nature genetics

SUPPLEMENTARY INFORMATION

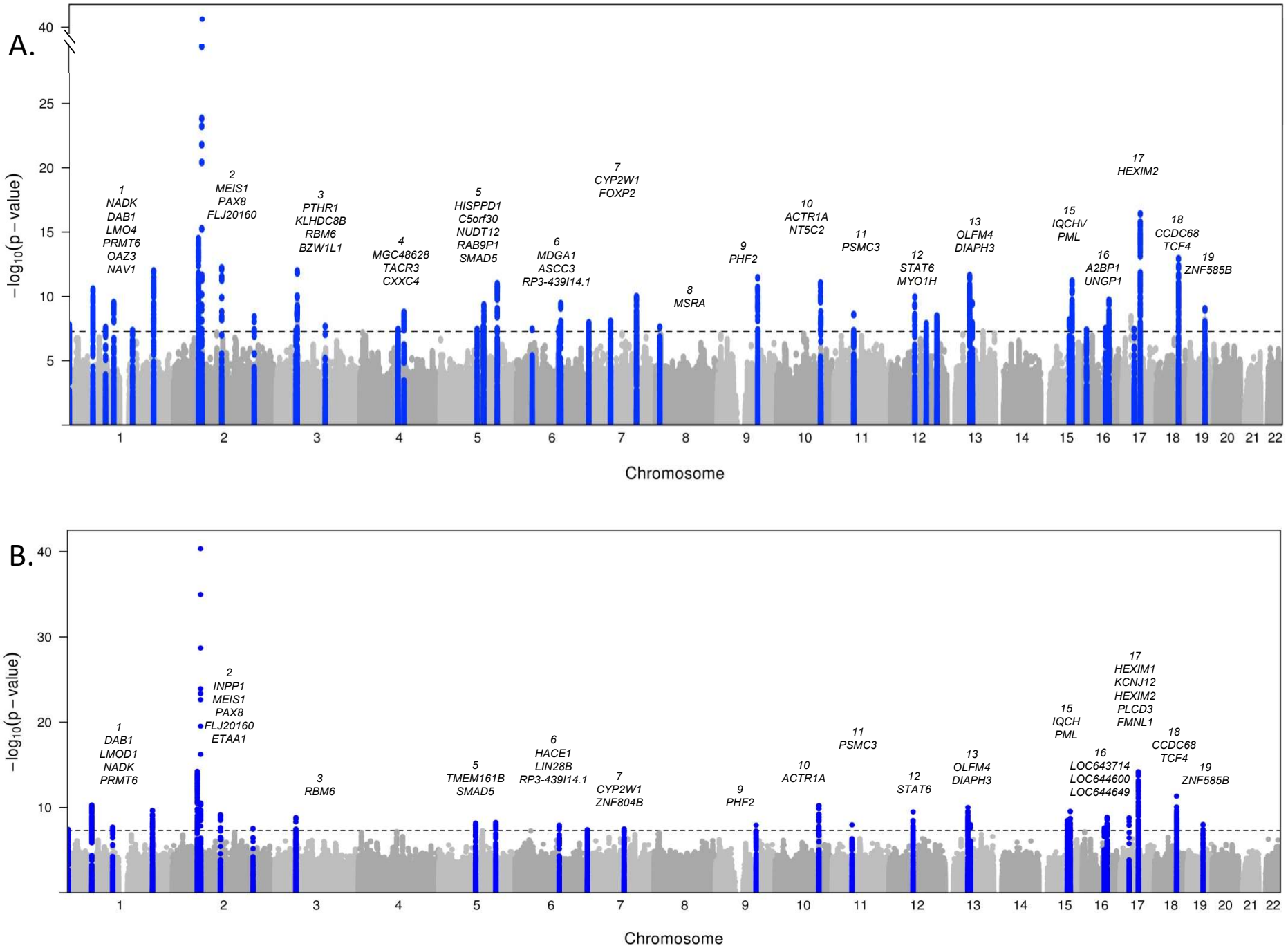
https://doi.org/10.1038/s41588-019-0361-7

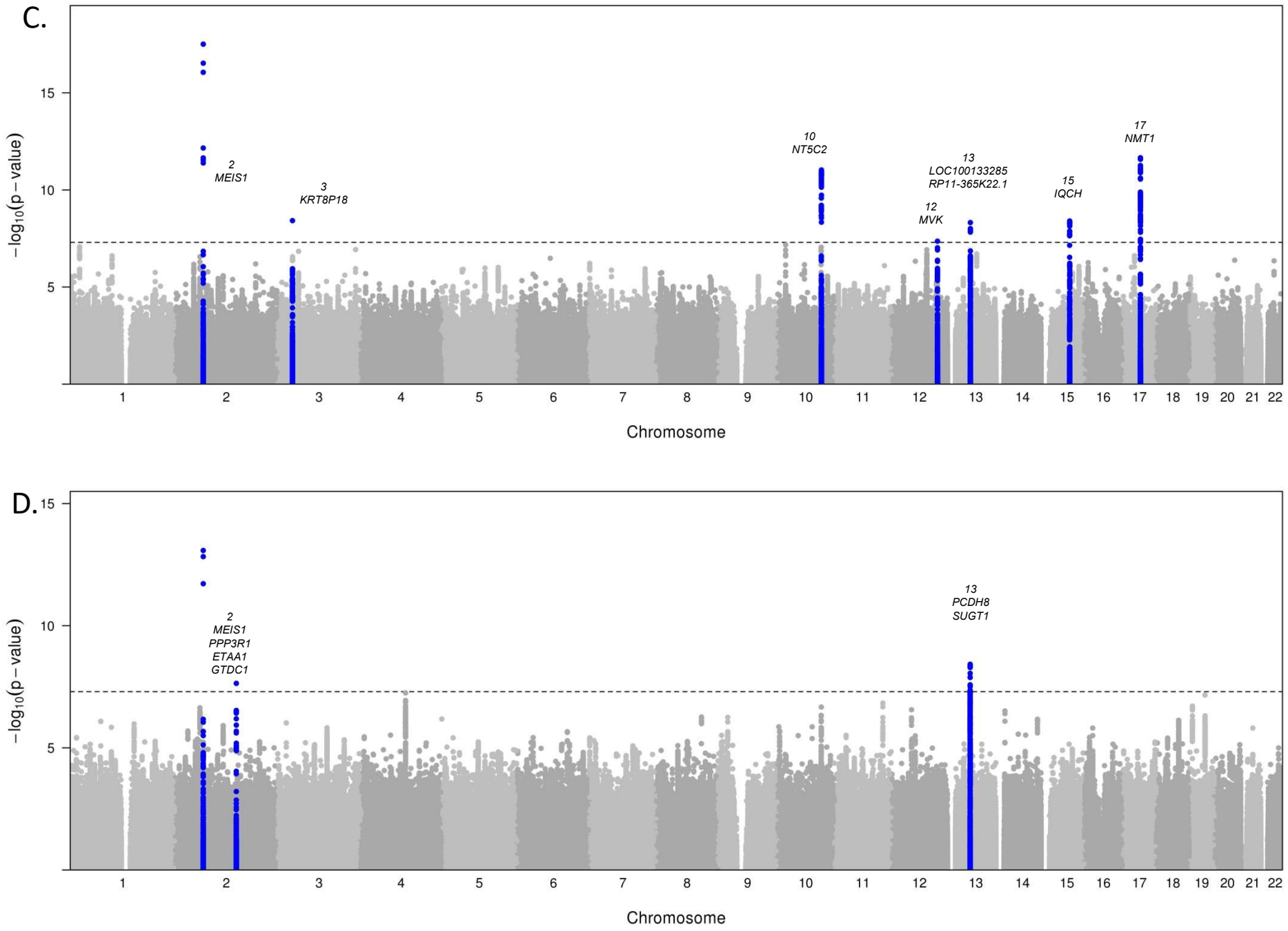
In the format provided by the authors and unedited.

