Genetic variants in *RBFOX3* are associated with sleep latency

Najaf Amin,[1] Karla V. Allebrandt,[2] Ashley van der Spek,[1] Bertram Müller-Myhsok,[3] Karin Hek,[4,5] Maris Teder-Laving,[6] Caroline Hayward,[7] Tõnu Esko,[6] Josine G. van Mill,[8] Hamdi Mbarek,[9] Nathaniel F. Watson,[10,11] Scott A. Melville,[12] Fabiola Del Greco M.,[13] Enda M. Byrne,[14,15] Edwin Oole,[16] Ivana Kolcic,[17] Ting-hsu Chen,[18] Daniel S. Evans,[19] Josef Coresh,[20] Nicole Vogelzangs,[8] Juha Karjalainen,[21] Gonneke Willemsen,[9] Sina A. Gharib,[11,22] Lina Zgaga,[7] Evelin Mihailov,[6] Katie L. Stone,[19] Harry Campbell,[23] Rutger Brouwer,[16] Ayse Demirkan,[1] Aaron Isaacs,[1] Zoran Dogas,[24] Kristin D. Marciante,[25] Susan Campbell,[7] Fran Borovecki,[26] Annemarie I. Luik,[4] Man Li,[27] Jouke Jan Hottenga,[9] Jennifer E. Huffman,[7] Mirjam C. G. N. van den Hout,[16] Steven R. Cummings,[19] Yurii S. Aulchenko,[1] Philip R. Gehrman,[28] Andre G. Uitterlinden,[4,29,30] H. Erich Wichmann,[31] Martina Müller-Nurasyid, [31] Rudolf S. N. Fehrmann,[21] Grant W. Montgomery,[15] Albert Hofman,[4] W.H. Linda Kao,[27] Ben A. Oostra,[1] Alan F. Wright,[7] Jacqueline M. Vink,[9] James F. Wilson,[23] Peter P. Pramstaller,[13,32,33] Andrew A. Hicks,[13] Ozren Polasek,[17,34] Naresh M. Punjabi,[35] Susan Redline,[36] Bruce M. Psaty,[37,38] Andrew C. Heath,[39] Martha Merrow,[2] Gregory J. Tranah,[19] Daniel J. Gottlieb,[40] Dorret I. Boomsma,[9] Nicholas G. Martin,[15] Igor Rudan,[23] Henning Tiemeier,[4,5,41,42] Wilfred F. J. van IJcken,[16] Brenda W. Penninx,[8] Andres Metspalu,[6] Thomas Meitinger,[43,44] Lude Franke,[21] Till Roenneberg,[2,*] Cornelia M. van Duijn,[1,30,45,*]

1. Unit of Genetic Epidemiology, Department of Epidemiology, Erasmus University Medical Centre, Rotterdam, the Netherlands;
2. Institute of Medical Psychology, Ludwig-Maximilians-University, Munich, Germany;
3. Max-Planck-Institute of Psychiatry, Munich, Germany;
4. Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands;
5. Department of Psychiatry, Erasmus Medical Center, Rotterdam, The Netherlands;
6. Estonian Genome Center, University of Tartu and Estonian Biocenter, Tartu, Estonia;
7. Medical Research Council, Human Genetics Unit, IGMM, Edinburgh, Scotland;
8. Department of Psychiatry, VU University Medical Center Amsterdam, Amsterdam, The Netherlands;
9. Department of Biological Psychology, VU University Amsterdam, The Netherlands;
10. Department of Neurology, University of Washington;
11. University of Washington Medicine Sleep Center;
12. Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, MA, USA;
13. Center for Biomedicine, European Academy of Bolzano, Bolzano, Italy - affiliated institute of the University of Lübeck, Germany;
14. Queensland Brain Institute, University of Queensland, Brisbane, Australia;
15. Queensland Institute of Medical Research, Brisbane, Australia;
16. Center for Biomics, Erasmus Medical Center, Rotterdam, the Netherlands;
17. School of Medicine, University of Split, Croatia;
18. VA Boston Healthcare System, Boston University;

[19.] California Pacific Medical Center Research Institute, San Francisco, CA, USA.
[20.] Departments of Epidemiology, Biostatistics, and Medicine, Johns Hopkins University, Baltimore, USA;
[21.] Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands;
[22.] Department of Medicine, Division of Pulmonary & Critical Care Medicine, University of Washington;
[23.] Centre For Population Health Sciences, University of Edinburgh, Edinburgh, Scotland;
[24.] Department of Neuroscience, Faculty of Medicine, University of Split, Split, Croatia;
[25.] Cardiovascular Health Research Unit, Department of Medicine, University of Washington;
[26.] Centre for Functional Genomics and Department of Neurology, Faculty of Medicine, University of Zagreb, Zagreb, Croatia;
[27.] Department of Epidemiology, Johns Hopkins University, Baltimore, USA;
[28.] Department of Psychiatry, University of Pennsylvania, Philadelphia, PA 19104;
[29.] Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands;
[30.] Netherlands Consortium on Health Aging and National Genomics Initiative, Leiden, the Netherlands;
[31.] Institute of Epidemiology I, Helmholtz Zentrum München, Neuherberg, Germany;
[32.] Department of Neurology, General Central Hospital, Bolzano, Italy;
[33.] Department of Neurology, University of Lübeck, Lübeck, Germany;
[34.] Centre for Global Health, University of Split, Croatia
[35.] Dept of Pulmonary Medicine and Epidemiology, Johns Hopkins University, Baltimore, USA;
[36.] Department of Medicine, Brigham and Women's Hospital and Beth Deaconess Medical Center, Harvard Medical School, Boston, MA, USA;
[37.] Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology and Health Services, University of Washington, Seattle, WA;
[38.] Group Health Research Institute, Group Health Cooperative, Seattle, WA;
[39.] Department of Psychiatry, Washington University, St.Louis, Missouri, USA;
[40.] VA Boston Healthcare System, Boston MA;
[41.] Department of Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, The Netherlands;
[42.] Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands;
[43.] Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, Germany;
[44.] Institute of Human Genetics, Techinsche Universität München, München, Germany;
[45.] Centre of Medical Systems Biology, Leiden, the Netherlands.
* These authors contributed equally to this work

Correspondence to:

Dr. Najaf Amin
Erasmus Medical Center Rotterdam
Department of Epidemiology
Wytemaweg 80, 3015 CN
Rotterdam, The Netherlands
E-mail: n.amin@erasmusmc.nl
Tel: +31 10 7038125

Conflict of Interest

Dr. Najaf Amin is supported by the Netherlands Brain Foundation (project number F2013(1)-28). Prof.dr. Gregory J. Tranah was supported by NIA grant R01AG030474. Prof.dr. Henning Tiemeier was supported by the Vidi Grant of ZonMw (the Netherlands Organization for Health Research and Development, 2009-017.106.370). All other authors report no biomedical financial interests or potential conflicts of interest.

ABSTRACT

Time to fall asleep (sleep latency) is a major determinant of sleep quality. Chronic, long sleep latency is a major characteristic of sleep-onset insomnia and/or delayed sleep phase syndrome. In this study we aimed to discover common polymorphisms that contribute to the genetics of sleep latency. We performed a meta-analysis of genome-wide association studies (GWAS) including 2 572 737 single nucleotide polymorphisms (SNPs) established in seven European cohorts including 4 242 individuals.

We found a cluster of three highly correlated variants (rs9900428, rs9907432 and rs7211029) in the RNA binding protein fox-1 homolog 3 gene (*RBFOX3*) associated with sleep latency (*p-values* = 5.77 *10^{-08}, 6.59*10^{-08} and 9.17*10^{-08}). These SNPs were replicated in up to 12 independent populations including 30 377 individuals (*p-values* = 1.5*10^{-02}, 7.0*10^{-03} and 2.5*10^{-03}; combined meta-analysis *p-values* = 5.5*10^{-07}, 5.4*10^{-07} and 1.0*10^{-07}). A functional prediction of *RBFOX3* based on co-expression with other genes shows that this gene is predominantly expressed in brain (*p-value* = 1.4*10^{-316}) and the central nervous system (*p-value* = 7.5*10^{-321}). The predicted function of *RBFOX3* based on co-expression analysis with other genes shows that this gene is significantly involved in the release cycle of neurotransmitters including GABA and various monoamines (*p-values* < 2.9*10^{-11}) that are crucial in triggering the onset of sleep. To conclude, in this first large-scale GWAS of sleep latency we report a novel association of variants in *RBFOX3* gene. Further, a functional prediction of *RBFOX3* supports the involvement of *RBFOX3* with sleep latency.

