

ARTICLE

Received 22 Oct 2015 | Accepted 29 Jan 2016 | Published 9 Mar 2016

DOI: 10.1038/ncomms10889

OPEN

# Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank

Jacqueline M. Lane<sup>1,2,3</sup>, Irma Vlasac<sup>1,3</sup>, Simon G. Anderson<sup>4</sup>, Simon D. Kyle<sup>5</sup>, William G. Dixon<sup>6</sup>, David A. Bechtold<sup>7</sup>, Shubhroz Gill<sup>8</sup>, Max A. Little<sup>9</sup>, Annemarie Luik<sup>5</sup>, Andrew Loudon<sup>7</sup>, Richard Emsley<sup>10</sup>, Frank A.J.L. Scheer<sup>11,12</sup>, Deborah A. Lawlor<sup>13,14</sup>, Susan Redline<sup>11,12</sup>, David W. Ray<sup>15</sup>, Martin K. Rutter<sup>15,16,\*</sup> & Richa Saxena<sup>1,2,3,11,\*</sup>

Our sleep timing preference, or chronotype, is a manifestation of our internal biological clock. Variation in chronotype has been linked to sleep disorders, cognitive and physical performance, and chronic disease. Here we perform a genome-wide association study of self-reported chronotype within the UK Biobank cohort ( $n=100,420$ ). We identify 12 new genetic loci that implicate known components of the circadian clock machinery and point to previously unstudied genetic variants and candidate genes that might modulate core circadian rhythms or light-sensing pathways. Pathway analyses highlight central nervous and ocular systems and fear-response-related processes. Genetic correlation analysis suggests chronotype shares underlying genetic pathways with schizophrenia, educational attainment and possibly BMI. Further, Mendelian randomization suggests that evening chronotype relates to higher educational attainment. These results not only expand our knowledge of the circadian system in humans but also expose the influence of circadian characteristics over human health and life-history variables such as educational attainment.

<sup>1</sup> Center for Human Genetic Research Massachusetts General Hospital, Boston, Massachusetts 02114, USA. <sup>2</sup> Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. <sup>3</sup> Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. <sup>4</sup> Cardiovascular Research Group, Institute of Cardiovascular Sciences, The University of Manchester, Manchester M139PL, UK. <sup>5</sup> Sleep and Circadian Neuroscience Institute (SCNi), Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford OX12JD, UK. <sup>6</sup> Centre for Musculoskeletal Research Institute of Inflammation and Repair, The University of Manchester, Manchester M139PL, UK. <sup>7</sup> Faculty of Life Sciences, The University of Manchester, Manchester M139PL, UK. <sup>8</sup> Chemical Biology Program, Broad Institute, Cambridge, Massachusetts 02142, USA.

<sup>9</sup> Department of Mathematics, Engineering and Applied Science, Aston University, Birmingham B47ET, UK. <sup>10</sup> Institute of Population Health, The University of Manchester, Manchester M139PL, UK. <sup>11</sup> Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

<sup>12</sup> Division of Sleep Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>13</sup> MRC Integrative Epidemiology Unit at the University of Bristol, Bristol BS81TH, UK. <sup>14</sup> School of Social and Community Medicine, University of Bristol, Bristol BS81TH, UK. <sup>15</sup> Centre for Endocrinology and Diabetes, Institute of Human Development, The University of Manchester, Manchester M139PL, UK. <sup>16</sup> Manchester Diabetes Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester M139PL, UK. \* These authors contributed equally to this work.

Correspondence and requests for materials should be addressed to R.S. (email: rsaxena@partners.org).

**C**hronotype is a behavioural manifestation of our internal timing system, the circadian clock. Individual variation within our biological clock drives our morning or evening preferences, thereby making us into 'morning larks' or 'night owls'. Chronotype is influenced by many factors, including age, sex, social constraints and environmental factors, among others<sup>1</sup>. Chronotype has been associated with sleep disorders, cognitive and physical performance, chronic metabolic and neurologic disease, cancer and premature aging<sup>2</sup>, in particular when there is desynchrony between internal chronotype and external environment increasing disease risk<sup>3</sup>. Despite the importance of circadian rhythms to human health and their fundamental role demonstrated in model organisms<sup>4,5</sup>, little is known about the biological mechanisms underlying inter-individual variation in human chronotype or how it impacts our health and physiology.

Genes that encode molecular components of the core circadian clock (*PER2*, *PER3*) or regulate the pace of the clock (*CSNK1D*) are disrupted in Advanced Sleep Phase Syndrome (ASPS) and Delayed Sleep Phase Syndrome both of which are monogenic circadian rhythm disorders causing extreme advance or delay in sleep onset<sup>6</sup>. ASPS mutations shorten circadian period in humans and mice<sup>7,8</sup>, linking the change in pace of the clock with sleep timing preference. More detailed biochemical and functional characterization of these mutations have greatly enhanced our understanding of the mechanisms regulating the circadian clock. Emerging evidence suggests that subjects with ASPS may be at increased risk for chronic disease, such as cardio-metabolic disease<sup>9</sup> or migraine<sup>10</sup>.

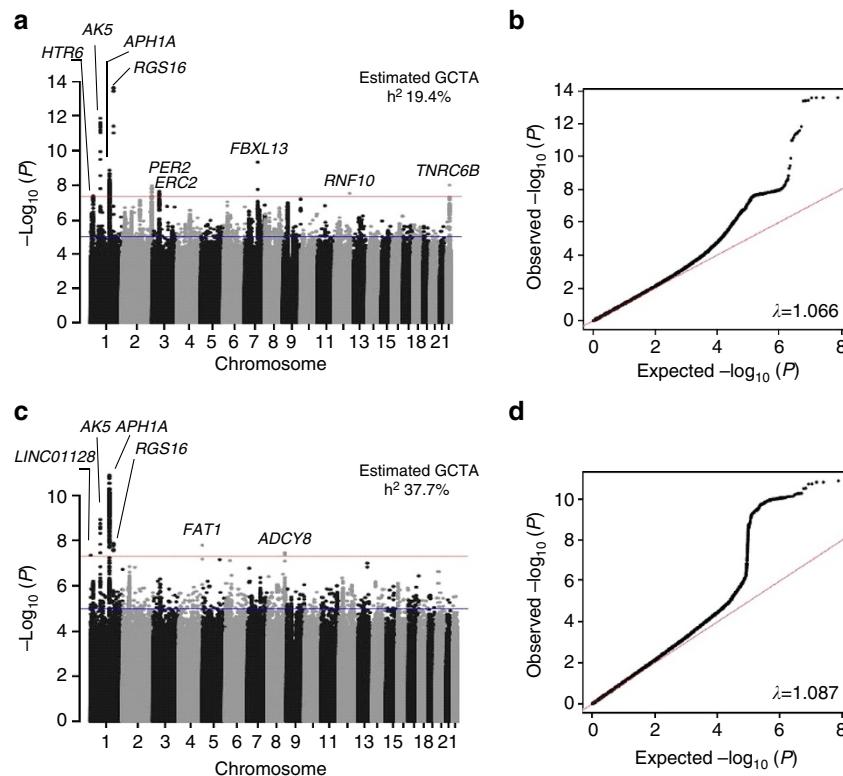
In addition to monogenic sleep phase disorders, pronounced inter-individual variation in chronotype exists within the general population<sup>5</sup>, and epidemiologic associations with adverse health outcomes have been reported<sup>2,11</sup>. Chronotype is heritable as estimated by twin and family studies (12–42%)<sup>12–14</sup> but its

genetic basis has not yet been well defined. Candidate gene association studies have reported variation associated with morningness or eveningness preference in the *CLOCK*, *PER1*, *PER2* and *PER3* genes<sup>15</sup>; however, these studies have often had limited reproducibility, suffering from small sample sizes, heterogeneity in chronotype assessment and inadequate correction for population structure. Recently, a genome-wide association study (GWAS) for self-reported habitual bedtime identified variation in *NPSR1*<sup>12</sup>, but again robust replication of this finding has not been reported. Nonetheless, these studies suggest that novel genetic loci for chronotype, like for other complex traits, may be identified by GWAS provided that sufficiently large cohorts are used.