Biological and clinical insights from genetics of insomnia symptoms

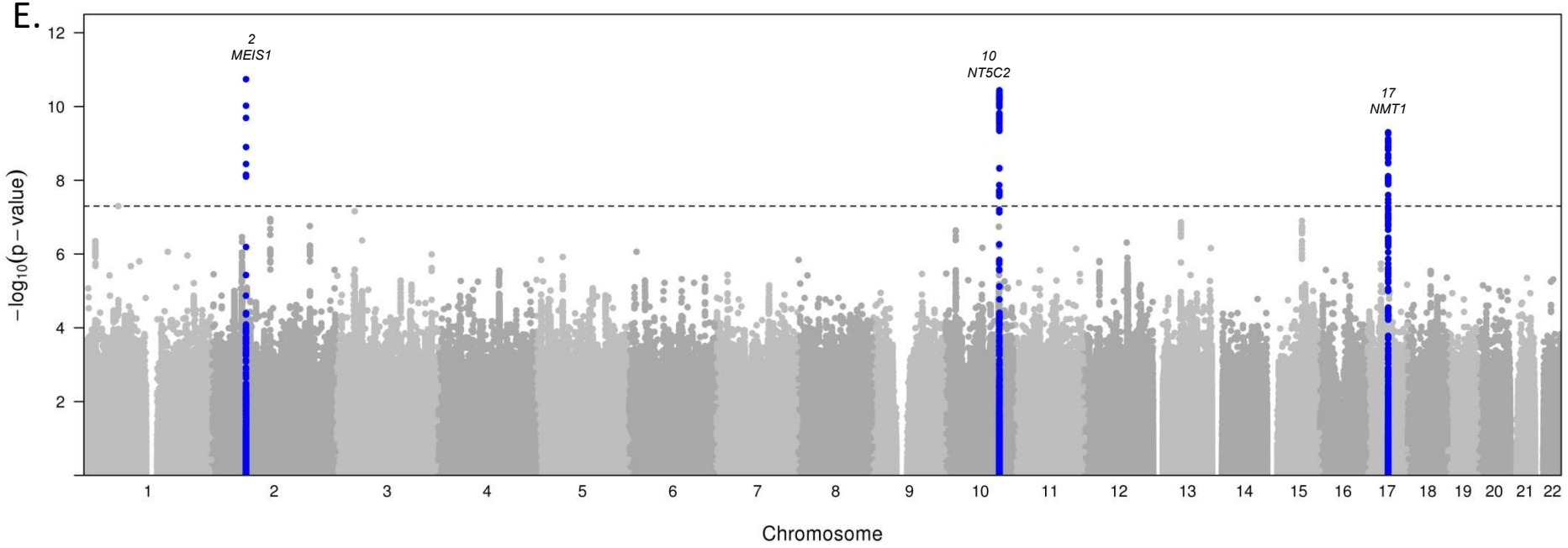
Jacqueline M. Lane[1,2,3,40], Samuel E. Jones[4,40], Hassan S. Dashti[1,3], Andrew R. Wood[4], Krishna G. Aragam[1,3,5,6], Vincent T. van Hees[7], Linn B. Strand[8], Bendik S. Winsvold[8,9,10], Heming Wang[3,11,12], Jack Bowden[13,14], Yanwei Song[1,3,15], Krunal Patel[1,15], Simon G. Anderson[16,17],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

[1]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [2]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [3]Broad Institute, Cambridge, MA, USA. [4]Genetics of Complex Traits, University of Exeter Medical School, Exeter, UK. [5]Cardiology Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [6]Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA. [7]Netherlands eScience Center, Amsterdam, the Netherlands. [8]K.G. Jebsen Centre for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. [9]FORMI and Department of Neurology, Oslo University Hospital, Oslo, Norway. [10]Division of Clinical Neuroscience, Oslo University Hospital and University of Oslo, Oslo, Norway. [11]Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA, USA. [12]Division of Sleep Medicine, Department of Medicine, Harvard Medical School, Boston, MA, USA. [13]MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK. [14]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. [15]College of Science, Northeastern University, Boston, MA, USA. [16]Division of Cardiovascular Sciences, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. [17]Farr Institute of Health Informatics Research, University College London, London, UK. [18]Division of Endocrinology, Diabetes & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. [19]Department of Mathematics, Aston University, Birmingham, UK. [20]Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [21]Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [22]Department of Epidemiology, Erasmus MC University Medical Center, Rotterdam, the Netherlands. [23]Department of Psychiatry, Brigham & Women's Hospital, Harvard Medical School, Boston, MA, USA. [24]School of Social and Community Medicine, University of Bristol, Bristol, UK. [25]Medical Research Council Integrative Epidemiology Unit, University of Bristol, Bristol, UK. [26]Institute of Neurogenomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany. [27]Departments of Psychiatry and Neurology, Massachusetts General Hospital, Boston, MA, USA. [28]Cluster for Systems Neurology (SyNergy), Munich, Germany. [29]Institute of Human Genetics, Technische Universität München, Munich, Germany. [30]Neurogenetics, Technische Universität München, Munich, Germany. [31]A list of members and affiliations appears at the end of the paper. [32]Division of Cardiovascular Medicine, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. [33]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [34]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [35]Departments of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. [36]Clinic for Psychiatry and Psychotherapy, Medical Centre—University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [37]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, OX37LE/NIHR Oxford Biomedical Research Centre, John Radcliffe Hospital, Oxford, UK. [38]Department of Thoracic and Occupational Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. [39]Manchester Diabetes Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK. [40]These authors contributed equally: Jacqueline M. Lane, Samuel E. Jones. [41]These authors jointly supervised this work: Deborah A. Lawlor, Martin K. Rutter, Michael N. Weedon, Richa Saxena. *e-mail: rsaxena@partners.org[A full list of affiliations appears at the end of the paper.]

