bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Genome-wide association analyses of chronotype in 697,828 individuals provides new insights into circadian rhythms in humans and links to disease

Samuel E. Jones[1]*, Jacqueline M. Lane[2,3,4]*, Andrew R. Wood[1]*, Vincent T. van Hees[5], Jessica Tyrrell[1], Robin N. Beaumont[1], Aaron Jefferies[1], Hassan S. Dashti[2,4], Melvyn Hillsdon[6], Katherine S. Ruth[1], Marcus A. Tuke[1], Hanieh Yaghootkar[1], Seth Sharp[1], Yingjie Jie[1], William D. Thompson[1], Jamie W. Harrison[1], Amy Dawes[1], Enda M. Byrne[7], Henning Tiemeier[8,9], Karla V. Allebrandt[10], Jack Bowden[11,12], David W. Ray[13,14], Rachel M. Freathy[1], Anna Murray[1], Diego R. Mazzotti[15], Philip R. Gehrman[16], the 23andMe Research Team[17], Debbie A. Lawlor[11,12]*, Timothy M. Frayling[1]*, Martin K. Rutter[13,14,18]*, David Hinds[17]*, Richa Saxena[1,2,3,19]*, Michael N. Weedon[1]*

*Equal Contributions

Affiliations

[1]Genetics of Complex Traits, University of Exeter Medical School, Exeter, United Kingdom
[2]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
[3]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[4]Broad Institute, Cambridge, MA, USA
[5]Netherlands eScience Center, Amsterdam, The Netherlands
[6]Sport and Health Sciences, College of Life and Environmental Sciences, University of Exeter, Exeter, UK
[7]The University of Queensland, Institute for Molecular Bioscience, Brisbane, Australia.
[8]Department of Epidemiology, Erasmus Medical Center, Rotterdam, Netherlands.
[9]Department of Psychiatry, Erasmus Medical Center, Rotterdam, Netherlands.
[10]Department of Translational Informatics, Translational Medicine Early Development, Sanofi-Aventis Deutschland GmbH, Industriepark Höchst, Frankfurt, Germany
[11]MRC Integrative Epidemiology Unit at the University of Bristol, UK
[12]Population Health Sciences, Bristol Medical School, University of Bristol, UK
[13]Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK
[14]Division of Endocrinology, Diabetes & Gastroenterology, School of Medical Sciences, Faculty of Biology, Medicine and Health, University of Manchester, UK
[15]Center for Sleep and Circadian Neurobiology, University of Pennsylvania, Philadelphia, PA, United States
[16]Perelman School of Medicine of the University of Pennsylvania, Philadelphia, Pennsylvania, United States of America
[17]23andMe Inc., 899 W. Evelyn Avenue, Mountain View, California 94041, USA
[18]Manchester Diabetes Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
[19]Departments of Medicine, Brigham and Women's Hospital and Beth Israel Deaconess Medical Center, Harvard Medical School, Boston

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Using genome-wide data from 697,828 research participants from 23andMe and UK Biobank, we increase the number of identified loci associated with being a morning person, a behavioural indicator of a person's underlying circadian rhythm, from 24 to 351. Using data from 85,760 individuals with activity-monitor derived measures of sleep timing we show that the chronotype loci influence sleep timing: the mean sleep timing of the 5% of individuals carrying the most "morningness" alleles was 25 minutes earlier than the 5% carrying the fewest. The loci were enriched for genes involved in circadian regulation, cAMP, glutamate and insulin signalling pathways, and those expressed in the retina, hindbrain, hypothalamus, and pituitary. We provide evidence that being a morning person is causally associated with better mental health but does not appear to affect BMI or Type 2 diabetes. This study offers new insights into the biology of circadian rhythms and links to disease in humans.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Circadian rhythms are fundamental cyclical processes that occur in most living organisms, including humans. These daily cycles affect a wide range of molecular and behavioural processes, including hormone levels, core body temperature and sleep-wake patterns[1]. Chronotype, often referred to as circadian preference, describes an individual's proclivity for earlier or later sleep timing and is a physical and behavioural manifestation of the coupling between internal circadian cycles and the need for sleep, driven by sleep homeostasis. Significant natural variation exists amongst the human population with chronotype often measured on a continuous scale[2], though individuals are often separated into "morning people" (or "larks") who prefer going to bed and waking earlier, "evening people" (or "owls") who prefer a later bedtime and later rising time, and "intermediates" who lie between the two extremes[3,4]. Age and gender, as well as environmental light levels explain a substantial proportion of variation in chronotype, but genetic variation is also an important contributor[5,6,7,8].