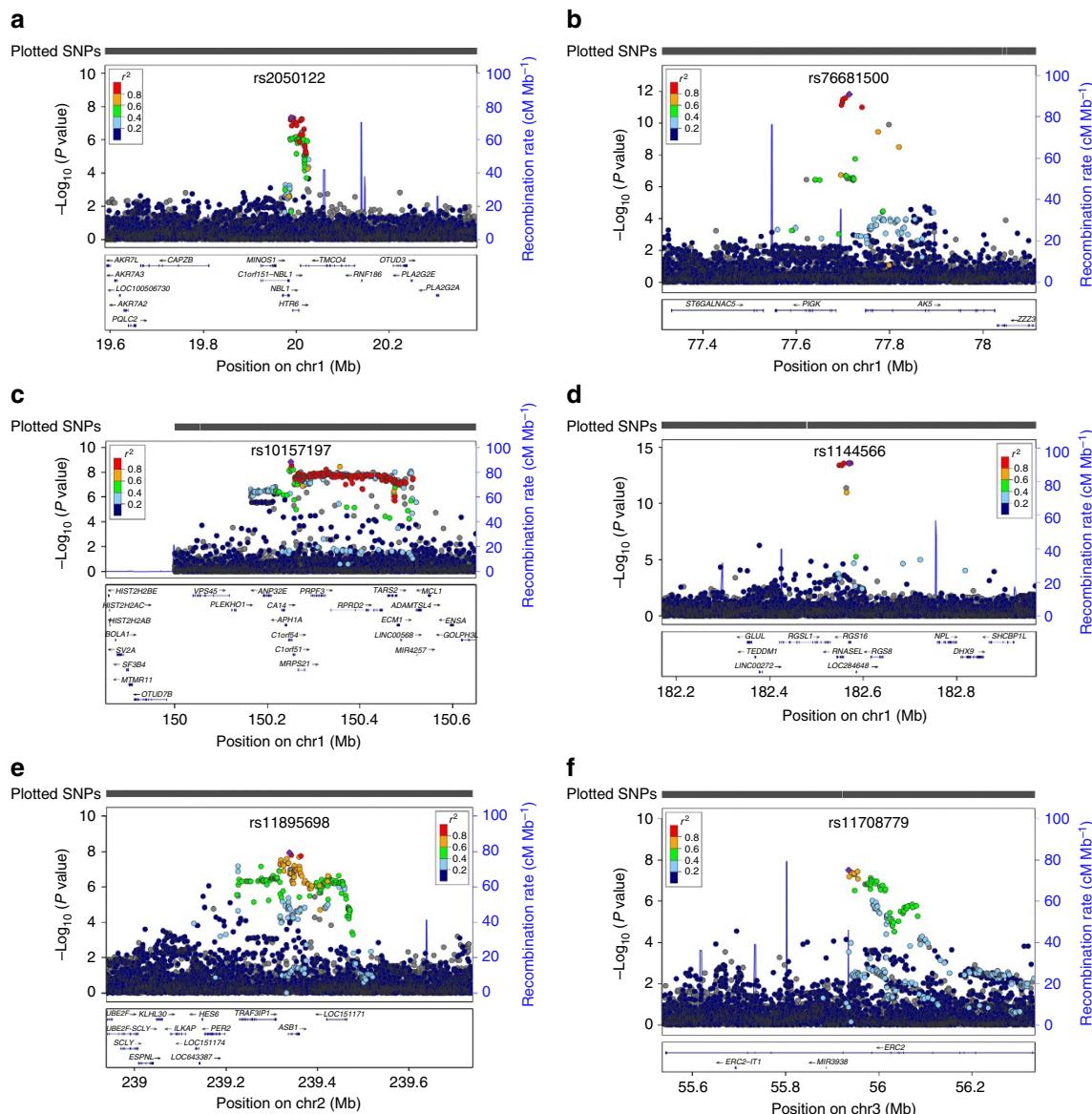
To define the spectrum of genetic variation contributing to variation in human circadian phenotype, and identify associative or causal links between chronotype and other health indices, we perform the largest GWAS of self-reported chronotype to date, within the UK Biobank cohort ( $n = 100,420$ ), a unique resource with an extensive set of individual life history parameters. Self-reported chronotype has been validated in previous studies, and correlates significantly with objectively measured physiological rhythms<sup>16</sup>. Our work identifies several novel genetic loci that associate significantly with chronotype, and importantly reveals a significant genetic correlation between chronotype and schizophrenia risk, BMI and educational attainment.

## Results

**Twelve genome-wide significant association signals.** Variation in chronotype associated significantly with age, sex, sleep duration, depression and psychiatric medication use, with 'eveningness' being associated with younger age, being male, having a longer sleep duration, being more likely to be depressed or using psychiatric medication (Supplementary Table 1).



**Figure 1 | Manhattan and Q-Q plots for genome-wide association analysis of both continuous and extreme chronotype.** Manhattan plots (a,c). Red line is genome-wide significant ( $5 \times 10^{-8}$ ) and blue line is suggestive ( $1 \times 10^{-6}$ ). Q-Q plots (b,d). Nearest gene name is annotated. Heritability estimates were calculated using BOLT-REML and lambda inflation values using GenABEL in R.



**Figure 2 | Regional association plots for genome-wide significant chronotype loci.** (a-i) show loci associated with continuous chronotype, (j-l) show loci associated with extreme chronotype. Genes within the region are shown in the lower panel. The blue line indicates the recombination rate. Filled circles show the  $-\log_{10} P$  value for each SNP, with the lead SNP shown in purple. Additional SNPs in the locus are coloured according to correlation ( $r^2$ ) with the lead SNP (estimated by LocusZoom based on the CEU HapMap haplotypes). \*chr7 rs372229746 is not in the reference panel, therefore LD data is unavailable for this SNP.

These characteristics together explained 1.4% of variation in chronotype.

Two parallel primary GWAS analyses of genotyped and imputed single nucleotide polymorphisms (SNPs) were performed using regression models adjusting for age, sex, 10 principal components of ancestry and genotyping array: an ordinal score of chronotype based on four categories from 'definite morning' to 'definite evening' treated as a continuous trait, using the whole population ( $n=100,420$ ) and a binary variable of chronotype extremes (8,724 definite evening-type cases versus 26,948 definite morning-type controls), to enrich for rarer variants expected to have stronger effects. In total, 12 genome-wide significant loci were identified (Figs 1 and 2; Table 1; and Supplementary Fig. 1;  $P<5\times 10^{-8}$ ) of which three surpassed genome-wide significance in both analyses (Table 1). Association was observed near *PER2*, an ASPS gene, and three other association signals were found in or near genes with a well-

known role in circadian rhythms (*APH1A*, *RGS16* and *FBXL13*), consistent with the hypothesis that circadian clock biology contributes to variation in chronotype. Conditional analyses at the 12 loci implicated one suggestive secondary association signal, a missense variant (V903I) in the core circadian clock gene *PER2* ( $P=8.43\times 10^{-8}$ ) predicted to be damaging (Polyphen 0.984, CADD scaled 16.21; Supplementary Table 2); thus, confirming that core circadian clock genes disrupted in ASPS harbour common variants that contribute to variation in chronotype. Together, in the discovery sample, the 12 loci explain 4.3% of variance in chronotype. Credible set analyses<sup>17</sup> highlight a limited number of potential causal variants at each locus (Table 1).