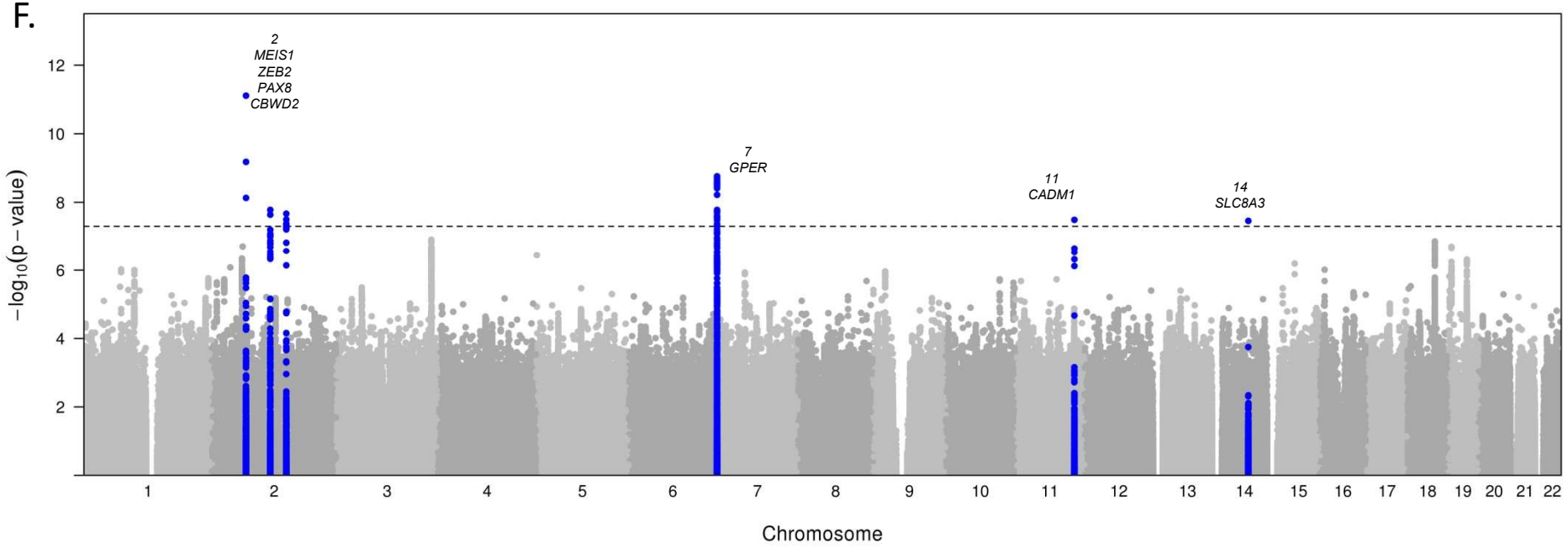

A.
−log10(p−value)
40
25
20
15
10
5
1
NADK
DAB1
LMO4
PRMT6
OAZ3
NAV1
2
MEIS1
PAX8
FLJ20160
3
PTHR1
KLHDC8B
RBM6
BZW1L1
4
MGC48628
TACR3
CXXC4
5
HISPPD1
C5orf30
NUDT12
RAB9P1
SMAD5
6
MDGA1
ASCC3
RP3-439I14.1
7
CYP2W1
FOXP2
8
MSRA
9
PHF2
10
ACTR1A
NT5C2
11
PSMC3
12
STAT6
MYO1H
13
OLFM4
DIAPH3
15
IQCHV
PML
16
A2BP1
UNGP1
17
HEXIM2
18
CCDC68
TCF4
19
ZNF585B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosome
B.
−log10(p−value)
40
30
20
10
1
DAB1
LMOD1
NADK
PRMT6
2
INPP1
MEIS1
PAX8
FLJ20160
ETAA1
3
RBM6
5
TMEM161B
SMAD5
6
HACE1
LIN28B
RP3-439I14.1
7
CYP2W1
ZNF804B
9
PHF2
10
ACTR1A
11
PSMC3
12
STAT6
13
OLFM4
DIAPH3
15
IQCH
PML
16
LOC643714
LOC644600
LOC644649
17
HEXIM1
KCNJ12
HEXIM2
PLCD3
FMNL1
18
CCDC68
TCF4
19
ZNF585B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosome


C.
−log10(p − value)
15
10
5
2
MEIS1
3
KRT8P18
10
NT5C2
12
MVK
13
LOC100133285
RP11-365K22.1
15
IQCH
17
NMT1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosome
D.
−log10(p − value)
15
10
5
2
MEIS1
PPP3R1
ETAA1
GTDC1
13
PCDH8
SUGT1
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22
Chromosome

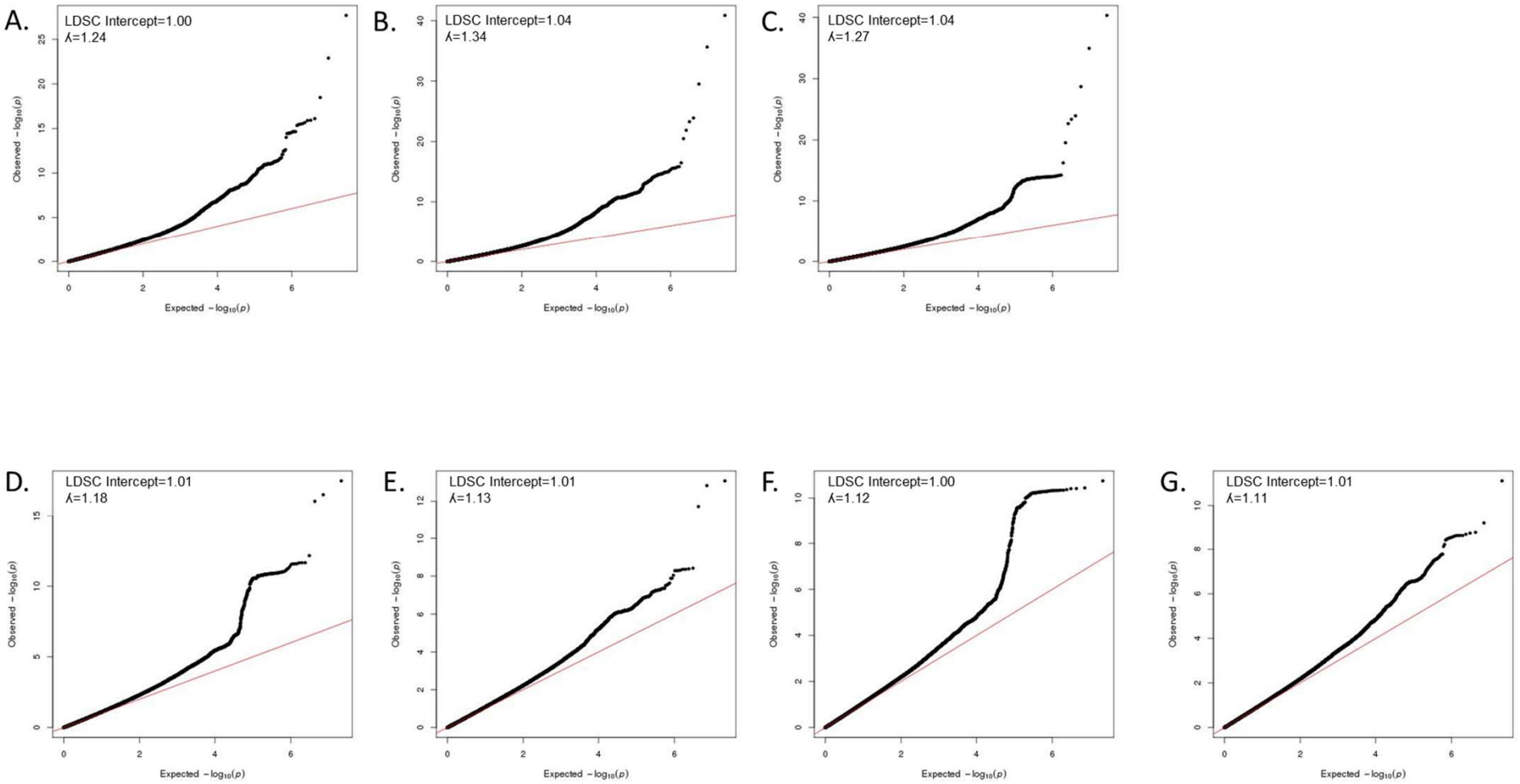
E.


