-67, 77-75, 95-98, 96-102, 102-110, 109-116, 118-128, 128-136, 130-139, 163-176, 165-177, 173-183, 177-185, 186-189, 187-190, 196-195).

Positional, eQTL and chromatin interaction mapping in FUMA connected SNPs in LD with the lead SNPs in these loci to 734 genes (**Supplementary Table 19**). Gene-based analyses using MAGMA led to 492 GWS significant genes (**Extended Data Fig. 9a**), of which 72 were also GWS for insomnia, and 297 were also indicated by FUMA-mapping (**Supplementary Table 24**)**.** The total number of uniquely implicated genes by any of the four gene-mapping strategies is 928**.** Gene-set analyses resulted in 7 gene-sets (*GO:negative regulation of nitrogen compound metabolic process*, *GO:negative regulation of gene expression*, *KEGG: circadian rhythm mammal*, *GO:mRNA ‘3-UTR binding*, *GO:positive regulation of dendrite development*, *GO:RNA destabilization*, *GO:negative regulation of transcription from RNA polymerase II promotor*) from the MSigDB that survived multiple testing, none of these overlapped with sets found for insomnia (**Supplementary Table 25**). Tissue-specific gene-set analyses implicated 6 brain tissues (cortical and basal ganglia structures) and single cell gene-set analyses implicated MSNs and pyramidal CA1 neurons. In level 2 single cell analyses, morningness was associated with a large number of cell types, with the strongest hit being hypothalamic Vglut2 A930013F10Rik Pou2f2 neurons (*P*=3.85×10-10). (**Supplementary Table 25**).

*2.5.2 Sleep duration*

The genome-wide analysis on sleep duration included 384,317 subjects and 10,862,567 SNPs. All subjects were derived from the UKB sample. The QQ-plot of the genome-wide analysis showed some inflation (λ=1.41) and mean χ2 statistic (1.537) (**Fig. 3c**). The LDSC intercept (1.012; SE=0.008) was consistent with inflation due to true polygenicity and large sample size. The LDSC SNP-based heritability (*h2*SNP) of sleep duration was 0.070 (SE=0.003).

The sleep duration GWAS analysis identified 3,886 GWS SNPs (P<5×10-8) (**Supplementary Table 17**), represented by 53 independent lead SNPs, which were mapped to 49 independent genomic loci, of which 14 overlapped with risk loci for insomnia (sleep duration locus id - insomnia locus id: 3-2, 4-3, 6-18, 7-25, 17-54, 19-56, 21-66, 24-75, 28-92, 29-99, 30-102, 43-173, 45-177, 49-189).

One locus stood out (locus 7 on chromosome 2, **Fig. 3b**) as the top lead SNP (rs62158206) and had a *P*-value of 3.00×10-43 (and several surrounding SNPs with similarly low *P*-values) which was much lower that the second most associated lead SNP (locus 6, *P*=4.75×10-14). Implicated genes from locus 7 were *PSD4, CBWD2, IL1RN, FOCD4L1*, and *PAX8*. *PAX8* and *PSD4* were positionally mapped, all of them were implicated by eQTL associations in multiple tissues, and *PAX8* was additionally implicated by chromatin interactions in spleen. This suggests that the strong association of locus 7 with sleep duration is likely due to regulatory effects of the GWS SNPs on these genes.

Linking of all SNPs in LD with the independent significant SNPs to genes, using positional, eQTL and chromatin interaction mapping in FUMA, led to 170 implicated genes (**Supplementary Table 20**). Gene-based analyses using MAGMA led to 135 GWS significant genes (**Extended Data Fig. 9b**) (37 overlapped with insomnia), of which 56 were also indicated by FUMA-mapping (**Supplementary Table 24**)**.**

Gene-set analyses resulted in 1 gene-sets (*REACTOME: neuronal system)* from the MSigDB that survived multiple testing, which did not overlap with sets found for insomnia (**Supplementary Table 25**). Tissue-specific gene-set analyses implicated 10 brain tissues, in single cell gene-set analyses sleep duration was associated with the D1R-type medium spiny neurons (*P*=3.39×10-6; **Supplementary Table 25**).