Robustness of the self-reported chronotype trait and genetic loci identified here was further validated by an independent GWAS of extreme chronotype from Hu *et al.*<sup>18</sup> In total, 8 of 15 reported loci replicated in our study, and all 15 showed a consistent direction of effect in our study. Three additional loci

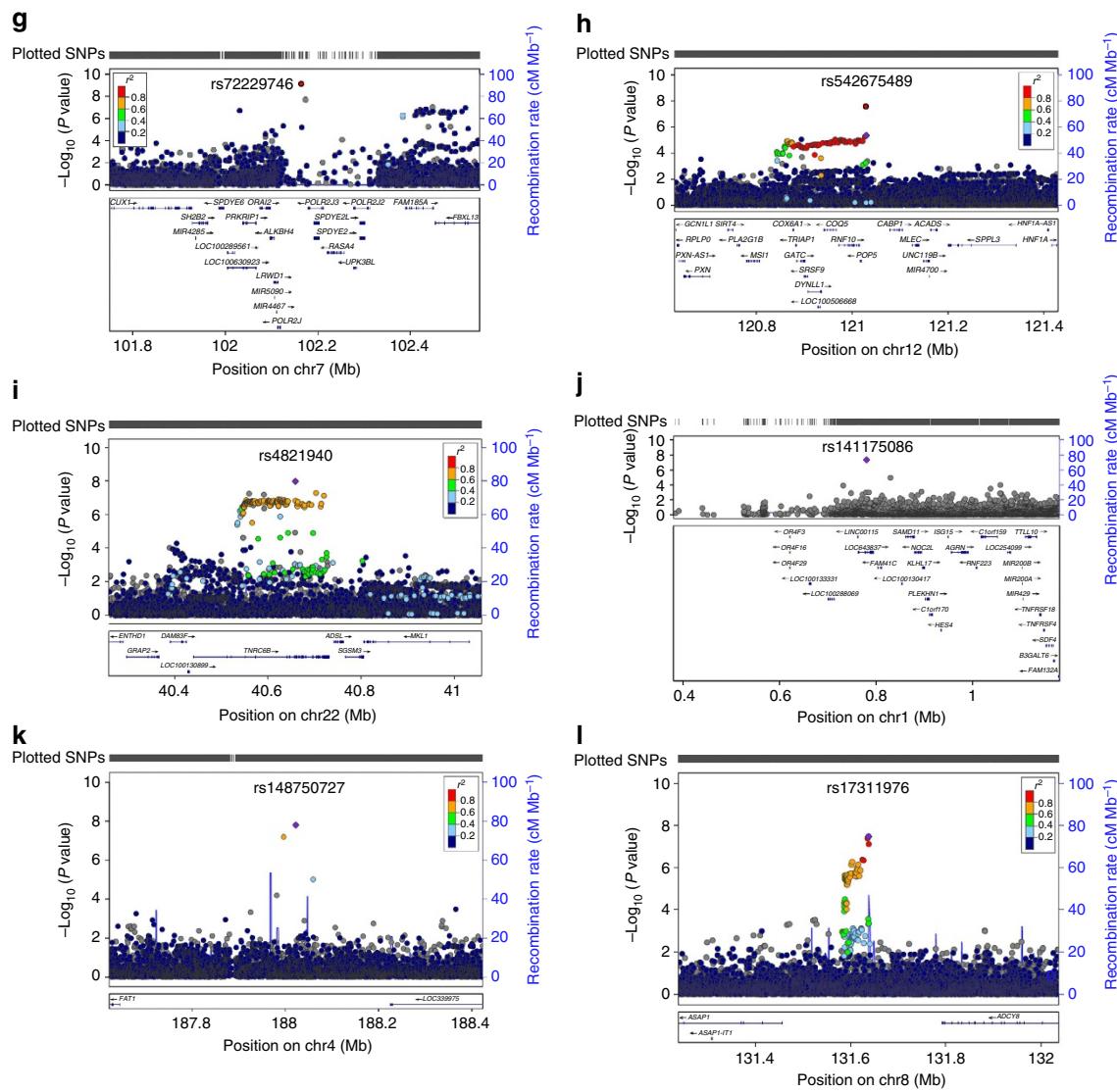


Figure 2 | Continued.

attain genome-wide significance in meta-analysis of both studies using publicly available results for the 15 SNPs from Hu *et al.* (near genes *PER3*, *VIP* and *TOX3*; Supplementary Table 3).

No evidence of association was observed for previously reported SNPs from other candidate gene or GWA studies (Supplementary Table 4). The *PER3* VNTR (rs57875989) previously associated with chronotype<sup>19</sup> was not directly genotyped or imputed in this study; nonetheless, a suggestive association signal was observed encompassing this region of *PER3* (lead SNP: rs7545893  $P = 6.5 \times 10^{-8}$ ; 33 kb from *PER3* VNTR) and largely independent from the lead 23andMe *PER3* region SNP ( $r^2 = 0.186$  in 1KG CEU<sup>18</sup>; Supplementary Fig. 2).

Secondary analyses were performed on the 12 lead SNPs within the chronotype loci, including (1) separate comparison of effects on morningness and eveningness, (2) sex-specific analysis, (3) pair-wise genetic interaction analysis and (4) regression models including additional covariates. Comparison of case extremes (8,724 evening or 26,948 morning) to the collapsed middle group ( $n = 64,748$ ) revealed three loci (*LINC01128*, *APH1A*, *FAT1*) with stronger effects in the eveningness case-control analysis as opposed to morningness analysis. A rare variant (0.2%) at the *LINC01128* locus (rs141175086C) exhibited the most striking protective effect for eveningness (odds ratio (OR) = 0.22

(0.10–0.50),  $P = 1.7 \times 10^{-5}$ ) but only a small risk effect for morningness (OR = 1.30 (0.97–1.75),  $P = 0.08$ ; Supplementary Table 5). No significant sex-specific effects (Supplementary Table 6) or epistasis between loci (Supplementary Table 7) were detected. Similarly, sensitivity analyses adjusting for factors known to be associated with chronotype, including sleep duration and disorders, depression and psychiatric medication use did not significantly alter the effect estimates or strength of the associations (Supplementary Table 8).

Candidate causal genes at these loci are highlighted in Supplementary Note 1. The 12 loci encompass 72 candidate genes enriched in pathways for circadian rhythms ( $P_{\text{adj}} = 0.014$ ), mental disorders ( $P_{\text{adj}} = 0.001$ ), sleep disorders ( $P_{\text{adj}} = 0.005$ ), the spliceosome ( $P_{\text{adj}} = 0.020$ ) and Alzheimer's disease ( $P_{\text{adj}} = 0.030$ ) among others (Supplementary Table 9). In addition, four loci are located in or near genes with a well-known role in circadian rhythms (*PER2*, *APH1A*, *RGS16* and *FBXL13*), however whether these genes are responsible for the association signals observed remains to be established. The remaining eight loci offer the potential of novel biological insights into circadian rhythms (Supplementary Note 1). Several candidate causal genes have been implicated in circadian rhythms. *TNRC6B* controls circadian behaviour in flies<sup>20</sup> and is bound by known circadian

**Table 1 | Genome-wide significant loci associated with chronotype in subjects of European ancestry in the UK Biobank.**