2
MEIS1
10
NT5C2
17
NMT1
−log10(p−value)
Chromosome

F.


2
MEIS1
ZEB2
PAX8
CBWD2
7
GPER
11
CADM1
14
SLC8A3
−log10(p−value)
Chromosome

Supplementary Figure 1. Manhattan plots for genome-wide association analysis of frequent insomnia symptoms (n_{cases}=129,270, $n_{controls}$=108,352) (A), frequent insomnia symptoms with exclusions (n_{cases}=99,383, $n_{controls}$=95,668) (B), frequent insomnia symptoms stratified by sex, females (n_{cases}=58,432, $n_{controls}$=33,210) (C), males (n_{cases}=37,265, $n_{controls}$=47,309) (D), any insomnia symptoms stratified by sex, females (n_{cases}=147,984, $n_{controls}$=33,210) (E), males (n_{cases}=108,805, $n_{controls}$=47,309) (F) . Dotted line is genome-wide significant (5x10-8) results of linear mixed models (A,B) or logistic regression (C-F). SNP-based heritability estimates were calculated using BOLT-REML variance components analysis. Chromosomes are annotated with the nearest gene to each association signal.

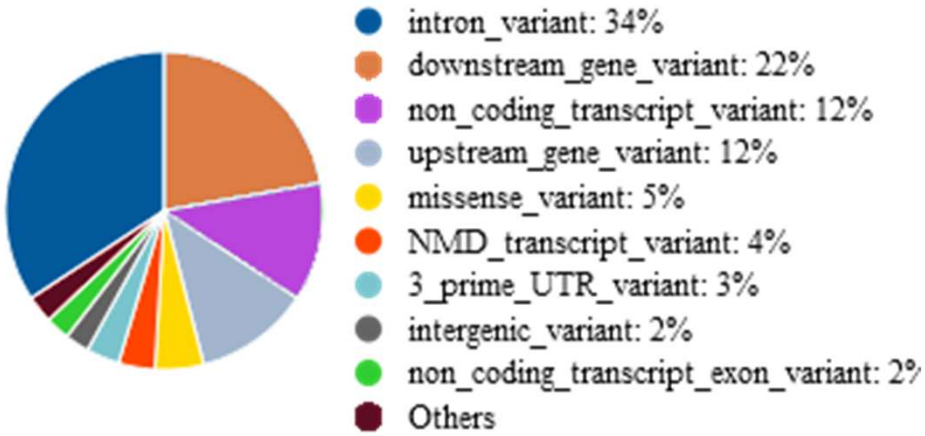

A.
LDSC Intercept=1.00
λ=1.24
B.
LDSC Intercept=1.04
λ=1.34
C.
LDSC Intercept=1.04
λ=1.27
D.
LDSC Intercept=1.01
λ=1.18
E.
LDSC Intercept=1.01
λ=1.13
F.
LDSC Intercept=1.00
λ=1.12
G.
LDSC Intercept=1.01
λ=1.11
Observed −log10(p)
Expected −log10(p)

Supplementary Figure 2. QQ plot for genome-wide association analysis of insomnia symptoms. Plots A-G show the expected verses observed P values from our association analysis of any insomnia symptoms (A), frequent insomnia symptoms (B), frequent insomnia symptoms with exclusions (C), sex stratified frequent insomnia symptoms, females (D), males (E), sex stratified any insomnia symptoms, females (F), males (G). Lambda inflation values were calculated using GenABEL in R and the intercept using LDSC.

		Frequent Insomnia Symptoms				Any Insomnia Symptoms			
		all	all with exclusions	male	female	all	all with exclusions	male	female
Frequent Insomnia Symptoms	all	1							
	all with exclusions	1.028	1						
	male	0.943	0.971	1					
	female	0.957	0.981	0.807	1				
Any Insomnia Symptoms	all	0.977	1.018	0.924	0.933	1			
	all with exclusions	0.966	0.931	0.877	0.951	1.016	1		
	male	0.916	0.953	0.962	0.792	0.952	0.932	1	
	female	0.937	0.978	0.787	0.984	0.945	1.001	0.802	1

Supplementary Figure 3. Genetic correlation between the reported insomnia symptoms GWAS shown, as measured by LD Score regression using LDSC. Color scale represents the strength of the correlation. Sample size of each GWAS as follows: Frequent insomnia symptoms all (n_{cases}=129,270, $n_{controls}$=108,352), with exclusions (n_{cases}=99,383, $n_{controls}$=95,668), females (n_{cases}=58,432, $n_{controls}$=33,210), males (n_{cases}=37,265, $n_{controls}$=47,309); any insomnia symptoms all (n_{cases}=345,022, $n_{controls}$=108,352), with exclusions (n_{cases}=96,298, $n_{controls}$=25,160), females (n_{cases}=147,984, $n_{controls}$=33,210), males (n_{cases}=108,805, $n_{controls}$=47,309).

Consequences (all)



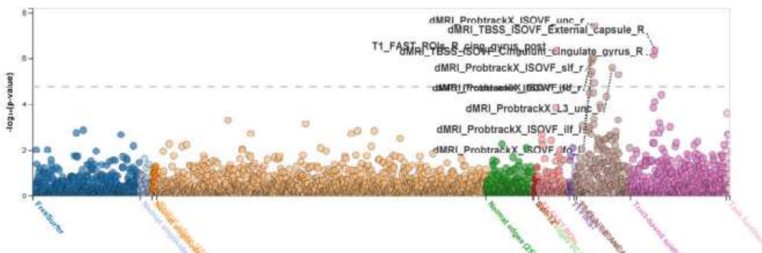
Variant Effect Predictor results for all PICS variants >=.2

Supplementary Figure 4. Summary of variant annotation. Variants within the credible set for each locus were mapped based on functional annotation of each SNP. NMD=nonsense mediated decay, UTR=untranslated region.

