

## **Genome-wide association analyses in >119,000 individuals identifies thirteen morningness and two sleep duration loci**

**Samuel E. Jones<sup>1</sup>, Jessica Tyrrell<sup>1</sup>, Andrew R. Wood<sup>1</sup>, Robin N. Beaumont<sup>1</sup>, Katherine S. Ruth<sup>1</sup>, Marcus A. Tuke<sup>1</sup>, Hanieh Yaghootkar<sup>1</sup>, Youna Hu<sup>2,3</sup>, Maris Teder-Laving<sup>4</sup>, Caroline Hayward<sup>5</sup>, Till Roenneberg<sup>6</sup>, James F. Wilson<sup>5,7</sup>, Fabiola Del Greco<sup>8</sup>, Andrew A. Hicks<sup>8</sup>, Chol Shin<sup>9,10</sup>, Chang-Ho Yun<sup>11</sup>, Seung Ku Lee<sup>12</sup>, Andres Metspalu<sup>4</sup>, Enda M. Byrne<sup>13</sup>, Philip R. Gehrman<sup>14</sup>, Henning Tiemeier<sup>15,16</sup>, Karla V. Allebrandt<sup>6</sup>, Rachel M. Freathy<sup>1</sup>, Anna Murray<sup>1</sup>, David A. Hinds<sup>2</sup>, Timothy M. Frayling<sup>1\*</sup>, Michael N. Weedon<sup>1\*</sup>**

1. Genetics of Complex Traits, University of Exeter Medical School, Exeter, U.K
2. 23andMe Inc., 899 W. Evelyn Avenue, Mountain View, California 94041, USA
3. A9.com Inc, Palo Alto, California 94301, USA
4. Estonian Genome Center and Institute of Molecular and Cell Biology of University of Tartu, Estonian Biocentre, Tartu, Estonia
5. Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, Edinburgh, Scotland
6. Institute of Medical Psychology, Ludwig-Maximilians-University, Munich, Germany
7. Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, Scotland
8. Center for Biomedicine, European Academy of Bolzano, Bozen, Italy – affiliated Institute of the University of Lübeck, Germany
9. Division of Pulmonary, Sleep and Critical Care Medicine, Department of Internal Medicine, Korea University Ansan Hospital, Ansan, Republic of Korea
10. Institute of Human Genomic Study, College of Medicine, Korea University Ansan Hospital, Ansan, Republic of Korea
11. Department of Neurology, Bundang Clinical Neuroscience Institute, Seoul National University Bundang Hospital, Seongnam, Republic of Korea
12. Institute of Human Genomic Study, College of Medicine, Korea University Ansan Hospital, Ansan, Republic of Korea
13. The University of Queensland, Queensland Brain Institute, Brisbane, Australia
14. Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA, U.S.A.
15. Department of Epidemiology, Erasmus Medical Center, Rotterdam, Netherlands
16. Department of Psychiatry, Erasmus Medical Center, Rotterdam, Netherlands

\*Joint contribution

**Corresponding author:**

Michael N Weedon  
University of Exeter Medical School  
RILD building Level 3  
Royal Devon & Exeter Hospital  
Barrack Road  
Exeter  
EX2 5DW

E-mail: [M.N.Weedon@exeter.ac.uk](mailto:M.N.Weedon@exeter.ac.uk)

## Abstract

Disrupted circadian rhythms and reduced sleep duration are associated with several human diseases, particularly obesity and type 2 diabetes, but little is known about the genetic factors influencing these heritable traits. We performed genome-wide association studies of self-reported chronotype (morning/evening person) and self-reported sleep duration in 128,266 White British individuals from the UK Biobank study. Sixteen variants were associated with chronotype ( $P < 5 \times 10^{-8}$ ), including variants near the known circadian rhythm genes *RGS16* (1.21 odds of morningness [95%CI 1.15, 1.27],  $P = 3 \times 10^{-12}$ ) and *PER2* (1.09 odds of morningness [95%CI 1.06, 1.12],  $P = 4 \times 10^{-10}$ ). The *PER2* signal has previously been associated with iris function. We sought replication using self-reported data from 89,823 23andMe participants; thirteen of the chronotype signals remained significant at  $P < 5 \times 10^{-8}$  on meta-analysis and eleven of these reached  $P < 0.05$  in the same direction in the 23andMe study. For sleep duration, we replicated one known signal in *PAX8* (2.6 [95%CI 1.9, 3.2] minutes per allele  $P = 5.7 \times 10^{-16}$ ) and identified and replicated two novel associations at *VRK2* (2.0 [95% CI: 1.3, 2.7] minutes per allele,  $P = 1.2 \times 10^{-9}$ ; and 1.6 [95% CI: 1.1, 2.2] minutes per allele,  $P = 7.6 \times 10^{-9}$ ). Although we found genetic correlation between chronotype and BMI ( $r_G = 0.056$ ,  $P = 0.048$ ); undersleeping and BMI ( $r_G = 0.147$ ,  $P = 1 \times 10^{-5}$ ) and oversleeping and BMI ( $r_G = 0.097$ ,  $P = 0.039$ ), Mendelian Randomisation analyses provided no consistent evidence of causal associations between BMI or type 2 diabetes and chronotype or sleep duration. Our study provides new insights into the biology of sleep and circadian rhythms in humans.

## Introduction

There are strong epidemiological associations among disrupted circadian rhythms, sleep duration and disease. A circadian rhythm refers to an underlying 24-hour physiological cycle that occurs in most living organisms. In humans, there are clear daily cyclical patterns in core body temperature, hormonal and most other biological systems<sup>1</sup>. These cycles are important for many molecular and behavioural processes. In particular, circadian rhythms are important in regulating sleeping patterns. While each individual has an endogenous circadian rhythm, the timing of these rhythms varies across individuals. Those with later circadian rhythms tend to sleep best with a late bedtime and late rising time and are often referred to as an “owl” or as an “evening” person. Those with earlier rhythms tend to feel sleepy earlier in the night and wake up early in the morning and are referred to as a “lark” or “morning” person. The remainder of the population falls in between these extremes. This dimension of circadian timing, or chronotype, is one behavioural consequence of these underlying cycles. Chronotype can be simply assessed by questionnaire and is considered a useful tool for studying circadian rhythms<sup>2,3</sup>.

There is substantial evidence for a relationship between short sleep duration, poor quality sleep and obesity and type 2 diabetes<sup>4,5</sup>. Eveningness has been associated with poor glycaemic control in patients with type 2 diabetes independently of sleep disturbance<sup>6</sup> and with metabolic disorders and body composition in middle-aged adults<sup>7</sup>. There is evidence from animal models that disruption to circadian rhythms and sleep patterns can cause various metabolic disorders<sup>8-10</sup>. For example, mice homozygous for dominant negative mutations in the essential circadian gene, *Clock*, develop obesity and hyperglycaemia<sup>10</sup> and conditional ablation of the *Bmal1* and *Clock* genes in pancreatic islets causes diabetes mellitus due to defective  $\beta$ -cell function<sup>9</sup>. Despite this evidence, in humans the causal nature of the epidemiological associations between sleep patterns, circadian rhythms and obesity and type 2 diabetes is unknown. Identifying genetic variants associated with sleep duration and chronotype will provide instruments to help test the causality of epidemiological associations<sup>11</sup>.

A previous genome-wide association study (GWAS) in 4251 individuals identified a single genetic variant in *ABCC9* associated with sleep duration<sup>12</sup>. A subsequent GWAS meta-analysis including 47,180 individuals identified a single locus for sleep duration near *PAX8*<sup>13</sup>. There have been no published reports of variants influencing chronotype. The UK Biobank is a study of 500,000 individuals from the UK aged between 37 and 73 years with genome-wide SNP analysis and detailed phenotypic information, including chronotype and sleep duration (<http://www.ukbiobank.ac.uk/>). The UK Biobank study provides an excellent

opportunity to identify novel genetic variants influencing chronotype and sleep duration which will provide insights into the biology of circadian rhythms and sleep and help test causal relationships between circadian rhythm and metabolic traits including obesity.

## Results

### ***Sixteen loci associated with chronotype in UK Biobank***

Using self-reported “morningness”, we generated a binary and a continuous chronotype score. We performed genome-wide association studies on 16,760,980 imputed autosomal variants. **Figure 1** presents the overall results for these GWAS. **Table 1** presents details of all 16 genome-wide significant chronotype-associated loci.