Keywords: sleep latency, GWAS, RBFOX3, GABA, methylation, monoamine.

INTRODUCTION

Sleep occurs during defined periods within the 24 hour cycle [1]. Its timing and duration is considered to be regulated by two processes; (I) a circadian process (*i.e*., being awake during the day and asleep during night) and (II) a homeostatic process that represents the sleep pressure accumulating during wakefulness, *i.e*. the longer one is awake, the greater sleep pressure and the greater will be the duration of sleep when permitted [2]. Sleep latency, *i.e*., the duration it takes to fall asleep, is a measure of sleep quality computed as the time interval between "lights out" until the onset of sleep. Photic information from the retina is projected to the suprachiasmatic nucleus (SCN) via the retino-hypothalamic tract. In humans, the SCN is considered the pacemaker for the timing of daily sleep-wake behavior and consists of approximately 10 000 neurons located on both sides of the midline above the optic chiasma, about 3 cm behind the eyes [3,4]. At night, the SCN signals the release of melatonin – a hormone produced by the pineal gland that promotes sleep. Sleep latency may be assessed using self-reports, actigraphy, or with polysomnography. Normal sleepers are known to provide estimates of sleep latency that correlate well with polysomnographic measures [5,6]. Previous studies show that sleep latency is largely independent of an individual's chronotype [7].

Inter-individual differences in sleep latency are caused by both genetic and non-genetic factors. The latter include gender [8,9], age [8], consumption of stimulants [10], dietary intake [11], sedentary life [12] and illnesses, such as depression [13]. Persistent increased sleep latency is a major characteristic of delayed sleep phase syndrome [14] and/or sleep-onset insomnia [15-18]. Prolonged sleep latency may shorten sleep duration and may lead to a wide range of problems including irritability, cognitive impairment, depression and loss of productivity as well as accident rates due to sleepiness. Increased sleep latency has

158 also been associated with poor academic performance in children and adolescents [19].
159 While very long sleep latency can lead to sleep deprivation, very short sleep latency can
160 reflect sleep deprivation (*i.e.*, 'sleep debt' due to insufficient sleep normally shortens
161 sleep latency). Unusually short sleep latency also may indicate disorders of excessive
162 sleepiness (e.g., narcolepsy). When giving individuals the opportunity to sleep during
163 the day, sleep latency is used as an objective measure of daytime sleepiness in the
164 diagnosis of sleep disorders.

165 Heritability of sleep latency is estimated to be around 17-44% [20-22]. Although
166 much is known about the environmental factors that can prolong sleep latency, we know
167 little about the genetic influences on sleep latency. In this study, we performed a meta-
168 analysis of unpublished genome-wide association studies (GWAS) on sleep latency in
169 order to elucidate genetic associations with this trait.
170

171 **MATERIALS AND METHODS**:

172 *Study populations (Stage 1/ Discovery cohorts)*

173 We meta-analyzed data from seven GWAS comprising of 4 242 subjects with European
174 ancestry (Tables S1 & S2). The participating cohorts in the gene discovery phase
175 included the Erasmus Rucphen Family (ERF), Estonian Genome Center
176 (EGP/EGCUT), CROATIA-Korcula, the Micro-isolates in South Tyrol Study
177 (MICROS), Cooperative health research in the Region of Augsburg (KORA), the
178 Netherlands Study of Depression and Anxiety (NESDA) and the Orkney Complex
179 Disease Study (ORCADES) (Tables S1 & S2). A detailed description of the included
180 studies is provided in the supplemental text. All studies in the discovery cohort used the
181 Munich Chronotype Questionnaire (MCTQ)[23] to assess sleep latency. Subjects were

asked to report how long they take to fall asleep on free and workdays. Free days sleep latency was used in the analyses of those cohorts, where a person's sleep pattern is not influenced by professional duties [24] (Figure S1). The question from the MCTQ used to assess sleep latency (in minutes): "I need … minutes to fall asleep". Sleep duration was calculated by subtracting sleep onset from sleep end. Mid-sleep was calculated as the midpoint between sleep onset and waking on free days [25]. The quality control was centralized and the inclusion criteria were: (i) no use of an alarm clock on free days; (ii) no shift-work during the last three months; and (iii) no use of sleep medication (benzodiazepines and other pharmacological agents that influence sleep; see Table S3). Informed consent was obtained from all study participants and an appropriate local committee approved the study protocols.

Study populations (Stage 2/ Replication cohorts)

The replication stage included 12 independent cohorts (see supplementary text; Figure S1). The descriptive statistics are provided in Table S1. Inclusion criteria were: 1) sleep assessment with any available tool, and 2) No use of sleep medication (Figure S1). Five replication cohorts (MrOS, RS-I, RS-II, RS-III and SOF) used the Pittsburg Sleep Quality Index (PSQI) to assess sleep (Table S2), which uses only a single measure for sleep assessment and does not distinguish between sleep on free and working days.

Genotyping & Imputation:

Both discovery and replication cohorts were genotyped on a variety of platforms (Affymetrix 250K, Illumina 318K, Illumina 370K, Illumina 610k; Perlegen 600K; Affymetrix 1000K). Quality control was done in each group separately. The overall

criteria were to exclude individuals with low call rate, excess heterozygosity, and gender mismatch, and exclude variants that were out of Hardy-Weinberg equilibrium, had low minor allele frequency (MAF) or low call rate (Table S2). In EGCUT1 study where the genome-wide data was not available, the 2 most significant SNPs for which a TaqMan assay was available were genotyped (Table S4). Imputations of non-genotyped SNPs in the discovery cohorts were carried out within each study using either MACH [26,27] or IMPUTE [28,29], and HapMap CEU v21a or v22 as reference (Table S2). Genetic imputations in the replication cohorts were performed using MACH, IMPUTE, minimac or BimBam (Table S2). Of the three SNPS rs9907432 was genotyped in most replication cohorts while rs9900428 was imputed in all replication cohorts (Table S4). The data are available in the GWAS Central database, under the accession number HGVST1836 (http://www.gwascentral.org/study/HGVST1836).

Methylation & Expression

The Infinium Human Methylation 450 array (Illumina, San Diego, CA) was used to quantify genome-wide DNA methylation for 748 samples from the Rotterdam study (RS-III) covering 485 577 CpG sites in the genome. Bisulfite conversion followed by amplification, hybridization and imaging were performed according to standard protocols. Illumina GenomeStudio software was used to estimate β score from intensity. Quality-control (QC) steps included removal of signal probes with a detection of P-value > 0.01; and exclusion of probes with low intensity. Samples with a call rate < 99% were removed from the dataset. SWAN package for R software was used to normalize remaining CpG sites and correct for batch effects.

RNA from the same samples was obtained (PAXgene) and hybridized to Illumina HumanHT-12 arrays. Raw probe intensity was obtained using BeadStudio. Gene expression data was quantile-normalized to the median distribution, and subsequently log2 transformed. The probe and sample means were centered to zero. Probes that had a detection p-value < 0.05 in > 10% of the samples were removed from the analysis. The final analysis included 21 328 probes, which were significantly expressed in blood.

Statistical analysis

Heritability Analysis

Heritability analysis of sleep latency was performed in the family based ERF cohort using Sequential Oligogenic Linkage Analysis Routines (SOLAR) [30]. SOLAR uses likelihood ratio tests to evaluate heritability by comparing a purely polygenic model with a sporadic model in the case of testing heritability. The ERF cohort forms one large family, which consists of more than 23 000 members spanning over 23 generations. Since this uniquely large family is too large to be analyzed with SOLAR, we cut it into smaller pedigrees (3 to 5 generations) using the Pedcut software [31] for the heritability analysis. A natural log transformation was applied to the trait before estimating the heritability. The heritability was estimated with the 'polygenic screen' option and using age and sex as covariates in the model. The estimate was then compared to the heritability estimate derived from the polygenic analysis of GenABEL that uses genome-wide genotype data instead of the pedigree to estimate heritability [32].

Genome-wide Association Analysis

Individual GWAS was performed using linear regression (under additive model), natural log of sleep latency as the dependent variable, SNP allele dosage as predictor and age and sex as covariates. CROATIA-Korcula and ORCADES additionally used first three principal components as covariates in the association model. The association analyses were conducted in ProbABEL [32-34] or SNPTEST [35]. For cohorts with related individuals (ERF, MICROS, CROATIA-Korcula, ORCADES), a linear mixed model in ProbABEL using the "mmscore" option was used to account for familial relationships. The mmscore option performs the score test that uses the inverse variance-covariance object estimated from the genetic data and returned from the "polygenic" function of GenABEL [32] to correct for familial relationships. This is a slightly modified FASTA method developed by Abecasis et al. [36].