16 : 51,484,837 T / C (rs1544637)

Nearest gene: *SALL1*
MAF ranges from 4.8e-1 to 4.8e-1
View on UCSC (http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&highlight=hg19.chr16%3A{ variant.pos }}-51484837&position=chr16%3A51284837-51684837) , GWAS Catalog (https://www.ebi.ac.uk/gwas/search?query=rs1544637) , dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?searchType=adhoc_search&type=rs&rs=rs1544637)

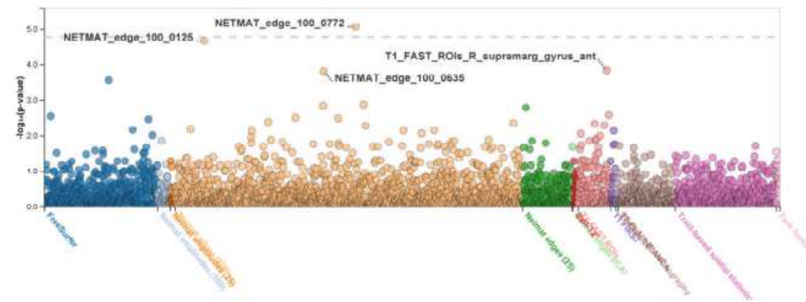
Download Image



3 : 49,980,596 C / T (rs4688760)

Nearest gene: *RBM6*
MAF ranges from 3.0e-1 to 3.0e-1
View on UCSC (http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&highlight=hg19.chr3%3A{ variant.pos }}-49980596&position=chr3%3A49780596-50180596) , GWAS Catalog (https://www.ebi.ac.uk/gwas/search?query=rs4688760) , dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?searchType=adhoc_search&type=rs&rs=rs4688760)

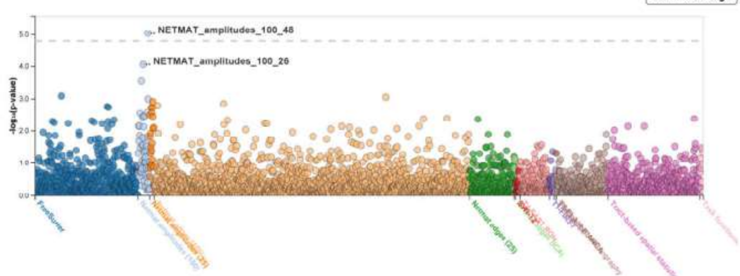
Download Image



15 : 74,340,336 G / C (rs4886860)

Nearest gene: *PML*
MAF ranges from 2.3e-1 to 2.3e-1
View on UCSC (http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&highlight=hg19.chr15%3A{ variant.pos }}-74340336&position=chr15%3A74140336-74540336) , GWAS Catalog (https://www.ebi.ac.uk/gwas/search?query=rs4886860) , dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?searchType=adhoc_search&type=rs&rs=rs4886860)

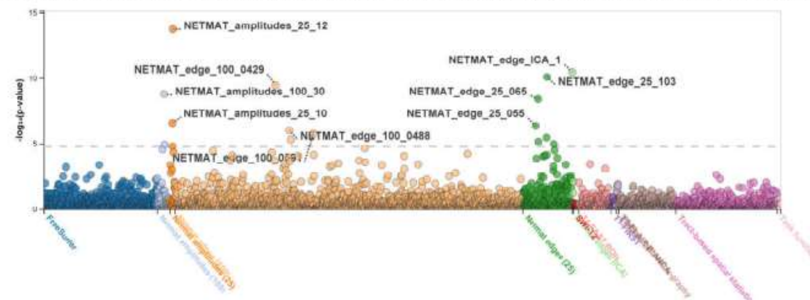
Download Image



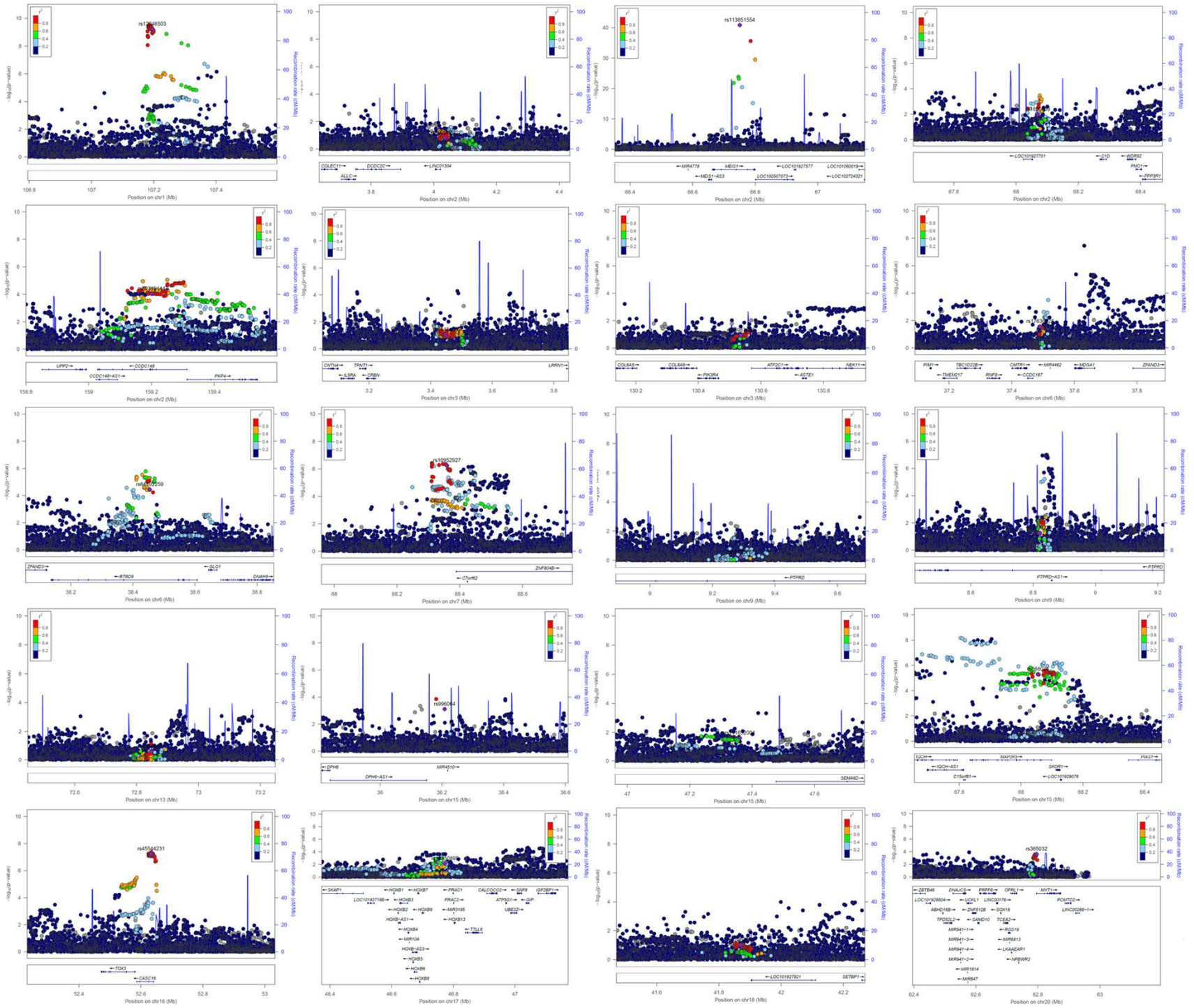
2 : 114,082,175 A / G (rs62158170)

Nearest gene: *PAX8*
MAF ranges from 2.0e-1 to 2.0e-1
View on UCSC (http://genome.ucsc.edu/cgi-bin/hgTracks?db=hg19&highlight=hg19.chr2%3A{ variant.pos }}-114082175&position=chr2%3A113882175-114282175) , GWAS Catalog (https://www.ebi.ac.uk/gwas/search?query=rs62158170) , dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?searchType=adhoc_search&type=rs&rs=rs62158170)