There is evidence that alterations to circadian timing are linked to disease development, particularly metabolic and psychiatric disorders[9,10]. Animal model studies have shown that mutations in, and altered expression of, key circadian rhythm genes can cause obesity, hyperglycaemia and defective beta-cell function leading to diabetes[11–13]. In humans, there are many reported associations between disrupted circadian rhythms and disease[14,15], but the evidence for a causal role of chronotype on disease is limited[16]. For example, evening people have an increased frequency of obesity[17], Type 2 diabetes[18] and depression[19] independent of sleep disturbance, and studies of shift workers show an increased risk of diabetes, depression and other diseases[20]. However, these associations could be explained by reverse causality (diseases affecting sleep patterns or dictating job options) or confounding (common risk factors influencing both chronotype and disease). Genetic analyses identifying variants robustly associated with putative risk factors, such as chronotype, can improve causal understanding by providing genetic instruments for use in Mendelian Randomization (MR) analyses[21–23], which minimise the effect of both reverse causality and bias caused by confounding. Identifying genetic variants associated with chronotype and sleep timing will also provide new insights into the biological processes underlying circadian rhythms and sleep homeostasis.

Three previous genome-wide association studies (GWAS)[24–26], using a maximum of 128,286 individuals, identified a total of 24 independent variants associated with self-report chronotype. In this study, we performed a GWAS meta-analysis of a substantially expanded set of 697,828 individuals, including 248,098 participants from 23andMe Inc., a personal

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genetics company, and 449,734 participants from UK Biobank[27,28]. In addition to confirming an enrichment of circadian rhythm and brain expressed genes at chronotype-associated loci and genetic correlation with mental health disorders[25,26], in this substantially enlarged GWAS study we identify 327 additional chronotype-associated loci and demonstrate that the chronotype-associated variants are associated with objective measures of sleep timing, but not sleep duration or quality, in 85,760 UK Biobank participants. By fine-mapping the genetic associations at all loci, we identify 10 coding variants with a high likelihood of being the causal variant, providing new targets for biological investigation of chronobiology and go on to provide evidence of a causal link between chronotype and mental health.

Results

***351 loci associated with morning chronotype from a GWAS meta-analysis including 697,828 individuals*.** We performed a GWAS of self-report chronotype (**Table 1**) using 11,977,111 imputed variants in 449,734 individuals of European ancestry from the UK Biobank and meta-analysed with summary statistics from a self-report morningness GWAS using 11,947,421 variants in 248,098 European-ancestry 23andMe research participants. We identified 351 independent loci at $P<5\times10^{-8}$, of which 258 reached $P<6\times10^{-9}$, a correction for the significance threshold based on permutation testing (**Supplementary Methods**). Of the 351 loci, 24 had been previously reported in earlier GWAS of chronotype[24–26] and 327 were novel associations. The primary meta-analysis, based on sample size, and individual study results are shown in **Figure 1** and **Supplementary Table 1**. Conditional analysis identified 49 loci with multiple independent signals (**Supplementary Table 2**). A sensitivity analysis was performed in the UK Biobank data alone, excluding shift workers and those either on medication or with disorders affecting sleep (see the **Methods** section and **Supplementary Methods** for details). Effect sizes were similar to those in the full UK Biobank GWAS (**Supplementary Table 1** and **Supplementary Figure 1**).