*2.5.3. Ease of getting up*

The genome-wide analysis for ease of getting up included 385,949 subjects and 10,862,568

SNPs. All subjects were derived from the UKB sample. The Q-Q plot of the genome-wide analysis showed some inflation (λ=1.446) and mean χ2 statistic (1.586) (**Fig. 3c**). The LDSC intercept (1.041; SE=0.010) was consistent with inflation due to true polygenicity and large sample size. The LDSC SNP-based heritability (*h2*SNP) of ease of getting up was 0.071 (SE=0.003).

The ease of getting up GWAS analysis identified 7,248 GWS SNPs (*P*<5×10-8) (**Supplementary Table 17**), represented by 70 independent lead SNPs, which were mapped to 62 independent genomic loci. Of these identified loci, 10 were in common with risk loci for insomnia (locus id ease of getting up - locus id insomnia: 6-9, 12-18, 18-36, 24-63, 28-89, 33-128, 35-137, 55-183, 56-184, 59-195).

Positional, eQTL and chromatin interaction mapping in FUMA connected SNPs in LD with the independent GWS SNPs to 250 genes (**Supplementary Table 21**). Gene-based analyses using MAGMA led to 171 GWS significant genes (**Extended Data Fig. 9c**) (30 overlapped with insomnia), of which 104 were also indicated by FUMA-mapping (**Supplementary Table 24**)**.**

Gene-set analyses resulted in 2 gene-sets (*GO:vocalization behavior, GO:negative regulation of circadian rhythm)* from the MSigDB that survived multiple testing, which did not overlap with sets found for insomnia (**Supplementary Table 25**). Tissue-specific gene-set analyses and single cell gene-set analyses did not result in enrichment in any specific tissue or cell type (**Supplementary Table 25**).

*2.5.4. Daytime Napping*

The genome-wide analysis on daytime napping included 386,577 subjects and 10,858,887 SNPs. All subjects were derived from the UKB sample. The QQ-plot of the genome-wide analysis showed some inflation (λ=1.159) and mean χ2 statistic (1.178) (**Fig. 3d**). The LD score regression (LDSC) intercept (0.995; SE=0.007) was consistent with inflation due to true polygenicity and large sample size. The LDSC SNP-based heritability (*h2*SNP) of napping was 0.105 (SE=0.008).

The daytime napping GWAS analysis identified 2,339 GWS SNPs (*P*<5×10-8) (**Supplementary Table 17**), represented by 7 independent lead SNPs, which were mapped to 7 independent genomic loci. Of these, 2 overlapped with risk loci for insomnia (risk locus id napping - risk locus id insomnia: 1-11, 4-121). The strongest association (locus 6) was on chromosome 17 (lead SNP rs117124984, *P*=2.76×10-13), in an inversion which has previously been associated with neuroticism 23,24.

Positional, eQTL and chromatin interaction mapping in FUMA connected SNPs in LD with the independent GWS SNPs to 29 genes, of which 19 were located in the inversion on chromosome 17 (**Supplementary Table 22**). Gene-based analyses using MAGMA led to 21 GWS significant genes (**Extended Data Fig. 9d**)(no overlap with GWS genes for insomnia), of which 15 were also indicated by FUMA-mapping unique genes (**Supplementary Tables 24**)**.**

There were no gene-sets from the MSigDB, tissue types or the single cell analysis that survived multiple testing. (**Supplementary Table 25**).

*2.5.5. Daytime Sleepiness/Dozing*

The genome-wide analysis on *dozing* included 385,333 subjects and 10,820,725 SNPs. All subjects were derived from the UKB sample. The Q-Q plot of the genome-wide analysis showed some inflation (λ=1.105) and mean χ2 statistic (1.107) (**Extended Data Fig. 3e**). The LDSC intercept (1.007; SE=0.007) was consistent with inflation due to true polygenicity and large sample size. The LDSC SNP-based heritability (*h2*SNP) of dozing was 0.091 (SE=0.010).