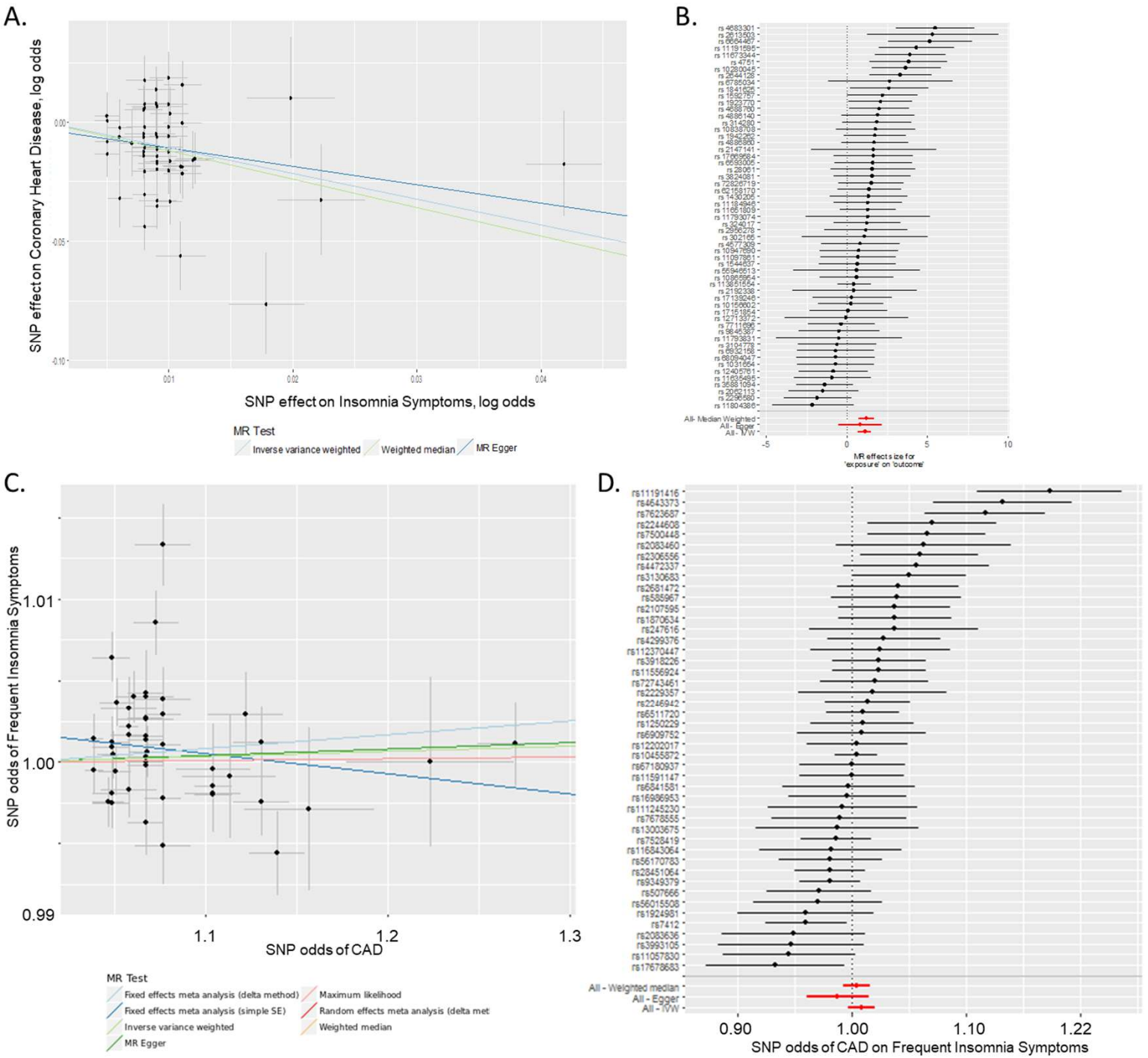
iAycHgsIDJweDsgc3Ryb2tlLXdpZHRoOiAycHg7IiB4MT0iMTY3LjM2MSIgeTE9IjIzLjI3MjdlIHgyPSIxNzguNzI3IiB5MjOiMjMuMjcyNyIgLz48L2c+PGcgY2xhc3M9Imx6LWRhdGFfbGF5ZXIic2NhdHRlci1sYWJlbCI+P)



Supplementary Figure 5. Association of frequent insomnia symptom loci with brain imaging phenotypes in the UK Biobank (images from http://big.stats.ox.ac.uk/). PheWAS results are from 3,144 GWAS of brain imaging phenotypes on 9,707 subjects in the UK Biobank by fitting an additive model of association at each variant.


rs12446503
rs113851554
rs10952927
rs9960064
rs45544231
rs365032
Position on chr1 (Mb)
Position on chr2 (Mb)
Position on chr3 (Mb)
Position on chr6 (Mb)
Position on chr7 (Mb)
Position on chr9 (Mb)
Position on chr13 (Mb)
Position on chr15 (Mb)
Position on chr16 (Mb)
Position on chr17 (Mb)
Position on chr18 (Mb)
Position on chr20 (Mb)
Recombination rate (cM/Mb)

Supplementary Figure 6. Regional association plots for RLS loci in frequent insomnia symptoms GWAS using linear mixed models (n_{cases}=129,270, $n_{controls}$=108,352). Panels highlight loci previously identified to associate with RLS. Genes within the region are shown in the lower panel. The blue line indicates the recombination rate. Filled circles show the log10 P value for each SNP, with the RLS SNP shown in purple. Additional SNPs in the locus are colored according to correlation (r^2) with the RLS SNP (estimated by LocusZoom based on the CEU HapMap haplotypes).


A.
SNP effect on Coronary Heart Disease, log odds
SNP effect on Insomnia Symptoms, log odds
MR Test
Inverse variance weighted
Weighted median
MR Egger
B.
All- Median Weighted
All - Egger
All - IVW
MR effect size for 'exposure' on 'outcome'
C.
SNP odds of Frequent Insomnia Symptoms
SNP odds of CAD
MR Test
Fixed effects meta analysis (delta method)
Fixed effects meta analysis (simple SE)
Inverse variance weighted
MR Egger
Maximum likelihood
Random effects meta analysis (delta met
Weighted median
D.
All - Weighted median
All - Egger
All - IVW
SNP odds of CAD on Frequent Insomnia Symptoms

Supplementary Figure 7. Causal relationship of insomnia symptoms with CAD in the UK Biobank. Method comparison plot of the association between single nucleotide polymorphisms associated with frequent insomnia symptoms and CAD (A) and forest plot shows the estimate of the effect of genetically increased insomnia risk on CAD (B). Association between single nucleotide polymorphisms associated with CAD and insomnia symptoms (C) and forest plot shows the estimate of the effect of genetically increased CAD risk on insomnia symptoms (D). Results are shown for multiple MR association tests. Method comparison plot shows SNP effects on the outcome plotted against SNP effects on the exposure, with each dot representing the effect of the SNP on the exposure and outcome with the 95% confidence interval shown in gray. The slope of each line represents the causal association for each method. Forest plots show each SNP with the 95% confidence interval (gray line segment) of the estimate and the Inverse Variance MR, MR-Egger, and Weighted Median MR results in red. Frequent insomnia symptoms GWAS n_{cases}=129,270, $n_{controls}$=108,352, CAD GWAS n_{cases}=23,980, $n_{controls}$=361,706.