Since all three SNPs show no heterogeneity across the cohorts (*p-values* > 0.7) (Table 1), a fixed effects meta-analysis was conducted using the inverse variance weighted method as implemented in METAL (http://www.sph.umich.edu/csg/abecasis/metal/). All SNPs that had a MAF < 0.01 and low imputation quality (Rsq/proper_info < 0.3) were dropped from the meta-analysis. Genomic control correction was also applied to all cohorts prior to the meta-analysis.

For the SNPs that approached genome-wide significance (*p-value* < $5x10^{-08}$) we performed replication in up to 12 independent cohorts using the same model as in the initial GWAS. In the replication cohorts, among the family based studies, ERF_ext used SOLAR, FHS used LMEKIN package of R software (http://www.r-project.org/), CROATIA-Split used the "mmscore" option in ProbABEL while QIMR used MERLIN to account for family structure. Other population based cohorts used SPSS, PLINK

(http://pngu.mgh.harvard.edu/purcell/plink/) [37] or R software to perform association analysis (Table S2). Meta-analysis was performed using square-root of the sample size as weights [38].

Methylome-wide association analysis

Top SNPs from the meta-analysis of GWAS were tested for association with genome-wide methylation (CpG) sites by performing linear regression analyses of methylation sites on each of the three SNPs while adjusting for age, sex, technical covariates including batch effects and blood cell counts (granulocytes, lymphocytes, monocytes, erythrocytes and platelets) in R software.

Gene functional prediction and network analysis

Prediction of gene function can be conducted using a guilt-by-association approach: e.g., if there are 100 genes that are known to be involved in apoptosis, identification of a gene that is strongly co-expressed with these 100 genes suggests that this gene is likely to be involved in apoptosis as well. As such co-expression data can be used to predict likely functions for genes. However, important to realize is that some phenomena exert very strong transcriptomic effects and therefore will overshadow more subtle effects. In order to be able to identify such subtle relationships as well, we conducted a principal component analysis on an unprecedented scale [39]: We collected gene expression data for three different species (homo sapiens, mus musculus and rattus norvegicus) from the Gene Expression Omnibus. We confined analyses to four different Affymetrix expression platforms (Affymetrix Human Genome U133A Array, Affymetrix Human Genome U133 Plus 2.0 Array, Affymetrix Mouse Genome 430 2.0

Array and Affymetrix Rat Genome 230 2.0 Array). For each of these platforms we downloaded the raw CEL files (20 108, 43 278, 18 639 and 6 124 arrays, respectively), and used RMA for normalization. We could run RMA on all samples at once for the 20 108 Human Genome U133A Array, 18 639 Mouse Genome 430 2.0 Array and 6 123 Rat Genome 230 2.0 Array. For the 43 278 Human Genome U133 Plus 2.0 Array samples we ran RMA in eight batches due to its size, by randomly assigning the samples to one of these batches. We subsequently conducted quality control (QC) on the data. We first removed duplicate samples, and then conducted a principal component analysis (PCA) on the sample correlation matrix. The first principal component (PCqc) on such a matrix describes nearly always a constant pattern (dominating the data), which explains around 80-90% of the total variance [40,41]. This pattern can be regarded as probe-specific variance, independent of the biological sample hybridized to the array. The correlation of each individual microarray with this PCqc can be used to detect outliers, as arrays of lesser quality will have a lower correlation with the PCqc. We removed samples that had a correlation $R < 0.75$. After QC in total 77 840 different samples remained for downstream analysis (54 736 human samples, 17 081 mouse samples, 6 023 rat samples). Although this QCed dataset can be well used for the aforementioned guilt-by-association co-expression analysis, we reasoned that the presence of profound effects on many genes will make it difficult to identify the more subtle relationships that exist between genes. Therefore, we conducted a PCA on the probe correlation matrix, resulting in the identification of in total 2 206 robustly estimated principal components (377 for Human Genome U133A, 777 for Human Genome U133 Plus 2.0, 677 for Mouse Genome 430 2.0 and 375 for Rat Genome 230 2.0) by requiring a Cronbach's alpha > 0.70 for each individual principal component.

Jointly these components explain between 79% and 90% of the variance in the data per Affymetrix expression platform, and many of these are well conserved across the three species.

Subsequent Gene Set Enrichment Analysis (GSEA) revealed that each of these 2 206 components are significantly enriched (False discovery rate < 0.05) for at least one GO term, KEGG, BioCarta or Reactome pathway, indicating that these components are describing biologically relevant but often diverse phenomena. While per species the very first components describe profound effects on expression (i.e. many enriched pathways and GO terms), the other components are potentially equally biologically relevant, as each of the components describe certain biological phenomena. We therefore used the individual components and integrated the different platforms and species by collapsing the probe identifiers to human Ensembl genes and used orthology information from Ensembl for the mouse and rat platform, resulting in a harmonized matrix of 19 997 unique Ensembl genes x 2 206 principal components.

We subsequently predicted the most likely Gene Ontology (GO) biological process using the following strategy: (i) we first ascertained each individual GO term and assessed per PC whether the genes that were explicitly annotated with this GO term showed a significant difference from the genes that were not annotated with this GO term using a T-Test; (ii) we converted the resulting P-Value into an 'enrichment' Z-Score (to ensure normality); and (iii) we subsequently investigated *RBFOX3* and correlated the 2 206 PC eigenvector coefficients of *RBFOX3* with each GO term by taking the 2 206 'enrichment' Z-Scores as the expression profile for that GO term. A significant positive correlation means *RBFOX3* has an expression profile that is comparable to the GO term. We have visualized this method at

www.genenetwork.nl/genenetwork (click on "Method"). In order to correct for multiple testing, we permuted Ensembl gene identifiers: using permuted data we redid the 'enrichment' Z-score calculation and investigated how strong *RBFOX3* correlated with permuted pathway. We repeated this analysis 100 times, allowing us to determine which of these predictions were significant (controlling to the false discovery rate of 0.05). We used the same procedure to predict in which BioCarta and Reactome pathways *RBFOX3* is involved.

RESULTS

A basic description of the study populations is given in Table S1. The heritability of sleep latency using the pedigree data of the Erasmus Rucphen Family (ERF) study was estimated to be 0.18 (SE = 0.112, *p-value* = 0.05) (see methods section), which is consistent with earlier findings of heritability for this trait [21] and also comparable to the heritability estimate (h^2 = 0.20) derived from the genome-wide genotype data (see methods section for details) of the same cohort. The quantile-quantile plot of the meta-analysis shows no inflation of the chi-square statistic since the genomic control inflation factor (λ) is 1.01 (Figure S2). The meta-analysis of GWAS produced a cluster of three borderline genome-wide significant SNPs (Table 1; Figure 1, and Figure S3) on chromosome 17q25. The highest-ranking SNP rs9900428 hg18.chr17:g.74651323G>A) (*p-value* = $5.7*10^{-08}$) had a minor-allele frequency (MAF) of 0.20-0.33, with the imputation quality ranging from 0.85 to 0.99 across all cohorts. The effect size (β = -0.094) and the direction of the risk were consistent across all cohorts (Table 1). The other two SNPs rs9907432 (hg18.chr17:g.74651967:G>A) (*p-value* = $6.5*10^{-08}$) and rs7211029 (hg18.chr17:g.74652903C>T) (*p-value* = $9.1*10^{-08}$) were in linkage

disequilibrium (LD; r^2 = 0.95) with rs9900428 (Figure 1, Figure S4). Figure 2 shows the mean sleep latency per genotype category for all the discovery cohorts for the three SNPs unadjusted for age and gender. Compared to the homozygote carriers of the reference allele (pooled average sleep latency = 16.4 min), the homozygote carriers (pooled average sleep latency = 13.5 minutes) of the minor allele are estimated to have a sleep latency of on average approximately three minutes less (a reduction of 18%) and the heterozygote carriers (pooled average sleep latency = 15.5 minutes) about 1 minute less (a reduction of about 6%) (Figure 2).