The dozing GWAS analysis identified 9 GWS SNPs (*P*<5×10-8) (**Supplementary Table 17**), represented by a single independent lead SNP (rs28600082, *P*=1.08×10-8) on chromosome 4, with surrounding SNPs located in non-coding RNA regions. The single risk locus for dozing was not shared with insomnia.

Positional, eQTL and chromatin interaction mapping in FUMA did not link the GWS SNPs to any gene, and gene-based analyses using MAGMA also did not yield any GWS significant genes (**Extended Data Fig. 9e**)**,** and there were no gene-sets from the MSigDB, tissue types or the single cell analysis that survived multiple testing (**Supplementary Table 24**).

Although there were relatively few GWS findings for dozing and therefore little overlap in GWS findings with insomnia, dozing did show the second most strongest *rg* with insomnia (0.29, SE=0.04) (**Supplementary Table 18**). The *rg* of 0.29 was thus likely due to a set of SNPs that did not reach GWS in the GWAS for dozing, yet whose effects were moderately correlated with the effects of those SNPs on insomnia.

*2.5.6. Snoring*

The genome-wide analysis on snoring included 359,916 subjects and 10,862,568 SNPs. All subjects were derived from the UKB sample. The QQ-plot of the genome-wide analysis showed some inflation (λ=1.358) and mean χ2 statistic (1.443) (**Fig. 3f**). The LDSC intercept (1.010; SE=0.009) was consistent with inflation due to true polygenicity and large sample size. The LDSC SNP-based heritability (*h2*SNP) of snoringwas 0.101 (SE=0.004).

The snoring GWAS analysis identified 3,416 GWS SNPs (*P*<5×10-8) (**Supplementary Table 17**), represented by 41 independent lead SNPs, which were mapped to 35 independent genomic loci. Only one locus was in common with risk loci for insomnia (locus id snoring - locus id insomnia: 2-6).

Positional, eQTL and chromatin interaction mapping in FUMA connected SNPs in LD with the independent GWS SNPs to 99 genes (**Supplementary Table 23**). Gene-based analyses using MAGMA led to 93 GWS significant genes (**Extended Data Fig. 9f**)(9 overlapped with GWS genes for insomnia), of which 42 were also indicated by FUMA-mapping unique genes (**Supplementary Table 24**)**.**

There was one gene-set (*GO:negative chemotaxis*) from the MSigDB that survived multiple testing, which did not overlap with gene-sets for insomnia. There were no tissue-specific gene-set or gene-sets from the single cell analysis that survived multiple testing (**Supplementary Table 25**).

**2.6 Mendelian Randomization**

To test whether directional causal effects were underlying the observed genetic correlations between other trait, we performed Generalized Sumstats-based Mendelian Randomization25 between insomnia and previous GWAS studies (GSMR). Reverse direction of causation analyses (direct effects of other traits on insomnia) were hampered by a lack of enough significant SNPs for subjective well-being, age of having first child, longevity, former smoker, head circumference, intracranial volume, number of children, ever smoker, asthma, cigarettes, anxiety (factor score), ADHD, major depression, anxiety (case-control) and depressive symptoms. We therefore lowered the significance threshold for selecting causal variants as instrumental variables to *P* < 10-5 for these traits and were able to test reverse causal effects of ADHD, which did not show to be significant. For the other traits, reverse causation could not be tested even with this lowered threshold.