### ***Replication and validation of chronotype associations***

Analysing UK Biobank data with that from 23andMe provides strong evidence that at least 13 of the 16 are robustly associated with chronotype. Thirteen of the chronotype signals remained at  $P < 5 \times 10^{-8}$  in a meta-analysis including UK Biobank and 89,283 individuals from 23andMe (Hu *et al.* Nature Communications, *In Press*), of which eleven reached  $P < 0.05$  in the same direction in 23andMe alone, and 15 of the 16 UK Biobank signals were in the same direction (binomial  $P = 0.0002$ ) (**Table 1**). We also attempted to validate the associations in 6,191 European-Ancestry from the Chronogen consortium and 2,532 Korean Ancestry individuals from the Insomnia, Chronotype and sleep EEG (ICE) consortium that used “Gold standard” chronotype questionnaire (Munich Chronotype Questionnaire – MCTQ and Morningness-Eveningness Questionnaire - MEQ). Given the sample size of 5% of the discovery UK Biobank study we assessed directional consistency rather than testing for replication P-values  $< 0.05$  or  $0.05/16$ . In the European-Ancestry individuals 11 of the 16 signals were represented. Nine of these 11 variants had the same direction of effect as the discovery UK Biobank cohort (binomial test  $P = 0.03$ ) and one replicated at Bonferroni significance (rs12140153,  $P = 0.003$ ). In the Korean study, 9 signals were represented, four of which had the same direction of effect as the discovery UK Biobank cohort (binomial test  $P = 1.00$ ). The level of directional consistency in these two smaller studies is consistent with what would be expected in cohorts  $< 5\%$  the size of our discovery cohort.

### ***The chronotype-associated variants occur near genes known to be important in photoreception and circadian rhythms***

The variant most strongly associated with chronotype, rs516134 (OR for morningness=1.21, [95% CI: 1.16,1.26], binary  $P = 3.7 \times 10^{-12}$ , continuous  $P = 8.9 \times 10^{-13}$ ) occurs near *RGS16*, which is a regulator of G-protein signalling and has a known role in circadian rhythms<sup>14</sup> (**Table 1 and Figure 2**). Another signal occurs near *PER2* (lead variant rs75804782, odds ratio=1.09,

[95% CI: 1.06, 1.12], binary  $P=7.2 \times 10^{-10}$ , continuous  $P=3.2 \times 10^{-7}$ ; **Figure 3**). *PER2* is a well-known regulator of circadian rhythms<sup>15-20</sup> and contains a variant, rs75804782, recently shown to be associated with iris formation<sup>21</sup> that is in LD ( $r^2 = 0.65$ ,  $D' = 0.97$ ) with our reported lead SNP. We also identified an association with a missense variant (rs12140153, OR=1.07 (95% CI: 1.04, 1.11), binary  $P=5 \times 10^{-6}$ , continuous  $P=7 \times 10^{-9}$ ) in *INADL* (*InaD-like*) that encodes a protein thought to be important in organising and maintaining the “intrinsically photosensitive retinal ganglion cells”, cells that are known to communicate directly with the suprachiasmatic nucleus; the primary circadian pacemaker in mammals<sup>22</sup>. As there is a reported link between season and reported chronotype<sup>23</sup>, we carried out a sensitivity analysis in which we adjusted for month of attendance (to assessment centre); all associations remained genome-wide significant for the reported variants. We tested for enrichment of specific biological and molecular pathways using MAGENTA<sup>24</sup> but none had a clear link to circadian rhythms (**Supplementary Table 1**).

### **Three loci associated with sleep duration**

We performed genome-wide association studies on a binary sleep phenotype and a continuous sleep duration score for 16,761,225 imputed variants. **Figure 4** presents the overall results for these GWAS. Three loci reached genome-wide significance. The most strongly associated variant was rs62158211 with an average 2.6 minute (95% CI: 1.9 to 3.2 minutes,  $P=5.7 \times 10^{-16}$ ) per-allele change in sleep duration and occurs at the previously reported association signal near *PAX8*<sup>13</sup>. We identified two, novel, conditionally independent, signals that were located ~900kb apart, one upstream and the other downstream of *VRK2*. The downstream variant, rs17190618, has an average per allele effect of 2.0 minutes (95% CI: 1.3 to 2.7 minutes),  $P=1.2 \times 10^{-9}$ , on sleep duration. The upstream variant, rs1380703 (which is not correlated with rs17190618,  $r^2=0.002$ ), has an average per allele effect of 1.6 minutes (95% CI: 1.1 to 2.2 minutes),  $P=7.6 \times 10^{-9}$ , on sleep duration. On adjusting for month of assessment, we saw marginally stronger associations for both rs62158211 ( $P=3 \times 10^{-16}$ ) and rs1380703 ( $P=6 \times 10^{-9}$ ), with no change for rs17190618. **Table 2** shows the three sleep duration loci and their lead variants. **Figure 5** shows locus zoom plots of the *VRK2* association signals. We did not replicate the association of a previously reported variant in *ABCC9*<sup>12</sup> with sleep duration (rs11046205, 0.1mins [95% CI: -0.6 to 0.7 minutes],  $P=0.83$ ).

### **Replication of novel sleep duration hits**

To replicate the two novel sleep duration hits we used data from 47,180 individuals from a published study<sup>13</sup>. The variant rs17190618 replicated with effect size=2.1 minutes (95% CI:

0.8 to 3.3),  $P=0.001$ , meta-analysis  $P=5 \times 10^{-12}$ . The variant rs1380703 replicated with effect size=1.3 minutes (95% CI: 0.3 to 2.2),  $P=0.01$ , meta-analysis  $P=3 \times 10^{-10}$ ).

### ***Sleep duration and chronotype are heritable and genetically correlated with BMI, Type 2 diabetes and psychiatric disease***

Using LD-score regression we estimated the heritability of chronotype and sleep duration within UK Biobank to be 0.12 (0.007), and 0.07 (0.007), respectively. There was no significant genetic correlation between sleep duration and chronotype ( $r_G=0.0177$ ,  $P=0.70$ ). Chronotype was nominally genetically correlated with BMI ( $r_G=0.056$ ,  $P=0.048$ ), but not Type 2 diabetes ( $r_G=0.004$ ,  $P=0.99$ ). As the relationship between sleep duration with BMI and risk of T2D is U-shaped (see **Supplementary Figure 1**), we defined two further binary phenotypes; undersleepers (<7 vs. 7-8 hours) and oversleepers (>8 vs. 7-8 hours). There was a strong genetic correlation between undersleeping and BMI ( $r_G=0.147$ ,  $P=1 \times 10^{-5}$ ), but not T2D ( $r_G=0.022$ ,  $P=0.79$ ). There was also a genetic correlation between oversleeping and both BMI ( $r_G=0.097$ ,  $P=0.039$ ) and T2D ( $r_G=0.336$ ,  $P=0.001$ ). We also performed LD-score regression analyses against a range of other diseases and traits where GWAS summary statistics are publically available (**Supplementary Table 2**). Schizophrenia was genetically correlated (after adjusting for the number of tests) with hours slept ( $r_G=0.25$ ,  $P=1 \times 10^{-4}$ ), oversleeping ( $r_G=0.32$ ,  $P=9 \times 10^{-4}$ ), but not significantly correlated with undersleeping ( $r_G=-0.11$ ,  $P=0.097$ ).

### ***Mendelian randomisation analyses provide no consistent evidence that higher BMI affects self-reported morningness or vice-versa***

Using a genetic risk score of 69 known BMI variants<sup>25</sup> (listed in **Supplementary Table 3**) as an instrumental variable, we next performed Mendelian randomisation analyses in the UK Biobank study to test the potential causal role of BMI in chronotype and sleep. Instrumental variables analyses using variants and their effect sizes identified by previous studies<sup>25</sup> provided no consistent evidence that self-reported “morningness” causally affects BMI or risk of type 2 diabetes (**Supplementary Table 4**). Association statistics of the BMI variants with chronotype are given in **Supplementary Table 3**. We repeated these analyses using a genetic risk score consisting of 55 type 2 diabetes SNPs<sup>26</sup> and did not find any evidence of causality. Association of the chronotype-associated variants with BMI are given in **Supplementary Table 5**. Performing the reciprocal Mendelian randomization analysis using a genetic risk score of the 13 replicated chronotype variants, with effect sizes obtained from 23andMe, we found no consistent evidence in the UK Biobank data that morningness or eveningness leads to higher BMI (**Supplementary Table 4**).



### **No evidence that BMI and Type 2 diabetes are causally associated with sleep duration**

Using the same genetic risk score of 69 known BMI variants as an instrument, we saw no consistent evidence that higher BMI increased an individual's likelihood of being an undersleeper (IVreg2  $P=0.95$ , IVW  $P=0.05$ ) or an oversleeper (IVreg2  $P=0.29$ , IVW  $P=0.62$ ) in the UK Biobank data (**Supplementary Table 4**). Because there were only three genetic variants of small effect associated with sleep duration, we did not perform any Mendelian Randomisation analyses of sleep on BMI or type 2 diabetes risk.

### **Discussion**

We performed a genome-wide association study of sleep duration and morningness in 128,266 individuals from the UK Biobank study. We discovered and replicated two novel loci associated with sleep duration. Through replication in a study of 89,823 individuals from 23andMe we found 13 genome-wide significant loci for chronotype. These loci occur in or near circadian rhythm and photoreception genes and provide new insights into circadian rhythm and sleep biology and their links to disease.