We attempted to replicate these three SNPs in up to 12 independent cohorts (Tables S2 & S4). SNP rs9900428 was available *in-silico* in 11 cohorts, while rs9907432 and rs7211029 were available in all 12 cohorts. The replication analyses showed consistency in the direction of the effect across most replication cohorts (Figure 3). The meta-analysis of the replication cohorts yielded significant evidence of association of rs9900428 (*p-value* = $1.5*10^{-02}$), rs9907432 (*p-value* = $7.1*10^{-03}$) and rs7211029 (*p-value* = $2.5*10^{-03}$) with sleep latency (Table 2). The three SNPs are intronic to the RNA binding protein fox-1 homolog 3 (*RBFOX3* also known as *HRNBP3*; Figure 1) and lie in the region with active regulatory elements (ENCODE) containing the H3K27ac mark and methylation marks. H3K27ac is an important enhancer mark that can distinguish between active and poised enhancer elements. Such enhancer elements are known to affect the expression of proximal genes and cluster near the genes they regulate [42]. We investigated the 3 SNPs further using the methylome data of RS. The 3 SNPs were methylome-wide significantly associated (p-value = $8.1*10^{-9}$, FDR = 0.004) with the CpG site: cg16185152 in the RBFOX3 gene. We attempted to investigate the effect of methylation on gene expression, there was,

however, only one probe for the RBFOX3 gene in the HumanHT-12_V4_Illumina 450K RNA expression array, which was removed in the quality control.

A gene-network analysis of *RBFOX3* using gene network tool (http://genenetwork.nl:8080/GeneNetwork/) shows strongest co-expression with the hippocalcin gene *HPCA* followed by *SNCB, CABP1, JPH3, CPLX2, GABRA6, GABRD, NRXN3, RBFOX1, RTN4R, CNTN2* and *WSCD2* among others (Figure S5). A functional prediction of the gene showed involvement in the biological processes of synaptic functioning, membrane depolarization, gamma-aminobutyric acid (GABA) signaling and nervous system development (Tables S5 & S6) and in dopamine, serotonin and glutamate neurotransmitters release cycle (Table S7). *RBFOX3* is expressed most strongly in various parts of the brain (brain, *p-value* = $1.4*10^{-316}$; central nervous system, *p-value* = $7.6*10^{-321}$; cerebral cortex, *p-value* = $1.0*10^{-174}$) including the hypothalamus (*p-value* = $9.7*10^{-10}$, AUC = 0.96) (Table S8), the locale of the SCN and a central part of the circadian clock.

None of the three SNPs showed strong association with sleep duration or mid-sleep, which is an indicator of the chronotype of an individual, Table S9.

DISCUSSION

In this first large-scale genome-wide association study of sleep latency, we report the association of novel variants located in the gene *RBFOX3* with sleep latency. Our gene discovery phase consisted of a sample of 4 242 individuals from seven European populations, where sleep latency was assessed according to a common protocol. With this sample we discovered a cluster of three borderline genome-wide significant SNPs that were intronic to the gene *RBFOX3*. The replication of the three SNPs in up to 30

377 individuals from 12 independent cohorts showed significant association of the three variants and consistency in the direction of the effect estimates across most cohorts. A functional prediction of *RBFOX3* based on a gene network analysis suggests significant involvement in the release cycle of neurotransmitters including GABA and various monoamines that are core to the human circadian clock, thus supporting the involvement of *RBFOX3* with sleep latency.

The *RBFOX3* gene (also called *Fox-3,Hrnbp3, Neun*) is located on the long arm of chromosome 17 (17q25). It belongs to the Fox-1 family of genes and shows high homology to *RBFOX1* (also called *Fox-1, A2bp1, Hrnbp1*) and *RBFOX2* (also called *Fox-2, Rbm9, Hrnbp2*). *RBFOX3* is a relatively new member of this family and was recently identified to code for neuronal nuclei (NeuN) protein [43]. The Fox proteins are a highly conserved family of tissue specific splicing regulators [44]. While *RBFOX1* is expressed in neurons, muscles and heart, and *RBFOX2* in ovary, whole embryo, neurons and muscles [45,46], the expression of *RBFOX3* was detected exclusively in the post-mitotic regions of embryonic mouse brain [44,47]. *RBFOX3* is believed to play a role in neuron-specific alternative splicing [43]. Alternative splicing occurs most frequently in the brain [48,49], presumably to generate large numbers of neuronal cell types and to support their diverse functions [43]. A recent analysis of rodent SCN (the biological clock) anatomy using antibody against NeuN protein shows that *RBFOX3* has a unique distribution which is limited to a particular sector of the SCN [50]. Mutations in the *FOX-1* genes lead to severe neuro-developmental phenotypes exhibiting mental retardation, epilepsy, and autism spectrum disorder [51-54]. Further a very recent study on patients with developmental delay detected a translocation disrupting the intron 2 of the *RBFOX3* gene [55]. Interestingly this patient additionally had sleeping difficulties.

Gene-network analysis of *RBFOX3* showed strong co-expression with genes involved in calcium channel activity and GABA signaling. GABA-ergic sleep neurons of the ventro-lateral pre-optic nucleus are activated by the circadian clock and adenosine, which progressively accumulates in the brain during waking. In turn these sleep-active neurons begin to inhibit the wake-active neurons via the neurotransmitter GABA [56]. *GABA-A* receptors are the site of action of a number of important pharmacologic agents including barbiturates, benzodiazepines (sleep inducing drugs), and ethanol [57,58]. Polymorphisms in *GABA-A* receptors have been associated with insomnia [59]. Functional prediction based on gene-network analysis shows significant involvement of *RBFOX3* in the release cycle of various neurotransmitters including dopamine, serotonin and glutamate, *GABA-A* receptor activation and Ras-activation upon Ca^{2+} influx through the NMDA receptor. Photic information is communicated to the molecular clockworks by release of glutamate from retino-hypothalamic nerve terminals and stimulation of glutamate receptors on SCN neurons [60]. Glutamate stimulation is followed by intra-cellular increases of cyclic adenosine monophosphate and Ca^{2+} and activation of a Ras-dependent signal cascade in the circadian clockwork. Non-photic signaling to the core subdivision of the SCN is conveyed through two major pathways including GABA-containing neurons derived from the thalamus and serotonin or 5-hydroxytryptamine **(**5-HT)–containing neurons derived from the midbrain. Serotonergic input to the SCN shifts the timing of the clock [60].

Our study shows strong evidence of association of sleep latency with *RBFOX3*. The fact that the association signal was stronger in the discovery sample compared to the replication sample, even though the replication sample was six-fold larger, may be explained by the differences in phenotyping as well as by different cohort characteristics

(e.g., age). Notably, all discovery cohorts were European, whereas the replication cohorts were drawn additionally from North America and Asia; it is possible that socio-cultural factors may have influenced sleep latency or how sleep latency was reported across cohorts. For instance, the phenotyping and quality control for the discovery cohorts was synchronized and centralized; all cohorts were assessed with Munich Chronotype Questionnaire (MCTQ), which, unlike other such instruments, assesses information separately for free days and working days. Our GWAS was based on sleep latency on free days, as sleep latency on workdays is heavily influenced by daily professional activities. Unfortunately, this distinction was not available for most of the replication cohorts. Moreover, exclusions in the GWAS discovery cohorts were based not only on sleep medication use but also on other drugs that are known to influence sleep. Sleep medication or for that matter any medication that has a sleep inducing effect reduces sleep latency [24], thereby introducing a potential bias on genetic studies of sleep latency. We also removed shift-workers and those using alarm clocks from the analysis in the discovery phase, which was not done in most replication cohorts. All of these factors likely affected the results in the replication phase. Nevertheless, despite a small effect size, the direction of association signal was consistent in most replication cohorts indicating the robustness of our finding. Furthermore, gene functional prediction and network analysis support the association of *RBFOX3* variants with sleep latency. The predicted functioning of *RBFOX3*, including neurotransmitters' release cycle and GABA-receptor activation strongly implicates a chronobiological explanation. However, further tests of association with various sleep disorders and functional analyses will provide a better insight into the relationship between *RBFOX3* and sleep.

Acknowledgements:

We would like to thank all the participants and staff of all the studies for their co-operation and contribution. We are grateful to all participants and their relatives, general practitioners and neurologists for their contributions and to P. Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and P. Snijders for his help in data collection for the Erasmus Rucphen Family study. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters for their help in creating the GWAS database, and Karol Estrada and Maksim V. Struchalin for their support in creation and analysis of imputed data for the Rotterdam Study. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. We would like to thank Dr. Fernando Rivadeneira, Dr. Tobias A. Knoch, Anis Abuseiris, Luc V. de Zeeuw, and Rob de Graaf for their help in creating GRIMP, and BigGRID, MediGRID, and Services@MediGRID/D-Grid, We would like to acknowledge the invaluable contributions of the recruitment team from the Croatian Centre for Global Health, University of Split, the administrative teams in Croatia and Edinburgh and the people of Split.