SNP	Chr:position NCBI 37	Nearest Gene	Alleles (E/A)	EAF	Continuous chronotype (n = 100,420) (1–4 ranging from 'definite morning' to 'definite evening')		Extreme chronotype (n = 8,724 evening-type cases, n = 26,948 morning-type controls)		Most likely causal SNPs (probability)
					Beta (s.e.)	P value	OR (95% CI)	P value	
rs141175086	1:780,397	LINC01128	<b>C/T</b>	0.998	0.266 (0.070)	$1.42 \times 10^{-4}$	2.16 (1.34–3.49)	<b><math>4.38 \times 10^{-8}</math></b>	rs141175086 (1.00)
rs2050122	1:19,989,205	HTR6	<b>C/T</b>	0.80	0.031 (0.006)	<b><math>4.61 \times 10^{-8}</math></b>	1.12 (1.07–1.17)	$7.94 \times 10^{-7}$	rs2050122 (0.385)
rs76681500	1:77,713,434	AK5	<b>G/A</b>	0.84	0.043 (0.006)	<b><math>1.50 \times 10^{-12}</math></b>	1.16 (1.10–1.22)	<b><math>1.77 \times 10^{-9}</math></b>	rs76681500 (0.665)
rs10157197	1:150,250,636	APH1A	<b>A/G</b>	0.40	0.028 (0.005)	<b><math>1.48 \times 10^{-9}</math></b>	1.13 (1.09–1.17)	<b><math>1.27 \times 10^{-11}</math></b>	rs10157197 (0.085), rs11205355 (0.085)
rs1144566	1:182,569,626	RGS16	<b>C/T</b>	0.97	0.099 (0.013)	<b><math>2.62 \times 10^{-14}</math></b>	1.35 (1.22–1.50)	<b><math>1.29 \times 10^{-8}</math></b>	rs1144566 (0.213), rs694383 (0.213), rs12743617 (0.213), rs509476 (0.213)
rs11895698	2:239,338,495	ASB1	<b>T/C</b>	0.14	0.035 (0.006)	<b><math>1.15 \times 10^{-8}</math></b>	1.10 (1.05–1.15)	$1.30 \times 10^{-4}$	rs11895698 (0.279), rs3769118 (0.279)
rs11708779	3:55,934,939	ERC2	<b>G/A</b>	0.65	0.026 (0.005)	<b><math>3.08 \times 10^{-8}</math></b>	1.09 (1.06–1.13)	$1.20 \times 10^{-6}$	rs11708779 (0.769)
rs148750727	4:188,022,952	FAT1	<b>T/G</b>	0.995	0.154 (0.033)	$3.61 \times 10^{-6}$	2.34 (1.69–3.23)	<b><math>1.58 \times 10^{-8}</math></b>	rs148750727 (1.00)
rs372229746	7:102,158,815	FBXL13	A/G	0.45	0.034 (0.005)	<b><math>5.18 \times 10^{-10}</math></b>	1.12 (1.07–1.16)	$4.29 \times 10^{-7}$	rs372229746 (1.00)
rs17311976	8:131,637,337	ADCY8	<b>C/T</b>	0.19	0.028 (0.006)	$1.08 \times 10^{-6}$	1.13 (1.08–1.18)	<b><math>3.37 \times 10^{-8}</math></b>	rs17311976 (0.346), rs17311976 (0.165), rs17311976 (0.107)
rs542675489	12:120,994,888	RNF10	<b>C/CA</b>	0.60	0.027 (0.005)	<b><math>3.29 \times 10^{-9}</math></b>	1.09 (1.05–1.13)	$1.38 \times 10^{-5}$	rs542675489 (1.00)
rs4821940	22:40,659,573	TNRC6B	<b>C/T</b>	0.55	0.026 (0.004)	<b><math>1.05 \times 10^{-8}</math></b>	1.07 (1.04–1.11)	$8.56 \times 10^{-5}$	rs4821940 (0.172)
<i>Suggestive secondary signal</i>									
rs35333999	2:239161957	PER2	<b>T/C</b>	0.043	0.059 (0.011)	$8.43 \times 10^{-8}$	1.21 (1.11–1.32)	$9.01 \times 10^{-6}$	rs35333999 (0.895)

A, alternative allele; Chr, chromosome; CI, confidence interval; E, effect allele; EAF, effect allele frequency; OR, odds ratio; TSLS, two-stage-least-squares method.

Ancestral allele is indicated in bold. Candidate putative causal SNPs were identified by PIC3<sup>17</sup>. Note, increasing beta and odds ratio indicate greater evening type preference.

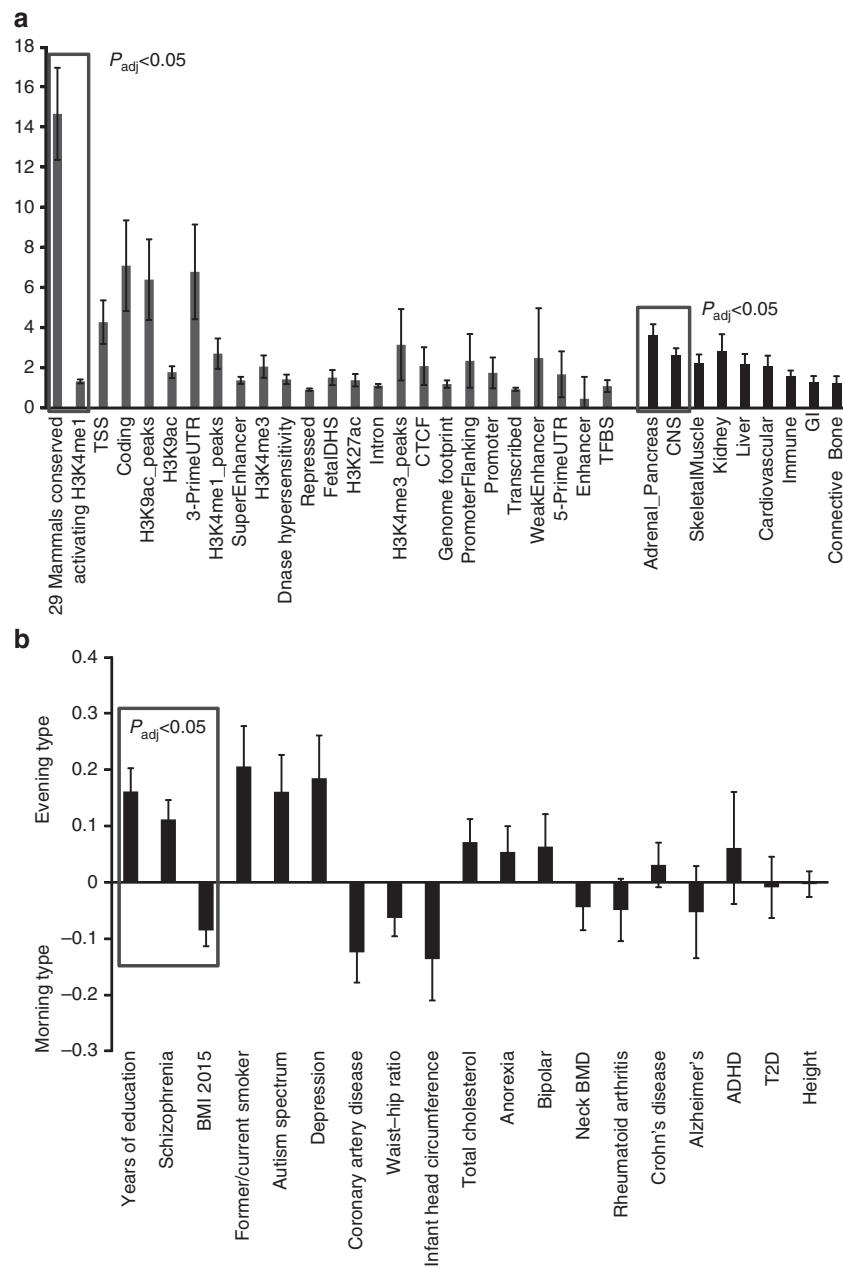
transcription factors in mouse liver<sup>21</sup>. *MCL1* has rhythmically expressed mRNA in liver<sup>22</sup>, disrupts circadian rhythms in an RNAi screen using a human osteosarcoma cell line<sup>23</sup>, and is bound by known circadian transcription factors in mouse liver<sup>21</sup>. *HTR6* is a G-protein-coupled receptor known to regulate the sleep–wake cycle<sup>24–26</sup>.

Fine-mapping, sequencing and experimental studies are necessary to identify the causal gene(s) and variant(s) at each locus to understand mechanisms by which DNA variation influences variation in chronotype. However, clues may emerge from exploration of bioinformatic annotations of candidate regulatory variants and ENCODE analyses of chromatin states and bound proteins<sup>27</sup>. For example, rare variant rs141175086 is predicted to disrupt a binding site for the known circadian transcription factor DEC1 in an enhancer element within or upstream of previously uncharacterized lncRNAs (*LOC643837*, *LINC01128*).