The two novel sleep duration association signals that we have discovered and replicated in this study occur upstream and downstream of *VRK2* (*vaccinia related kinase 2*). *VRK2* is a serine/threonine kinase important in several signal transduction cascades, and variants near *VRK2* are associated with schizophrenia<sup>27</sup> and epilepsy<sup>28</sup>. The two sleep duration variants we identified do not represent the same signals as those associated with schizophrenia at genome wide significance but one is associated with schizophrenia (based on publically available data from the schizophrenia genetics consortium (rs1380703  $P=2 \times 10^{-5}$ ), with the allele associated with more sleep being associated with higher risk of schizophrenia). Furthermore, the variants associated with epilepsy and schizophrenia at genome wide significance are associated with sleep duration in UK Biobank (epilepsy lead variant rs2947349<sup>28</sup>,  $P=2 \times 10^{-5}$  and schizophrenia lead variant<sup>27</sup> rs11682175  $P=3 \times 10^{-5}$ ) but did not reach genome wide significance. We also observed genetic correlation between sleep duration and schizophrenia using LD-score regression ( $r_G=0.25$ ,  $P=1 \times 10^{-4}$ ). Further work is required to determine whether variation in *VRK2* either has independent associations with both sleep and schizophrenia or whether there is some causal link between sleep duration and schizophrenia and epilepsy.

Several of the loci that we identified as associated with chronotype contain genes that have a known role in circadian rhythms. The most strongly associated variant, rs516134, occurs 20kb downstream of *RGS16* (regulator of G protein signalling 16). *RGS16* has recently been



shown to have a key role in defining 24 hour rhythms in behaviour<sup>14</sup>. In mice, gene ablation of *Rgs16* lengthens the circadian period of behavioural rhythm<sup>14</sup>. By temporally regulating cAMP signalling, *Rgs16* has been shown to be a key factor in synchronising intercellular communication between pacemaker neurons in the suprachiasmatic nucleus (SCN), the centre for circadian rhythm control in humans.

The association signal with lead SNP rs75804782 occurs ~100kb upstream of *PER2* (*Period 2*). *Per2* is a key regulator of circadian rhythms and is considered one of the most important clock genes, and, under constant darkness, *Per2* knockout mice show arrhythmic locomotor activity<sup>15-20</sup>. This locus also contains a variant that has recently been shown to be associated with iris furrow contractions<sup>21</sup>. Our signal is very likely to represent the same association and suggests a link between iris function and chronotype (rs75804782 has an LD  $r^2 = 0.65$  and  $D' = 0.97$  with the reported lead SNP, rs3739070). Larsson *et al.*<sup>21</sup> suggest *TRAF3IP1* as the most likely candidate gene at the locus because of its critical role in the cytoskeleton and neurogenesis. Further work is needed to elucidate whether the chronotype association at this locus acts through *PER2* or *TRAF3IP1*.

Several of the variants associated with chronotype are also associated with BMI and we found genetic correlation between chronotype and sleep duration and BMI. There is substantial evidence for a role of sleep disruption and circadian rhythms in metabolic disease<sup>1</sup>. Data from animal models and epidemiology provide strong evidence that sleep quality or disrupted circadian rhythms can cause metabolic diseases including obesity and type 2 diabetes<sup>4-6,8-10</sup>. Our Mendelian Randomisation analyses provided no consistent evidence for a role of higher BMI leading to increased self-reported morningness.

There are some important limitations to our study. First, chronotype and sleep duration were self-reported and are subject to reporting bias (e.g. obese individuals may be more likely to falsely claim to be morning people). Second, whilst we did not find any evidence that overall chronotype or sleep duration causally lead to obesity or type 2 diabetes, it is possible that sub-pathways of genes involved in, for example, feeding behaviour may be important in both obesity and chronotype regulation. The availability of the full UK Biobank study of 500,000 will provide further insight into this relationship.

In conclusion, we have identified novel genetic associations for chronotype and sleep duration. The chronotype loci cluster near genes known to be important in determining circadian rhythms and will provide new insights into circadian regulation. Our results provide new insights into circadian rhythm and sleep biology and their links to disease.

## **Materials and Methods**

### ***Discovery Samples***

We used 128,266 individuals of British descent from the first UK Biobank genetic data release (see <http://biobank.ctsu.ox.ac.uk>). British-descent was defined as individuals who both self-identified as white British and were confirmed as ancestrally Caucasian using principal components analyses (<http://biobank.ctsu.ox.ac.uk>). Of these individuals, 120,286 were classified as unrelated, with a further 7,980 first- to third-degree relatives of these. As the association tests were carried out in BOLT-LMM<sup>29</sup>, which adjusts for relationships between individuals and corrects for population structure, we included all 128,266 related white British individuals in the association analyses.

### ***Genotyping and quality control***

We used imputed variants provided by the UK Biobank. Details of the imputation process are provided at the UK Biobank website (see <http://biobank.ctsu.ox.ac.uk>). For this study we only included the ~16.7M imputed variants with an imputation  $R^2 \geq 0.4$ , MAF  $\geq 0.001$  and with a Hardy–Weinberg equilibrium  $P > 1 \times 10^{-5}$ .

### ***Phenotypes***

#### ***Chronotype***

UK Biobank provides a single measure of Chronotype, from which we produced a continuous and a dichotomous phenotype. Chronotype (or morningness) is a self-reported measure and asks individuals to categorise themselves as “Definitely a ‘morning’ person”, “More a ‘morning’ than ‘evening’ person”, “More an ‘evening’ than a ‘morning’ person”, “Definitely an ‘evening’ person” or “Do not know”, which we coded as 2, 1, -1, -2 and 0 respectively, in our raw continuous “score”. Individuals had the option not to answer; these individuals were set to missing. We then produced a normally distributed phenotype by adjusting the raw phenotype for age, gender and study centre (categorical) and inverse normalising the resulting residuals. The dichotomous chronotype trait defines morning people (“Definitely a ‘morning’ person” and “More a ‘morning’ than ‘evening’ person”) as cases and evening people (“Definitely an ‘evening’ person” and “More an ‘evening’ than a ‘morning’ person”) as controls. All other individuals are coded as missing. All results reported for continuous chronotype refer to the inverse-normalised residualised chronotype “score”. For interpretable results, however, we report effect sizes using the odds ratios of the dichotomous chronotype phenotype. A total number of 127,898 and 114,765 individuals were available with non-missing continuous and binary chronotype phenotypes, respectively, for the association tests; for the Mendelian Randomisation this became 119,935 and 107,634 respectively.

### *Sleep duration*

The UK Biobank also provides self-reported “sleep duration”, in which individuals were asked to provide the average number of hours slept in a 24-hour period. The phenotype was derived by first excluding individuals reporting greater than 18 hours sleep, then adjusting for age, gender and study centre (categorical) and obtaining the model residuals and finally inverse-normalising to assure a normally distributed phenotype. When reporting results for the continuous sleep duration phenotype, we are referring to the inverse-normalised phenotype, though we report effect sizes of the residualised phenotype to allow easier interpretation of results. There were 127,573 individuals with reported sleep duration available for the association tests, with 119,647 available for the MR analyses.

### *“Oversleepers” and “Undersleepers”*

These two dichotomous phenotypes share the same set of controls; those individuals that reported sleeping either 7 or 8 hours (81,204 individuals). In oversleepers, cases (10,102 individuals) are those reporting 9 or more hours sleep on average, whereas undersleeper cases (28,980 individuals) are those reporting 6 or fewer hours.

### *BMI*

The UK Biobank provided a BMI (weight (kg)/height<sup>2</sup>) measurement and an estimate based on electrical impedance analyses. To help avoid reporting error we excluded individuals with significant differences (>4.56 SDs) between these two variables where both were available. If only one of these measurements was available this was used. We corrected BMI by regressing age, sex, study centre, and the first 5 within-British principal components and taking residual values. We then inverse normalised the residuals. A total of 119,684 white-British individuals with BMI and genetic data were available for the Mendelian Randomisation analyses.

### *Type 2 diabetes*

Individuals were defined as having T2D if they reported either T2D or generic diabetes at the interview stage of the UK Biobank study. Individuals were excluded if they reported insulin use within the first year of diagnosis. Individuals reportedly diagnosed under the age of 35 years or with no known age of diagnosis were excluded, to limit the numbers of individuals with slow-progressing autoimmune diabetes or monogenic forms. Individuals diagnosed with diabetes within the last year of this study were also excluded as we were unable to determine whether they were using insulin within this time frame. A total of 4,040 cases and 113,735 controls within the white British subset of UK Biobank were identified with genetic data available.

### **Genome-wide association analysis**

To perform the association tests, we used BOLT-LMM<sup>29</sup> to perform linear mixed models (LMMs) in the 128,266 individuals. We used BOLT-LMM as it adjusts for population structure and relatedness between individuals whilst performing the association tests with feasible computing resources. As it adjusts for population structure and relatedness between individuals whilst performing the association tests, it allowed us to include the additional 7,980 related individuals and therefore improved our power to detect associations. To calculate the relationships between individuals, we provided BOLT-LMM a list of 328,928 genotyped SNPs (MAF>5%; HWE  $P > 1 \times 10^{-6}$ ; missingness<0.015) for the individuals included in the association analysis and used the 1000 Genomes LD-Score table provided with the software.

As the continuous phenotypes were derived by adjusting for age, gender and study centre, the LMM only included chip (BiLEVE vs. UKBiobank arrays) as a covariate at run-time (see [http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank\\_genotyping\\_QC\\_documentation-web.pdf](http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web.pdf)). The binary phenotypes were unadjusted and so included age, gender and chip at run-time. BOLT-LMM reported no improvement of the non-infinitesimal mixed model test over the standard infinitesimal test and so all association results reported in this paper are for the infinitesimal model<sup>29</sup>.