Our work was supported by the FP6 programme EUCLOCK, the Dutch Science Foundation (the NWO), the Hersenstichting Nederland, the Rosalind Franklin Fellowships of the University of Groningen, targeted Financing from the Estonian Government, the European Union through the European Regional Development Fund in the frame of Centre of Excellence in Genomics, FP7 Projects ECOGENE, BBMRI, ENGAGE, and OPENGENE, the Geestkracht programme of the Dutch Scientific

Organization (ZON-MW) and matching funds from participating universities and mental health care organizations, the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, the EUROSPAN (European Special Populations Research Network) through the European Commission FP6 STRP grant, the Chief Scientist Office of the Scottish Government, the Royal Society , the Medical Research Council, Erasmus University, Erasmus MC, the Centre for Medical Systems Biology (CMSB1 and CMSB2) and the Netherlands Genomics Initiative (NGI), and ALBAN. We acknowledge the High Performance Computing Center of University of Tartu, for space and facilities to conduct part of the EGCUT analysis. We also acknowledge the Netherlands Organization for Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), The Ministry of Education, Culture and Science, the Ministry of Health, Welfare and Sports of the Netherlands, the European Commission (DG XII), and the Municipality of Rotterdam. This study was supported by grants from the Netherlands Organization for Scientific Research (NWO-VENI grant 916.10.135 to L.F.) and a Horizon Breakthrough grant from the Netherlands Genomics Initiative (grant 92519031 to L.F.). The research leading to these results has received funding from the European Community's Health Seventh Framework Programme (FP7/2007-2013) under grant agreement 259867.

Study specific acknowledgments and the ethical statement are provided in the supplemental text.

Conflict of Interest

Dr. Najaf Amin is supported by the Netherlands Brain Foundation (project number F2013(1)-28). Prof.dr. Gregory J. Tranah was supported by NIA grant R01AG030474. Prof.dr. Henning Tiemeier was supported by the Vidi Grant of ZonMw (the Netherlands Organization for Health Research and Development, 2009-017.106.370). All other authors report no biomedical financial interests or potential conflicts of interest.

Supplementary information is available at European Journal of Human Genetics website (http://www.nature.com/ejhg).

References

1 Krauchi K, Wirz-Justice A. Circadian clues to sleep onset mechanisms. *Neuropsychopharmacology* 2001; **25**: S92-96.

2 Cirelli C. The genetic and molecular regulation of sleep: from fruit flies to humans. *Nat Rev Neurosci* 2009; **10**: 549-560.

3 Hastings M. The brain, circadian rhythms, and clock genes. *BMJ* 1998; **317**: 1704-1707.

4 Hastings MH. Central clocking. *Trends Neurosci* 1997; **20**: 459-464.

5 Baekeland F, Hoy P. Reported vs recorded sleep characteristics. *Arch Gen Psychiatry* 1971; **24**: 548-551.

6 Hoch CC, Reynolds CF, 3rd, Kupfer DJ, Berman SR, Houck PR, Stack JA. Empirical note: self-report versus recorded sleep in healthy seniors. *Psychophysiology* 1987; **24**: 293-299.

7 Lehnkering H, Siegmund R. Influence of chronotype, season, and sex of subject on sleep behavior of young adults. *Chronobiol Int* 2007; **24**: 875-888.

8 Geisler P, Tracik F, Cronlein T *et al*. The influence of age and sex on sleep latency in the MSLT-30--a normative study. *Sleep* 2006; **29**: 687-692.

9 Krishnan V, Collop NA. Gender differences in sleep disorders. *Curr Opin Pulm Med* 2006; **12**: 383-389.

10 Bonnet MH, Webb WB. The return to sleep. *Biol Psychol* 1979; **8**: 225-233.

11 Afaghi A, O'Connor H, Chow CM. High-glycemic-index carbohydrate meals shorten sleep onset. *Am J Clin Nutr* 2007; **85**: 426-430.

12 Nixon GM, Thompson JM, Han DY *et al*. Falling asleep: the determinants of sleep latency. *Arch Dis Child* 2009; **94**: 686-689.

13 Gupta R, Dahiya S, Bhatia MS. Effect of depression on sleep: Qualitative or quantitative? *Indian J Psychiatry* 2009; **51**: 117-121.

14 Blader JC, Koplewicz HS, Abikoff H, Foley C. Sleep problems of elementary school children. A community survey. *Arch Pediatr Adolesc Med* 1997; **151**: 473-480.

15 Eddy M, Walbroehl GS. Insomnia. *Am Fam Physician* 1999; **59**: 1911-1916, 1918.

16 Weitzman ED, Czeisler CA, Coleman RM *et al*. Delayed sleep phase syndrome. A chronobiological disorder with sleep-onset insomnia. *Arch Gen Psychiatry* 1981; **38**: 737-746.

17 Freedman RR. EEG power spectra in sleep-onset insomnia. *Electroencephalogr Clin Neurophysiol* 1986; **63**: 408-413.

18 van den Heuvel CJ, Lushington K. Chronobiology and insomnia: pathophysiology and treatment of circadian rhythm sleep disorders. *Expert Rev Neurother* 2002; **2**: 249-260.

19 Touchette E, Petit D, Seguin JR, Boivin M, Tremblay RE, Montplaisir JY. Associations between sleep duration patterns and behavioral/cognitive functioning at school entry. *Sleep* 2007; **30**: 1213-1219.

20 Heath AC, Kendler KS, Eaves LJ, Martin NG. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. *Sleep* 1990; **13**: 318-335.

21 Klei L, Reitz P, Miller M *et al*. Heritability of morningness-eveningness and self-report sleep measures in a family-based sample of 521 hutterites. *Chronobiol Int* 2005; **22**: 1041-1054.

22 Partinen M, Kaprio J, Koskenvuo M, Putkonen P, Langinvainio H. Genetic and environmental determination of human sleep. *Sleep* 1983; **6**: 179-185.

598 23 Kantermann T, Juda M, Merrow M, Roenneberg T. The human circadian clock's
599 seasonal adjustment is disrupted by daylight saving time. *Curr Biol* 2007; **17**: 1996-
600 2000.
601 24 DeMartinis NA, Winokur A. Effects of psychiatric medications on sleep and
602 sleep disorders. *CNS Neurol Disord Drug Targets* 2007; **6**: 17-29.
603 25 Wittmann M, Dinich J, Merrow M, Roenneberg T. Social jetlag: misalignment
604 of biological and social time. *Chronobiol Int* 2006; **23**: 497-509.
605 26 Li Y, Abecasis GR. Mach 1.0: Rapid Haplotype Reconstruction and Missing
606 Genotype Inference. *Am J Hum Genet* 2006; **S79**: 2290.
607 27 Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annu Rev Genomics*
608 *Hum Genet* 2009; **10**: 387-406.
609 28 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation
610 method for the next generation of genome-wide association studies. *PLoS Genet* 2009;
611 **5**: e1000529.
612 29 Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint
613 method for genome-wide association studies by imputation of genotypes. *Nat Genet*
614 2007; **39**: 906-913.
615 30 Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general
616 pedigrees. *Am J Hum Genet* 1998; **62**: 1198-1211.
617 31 Liu F, Kirichenko A, Axenovich TI, van Duijn CM, Aulchenko YS. An
618 approach for cutting large and complex pedigrees for linkage analysis. *Eur J Hum Genet*
619 2008; **16**: 854-860.
620 32 Aulchenko YS, Ripke S, Isaacs A, van Duijn CM. GenABEL: an R library for
621 genome-wide association analysis. *Bioinformatics* 2007; **23**: 1294-1296.

622 33 Amin N, van Duijn CM, Aulchenko YS. A genomic background based method
623 for association analysis in related individuals. *PLoS One* 2007; **2**: e1274.
624 34 Aulchenko YS, Struchalin MV, van Duijn CM. ProbABEL package for genome-
625 wide association analysis of imputed data. *BMC Bioinformatics* 2010; **11**: 134.
626 35 Wellcome Trust Case Control C. Genome-wide association study of 14,000
627 cases of seven common diseases and 3,000 shared controls. *Nature* 2007; **447**: 661-678.
628 36 Chen WM, Abecasis GR. Family-based association tests for genomewide
629 association scans. *Am J Hum Genet* 2007; **81**: 913-926.
630 37 Orwoll E, Blank JB, Barrett-Connor E *et al*. Design and baseline characteristics
631 of the osteoporotic fractures in men (MrOS) study--a large observational study of the
632 determinants of fracture in older men. *Contemp Clin Trials* 2005; **26**: 569-585.
633 38 Zaykin DV. Optimally weighted Z-test is a powerful method for combining
634 probabilities in meta-analysis. *J Evol Biol* 2011; **24**: 1836-1841.
635 39 Fehrmann RS, Karjalainen JM, Krajewska M *et al*. Gene expression analysis
636 identifies global gene dosage sensitivity in cancer. *Nat Genet* 2015; **47**: 115-125.
637 40 Alter O, Brown PO, Botstein D. Singular value decomposition for genome-wide
638 expression data processing and modeling. *Proc Natl Acad Sci U S A* 2000; **97**: 10101-
639 10106.
640 41 Sherlock G. Analysis of large-scale gene expression data. *Brief Bioinform* 2001;
641 **2**: 350-362.
642 42 Creyghton MP, Cheng AW, Welstead GG *et al*. Histone H3K27ac separates
643 active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S*
644 *A* 2010; **107**: 21931-21936.