Supplementary Note

Materials and Methods

UK Biobank

Phenotype and Covariate measures

Activity-monitor derived measures of sleep

Actigraphy devices (Axivity AX3) were worn 2.8 - 9.7 years after study baseline by 103,711 individuals from the UK Biobank for up to 7 days. Details are described elsewhere [REF: PMID: 28146576]. Of these 103,711 individuals, we excluded 11,067 based on accelerometer data quality. Participants were excluded if they satisfied at least one of the following conditions (see also http://biobank.ctsu.ox.ac.uk/crystal/label.cgi?id=1008): a non-zero or missing value in data field 90002 ("Data problem indicator"), "good wear time" flag (field 90015) set to 0 (No), "good calibration" flag (field 90016) set to 0 (No), "calibrated on own data" flag (field 90017) set to 0 (No) or overall wear duration (field 90051) less than 5 days. Additionally, samples with extreme values of mean sleep duration (<3 hours or >12 hours) or mean number of sleep periods (<5 or >30) were excluded. 85,502 samples remained after non-white ethnicity exclusions. Sleep measures were derived by processing raw accelerometer data (.cwa). First we converted .cwa files available from the UK Biobank to .wav files using Omconvert (https://github.com/digitalinteraction/openmovement/tree/master/Software/AX3/omconvert) for signal calibration to gravitational acceleration[1,2] and interpolation[2]. The .wav files were processed with the R package GGIR to infer activity monitor wear time[3], and extract the z-angle across 5-second epoch time-series data for subsequent use in estimating the sleep period time window (SPT-window)[4] and sleep episodes within it[5].

The SPT-window was estimated using an algorithm described in Van Hees et al.[4], implemented in the GGIR R package and validated using PSG in an external cohort. Briefly, for each individual, median values of the absolute change in z-angle (representing the dorsal-ventral direction when the wrist is in the anatomical position) across 5-minute rolling windows were calculated across a 24-hour period, chosen to make the algorithm insensitive to activity-monitor orientation. The 10th percentile was incorporated into the threshold distinguishing movement from non-movement. Bouts of inactivity lasting ≥30 minutes are recorded as inactivity bouts. Inactivity bouts that are <60 minutes apart are combined to form inactivity blocks. The start and end of longest block defines the start and end of the SPT-window [van Hees et al 2018, in press]. *Sleep duration*. Sleep episodes within the SPT-window were defined as periods of at least 5 minutes with no change larger than 5° associated with the z-axis of the accelerometer, as motivated and described in van Hees et al[5]. The summed duration of all sleep episodes was used as indicator of sleep duration. *Sleep efficiency.* This was calculated as sleep duration (defined above) divided by the time elapsed between the start of the first inactivity bout and the end of the last inactivity bout (which equals the SPT-window duration). *Number of sleep episodes within the SPT-window.* This is defined as the number of sleep episodes separated by last least 5 minutes of wakefulness within the SPT-window. The least-active five hours (*L5*) and the most-active ten hours (*M10*) of each day were defined using a five-hour and ten-hour daily period of minimum and maximum activity,

respectively. These periods were estimated using a rolling average of the respectively time window. L5 was defined as the number of hours elapsed from the previous midnight whereas M10 was defined as the number of hours elapsed from the previous midday. *Sleep midpoint* was calculated for each sleep period as the midpoint between the start of the first detected sleep episode and the end of the last sleep episode used to define the overall SPT-window (above). This variable is represented as the number of hours from the previous midnight, e.g. 2am = 26. *Diurnal inactivity duration* is the total daily duration of estimated bouts of inactivity that fall outside of the SPT-window. All activity-monitor phenotypes were adjusted for age at accelerometer wear, sex, season of wear, release (categorical; UK BiLeVe, UKB Axiom interim, release UKB Axiom full release) and number of valid recorded nights (or days for M10) when performing the association test in BOLT-LMM.

Coronary Artery Disease

Coronary artery disease (CAD) in the UK Biobank was defined as a diagnosis of myocardial infarction or coronary revascularization, as described previously[6]. Myocardial infarction was defined as a self-reported diagnosis of "heart attack" or "myocardial infarction" or hospitalization/death due to an ICD-10 code for acute myocardial infarction (I21.0, I21.1, I21.2, I21.4, I21.9). Coronary revascularization was defined based on a self-report of "coronary artery bypass grafting" or "coronary artery angioplasty," or hospitalization for an OPCS-4 code for coronary artery bypass grafting (K40.1-K40.4, K41.1-K41.4, K45.1-K45.5) or coronary artery angioplasty ± stenting (K49.1–49.2, K49.8–49.9, K50.2, K75.1–75.4, K75.8–75.9). Individuals with self-reported "angina" or hospitalization for an ICD-10 code of angina pectoris (I20) or chronic ischemic heart disease (I25.1, I25.5, I25.6, I25.8, I25.9) were excluded from analyses. All other individuals were defined as controls.

The following were all considered to be Sleep medications (note this list includes generic and propriatory names for the same medication so that we included all ways that sleeping tables may have been reported): : oxazepam, meprobamate, medazepam, bromazepam, lorazepam, clobazam, chlormezanone, temazepam, nitrazepam, lormetazepam, diazepam, zopiclone, triclofos, methyprylone, prazepam, triazolam, ketazolam, dichloralphenazone, clomethiazole, zaleplon, butobarbital. Antidepressants: amitriptyline, citalopram, fluoxetine, sertraline, venlafaxine, dosulepin, paroxetine, mirtazapine, escitalopram, trazodone, prozac, seroxat, cipralex, duloxetine, lofepramine, clomipramine, nortriptyline, imipramine, dothiepin, cipramil, amitriptyline, prothiaden, trimipramine, lustral, reboxetine, zispin, cymbalta, anafranil, doxepin, moclobemide, phenelzine, fluvoxamine, yentreve, triptafen, surmontil, tranylcypromine, allegron, edronax, molipaxin, mianserin, nardil, faverin, nefazodone, amitriptyline+chlordiazepoxide, isocarboxazid, manerix, maoi, sinequan, tranylcypromine+trifluoperazine, ludiomil, norval, tryptizol, and fluphenazine hydrochloride+nortriptyline. Antipsychotics: prochlorperazine, olanzapine, quetiapine, risperidone, chlorpromazine, trifluoperazine, amisulpride, sulpiride, seroquel, haloperidol, aripiprazole, stelazine, depixol, flupentixol, clozapine, promazine, risperdal, modecate, fluanxol, flupenthixol, zyprexa, zuclopenthixol, clopixol, largactil, abilify, fluphenazine, haldol, serenace, clozaril, cpz, perphenazine, levomepromazine, pericyazine, dolmatil, fentazin, fluphenazine, benperidol, pimozide, zaponex, denzapine, neulactil, thioridazine, dozic, fluspirilene, panadeine, and sertindole. Anxiolytics: zopiclone, diazepam, temazepam, zolpidem, nitrazepam, lorazepam, hydroxyzine, zimovane, phenergan, promethazine,

buspirone, atarax, oxazepam, loprazolam, chlordiazepoxide, lormetazepam, ucerax, stilnoct, diazepam, buspar, alprazolam, librium, xanax, meprate, dalmane, clomethiazole, meprobamate, welldorm, amitriptyline+chlordiazepoxide, flurazepam, heminevrin, medazepam, neulactil, sinequan, almazine, atensine, carisoma, chloractil, chloral, dichloralphenazone, dormonoct, methyprylone, mogadon, rohypnol and tryptizol.