### **Chronotype replication samples**

Participants were from the customer base of 23andMe, Inc. The descriptions of the samples, genotyping and imputation are in Hu *et al.* Nature Communications, *In Press*. Of the 16 chronotype-associated variants for which we attempted replication, 10 were available from imputation from the 1000 Genomes imputation panel phase 1 pilot. An additional 4 were imputed from the phase 1 version 3 1000 Genomes imputation panel. The final two could not be imputed. We used <http://analysistools.nci.nih.gov/LDlink/> to find proxies --the best available were rs4729854 for rs372229746 ( $r^2=0.33$ ), and rs12621152 for rs70944707 ( $r^2=0.33$ ). We meta-analysed P-values from the discovery and replication samples using sample size weighting implemented in METAL<sup>30</sup>.

### **Chronotype validation samples**

Genotypes consisting of both directly typed and imputed SNPs were used for the individual GWAS<sup>12</sup>. To avoid over-inflation of test statistics due to population structure or relatedness, we applied genomic control for the independent studies and meta-analysis. Linear

regression for associations with normalized chronotype was performed (see <sup>12</sup> for packages used) under an additive model, with SNP allele dosage as predictor and with age, age<sup>2</sup>, gender, normalized sleep duration, season of assessment (dichotomized based on time of the year, and day-light savings time – DST or standard zone time assessments) as covariates. A fixed-effects meta-analysis was conducted with GWAMA using the inverse-variance-weighted method and low imputation quality (Rsq/proper\_info < 0.3) were dropped from the meta-analysis.

### ***Pathway analyses***

Pathway analyses were carried out in MAGENTA using all available libraries provided with the software. We included all imputed variants with association  $P < 1 \times 10^{-5}$  from the continuous chronotype trait. For the results presented in **Supplementary Table 1**, we used gene upstream and downstream limits of 250Kb, excluded the HLA region (default setting) and set the number of permutations for GSEA estimation at 10,000 (default).

### ***Genetic correlation analyses***

Genetic correlations (see <sup>31</sup> for methodology) between traits were calculated using the LD Score Regression software LDSC (available at <https://github.com/bulik/ldsc/>) <sup>32</sup>. Summary statistics of our traits outputted by BOLT-LMM were first “munged”, a process that converts the summary statistics to a format that LDSC understands and aligns the alleles to the Hapmap 3 reference panel, removing structural variants and multi-allelic and strand-ambiguous SNPs. Genetic correlations were then calculated between our phenotypes and a set of 100 phenotypes for which summary statistics are publicly available (full list in **Supplementary Table 2**). We used precomputed LD structure data files specific to Europeans of HAPMAP 3 reference panel, obtained from ([http://www.broadinstitute.org/~bulik/eur\\_ldscores/](http://www.broadinstitute.org/~bulik/eur_ldscores/)) as suggested on the LDSC software page.

### ***Mendelian Randomisation IV analysis***

The 13 variants in **Table 1** which reached  $P < 5 \times 10^{-8}$  in combined analyses were used as chronotype instruments in the Mendelian Randomisation analyses. Where binary and continuous traits shared a locus, we selected the top variant of the continuous trait over that of the binary. For loci that reach GW-significance in the binary trait only, we selected the top variant but used the effect size from the continuous trait.

To test for a causal effect of BMI on chronotype and sleep-duration, we selected 69 of 76 common genetic variants that were associated with BMI at genome wide significance in the

GIANT consortium in studies of up to 339,224 individuals (**Supplementary Table 3**)<sup>25</sup>. We limited the BMI SNPs to those that were associated with BMI in the analysis of all European ancestry individuals and did not include those that only reached genome-wide levels of statistical confidence in one-sex only, or one stratum only. Variants were also excluded if known to be classified as a secondary signal within a locus. Three variants were excluded from the score due to potential pleiotropy (rs11030104 [*BDNF* reward phenotypes], rs13107325 [*SLC39A8* lipids, blood pressure], rs3888190 [*SH2B1* multiple traits]), three due to being out of HWE (rs17001654, rs2075650 and rs9925964) and the last variant due to not being present in the imputed data (rs2033529).

For testing reverse causality of type 2 diabetes on our sleep phenotypes, we used 55 of 65 common variants (listed in **Supplementary Table 3**) known to be associated with type 2 diabetes at genome wide significance in a meta-analysis of 34,840 cases and 114,981<sup>26</sup>, excluding those known or suspected to be pleiotropic.

We performed the Mendelian Randomisation analysis two ways; firstly using instrumental variables (IV) using STATA's "IVreg2" function<sup>33</sup> and secondly through the inverse-variance weighted (IVW) and MR-Egger methods described in<sup>34</sup>. Analyses were performed in STATA 13.1 (StataCorp. 2013. Stata Statistical Software: Release 13. College Station, TX: StataCorp LP.).

In the instrumental variables method, we generated genetic risk scores (GRS) for BMI and type 2 diabetes using the published list of associated variants and their respective betas. For Chronotype, we generated a GRS using the thirteen replicated variants and their respective betas from 23andMe summary statistics. Using the IVreg2 command, we performed two-stage least squares estimation to calculate the effect of predicted exposure (through the GRS) on the continuous outcome traits. For binary outcomes (type 2 diabetes, undersleeper and oversleeper), we manually carried out the two-stage process by regressing the exposure trait on its GRS and storing both predicted values and residuals. We then used these predicted values and residuals as independent variables in a logistic regression where the dependent variable was the binary outcome.

The inverse-variance weighted (IVW) method is equivalent to a meta-analysis of the associations of the individual instruments and uses associations between the instruments and both the exposure and the outcome to estimate the additive effect of the instruments combined<sup>34</sup>. The MR-Egger method is a modification to the IVW method that allows the inclusion of "invalid" instruments (i.e. those that don't satisfy all three conditions), by

performing Egger regression using the summary data of the variants. The IVW and Egger methods operate under the assumption that all instruments are valid, in that they satisfy the three IV conditions: the genetic variants are 1) independent of confounders, 2) associated with the exposure and 3) independent of the outcome. The MR-Egger method, however, accounts for the fact that genetic variants could be pleiotropic and may influence the outcome via pathways other than through the exposure and therefore the resulting association between genetic instruments and the outcome should not be biased by invalid instruments and pleiotropy. The MR-Egger method was used purely as a sensitivity test for the IVW method and so MR-Egger results were not considered if the IVW result did not reach nominal significance.

For the IVW and MR-Egger methods, associations of genetic instruments (variants) with both exposure and outcome phenotypes were generated in STATA by regressing the phenotype against the instrument while adjusting for covariates. As a further sensitivity test, we also repeated these analyses by replacing exposure phenotype-variant associations with their respective published betas and found only slight differences in betas and P-values, though all exposure-outcome associations remained non-significant.

### **Acknowledgements**

This research has been conducted using the UK Biobank Resource. We would like to thank the research participants and employees of 23andMe for making this work possible.

**Funding Information.** S.E.J. is funded by the Medical Research Council (grant: MR/M005070/1) M.A.T., M.N.W. and A.M. are supported by the Wellcome Trust Institutional Strategic Support Award (WT097835MF). A.R.W., T.M.F and H.Y. are supported by the European Research Council grants: SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC and 323195. R.M.F. is a Sir Henry Dale Fellow (Wellcome Trust and Royal Society grant: 104150/Z/14/Z). R.B. is funded by the Wellcome Trust and Royal Society grant: 104150/Z/14/Z. J.T. is funded by a Diabetes Research and Wellness Foundation Fellowship. This study was provided with biospecimens and data from the Korean Genome Analysis Project (4845-301), the Korean Genome and Epidemiology Study (2010-E71001-00, 2011-E71004-00, and 2011-E71008-00), and Korea Biobank Project (4851-307) that were supported by the Korea Centers for Disease Control & Prevention, Republic of Korea. The funders had no influence on study design, data collection and analysis, decision to publish, or preparation of the manuscript.



**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

**Table 1. Genetic variants associated with chronotype.** Sixteen loci associated with chronotype as either a continuous or binary trait in UK Biobank. Genes listed are candidate or nearest genes within 250Kb of the lead SNP. Odds ratios correspond to risk of morningness over eveningness. Beta, OR and frequency refers to A1. Replication data is based on continuous data and because the replication beta is in different units to the discovery GWAS beta a P-value meta-analysis was performed. \* Proxies used for replication cohort: rs4729854 for rs372229746 ( $r^2=0.33$ ), and rs12621152 for rs70944707 ( $r^2=0.33$ ).