645 43 Kim KK, Adelstein RS, Kawamoto S. Identification of neuronal nuclei (NeuN)
646 as Fox-3, a new member of the Fox-1 gene family of splicing factors. *J Biol Chem*
647 2009; **284**: 31052-31061.
648 44 Dredge BK, Jensen KB. NeuN/Rbfox3 nuclear and cytoplasmic isoforms
649 differentially regulate alternative splicing and nonsense-mediated decay of Rbfox2.
650 *PLoS One* 2011; **6**: e21585.
651 45 Kiehl TR, Shibata H, Vo T, Huynh DP, Pulst SM. Identification and expression
652 of a mouse ortholog of A2BP1. *Mamm Genome* 2001; **12**: 595-601.
653 46 Underwood JG, Boutz PL, Dougherty JD, Stoilov P, Black DL. Homologues of
654 the Caenorhabditis elegans Fox-1 protein are neuronal splicing regulators in mammals.
655 *Mol Cell Biol* 2005; **25**: 10005-10016.
656 47 McKee AE, Minet E, Stern C, Riahi S, Stiles CD, Silver PA. A genome-wide in
657 situ hybridization map of RNA-binding proteins reveals anatomically restricted
658 expression in the developing mouse brain. *BMC Dev Biol* 2005; **5**: 14.
659 48 Modrek B, Resch A, Grasso C, Lee C. Genome-wide detection of alternative
660 splicing in expressed sequences of human genes. *Nucleic Acids Res* 2001; **29**: 2850-
661 2859.
662 49 Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across
663 human tissues. *Genome Biol* 2004; **5**: R74.
664 50 Morin LP, Hefton S, Studholme KM. Neurons identified by NeuN/Fox-3
665 immunoreactivity have a novel distribution in the hamster and mouse suprachiasmatic
666 nucleus. *Brain Res* 2011; **1421**: 44-51.

667 51 Bhalla K, Phillips HA, Crawford J *et al*. The de novo chromosome 16
668 translocations of two patients with abnormal phenotypes (mental retardation and
669 epilepsy) disrupt the A2BP1 gene. *J Hum Genet* 2004; **49**: 308-311.
670 52 Barnby G, Abbott A, Sykes N *et al*. Candidate-gene screening and association
671 analysis at the autism-susceptibility locus on chromosome 16p: evidence of association
672 at GRIN2A and ABAT. *Am J Hum Genet* 2005; **76**: 950-966.
673 53 Martin CL, Duvall JA, Ilkin Y *et al*. Cytogenetic and molecular characterization
674 of A2BP1/FOX1 as a candidate gene for autism. *Am J Med Genet B Neuropsychiatr*
675 *Genet* 2007; **144B**: 869-876.
676 54 Sebat J, Lakshmi B, Malhotra D *et al*. Strong association of de novo copy
677 number mutations with autism. *Science* 2007; **316**: 445-449.
678 55 Utami KH, Hillmer AM, Aksoy I *et al*. Detection of chromosomal breakpoints in
679 patients with developmental delay and speech disorders. *PLoS One* 2014; **9**: e90852.
680 56 Luppi PH. Neurochemical aspects of sleep regulation with specific focus on
681 slow-wave sleep. *World J Biol Psychiatry* 2010; **11 Suppl 1**: 4-8.
682 57 Whiting PJ. The GABA-A receptor gene family: new targets for therapeutic
683 intervention. *Neurochem Int* 1999; **34**: 387-390.
684 58 Whiting PJ, Bonnert TP, McKernan RM *et al*. Molecular and functional
685 diversity of the expanding GABA-A receptor gene family. *Ann N Y Acad Sci* 1999; **868**:
686 645-653.
687 59 Buhr A, Bianchi MT, Baur R *et al*. Functional characterization of the new
688 human GABA(A) receptor mutation beta3(R192H). *Hum Genet* 2002; **111**: 154-160.
689 60 Charney DS, Nestler EJ. Neurobiology of mental illness. Oxford University
690 Press: Oxford; New York, 2004.
691

692 **Titles and legends to figures**

693 **Figure 1**

694 **Regional association plot of sleep latency for the region 17q25**. The plot was
695 constructed using Locus Zoom (http://csg.sph.umich.edu/locuszoom/). The most
696 significant SNP is depicted as a diamond and other SNPs in the region are depicted by
697 circles. Various colors represent the extent of linkage disequilibrium with most
698 significant SNP. The X-axis gives the position in mega bases and the Y-axis shows the
699 negative logarithm of the p-values from the meta-analysis.

700 **Figure 2**

701 **Means per genotype category for all the discovery cohorts for the three most**
702 **significant SNPs**. Black bars represent the mean sleep latency for the homozygous
703 carriers of the effect allele. Y-axis depicts the sleep latency time in minutes. Panels A, B
704 and C represent the effects of the SNPs in individual cohorts. Panel D represents the
705 pooled average sleep latency per genotype category across all cohorts for the three
706 SNPs (X-axis). Data used to generate this figure was not adjusted for age and sex.

707 **Figure 3**

708 **Forest plot for the three SNPs at chromosome 17q25.** On the left, the populations
709 including the populations in the GWAS and the replication phase. The boxes represent
710 the precision and horizontal lines representing the confidence intervals. Pooled estimate
711 is the effect estimate from the meta-analysis of all cohorts. The horizontal axis
712 represents the scale of effects.

713

714

Legends Supplementary Information

Supplemental text

1. Study Populations (Stage 1/GWAS cohorts)

2. Study Populations (Stage 2/Replication cohorts)

3. Study specific acknowledgements

4. Ethical statement

Table S1

Descriptive Statistics of stage1 (discovery) and stage2 (replication) cohorts.

Table S2

Genotyping and quality control in the stage 1 (discovery) and stage 2 (replication) cohorts.

Table S3

Pharmacological sleep agents. Drug groups and correspondent ATC codes used as exclusion criteria when selecting phenotyped subjects for the independent GWA studies.

Table S4

Summary of replicated SNPs in the replication cohorts.

Table S5

Results of Biological process prediction of *RBFOX3* based on Gene Ontology.

Table S6

Results of pathway analysis of *RBFOX3* based on BioCarta.

Table S7

Results of functional prediction of *RBFOX3* based on Reactome.

Table S8

Expression of *RBFOX3* in various tissues.

Table S9

Association results of the SNPs with sleep duration and mid-sleep/chronotype.

Figure S1

Inclusion criteria for studies in discovery and replication phase.

Figure S2

Quantile-Quantile Plot for the meta-analysis. Horizontal axis shows the expected chi-square distribution while the vertical axis shows the observed chi-square. Red line shows the distribution under the null hypothesis and the black dots represent a single point in the observed chi-square distribution.

Figure S3

Genome-wide association plot of sleep latency. The horizontal axis shows the whole autosomal genome divided into 22 chromosomes. The vertical axis shows the negative principal log of the p-values. Each dot represents a single nucleotide polymorphism.

Figure S4

Linkage disequilibrium (LD) plot for 17q25. The intensity of LD is depicted by the intensity of the color, i.e. the higher the intensity, the greater is the LD.

Figure S5

Gene network of RBFOX3 based on co-expression. The grey lines indicate positive co-expression and the blue lines indicate negative co-expression. The intensity of the color indicates the strength of the co-expression.