**Pathway analyses.** Heritability of chronotype, captured by genome-wide genotypes in this study, was estimated to be 19.4% (continuous) and 37.7% (extreme) using GCTA<sup>28</sup>. Heritability partitioning of continuous chronotype GWAS by tissue and functional category using LD-score regression<sup>29</sup> identified enrichment in the central nervous system (enrichment 2.63,  $P = 1.91 \times 10^{-6}$ ) and adrenal/pancreatic tissues (enrichment 3.63,  $P = 1.34 \times 10^{-8}$ ; Fig. 3a and Supplementary Table 10). Regions of the genome annotated as highly conserved across mammals<sup>30</sup> (enrichment 14.33,  $P = 1.75 \times 10^{-9}$ ), and in regions of histone 3 lysine 4 monomethylation that mark active/poised enhancer elements (enrichment 1.30,  $P = 0.0017$ ; Fig. 3a and Supplementary Table 10) were significantly enriched, supporting a key role of circadian rhythms throughout mammalian evolution.

Gene-based analysis<sup>31</sup> identified 23 genes significantly associated with chronotype ( $P < 2.8 \times 10^{-6}$ , Supplementary Table 11; Supplementary Fig. 3). Pathway analysis<sup>32</sup> shows a significant enrichment in this gene set for genes previously implicated in Alzheimer's disease ( $P_{adj} = 0.0176$ ) and dementia ( $P_{adj} = 0.0192$ ), eye abnormalities ( $P_{adj} = 0.0176$ ) and eye diseases ( $P_{adj} = 0.0253$ ), chromosomal deletions ( $P_{adj} = 0.0253$ ), brain diseases ( $P_{adj} = 0.0253$ ), central nervous system diseases ( $P_{adj} = 0.0253$ ) and mental disorders ( $P_{adj} = 0.0365$ ). In support, integrative analysis of signals with  $P < 1 \times 10^{-5}$  using DEPICT<sup>33</sup>, a tool that uses predicted gene functions to prioritize genes, gene sets and tissues, showed suggestive enrichment in gene sets associated with 'fear response' and 'behavioural defence response' (False Discovery Rate  $< 0.20$ ), and central nervous and hemic/immune system tissues (Supplementary Table 12). In total, pathway analyses link the genetics of chronotype to central nervous system function and neurological disorders including dementia and affective disorders.

**Genetic links with schizophrenia and educational attainment.** Given that circadian rhythms play a fundamental role in human physiology, a key question is the extent to which the genetics of chronotype is shared with other behavioural or disease states, and importantly whether genetic relationships between chronotype and other traits are causal. To address this, we tested for genetic correlation of chronotype with GWAS variants for 19 phenotypes spanning a range of cognitive, neuro-psychiatric, anthropometric, cardio-metabolic and auto-immune traits using LD score regression on chronotype GWAS and publicly available GWAS for each trait<sup>34</sup>. Genetic correlations suggested that tendency towards an evening chronotype is related to greater years of education ( $r_g$  (s.e.) 0.161 (0.041),  $P = 8.96 \times 10^{-5}$ ) and increased schizophrenia risk ( $r_g$  (s.e.) 0.112 (0.034),  $P = 0.0011$  (Fig. 3b;

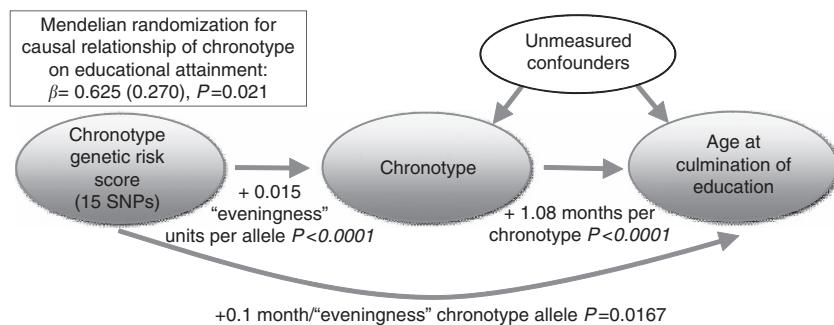


**Figure 3 | Overall genetic architecture of chronotype across tissues, functional categories and cross-trait genetic correlation.** (a) Enrichment estimates for the main annotations and tissues of LDSC. Error bars represent 95% confidence intervals around the estimate. Categories are sorted by  $P$  value, with boxes indicating annotations or tissues that pass the multiple testing significance threshold. (b) Chronotype regression estimates of genetic correlation with the summary statistics from 19 publicly available genome-wide association studies of psychiatric and metabolic disorders, immune diseases and other traits of natural variation. The horizontal axis indicates the phenotype compared with categorical chronotype and the vertical axis indicates genetic correlation. Error bars are s.e.'s. ADHD, attention deficit hyperactivity disorder; BMI, body mass index; CNS, central nervous system; CTCF, CCCTC-binding transcription factor; DHS, DNase hypersensitivity; GI, gastrointestinal; T2D, type 2 diabetes; TFBS, transcription factor binding site; Tss, transcription start site; UTR, untranslated region.

Supplementary Table 10). Genetic correlations also suggested that a morning chronotype may share underlying biology with increased BMI ( $r_g$  (s.e.) = 0.0851 (0.0281),  $P = 0.0025$ ; Fig. 3b; Supplementary Table 10).

**Mendelian randomization analyses.** To explore whether the relationship between chronotype and traits with significant genetic correlations might be causal, we tested for association of a risk score of genome-wide significant chronotype SNPs from 23andMe<sup>18</sup> with years of education, schizophrenia and BMI. SNPs can be used as instrument variables to test for a causal

relationship between two traits, and because genotypes are assigned randomly at meiosis, genetic association is not biased by confounding or reverse causation possible in observational epidemiology<sup>35,36</sup>. Since individuals do not know their genotype any phenotypic misclassification will be random with respect to genotype. In the UK Biobank, a significant association was observed between a chronotype genetic risk score of SNPs related to eveningness and increased educational attainment ( $P = 0.0167$ ), but not schizophrenia ( $P = 0.101$ ) or BMI ( $P = 0.285$ ; Supplementary Table 13). Further instrumental variable analyses suggested that for each increase in



**Figure 4 | Mendelian Randomization.** under the assumptions of instrumental variable Mendelian randomization analyses<sup>70</sup>, our results show that having an evening chronotype results in higher educational attainment. In this analysis, for the chronotype risk score (comprised of 15 SNPs from the 23andMe GWAS of chronotype, weighted by effect size), the  $\beta$  coefficient for the association with chronotype was regressed on the  $\beta$ -coefficient for the association with the main educational attainment trait in the UK Biobank ( $n = 68,718$ ) using TSLS. TSLS, two-stage-least-squares method.

'eveningness' category, educational attainment increased by 7.5 months ( $P = 0.021$ ) (Fig. 4; Supplementary Fig. 4; Supplementary Table 13). We then tested for reverse causation by assessing whether variation in education, schizophrenia or BMI might cause variation in chronotype by testing for association of risk scores for each of these traits obtained from prior large-scale GWAS studies with chronotype. No significant associations were observed (Supplementary Table 13).

## Discussion

In this largest GWAS of chronotype to date, we report the discovery of 12 genetic loci associated with chronotype, and pathway analysis suggests key roles of genes in the nervous and ocular systems. Further, we demonstrate shared biology of chronotype with schizophrenia, and possibly BMI, with a putative causal link to educational attainment.

Several lines of evidence support these association signals as true positives that may help to uncover new aspects of circadian biology in humans. First, we detect signals in or near known circadian genes at 4 of the 12 loci, including near and in *PER2*, a clock gene previously associated with *ASPS*<sup>6</sup>. Second, three of these signals have been observed in an independent GWAS<sup>18</sup> suggesting independent validation of our findings. Third, novel associated loci include candidate central circadian clock genes with rhythmic expression in the SCN or circadian behavioural phenotypes in model organisms. Fourth, genes under association peaks are enriched for central nervous system and ocular processes, both important for generation of circadian rhythms. Additional replication to confirm chronotype genetic associations and functional follow-up will be necessary to identify causal genes and circuits disrupted by causal variants at these loci.