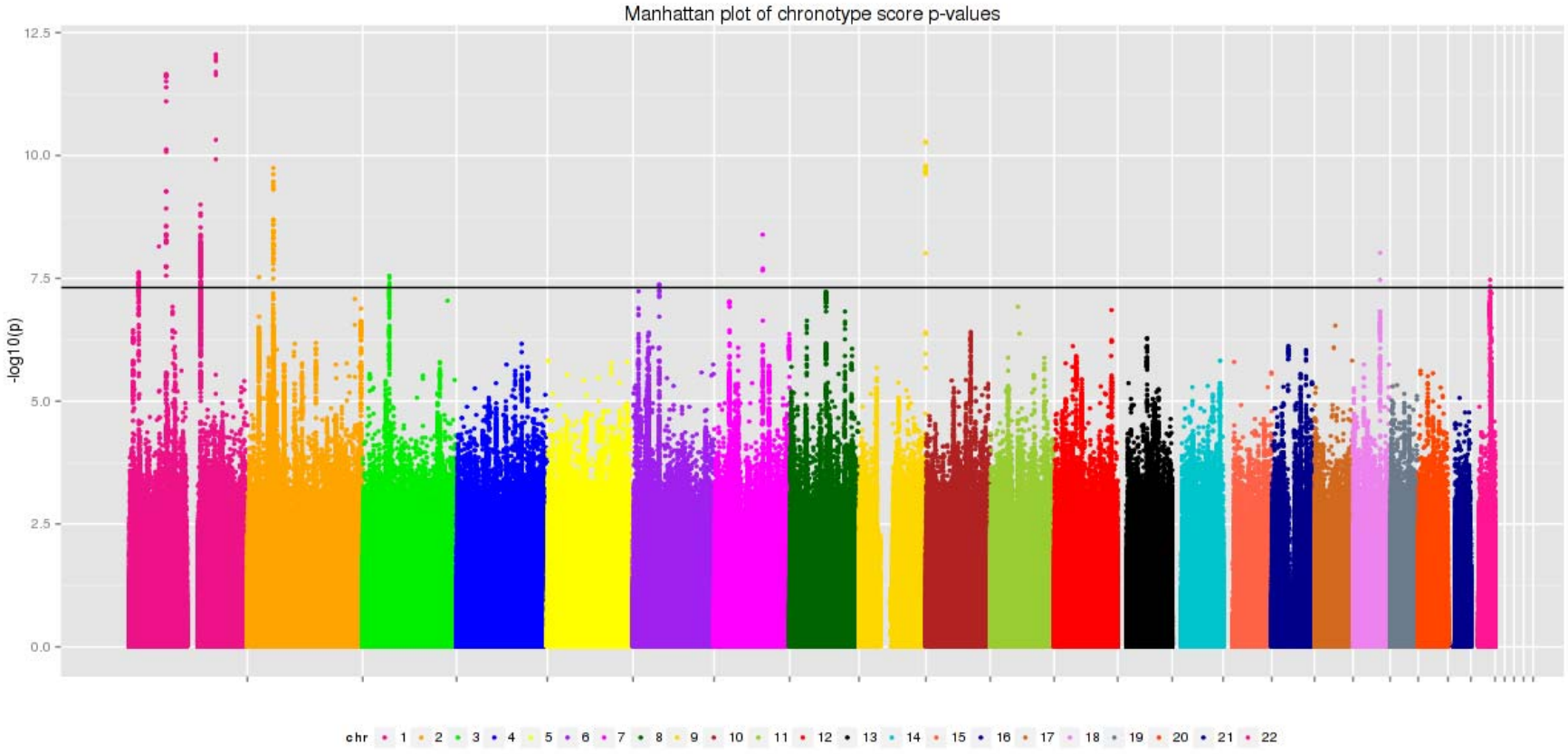
Variant	Chr:Pos	A1/A2	Freq	GWAS Continuous Beta (SE)	GWAS Continuous P	GWAS Binary OR (SE)	GWAS Binary P	Replication Beta (SE)	Replication P	Combined P	Genes
rs516134	1:182,553,693	C/T	0.03	0.081 (0.011)	9E-13	1.21 (0.032)	3E-12	0.295 (0.035)	2E-17	7E-28	<i>RGS16</i>
rs11162296	1:77,700,196	G/C	0.84	-0.037 (0.005)	2E-12	0.93 (0.011)	1E-12	-0.097 (0.015)	2E-10	2E-21	<i>PIGK, AK5</i>
rs77641763	9:140,265,782	C/T	0.88	0.039 (0.006)	5E-11	1.07 (0.015)	7E-09	0.065 (0.020)	2E-03	2E-12	<i>GRIN1, NRARP</i>
rs1075265	2:54,354,927	C/G	0.48	-0.025 (0.004)	2E-10	0.95 (0.009)	4E-08	-0.010 (0.011)	4E-01	4E-08	<i>PSME4, ACYP2</i>
rs75804782	2:239,316,043	T/C	0.88	0.030 (0.006)	3E-07	1.09 (0.015)	4E-10	0.106 (0.018)	4E-09	1E-14	<i>PER2</i>
rs10157197	1:150,250,636	G/A	0.6	0.025 (0.004)	1E-09	1.05 (0.010)	5E-07	0.064 (0.011)	1E-08	6E-17	<i>PRPF3, TARS2</i>
rs372229746 *	7:102,158,815	G/A	0.55	0.028 (0.005)	4E-09	1.06 (0.012)	7E-07	0.068 (0.013)	4E-07	8E-15	<i>ORAI2, RASA4</i>
rs12140153	1:62,579,891	G/T	0.9	0.039 (0.007)	7E-09	1.07 (0.017)	4E-06	0.043 (0.025)	8E-02	3E-08	<i>INADL</i>
rs9961653	18:56,767,671	T/C	0.42	0.023 (0.004)	1E-08	1.04 (0.010)	1E-06	0.032 (0.012)	6E-03	7E-10	<i>RAX, CPLX4, LMAN1</i>
rs192534763	8:36,202,946	T/C	0.99	0.100 (0.021)	3E-07	1.25 (0.057)	2E-08	-0.005 (0.055)	9E-01	1E-04	<i>UNC5D</i>
rs2050122	1:19,989,205	T/C	0.2	0.028 (0.005)	2E-08	1.06 (0.012)	3E-06	0.030 (0.014)	3E-02	1E-08	<i>HRT6</i>
rs12635074	3:55,982,416	T/G	0.68	-0.023 (0.004)	3E-08	0.96 (0.009)	2E-06	-0.002 (0.012)	8E-01	1E-05	<i>ERC2</i>
rs70944707 *	2:24,257,444	C/CT	0.23	0.030 (0.005)	3E-08	1.05 (0.013)	2E-05	0.035 (0.016)	3E-02	2E-08	<i>FKBP1B</i>
rs4821940	22:40,659,573	T/C	0.45	0.022 (0.004)	3E-08	1.05 (0.010)	4E-08	0.006 (0.011)	6E-01	5E-06	<i>SGSM3</i>
rs72720396	1:91,191,582	A/G	0.77	-0.025 (0.005)	1E-07	0.95 (0.010)	3E-08	-0.035 (0.014)	2E-02	2E-08	<i>CALB1</i>
rs76899638	6:55,147,508	A/ATG	0.22	0.026 (0.005)	4E-08	1.05 (0.012)	2E-07	0.067 (0.014)	4E-06	8E-13	<i>HCRTR2</i>

**Table 2. Three loci associated with sleep duration and their lead variants.** Genes listed are candidate genes at each locus. Beta, OR and frequency refers to A1. Because the replication beta is in different units to the discovery GWAS beta a P-value meta-analysis was performed. Beta units are in hrs.

Variant	Chr:Pos	A1/A2	A1 Freq	GWAS Continuous Beta (SE)	GWAS Continuous P	GWAS Binary OR (SE)	GWAS Binary P	Replication Beta (SE)	Replication P	Combined P	Gene
rs62158211	2:114,106,139	G/T	0.79	-0.039 (0.005)	6E-16	0.94 (0.011)	1E-07	-0.053 (0.009)	4E-9	2E-23	PAX8
rs17190618	2:58,882,765	A/T	0.84	-0.033 (0.005)	1E-09	0.96 (0.013)	3E-04	-0.035 (0.011)	1E-3	5E-12	VRK2
rs1380703	2:57,941,287	A/G	0.62	0.025 (0.004)	8E-09	1.06 (0.011)	8E-08	0.021 (0.008)	1E-2	3E-10	VRK2

Figure 1 a) Manhattan and b) quantile-quantile (QQ) plot of chronotype score (inverse-normalised) P-values

a)



b)