Replication Cohorts

The HUNT Study

Sample ascertainment and phenotype definition.

The Nord-Trøndelag Health Study (HUNT) consists of three different population-based health surveys conducted in the county of Nord-Trøndelag, Norway over approximately 20 years (HUNT1 [1984-1986], HUNT2 [1995-1997] and HUNT3 [2006-2008])[7]. At each survey, the entire adult population (≥ 20 years) was invited to participate by completing questionnaires, attending clinical examinations and interviews. Participation rates in HUNT1, HUNT2 and HUNT3 were 89.4% (n=77,212), 69.5% (n=65 237) and 54.1% (n=50 807), respectively[7]. Taken together, the study included more than 120,000 different individuals from Nord-Trøndelag County. Biological samples including DNA have been collected for approximately 70,000 participants. The HUNT Study has been described in more detail elsewhere[7].

Participants reporting insomnia in either HUNT1, HUNT2 or HUNT3 were classified as having insomnia for the present study. In HUNT1, participants who answered "often" or "almost every night" to the question "During the last month, have you had any problems falling asleep or sleep disorders?" were classified as having insomnia. Those who answered "almost every night" were additionally classified as having frequent insomnia. In HUNT2, participants who answered "often" or "almost every night" to either of the questions "Have you had difficulty falling asleep in the last month?" or "During the last month, have you woken too early and not been able to get back to sleep?", were classified as having insomnia. Those who answered "almost every night" to either question were additionally classified as having frequent insomnia. In HUNT3, participants were asked how often in the last 3 months they had "had difficulty falling asleep at night" and "woken up repeatedly during the night" with the response options "Never/seldom", "Sometimes" and "Several times a week". Those who answered "Several times a week" to either question were classified as having insomnia. Because of fewer response options on the insomnia questions in HUNT3 than in HUNT1 and HUNT2, we classified frequent insomnia differently than in the two previous studies, and only those who answered "Several times a week" to both questions were classified as having frequent insomnia.

Participants were classified as controls (no insomnia) if they answered the insomnia questions in at least one of the three HUNT studies, and did not qualify as cases.

Genotyping, quality control and imputation.

DNA from 71,860 HUNT samples was genotyped using one of three different Illumina HumanCoreExome arrays (HumanCoreExome12 v1.0, HumanCoreExome12 v1.1 and UM HUNT Biobank v1.0). Genotyping and quality control are described here[8] [PMID:29083406]. Imputation was performed on samples of recent European ancestry using Minimac3 (v2.0.1, http://genome.sph.umich.edu/wiki/Minimac3)[9] and a merged reference panel that was constructed

by combining the Haplotype Reference Consortium panel (release version 1.1)[10] and a local reference panel based on 2,202 whole-genome sequenced HUNT study participants. After restricting to the above, those with European ancestry and with available phenotype information, 62,533 individuals were included in the analysis sample.

Association analysis.

Association analyses were conducted using SAIGE [Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies[11], a generalized mixed effects model approach, to account for cryptic population structure and relatedness when modelling the association between genotype probabilities (dosages) and insomnia symptoms. Models were adjusted for sex, birth year, genotyping batch and four principal components (PCs). PCs were computed using PLINK. Additional filters applied to the analysis included minor allele count ≥ 5 and imputation r2 ≥ 0.3.

Partners Healthcare Biobank

The Partners Biobank[12] is an ongoing hospital-based cohort study launched in 2010 from the Partners HealthCare hospitals with electronic medical record (EMR) and genetic data supplemented with health surveys. Recruitment is from participating clinics at Brigham and Women's Hospital (BWH), Massachusetts General Hospital (MGH), Spaulding Rehabilitation Hospital (SRH), Faulkner Hospital (FH) and McLean Hospital (MCL), (NWH), (NSM) and electronically from previous patients. At the time of analysis, a total of 73,683 participants have provided consent, of which 20,087 have been genotyped. Cases of insomnia were ascertained from EMR using diagnoses of insomnia (and primary insomnia) (n =2,217). Participants without insomnia and restless leg syndrome were selected as controls (n =17,018). Participants were genotyped using the Illumina Multi-Ethnic GWAS/Exome SNP Array. Imputation was performed using Minimac3 (http://genome.sph.umich.edu/wiki/Minimac3) using the HRC (Version r1.1 2016) reference panel for imputation. This HRC panel consists of 64,940 haplotypes of predominantly European ancestry. Haplotype phasing was performed using SHAPEIT[13]. In total, 16,791 samples of self-reported European ancestry with high-quality genotyping, EMR data and covariate data were used for these analyses. We derived a genetic risk score (GRS) for frequent insomnia symptoms using 57 SNPs. Individual participant scores were created by summing the number of risk alleles at each genetic variant, which where weighted by the respective allelic effect sizes on frequent insomnia symptoms. We tested whether the GRSs for insomnia by estimating linear trends of the GRS adjusted for age, sex, and genotyping array.

1. van Hees, V. T. *et al.* Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J. Appl. Physiol. Bethesda Md 1985* **117**, 738–744 (2014).
2. Doherty, A. *et al.* Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank Study. *PloS One* **12**, e0169649 (2017).
3. van Hees, V. T. *et al.* Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PloS One* **8**, e61691 (2013).
4. Hees, V. T. van *et al.* Estimating sleep parameters using an accelerometer without sleep diary. *bioRxiv* 257972 (2018). doi:10.1101/257972
5. van Hees, V. T. *et al.* A novel, open access method to assess sleep duration using a wrist-worn accelerometer. *PloS One* **10**, e0142533 (2015).
6. Klarin, D. *et al.* Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat. Genet.* **49**, 1392–1397 (2017).
7. Krokstad, S. *et al.* Cohort profile: the HUNT Study, Norway. *Int. J. Epidemiol.* **42**, 968–977 (2013).
8. Ferreira, M. A. *et al.* Shared genetic origin of asthma, hay fever and eczema elucidates allergic disease biology. *Nat. Genet.* **49**, 1752–1757 (2017).
9. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
10. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
11. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *bioRxiv* 212357 (2018). doi:10.1101/212357
12. Karlson, E. W., Boutin, N. T., Hoffnagle, A. G. & Allen, N. L. Building the partners HealthCare Biobank at Partners Personalized Medicine: Informed consent, return of research results, recruitment lessons and operational considerations. *J. Pers. Med.* **6**, (2016).
13. O'Connell, J. *et al.* A general approach for haplotype phasing across the full spectrum of relatedness. *PLoS Genet* **10**, e1004234 (2014).