Our study also defines the genetic architecture of self-reported chronotype, revealing heritability estimates consistent with previous literature<sup>12–14</sup>, despite using a different questionnaire instrument than previous studies<sup>16</sup>. The 12 genome-wide significant loci appear to explain a large fraction of chronotype variance (4.3%) but this may be overestimated due to winners curse, or may reflect lower polygenicity of chronotype than seen for other complex traits, since variation in a limited number of biological processes (light-sensing, core circadian clock and limited downstream effectors) may be causal. Significant enrichment of heritability in highly conserved regions is consistent with the strong conservation of circadian rhythms throughout evolution<sup>37</sup> and may aid in fine mapping of causal variants and creation of faithful animal models for future experimental studies. Similarly, enrichment of heritability in activating enhancer sites and borderline enrichment in transcriptional start sites is consistent with the role of the

circadian molecular clock in fine tuning of transcriptional regulation<sup>23</sup>.

The association signals at loci identified by our study when combined with signals from 23andMe cover genes identified in GWAS for restless legs syndrome and Mendelian and model organism studies of narcolepsy, suggesting overlap with other sleep traits. Genetic variants in the region of *TOX3* have been previously associated with restless legs syndrome in a GWAS<sup>38</sup>. Although the chronotype-associated variant (rs12927162) is not in linkage disequilibrium with the lead restless legs syndrome variant (rs3104767;  $r^2 < 0.001$  in 1KG CEU population), it does suggest that *TOX3* may have a broader role in basic sleep/circadian physiology. In addition, rare forms of severe early onset narcolepsy in humans<sup>39</sup> and familial narcolepsy in canines<sup>40</sup> are caused by mutations in *HCRTR2*, again suggesting shared underlying biology.

Chronotype has previously been associated with many behaviours and diseases, such as cardiovascular disease, type 2 diabetes, metabolic disorders, risk-taking behaviour, cancer, psychiatric disorders and even creativity<sup>1,3</sup>. Comparing the genetic architecture of chronotype captured in this study with an initial series of select phenotypes with publicly available GWAS data, we identified significant genetic overlap between chronotype and schizophrenia, educational attainment and possibly BMI. Previous literature links evening chronotype with schizophrenia<sup>41–43</sup>, consistent with our findings. These studies also demonstrate severe circadian sleep/wake disruptions in people with schizophrenia, indicating that this relationship may be bidirectional. However, our Mendelian randomization analyses did not support causal relationships between these two. It is possible that even with our large sample size, we are under-powered to rule out an effect of schizophrenia on chronotype.

We detect a surprising putative genetic link between morning chronotype and higher BMI. Previous observational studies have shown association of evening chronotype with higher BMI, poorer dietary habits and decreased inhibitions<sup>44–47</sup>. Consistently, we noted an association between evening chronotype and BMI ( $\beta = 1.003$  BMI units/chronotype;  $P = 1 \times 10^{-4}$ ;  $r = 0.011$ ). Our genetic correlation analyses suggest the intriguing possibility that some underlying pathways contributing to morning chronotype might also contribute to increased BMI. However, independent replication and further large studies are required to fully understand the genetic and causal relationships between chronotype and BMI.

Until now, it has been difficult to discern causal relationships between chronotype and other traits because of potential biases in observational studies, for example, due to confounding or reverse causality, which are unlikely to affect genetic studies<sup>48</sup>. Our work

suggests that tendency to eveningness chronotype is potentially causally related to increased educational attainment, but replication of these findings, and more comprehensive assessment of potential sources of bias will require future investigation. Previous studies have reported that night owls earn a larger mean income than their earlier rising counterparts<sup>49</sup>. Another study, performed at a top-ranked business school, demonstrated higher GMAT scores in evening types even within a high achieving group<sup>50</sup>. However, it is possible these results are impacted by misclassification in our self-reported measurement of chronotype. Although the question clearly asks for preference, participants might have been influenced by the reality of their working lives. Those from more deprived socioeconomic positions might have occupations that are more restrictive in terms of working hours and hence less able to 'adhere' to their preference. If this results in a relationship between socioeconomic position and misclassification then socioeconomic position would confound any observational associations. However, since participants are extremely unlikely to know their genotype for the variants we have identified, any misclassification of chronotype by genotype will be random with the expectation that the genetic correlation and Mendelian randomization studies would be biased towards the null.

Our study is well-powered to detect genetic variants associated with chronotype, with previous studies demonstrating the power of a sample size  $>100,000$  for detecting genetic effects<sup>51</sup>. The study uses a single harmonized question across a large cohort, which is in contrast with previous studies that needed to harmonize data across several cohorts with varying measures of chronotype. Our measure of chronotype is based on self-identification, and may reflect timing preference more so than objective measures of chronotype and since it does not take weekday and weekend behaviour into account, any misclassification may be related to occupation and/or socioeconomic position. However, as noted above, for our genetic correlation and Mendelian randomization analyses this would be expected to bias findings towards the null. Our cohort is aged 40 to 69 years and of European ancestry, which reduces the likelihood of bias due to population structure, but means we cannot necessarily assume our results generalize to other groups. That said the distribution of chronotype is consistent with that found in previous studies<sup>52–54</sup>.

In summary, in a large-scale GWAS of chronotype, we identified 12 new genetic loci that implicate known components of the circadian clock machinery and point to previously unstudied genetic variants and candidate genes that might modulate core circadian rhythms or light-sensing pathways. Furthermore, genome-wide analysis suggests that chronotype shares underlying genetic pathways with educational attainment, schizophrenia and possibly BMI, and that evening chronotype might be causally related to higher educational attainment. This work should advance biological understanding of the molecular processes underlying circadian rhythms, and open avenues for future research in the potential of modulating circadian biology to aide prevention and treatment of associated diseases.

## Methods

**Population and study design.** Study participants were from the UK Biobank study, described in detail elsewhere<sup>55</sup>. In brief, the UK Biobank is a prospective study of  $>500,000$  people living in the United Kingdom. All people in the National Health Service registry who were aged 40–69 years and living  $<25$  miles from a study centre were invited to participate between 2006 and 2010. In total, 503,325 participants were recruited from over 9.2 million mailed invitations. Self-reported baseline data was collected by questionnaire and anthropometric assessments were performed. For the current analysis, individuals of non-white ethnicity were excluded to avoid confounding effects.

**Chronotype and covariate measures.** Study subjects self-reported chronotype, sleep duration, depression, medication use, age, and sex on a touch-screen questionnaire. Chronotype was derived from responses to a chronotype question that participants answered, along with other study questions, on a touch-screen computer at each assessment centre. The question was taken from the Morningness–Eveningness questionnaire; it is the question from that questionnaire that explains the highest fraction of variance in preferences in sleep–wake timing and is an accepted measure of chronotype<sup>54</sup>. The question asks: 'Do you consider yourself to be...' with response options 'Definitely a 'morning' person', 'More a 'morning' than 'evening' person', 'More an 'evening' than a 'morning' person', 'Definitely an 'evening' person', 'Do not know', 'Prefer not to answer'. This question specifically does not ask about actual sleeping pattern, nor does it distinguish between weekday and weekend behaviour and was accessed at the time of exam, which crosses days of the week and seasons across participants. In all, 498,450 subjects answered this question, but only the 153,000 with genetic data were considered for this analysis. Subjects who responded 'Do not know' or 'Prefer not to answer' were set to missing. Chronotype was treated both as a continuous trait, with chronotype coded 1–4, where 1 represents definite morning chronotype, and a dichotomous trait, with definite morning responders set to control ( $n = 26,948$ ) and definite evening responders set to case ( $n = 8,724$ ). Depression was reported in answer to the question 'How often did you feel down, depressed or hopeless mood in last 2 weeks?' (cases,  $n = 4,279$ ). Subjects with self-reported shift work ( $n = 22,165$ ) or sleep medication use ( $n = 4,575$ ) were excluded.