Table 1: Top SNPs from the genome-wide association analysis for Sleep latency

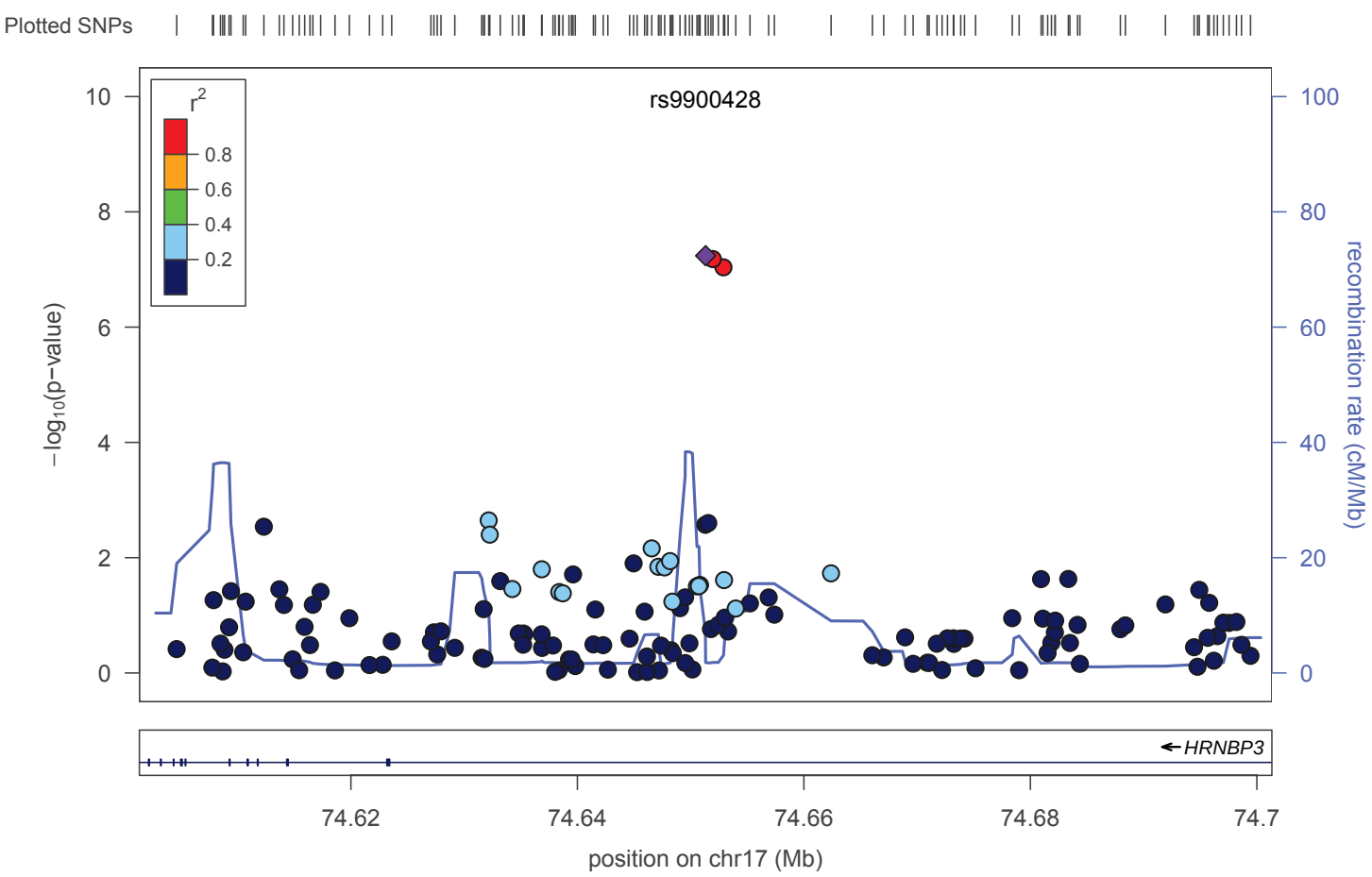
SNP	position(B36)	allele	gene	maf	chrom	cohort	EGP	ERF	KORA	KORCULA	MICROS	NESDA	ORKNEY	Meta-analysis
						N	933	740	548	610	693	540	206	4270
rs9900428	74651323	G>A	RBFOX3	0.20-0.33	17	*β*	-0.106	-0.085	-0.158	-0.133	-0.082	-0.03	-0.122	-0.094
						SE	0.051	0.024	0.057	0.067	0.055	0.064	0.104	0.017
						p-value	0.037	0.00036	0.0057	0.045	0.137	0.632	0.239	$5.77*10^{-08}$
						Rsq	0.97	0.90	0.90	0.97	0.96	0.85	0.99	
						P_{HET}								0.823
rs9907432	74651967	G>A	RBFOX3	0.20-0.34	17	*β*	-0.106	-0.083	-0.153	-0.127	-0.084	-0.0248	-0.1216	-0.092
						SE	0.0501	0.0234	0.0551	0.066	0.0543	0.0616	0.1031	0.017
						p-value	0.0354	0.00038	0.00568	0.053	0.1206	0.6874	0.2382	$6.59*10^{-08}$
						Rsq	1.0	0.94	0.95	1.0	0.99	0.90	1.0	
						P_{HET}								0.814
rs7211029	74652903	C>T	RBFOX3	0.21-0.34	17	*β*	-0.107	-0.083	-0.1499	-0.127	-0.083	-0.0119	-0.1215	-0.091
						SE	0.050	0.0234	0.0538	0.066	0.0544	0.0596	0.1034	0.017
						p-value	0.035	0.00038	0.0054	0.054	0.1249	0.8416	0.2397	$9.17*10^{-08}$
						Rsq	0.97	0.93	1.0	0.98	0.98	0.96	0.99	
						P_{HET}								0.739

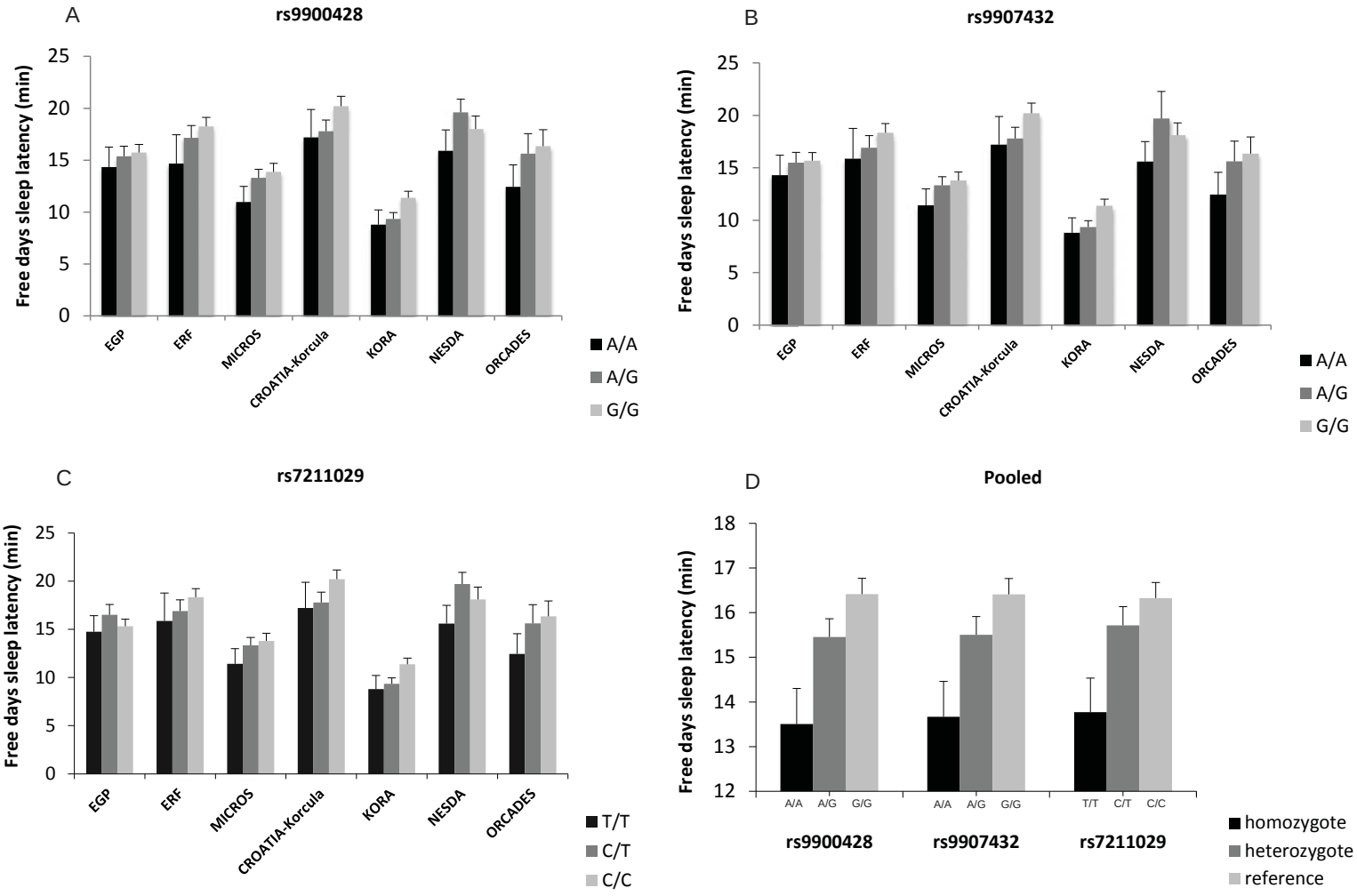
β Effect of the genetic variant
SE Standard error of the effect estimate
$\boldsymbol{P_{HET}}$ *p-value* for heterogeneity
Rsq *imputation quality*