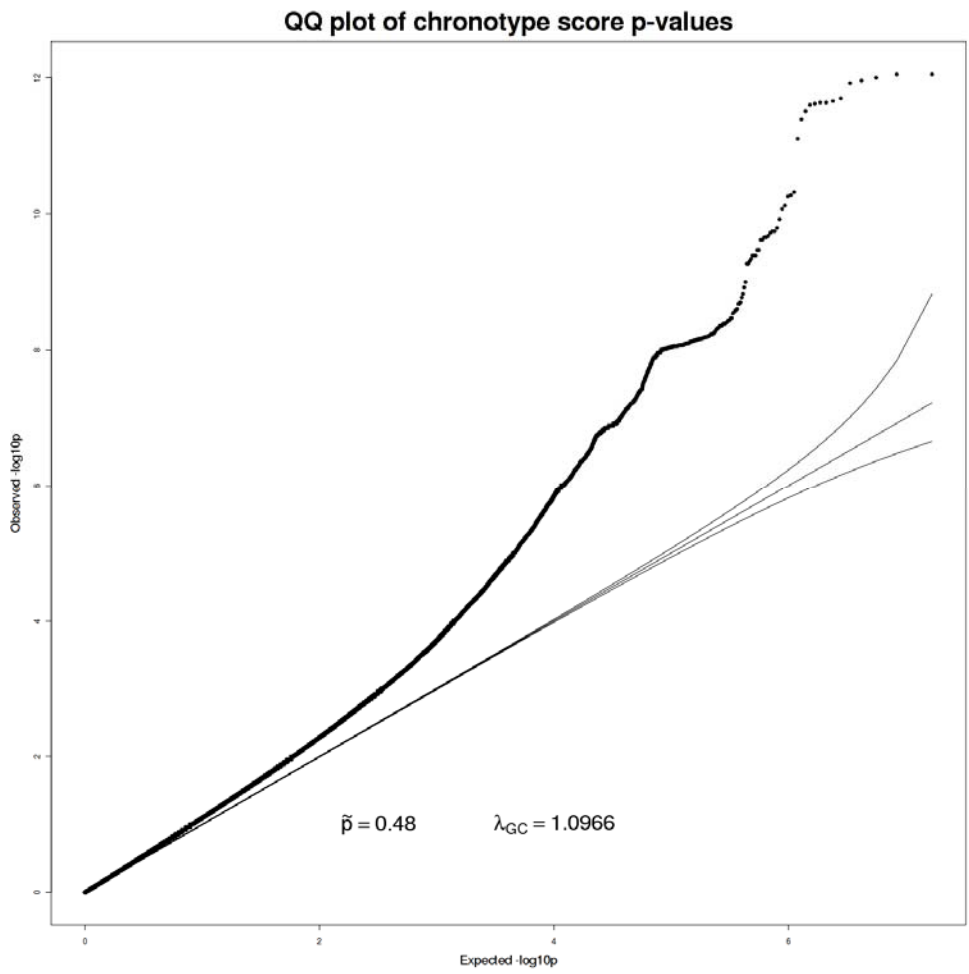


Figure 2: Locuszoom plot around *RGS16* locus

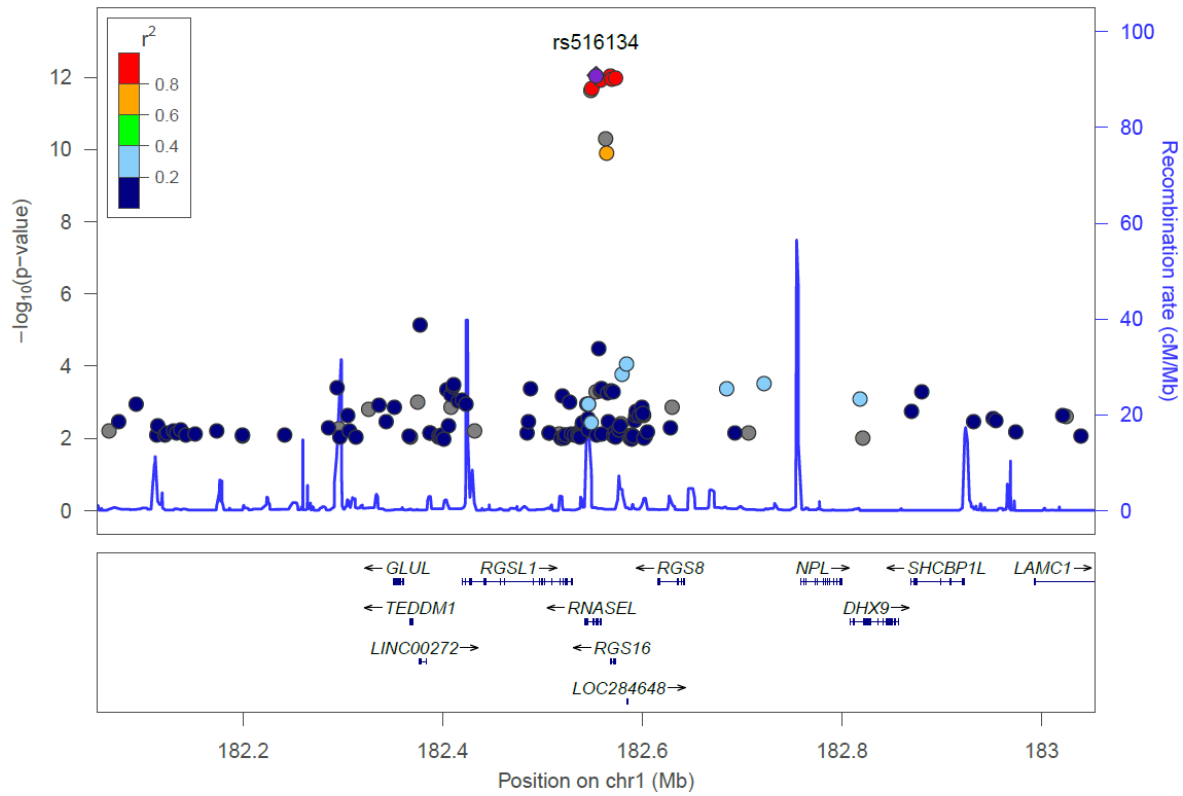


Figure 3: Locuszoom plot around *PER2*

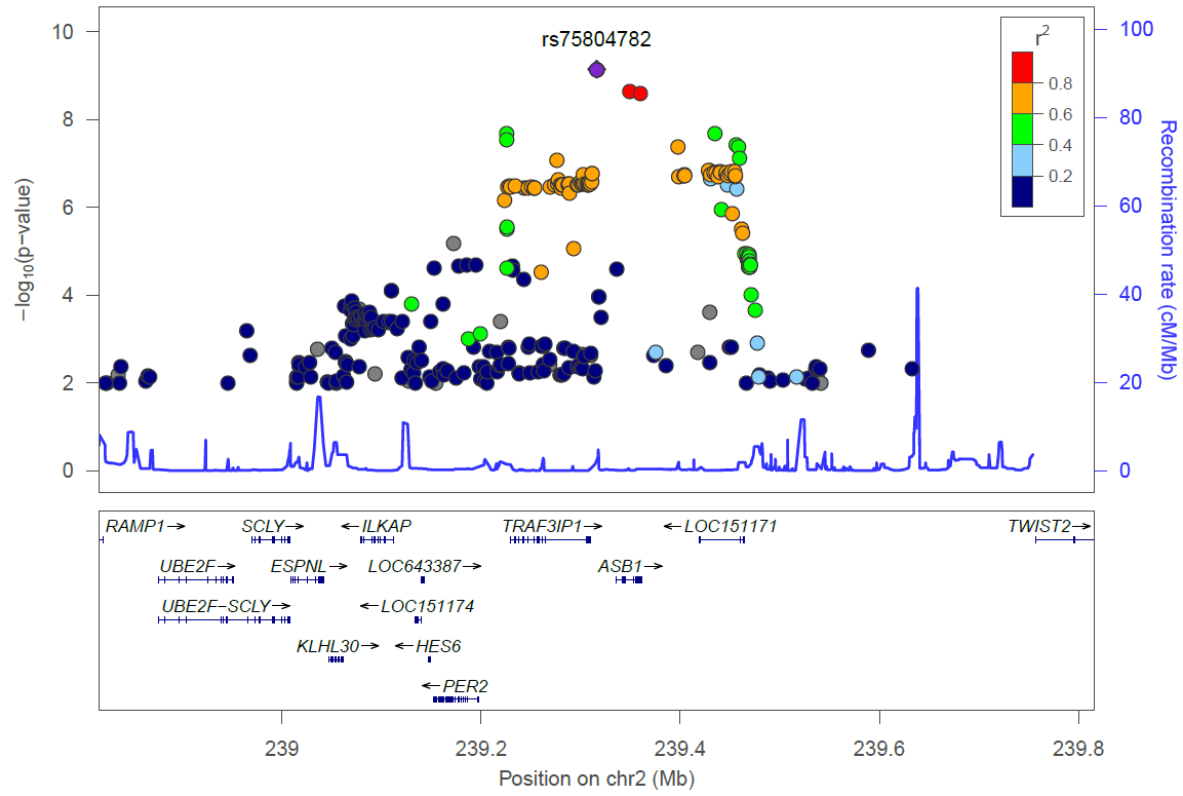
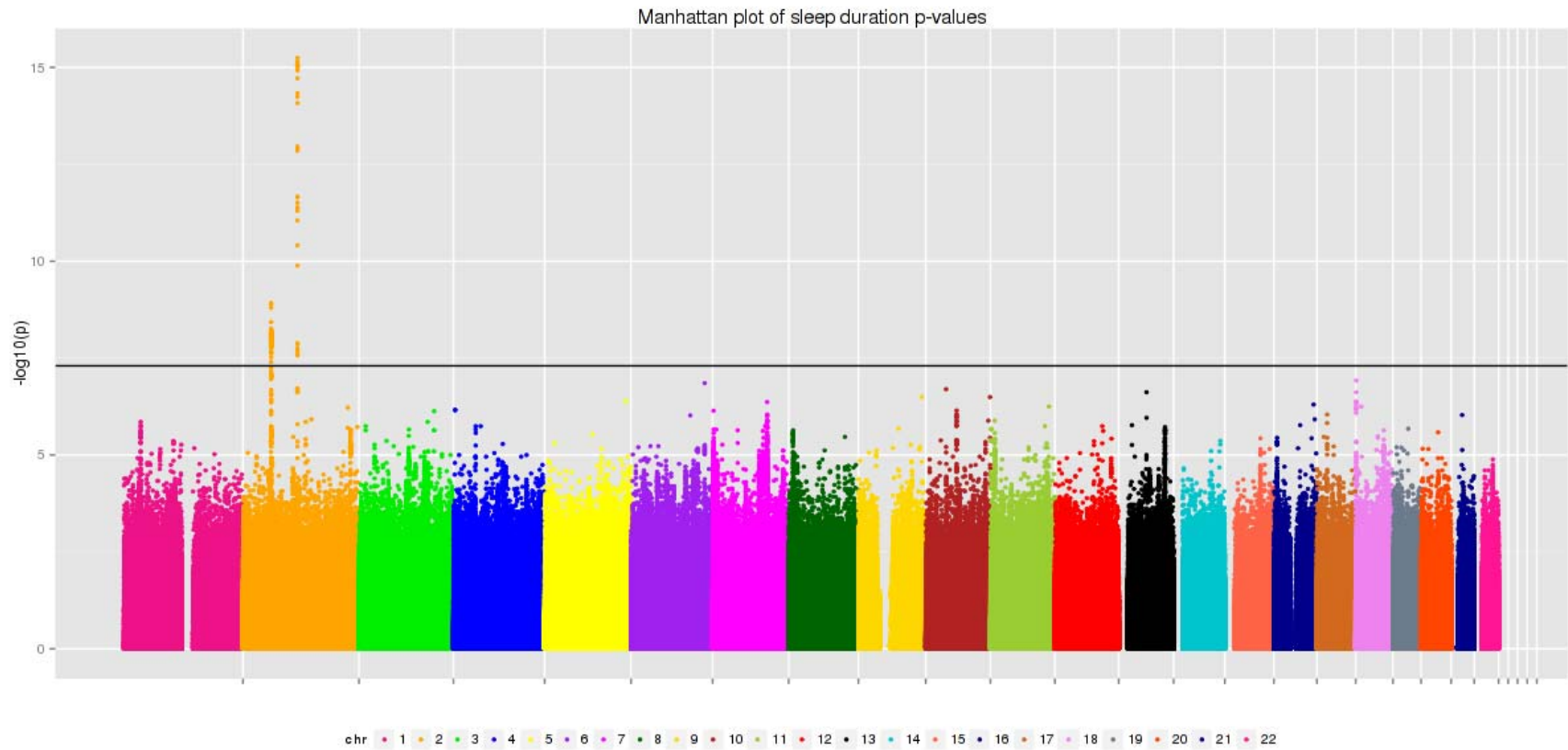