We observed suggested causal effects of insomnia on a variety of traits (**Supplementary Table 28**). Evidence for positive causal associations were observed with psychiatric traits, including ADHD (*bxy*=0.77, SE=0.06, *P*=2.50×10-45), schizophrenia (*bxy*=0.68, SE=0.10, *P*=5.12×10-11) and anxiety disorder (*bxy*=0.47, SE=0.10, *P*=4.11×10-6), while a negative association was observed with subjective well-being (*bxy*=−0.51, SE=0.31, *P*=5.02×10-11). In addition, we observed unidirectional effects of insomnia on traits associated with metabolic syndrome traits, including BMI (*bxy*=0.36, SE=0.05, *P*=1.25×10-12, Type 2 diabetes (*bxy*=0.62, SE=0.11, *P*=2.29×10-8) and coronary artery disease (*bxy*=0.61, SE=0.09, *P*=1.58×10-7). We then repeated the GSMR analyses using insomnia variants as the outcome and GWS SNPs from previously studied traits as instrumental variables. We note that reverse analyses were only possible for traits with sufficient GWS SNPs available (nSNPs ≥ 10), and lowered the significance threshold to *P*<1×10-5 when less than 10 GWS SNPs were available. Results showed bidirectional effects for educational attainment (*bxy*=−0.10, SE=0.01, *P*=2.27×10-23), intelligence (*bxy*=−0.05, SE=0.01, *P*=2.17×10-6), and neuroticism (*bxy*=0.09, SE=0.02, *P*=1.24×10-6) and unidirectional reverse effects only for height (*bxy*=−0.03, SE=0.002, *P*=4.90×10-36). Interestingly, the observed associations of insomnia were not observed in the reverse direction, including schizophrenia, ADHD, BMI, Type 2 Diabetes and coronary artery disease, suggesting an important causal role of insomnia for these traits.

**3. Supplementary Figures**

**Supplementary Fig. 1. Regional association plots of 202 genome-wide significant loci in the insomnia meta-analysis**.   
  
*External PDF file*

**Supplementary Fig. 2. Chromatin interactions maps based on significant loci in the insomnia complaints meta-analysis.** Circos plot of chromosome 1-22 showing genes associated with insomnia complaints that were implicated by the genomic risk loci (blue areas), eQTL (green), chromatin interaction (orange), or by both eQTL and chromatin interactions (red). The outer ring shows the Manhattan plot indicating the negative log10-transformed SNP *P*-value in the GWAS meta-analysis of insomnia.

*External PDF file*

**4. Supplementary Tables**

*Supplementary Tables 1 to 28 in separate excel file*

**References**

1. 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 Medicine* **12**, e1001779 (2015).

2. Bycroft, C. *et al.* Genome-wide genetic data on 500,000 uk biobank participants. *Bioarchive* (2017).

3. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279-1283 (2016).

4. 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).

5. Abraham, G., Qiu, Y. & Inouye, M. Flashpca2: Principal component analysis of biobank-scale genotype datasets. *Bioinformatics* (2017).

6. Eriksson, N. *et al.* Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genetics* **6**, e1000993 (2010).

7. Tung, J.Y. *et al.* Efficient replication of over 180 genetic associations with self-reported medical data. *PLoS One* **6**, e23473 (2011).

8. Hu, Y. *et al.* Gwas of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Comm.* **7**, 10448 (2016).

9. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210-223 (2009).

10. Genomes Project, C. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).

11. Fuchsberger, C., Abecasis, G.R. & Hinds, D.A. Minimac2: Faster genotype imputation. *Bioinformatics* **31**, 782-784 (2014).

12. Hammerschlag, A. *et al.* Genome-wide association analysis of insomnia identifies novel risk genes and genetic overlap with psychiatric and metabolic traits. *Nat. Genet.* **49**, 1584 (2017).

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

14. Yang, J. *et al.* Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807-812 (2011).

15. Vilhjalmsson, B.J. *et al.* Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576-592 (2015).

16. Euesden, J., Lewis, C.M. & O'Reilly, P.F. Prsice: Polygenic risk score software. *Bioinformatics* **31**, 1466-1468 (2015).

17. Wang, K., Li, M. & Hakonarson, H. Annovar: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164-e164 (2010).

18. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* **46**, 310-315 (2014).

19. Boyle, A.P. *et al.* Annotation of functional variation in personal genomes using regulomedb. *Genome Res.* **22**, 1790-1797 (2012).

20. Ernst, J. & Kellis, M. Chromhmm: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215-216 (2012).

21. Lane, J.M. *et al.* Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the uk biobank. *Nat. Comm.* **7**, 10889 (2016).

22. Watanabe, K., Taskesen, E., Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with fuma. *Nat. Comm.* **8**, 1826 (2017).

23. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624-633 (2016).

24. Nagel, M. *et al.* Gwas meta-analysis of neuroticism (n= 449,484) identifies novel genetic loci and pathways. *bioRxiv*, 184820 (2017).

25. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from gwas summary data. *bioRxiv*, 168674 (2017).