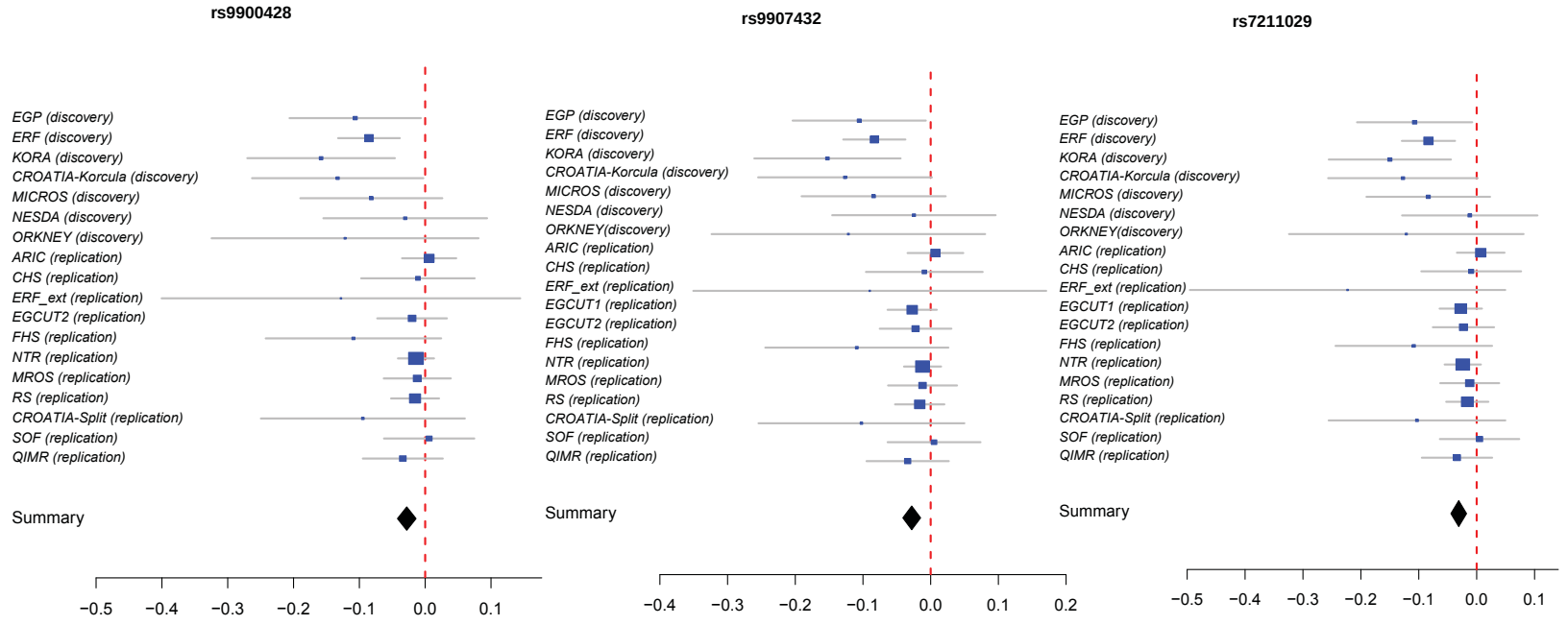
Table 2: Results of the replication of the top SNPs

	cohort	ARIC	CHS	EGCUT1	EGCUT2	ERF_ext	FHS	MROS	NTR	QIMR	RS	SOF	SPLIT	Meta-analysis Replication
SNP	*N*	3583	1533	5925	3540	143	2192	1849	1795	2280	5641	1480	416	30377
rs9900428	*β*	0.006	-0.011	NA	-0.02	-0.128	-0.109	-0.012	-0.014	-0.034	-0.015	0.006	-0.095	-0.026
	SE	0.021	0.044	NA	0.027	0.139	0.068	0.026	0.014	0.031	0.019	0.035	0.079	0.010
	p-value	0.775	0.802	NA	0.461	0.357	0.109	0.644	0.317	0.273	0.408	0.864	0.231	0.015
	P_{HET}													0.919
rs9907432	*β*	0.007	-0.009	-0.027	-0.022	-0.09	-0.109	-0.012	-0.012	-0.034	-0.016	0.005	-0.102	-0.026
	SE	0.021	0.044	0.019	0.027	0.133	0.069	0.026	0.014	0.031	0.0186	0.035	0.078	0.0097
	p-value	0.739	0.831	0.14	0.408	0.499	0.114	0.644	0.391	0.273	0.384	0.886	0.187	0.007
	P_{HET}													0.949
rs7211029	*β*	0.007	-0.009	-0.028	-0.023	-0.223	-0.109	-0.012	-.024	-0.034	-0.016	0.005	-0.103	-0.030
	SE	0.021	0.044	0.019	0.027	0.139	0.069	0.026	0.016	0.031	0.019	0.035	0.078	0.0098
	p-value	0.739	0.831	0.142	0.398	0.109	0.116	0.644	0.133	0.272	0.384	0.886	0.186	0.0025
	P_{HET}													0.792

β Effect of the genetic variant
SE Standard error of the effect estimate
P_{HET} *p-value* for heterogeneity


Plotted SNPs
rs9900428
r2
0.8
0.6
0.4
0.2
−log10(p−value)
recombination rate (cM/Mb)
HRNBP3
74.62
74.64
74.66
74.68
74.7
position on chr17 (Mb)


A
rs9900428
Free days sleep latency (min)
25
20
15
10
5
0
EGP
ERF
MICROS
CROATIA-Korcula
KORA
NESDA
ORCADES
A/A
A/G
G/G
B
rs9907432
Free days sleep latency (min)
25
20
15
10
5
0
EGP
ERF
MICROS
CROATIA-Korcula
KORA
NESDA
ORCADES
A/A
A/G
G/G
C
rs7211029
Free days sleep latency (min)
25
20
15
10
5
0
EGP
ERF
MICROS
CROATIA-Korcula
KORA
NESDA
ORCADES
T/T
C/T
C/C
D
Pooled
Free days sleep latency (min)
18
17
16
15
14
13
12
A/A
A/G
G/G
A/A
A/G
G/G
T/T
C/T
C/C
rs9900428
rs9907432
rs7211029
homozygote
heterozygote
reference


rs9900428
EGP (discovery)
ERF (discovery)
KORA (discovery)
CROATIA-Korcula (discovery)
MICROS (discovery)
NESDA (discovery)
ORKNEY (discovery)
ARIC (replication)
CHS (replication)
ERF_ext (replication)
EGCUT2 (replication)
FHS (replication)
NTR (replication)
MROS (replication)
RS (replication)
CROATIA-Split (replication)
SOF (replication)
QIMR (replication)
Summary
−0.5 −0.4 −0.3 −0.2 −0.1 0.0 0.1
rs9907432
EGP (discovery)
ERF (discovery)
KORA (discovery)
CROATIA-Korcula (discovery)
MICROS (discovery)
NESDA (discovery)
ORKNEY(discovery)
ARIC (replication)
CHS (replication)
ERF_ext (replication)
EGCUT1 (replication)
EGCUT2 (replication)
FHS (replication)
NTR (replication)
MROS (replication)
RS (replication)
CROATIA-Split (replication)
SOF (replication)
QIMR (replication)
Summary
−0.4 −0.3 −0.2 −0.1 0.0 0.1 0.2
rs7211029
EGP (discovery)
ERF (discovery)
KORA (discovery)
CROATIA-Korcula (discovery)
MICROS (discovery)
NESDA (discovery)
ORKNEY(discovery)
ARIC (replication)
CHS (replication)
ERF_ext (replication)
EGCUT1 (replication)
EGCUT2 (replication)
FHS (replication)
NTR (replication)
MROS (replication)
RS (replication)
CROATIA-Split (replication)
SOF (replication)
QIMR (replication)
Summary
−0.5 −0.4 −0.3 −0.2 −0.1 0.0 0.1
Effect estimate