**Genotyping and quality control.** Of the  $\sim 500,000$  subjects with phenotype data in the UK Biobank,  $\sim 153,000$  are currently genotyped. Genotyping was performed by the UK Biobank, and genotyping, quality control and imputation procedures are described in detail here<sup>56</sup>. In brief, blood, saliva and urine were collected from participants, and DNA was extracted from the buffy coat samples. Participant DNA was genotyped on two arrays, UK BiLEVE and UKB Axiom with  $>95\%$  common content and genotypes for  $\sim 800,000$  SNPs were imputed to the UK10K reference panel. Genotypes were called using Affymetrix Power Tools software. Sample and SNP quality control were performed. Samples were removed for high missingness or heterozygosity (480 samples), short runs of homozygosity (8 samples), related individuals (1,856 samples) and sex mismatches (191 samples). Genotypes for 152,736 samples passed sample quality control ( $\sim 99.9\%$  of total samples). SNPs were excluded if they did not pass quality control filters across all 33 genotyping batches, with a missingness threshold of 0.90. Batch effects were identified through frequency and Hardy–Weinberg equilibrium tests ( $P$  value  $<10^{-12}$ ). Before imputation, 806,466 SNPs pass quality control in at least one batch ( $>99\%$  of the array content). Population structure was captured by principal component analysis on the samples using a subset of high quality (missingness  $<1.5\%$ ), high-frequency SNPs ( $>2.5\%$ ) ( $\sim 100,000$  SNPs) and identified the sub-sample of European descent. Imputation of autosomal SNPs was performed to a merged reference panel of the Phase 3 1000 Genome Project and the UK10K using IMPUTE3 (ref. 57). Data was prephased using SHAPEIT3 (ref. 58). In total, 73,355,677 SNPs, short indels and large structural variants were imputed. Post-imputation quality control was performed as previously outlined and an info score cutoff of 0.1 was applied. For GWAS, we further excluded SNPs with minor allele frequency (MAF)  $<0.00016$ , a threshold that represents a minimum 50 counts of each genotype, a conservative threshold. In total, 100,400 samples of European descent with high-quality genotyping and complete phenotype/covariate data were used for these analyses. Genotyping quality of two significant rare SNPs (rs1144566 and rs35333999) was verified by examination of genotyping intensity cluster plots (Supplementary Fig. 5). In addition, for two significant imputed rare SNPs we checked Information Quality Scores (info) and found these to be above the standard threshold of 0.40 used to indicate good imputation quality<sup>59</sup> (rs141175086 info = 0.48 and rs148750727 info = 0.88). Considering the size of the genotyped UK Biobank cohort ( $N \sim 150,000$ ), an information measure of 0.4 on a sample of 150,000 individuals indicates that the amount of data at the imputed SNP is roughly equivalent to perfectly observed genotype data in a sample of  $N \sim 60,000$ .

**Statistical analysis.** Genetic association analysis was performed in SNPTEST<sup>60</sup> with the 'expected' method using an additive genetic model adjusted for age, sex, 10 principal components of ancestry and genotyping array. Genome-wide association analysis was performed separately for continuous chronotype and 'extreme' chronotype with a genome-wide significance threshold of  $5 \times 10^{-8}$ . Follow-up analyses on genome-wide significant loci included sex interaction testing using a linear regression model including a sex\*SNP interaction term, performed in R<sup>61</sup>, and conditional analysis using SNPTEST conditioning on the lead signal in each locus  $\pm 500$  kb, covariate sensitivity analysis individually adjusting for sleep duration, sleep disorders, insomnia and depression/psychiatric medication use. Heritability was calculated using BOLT-Reml<sup>62</sup>. Post-GWAS analysis of LD Score Regression (LDSC)<sup>29,34,63,64</sup> was conducted using all UK Biobank SNPs also found in HapMap3 and included publicly available data from 19 published genome-wide association studies, with a  $P$  value threshold of 0.0026 after Bonferroni correction for all 19 tests performed. Gene-based testing was performed using VEGAS<sup>31</sup> on GWAS summary statistics from SNPs and samples passing rigorous quality control, and gene-set enrichment of genes significant after Bonferroni correction was performed using Web-Gestalt<sup>32</sup>. Given that gene-based tests like VEGAS are

sensitive to missing data and may show inflation and low power if data for rare variants is missing<sup>65</sup>, we note low missingness rates (>80% of SNPs had over 99.5% genotyping call rate), and for <5% of SNPs that may have failed in a subset of 33 batches, imputation was used to infer missing genotypes. Furthermore, our gene-based testing included only single SNP association results for variants with over 50 minor allele counts. A Q-Q plot of inflation-adjusted gene-based results for 17,791 genes is shown in Supplementary Fig. 3. Pathway-based analysis to identify enrichment in biological processes, gene sets and tissues suggested by the top loci was performed in DEPICT<sup>33</sup> for all SNPs present in 1KG phase 3 (ref. 66).

For Mendelian randomization analyses, the weighted genetic risk score was calculated by summing the products of the chronotype risk allele count for 15 SNPs multiplied by the scaled chronotype effect reported by 23andMe<sup>18</sup> (that is, using weights from study independent from our own). The instrumental variable analyses were performed in R<sup>40</sup> using the two-stage-least-squares method (TSLs function in the SEM package). The risk scores for education, schizophrenia and BMI were constructed using the GWS SNPs and weights from previously published GWAS<sup>67–69</sup> and tested on chronotype using the summary statistics from our reported GWAS using the GTx package in R.