Figure 4: a) Manhattan and b) quantile-quantile plot of hours slept (inverse-normalised) P-values.

a)



b)

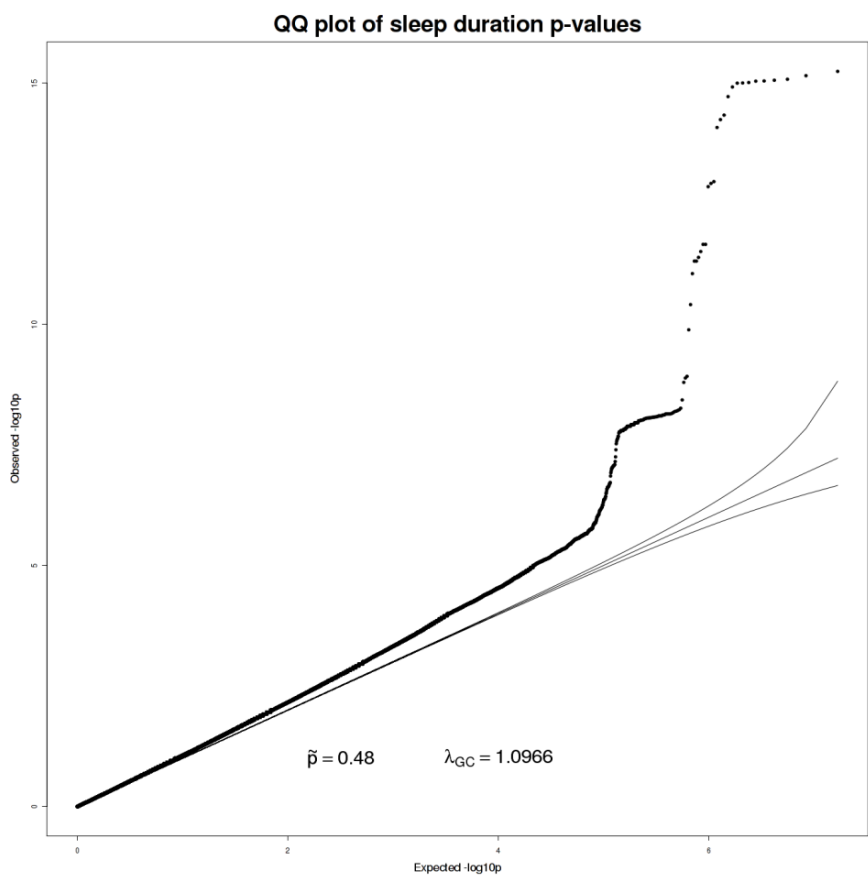
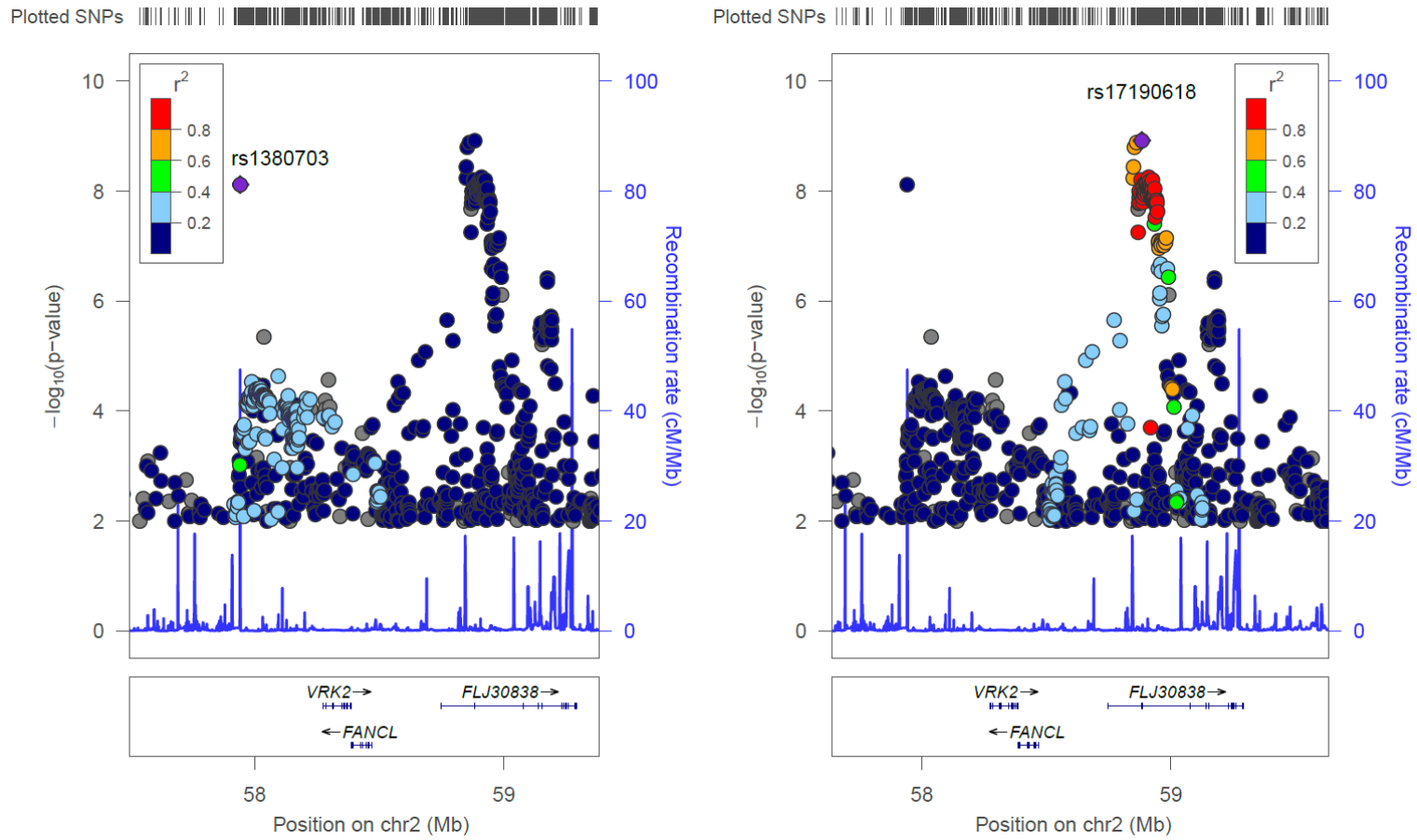
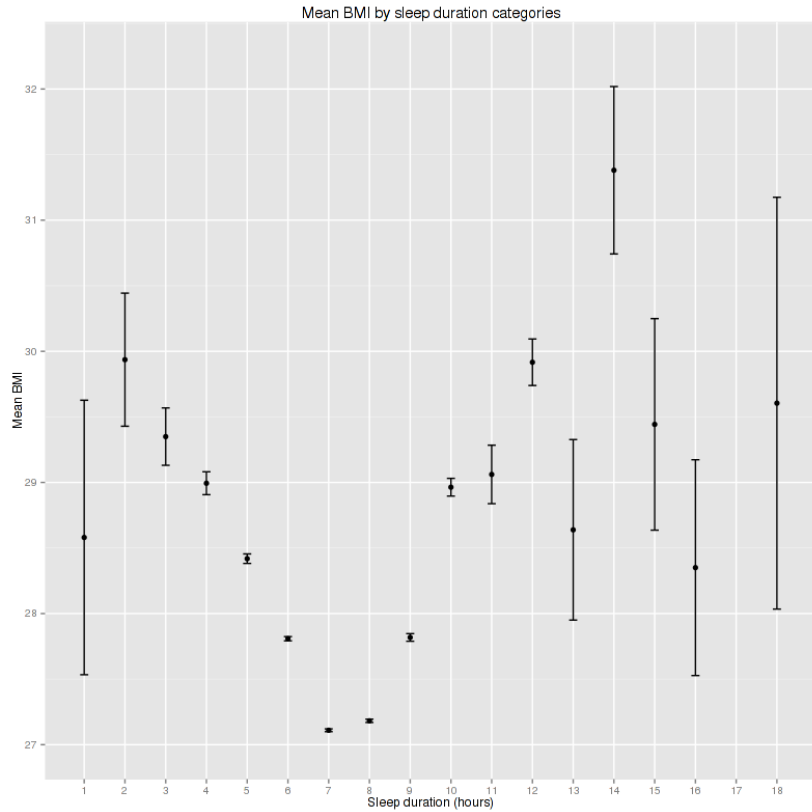