Known circadian genes amongst associated loci. Well-documented circadian rhythm genes were among the most strongly associated loci (**Supplementary Table 1**). These genes included the previously reported loci containing *RGS16*, *PER2*, *PER3*, *PIGK/AK5*, *INADL*, *FBXL3*, *HCRTR2* and *HTR6*[24–26], and newly associated loci containing known circadian rhythm genes *PER1, CRY1 and ARNTL* (**Supplementary Figure 2**). At the *PER3* locus, two highly correlated low frequency missense variants (rs150812083 and rs139315125, MAF=0.5%), previously reported to be a monogenic cause of familial advanced sleep phase syndrome[29], were associated with self-reported morningness (OR=1.44 for minor allele; $P=2\times10^{-38}$) but with a lower magnitude of effect on sleep timing than expected in the activity-monitor derived measures of chronotype, advancing sleep

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

timing (as measured by time of minimum activity) by only 8 minutes (95% CI: 4, 13, P=4.3x10^{-4}) as opposed to the average 4.2 hours reported in the previous study[29].

Chronotype associated variants affect objective sleep timing, but not quality or duration. Self-report assessments of sleep and chronotype can be subject to reporting bias[30–33]. To assess and quantify the effect of the genetic associations on objective measures of sleep timing, duration and quality, we tested the association of the chronotype-associated variants with sleep estimates derived from the UK Biobank activity monitor data. Derived phenotypes included sleep timing, efficiency and duration. Timing was determined by timings of midpoint of sleep, the least active 5 hours of the day (L5 timing) and midpoint of the most active 10 hours of the day (M10 timing). Summary statistics of these derived phenotypes and their associations with self-report morningness are presented in **Supplementary Table 3**, and their associations with the newly identified chronotype SNPs are provided in **Supplementary Table 4**. To avoid inflation of associations due to overlapping samples, we performed an additional GWAS meta-analysis of self-reported morningness excluding all UK Biobank individuals with activity monitor data. Of the 292 lead chronotype variants reaching P<5x10^{-8} from this meta-analysis that were available in the UK Biobank imputed genotype data, 258 had a consistent direction of effect for sleep midpoint (binomial test P=3.8x10^{-44}), 262 with L5 timing (binomial test P=9.3x10^{-48}) and 260 with M10 timing (binomial test P=6.4x10^{-46}). A genetic risk score (GRS) of these 292 variants was associated with earlier sleep midpoint, L5 timing and M10 timing (P=4x10^{-128}, P=1x10^{-182} and P=7x10^{-130} respectively). There was little evidence of association between the chronotype GRS and the activity monitor sleep phenotypes that estimate sleep duration and fragmentation (**Supplementary Table 5**), indicating a specific effect of the chronotype SNPs on sleep timing and circadian metrics. Limiting the analysis to the 109 lead variants identified from the independent 23andMe GWAS provided similar results (**Supplementary Table 5**). Using the activity-monitor derived estimates of sleeping timing, the 5% of individuals carrying the most morningness alleles at the 292 associated loci had L5 timing shifted earlier, on average, by 25.1 minutes (95% CI: 22.5, 27.6) compared to the 5% carrying the fewest morningness alleles: a mean L5 time of 03:06 rather than 03:32. The data suggests that self-report chronotype strongly relate to an individual's sleep timing and therefore provide valid instruments for MR and compelling insights into circadian biology.