## References

1. Adan, A. *et al.* Circadian typology: a comprehensive review. *Chronobiol. Int.* **29**, 1153–1175 (2012).
2. Takahashi, J. S., Hong, H. K., Ko, C. H. & McDearmon, E. L. The genetics of mammalian circadian order and disorder: implications for physiology and disease. *Nat. Rev. Genet.* **9**, 764–775 (2008).
3. Foster, R. G. *et al.* Sleep and circadian rhythm disruption in social jetlag and mental illness. *Prog. Mol. Biol. Transl. Sci.* **119**, 325–346 (2013).
4. Bell-Pedersen, D. *et al.* Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat. Rev. Genet.* **6**, 544–556 (2005).
5. Roenneberg, T. *et al.* Epidemiology of the human circadian clock. *Sleep Med. Rev.* **11**, 429–438 (2007).
6. Sack, R. L. *et al.* Circadian rhythm sleep disorders: part II, advanced sleep phase disorder, delayed sleep phase disorder, free-running disorder, and irregular sleep-wake rhythm. An American Academy of Sleep Medicine review. *Sleep* **30**, 1484–1501 (2007).
7. Jones, C. R. *et al.* Familial advanced sleep-phase syndrome: a short-period circadian rhythm variant in humans. *Nat. Med.* **5**, 1062–1065 (1999).
8. Xu, Y. *et al.* Modeling of a human circadian mutation yields insights into clock regulation by PER2. *Cell* **128**, 59–70 (2007).
9. Gamaldo, C. E., Chung, Y., Kang, Y. M. & Salas, R. M. Tick-tock-tick-tock: the impact of circadian rhythm disorders on cardiovascular health and wellness. *J. Am. Soc. Hypertens.* **8**, 921–929 (2014).
10. Brennan, K. C. *et al.* Casein kinase I delta mutations in familial migraine and advanced sleep phase. *Sci. Transl. Med.* **5**, 183ra56 (2013).
11. Saini, C., Brown, S. A. & Dibner, C. Human peripheral clocks: applications for studying circadian phenotypes in physiology and pathophysiology. *Front. Neurol.* **6**, 95 (2015).
12. Gottlieb, D. J., O'Connor, G. T. & Wilk, J. B. Genome-wide association of sleep and circadian phenotypes. *BMC Med. Genet.* **8**(Suppl 1): S9 (2007).
13. Heath, A. C., Kendler, K. S., Eaves, L. J. & Martin, N. G. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. *Sleep* **13**, 318–335 (1990).
14. Klei, L. *et al.* Heritability of morningness-eveningness and self-report sleep measures in a family-based sample of 521 sibships. *Chronobiol. Int.* **22**, 1041–1054 (2005).
15. Allebrandt, K. V. & Roenneberg, T. The search for circadian clock components in humans: new perspectives for association studies. *Braz. J. Med. Biol. Res.* **41**, 716–721 (2008).
16. Levandovski, R., Sasso, E. & Hidalgo, M. P. Chronotype: a review of the advances, limits and applicability of the main instruments used in the literature to assess human phenotype. *Trends Psychiatry Psychother.* **35**, 3–11 (2013).
17. Farh, K. K. *et al.* Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* **518**, 337–343 (2014).
18. Hu, Y. *et al.* GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* **7**, 10448 (2016).
19. Archer, S. N. *et al.* A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference. *Sleep* **26**, 413–415 (2003).
20. Zhang, Y. & Emery, P. GW182 controls Drosophila circadian behavior and PDF-receptor signaling. *Neuron* **78**, 152–165 (2013).
21. Koike, N. *et al.* Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* **338**, 349–354 (2012).
22. Vollmers, C. *et al.* Time of feeding and the intrinsic circadian clock drive rhythms in hepatic gene expression. *Proc. Natl. Acad. Sci. USA* **106**, 21453–21458 (2009).
23. Zhang, E. E. *et al.* A genome-wide RNAi screen for modifiers of the circadian clock in human cells. *Cell* **139**, 199–210 (2009).
24. Ly, S., Pishdari, B., Lok, L. L., Hajos, M. & Kocsis, B. Activation of 5-HT<sub>6</sub> receptors modulates sleep-wake activity and hippocampal theta oscillation. *ACS Chem. Neurosci.* **4**, 191–199 (2013).
25. Monti, J. M., Jantos, H. & Schechter, L. E. The effects of systemic and local microinjection into the central nervous system of the selective serotonin 5-HT<sub>6</sub> receptor agonist WAY-208466 on sleep and wakefulness in the rat. *Behav. Brain Res.* **249**, 65–74 (2013).
26. Nonogaki, K. Serotonin conflict in sleep-feeding. *Vitam. Horm.* **89**, 223–239 (2012).
27. Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
28. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
29. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
30. Lindblad-Toh, K. *et al.* A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).
31. Liu, J. Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am. J. Hum. Genet.* **87**, 139–145 (2010).
32. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res.* **41**, W77–W83 (2013).
33. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).
34. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
35. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
36. Ebrahim, S. & Davey Smith, G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum. Genet.* **123**, 15–33 (2008).
37. Panda, S., Hogenesch, J. B. & Kay, S. A. Circadian rhythms from flies to human. *Nature* **417**, 329–335 (2002).
38. Winkelmann, J. *et al.* Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet.* **7**, e1002171 (2011).
39. Peyron, C. *et al.* A mutation in a case of early onset narcolepsy and a generalized absence of hypocretin peptides in human narcoleptic brains. *Nat. Med.* **6**, 991–997 (2000).
40. Lin, L. *et al.* The sleep disorder canine narcolepsy is caused by a mutation in the hypocretin (orexin) receptor 2 gene. *Cell* **98**, 365–376 (1999).
41. Hofstetter, J. R., Mayeda, A. R., Happel, C. G. & Lysaker, P. H. Sleep and daily activity preferences in schizophrenia: associations with neurocognition and symptoms. *J. Nerv. Ment. Dis.* **191**, 408–410 (2003).
42. Wulff, K., Dijk, D. J., Middleton, B., Foster, R. G. & Joyce, E. M. Sleep and circadian rhythm disruption in schizophrenia. *Br. J. Psychiatry* **200**, 308–316 (2012).
43. Lemoine, P., Zawieja, P. & Ohayon, M. M. Associations between morningness/eveningness and psychopathology: an epidemiological survey in three in-patient psychiatric clinics. *J. Psychiatr. Res.* **47**, 1095–1098 (2013).
44. Yu, J. H. *et al.* Evening chronotype is associated with metabolic disorders and body composition in middle-aged adults. *J. Clin. Endocrinol. Metab.* **100**, 1494–1502 (2015).
45. Culnan, E., Kloss, J. D. & Grandner, M. A prospective study of weight gain associated with chronotype among college freshmen. *Chronobiol. Int.* **30**, 682–690 (2013).
46. Arora, T. & Taheri, S. Associations among late chronotype, body mass index and dietary behaviors in young adolescents. *Int. J. Obes. (Lond.)* **39**, 39–44 (2015).
47. Fleig, D. & Randler, C. Association between chronotype and diet in adolescents based on food logs. *Eat. Behav.* **10**, 115–118 (2009).
48. Wain, L. V. *et al.* Novel insights into the genetics of smoking behaviour, lung function, and chronic obstructive pulmonary disease (UK BiLEVE): a genetic association study in UK Biobank. *Lancet Respir. Med.* **3**, 769–781 (2015).
49. Gale, C. & Martyn, C. Larks and owls and health, wealth, and wisdom. *BMJ* **317**, 1675–1677 (1998).
50. Davide Piffra, D. P., Sapienzab, P., Zingalesc, L. & Maestripieri, D. Morningness-eveningness and intelligence among high-achieving US students: night owls have higher GMAT scores than early morning types in a top-ranked MBA program. *Intelligence* **47**, 107–112 (2014).
51. Macfarlane, T. V., Beasley, M. & Macfarlane, G. J. Self-Reported Facial Pain in UK Biobank Study: Prevalence and Associated Factors. *J. Oral Maxillofac. Res.* **5**, e2 (2014).
52. Merikanto, I. *et al.* Associations of chronotype and sleep with cardiovascular diseases and type 2 diabetes. *Chronobiol. Int.* **30**, 470–477 (2013).

53. Smith, C. *et al.* Investigation of morning-evening orientation in six countries using the preferences scale. *Pers. Individ. Differ.* **32**, 949–968 (2002).

54. Taillard, J., Philip, P., Chastang, J. F. & Bioulac, B. Validation of Horne and Ostberg morningness-eveningness questionnaire in a middle-aged population of French workers. *J. Biol. Rhythms* **19**, 76–86 (2004).

55. Allen, N. E., Sudlow, C., Peakman, T. & Collins, R. UK Biobank. UK biobank data: come and get it. *Sci. Transl. Med.* **6**, 224ed4 (2014).

56. UKBiobank. Genotyping and quality control of UK Biobank, a large-scale, extensively phenotyped prospective resource (2015) [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).

57. Yu, Z. & Schaid, D. J. Methods to impute missing genotypes for population data. *Hum. Genet.* **122**, 495–504 (2007).

58. Kanterakis, A. *et al.* Molgenis-impute: imputation pipeline in a box. *BMC Res. Notes* **8**, 359 (2015).

59. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).

60. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).

61. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2008).

62. Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).

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

64. International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).

65. Auer, P. L., Wang, G. & Leal, S. M. Testing for rare variant associations in the presence of missing data. *Genet. Epidemiol.* **37**, 529–538 (2013).

66. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).

67. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

68. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).

69. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).

70. Smith, G. D. *et al.* Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med.* **4**, e352 (2007).

## Acknowledgements

This work was supported by NIH grants R21HL121728-02 (R.S.), F32DK102323-01A1 (J.M.L.), R01HL113338-04 (J.M.L., S.R. and R.S.), the University of Manchester (Research Infrastructure Fund), the Wellcome Trust (salary support D.W.R. and A.L.) and UK Medical Research Council MC\_UU\_12013/5 (D.A.L.). Data on glycemic traits have been contributed by MAGIC investigators and have been downloaded from [www.magicinvestigators.org](http://www.magicinvestigators.org). Data on coronary artery disease/myocardial infarction have been contributed by CARDIo-GRAMplusC4D investigators and have been downloaded from [www.CARDIOGRAMPLUSC4D.ORG](http://www.CARDIOGRAMPLUSC4D.ORG). We thank the International Genomics of Alzheimer's Project (IGAP) for providing summary results data for these analyses.

## Author contributions

The study was designed by J.M.L., M.K.R. and R.S. J.M.L., I.V. and R.S. performed genetic analyses. J.M.L. and R.S. wrote the manuscript and all co-authors helped interpret data, reviewed and edited the manuscript, before approving its submission. R.S. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Lane, J. M. *et al.* Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nat. Commun.* **7**:10889 doi: 10.1038/ncomms10889 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>