Figure 5: Locuszoom plots for the novel sleep duration associations at the *VRK2* locus

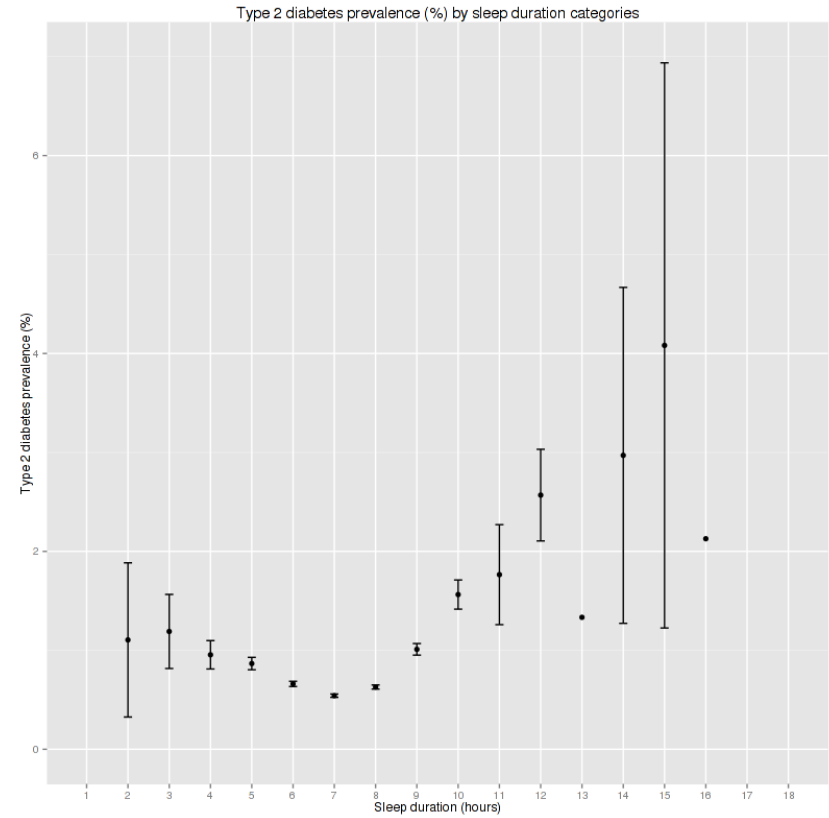


**Supplementary Figure 1: Distribution of a) self-reported BMI and b) proportion of individuals reporting type 2 diabetes in each of the possible sleep duration categories.**

**a)**



**b)**



## References

1. Dibner, C. & Schibler, U. Circadian timing of metabolism in animal models and humans. *J Intern Med* **277**, 513-27 (2015).
2. Duffy, J.F. & Czeisler, C.A. Age-related change in the relationship between circadian period, circadian phase, and diurnal preference in humans. *Neurosci Lett* **318**, 117-20 (2002).
3. von Schantz, M. *et al.* Distribution and heritability of diurnal preference (chronotype) in a rural Brazilian family-based cohort, the Baependi study. *Sci Rep* **5**, 9214 (2015).
4. Cappuccio, F.P. *et al.* Meta-analysis of short sleep duration and obesity in children and adults. *Sleep* **31**, 619-26 (2008).
5. Schmid, S.M., Hallschmid, M. & Schultes, B. The metabolic burden of sleep loss. *Lancet Diabetes Endocrinol* **3**, 52-62 (2015).
6. Reutrakul, S. *et al.* Chronotype is independently associated with glycemic control in type 2 diabetes. *Diabetes Care* **36**, 2523-9 (2013).
7. Yu, J.H. *et al.* Evening chronotype is associated with metabolic disorders and body composition in middle-aged adults. *J Clin Endocrinol Metab* **100**, 1494-502 (2015).
8. Kohsaka, A. *et al.* High-fat diet disrupts behavioral and molecular circadian rhythms in mice. *Cell Metab* **6**, 414-21 (2007).
9. Marcheva, B. *et al.* Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* **466**, 627-31 (2010).
10. Turek, F.W. *et al.* Obesity and metabolic syndrome in circadian Clock mutant mice. *Science* **308**, 1043-5 (2005).
11. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* **23**, R89-98 (2014).
12. Allebrandt, K.V. *et al.* A K(ATP) channel gene effect on sleep duration: from genome-wide association studies to function in *Drosophila*. *Mol Psychiatry* **18**, 122-32 (2013).
13. Gottlieb, D.J. *et al.* Novel loci associated with usual sleep duration: the CHARGE Consortium Genome-Wide Association Study. *Mol Psychiatry* (2014).
14. Doi, M. *et al.* Circadian regulation of intracellular G-protein signalling mediates intercellular synchrony and rhythmicity in the suprachiasmatic nucleus. *Nat Commun* **2**, 327 (2011).
15. van der Horst, G.T. *et al.* Mammalian Cry1 and Cry2 are essential for maintenance of circadian rhythms. *Nature* **398**, 627-30 (1999).
16. Zheng, B. *et al.* The mPer2 gene encodes a functional component of the mammalian circadian clock. *Nature* **400**, 169-73 (1999).
17. Bunger, M.K. *et al.* Mop3 is an essential component of the master circadian pacemaker in mammals. *Cell* **103**, 1009-17 (2000).
18. Shearman, L.P., Jin, X., Lee, C., Reppert, S.M. & Weaver, D.R. Targeted disruption of the mPer3 gene: subtle effects on circadian clock function. *Mol Cell Biol* **20**, 6269-75 (2000).
19. Zheng, B. *et al.* Nonredundant roles of the mPer1 and mPer2 genes in the mammalian circadian clock. *Cell* **105**, 683-94 (2001).
20. Preitner, N. *et al.* The orphan nuclear receptor REV-ERB $\alpha$  controls circadian transcription within the positive limb of the mammalian circadian oscillator. *Cell* **110**, 251-60 (2002).
21. Larsson, M. *et al.* GWAS Findings for Human Iris Patterns: Associations with Variants in Genes that Influence Normal Neuronal Pattern Development. *The American Journal of Human Genetics* **89**, 334-343 (2011).
22. Mazzotta, G. *et al.* Fly cryptochrome and the visual system. *Proc Natl Acad Sci U S A* **110**, 6163-8 (2013).
23. Allebrandt, K.V. *et al.* Chronotype and sleep duration: the influence of season of assessment. *Chronobiol Int* **31**, 731-40 (2014).
24. Segre, A.V. *et al.* Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet* **6**(2010).

25. Locke, A.E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
26. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44**, 981-90 (2012).
27. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-7 (2014).
28. International League Against Epilepsy Consortium on Complex Epilepsies. Electronic address, e.-a.u.e.a. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *Lancet Neurol* **13**, 893-903 (2014).
29. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
30. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-1 (2010).
31. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat Genet* **47**, 1236-1241 (2015).
32. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-295 (2015).
33. Christopher, F.B., Mark, E.S. & Steven, S. IVREG2: Stata module for extended instrumental variables/2SLS and GMM estimation. S425401 edn (Boston College Department of Economics, 2002).
34. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *International Journal of Epidemiology* **44**, 512-525 (2015).