Circadian rhythm, neuron development and cell signalling pathways, and brain and retina expressed genes are strongly enriched at the associated loci. To identify biological pathways and tissues enriched for genes at the associated loci, we used MAGMA[34] , implemented as part of the FUMA GWAS[35] platform (**Figures 2-3**.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Tables 6-7). Because of the variety of methods available and databases employed, we also performed secondary gene-set and tissue enrichment using the software packages PASCAL[36], MAGENTA[37] and DEPICT[38] (**Supplementary Tables 8-11**). We identified strong enrichment in circadian rhythm and circadian clock pathways as with previous morningness GWAS[24–26]. We also identified multiple pathways that correspond to (central) nervous system and brain development, components of neuronal cells such as synapses, axons and dendrites, as well as neurogenesis. There was clear enrichment in all types of brain tissue (**Figure 4**, **Supplementary Tables 7 and 11**), in behavioural pathways, containing genes responsible for mediating behavioural responses to internal and external stimuli, and in retinal tissue (**Supplementary Table 11**). The genes in the associated loci were also enriched in multiple pathways relating to the regulation and metabolism of cyclic nucleotides, such as cAMP and cGMP, as well as pathways involved in G-protein signalling and activation. The NMDA glutamate signalling pathway was also enriched and MAGMA mapped genes in this pathway include *NRXN1* and *RELN*, which have been shown to influence risk of schizophrenia[39,40], but for which there is limited evidence of a role in circadian rhythm regulation.

Fine-mapping identifies likely causal variants and genes

To highlight putative causal variants and genes, we fine-mapped the associated loci using FINEMAP[41]. FINEMAP uses a shotgun stochastic search to identify the most plausible causal variant configuration given the GWAS association statistics and local LD patterns and outputs posterior probabilities for each variant configuration being causal. Forty-two loci had a single variant with a probability of >50% of being causal (**Supplementary Table 12**). Annotation of these likely causal variants identified 10 coding variants. These include a low frequency missense variant in *RGS16* (MAF=3%, morningness OR=1.26 for minor allele), previously associated with chronotype[25] and the most strongly associated with morningness in this study, and missense variants in the *INADL*, *HCRTR2, PLCL1* and *CLN5* genes, all four genes having been identified in previous GWAS[24,26]. Fine-mapping also identified missense variants in *PCYOX1* and *SKOR2*, and a stop gain variant in the *MADD* gene, as likely causal variants in these loci, highlighting new candidate genes for chronotype. The *MADD* stop gain variant (rs35233100) has previously been associated with levels of proinsulin[42], providing a potential link between insulin secretion and chronotype. To gain further insight into additional genes that may play a role in determining chronotype, we annotated the putative causal variants using the GTEx eQTL database (**Supplementary Table 12**). There were 90 variants across 51 loci that were eQTLs for one or more genes, with a total of 208 mapped genes. As an example, this included a putative causal variant in the promoter of *FBXO3* which represents the strongest eQTL for *FBXO3*. *FBXO3* is in the

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ubiquitin-proteosome pathway; protein (de)ubiquitination has been shown to be involved with the degradation of several core clock genes[43,44], influencing the build-up up these proteins and the pace of the circadian clock[45,46]. *FBXO3* expression has been shown to be altered by light treatment and to demonstrate rhythmic expression[46].

Integration of GWAS and SCN-enrichment data highlights potential circadian clock genes at associated loci. The suprachiasmatic nucleus is a small region of the brain, consisting of around 20,000 neurons, that is integral to maintaining circadian rhythms in humans and is a likely mechanism of action for at least some of the associated genes and loci. Indeed, the associated loci included many key mammal SCN clock genes including *PER1*, *PER2*, *PER3*, *CRY1*, *FBXL3* and *ARNTL* (**Supplementary Table 13**). To identify new genes important in setting and modulating circadian rhythms in the SCN, we assessed expression, enrichment and fluctuation of proximal or eQTL-mapped genes using expression data from the mouse SCN. We cross-referenced all mapped genes at the fine-mapped loci against whether there was evidence for enrichment of expression in the SCN compared to other brain tissue[47,48] and whether the genes demonstrated evidence of fluctuation in expression over the 24-hour cycle[47] (**Supplementary Table 12**). We also annotated the genes against a set of 343 putative clock genes identified from RNAi knockdown experiments a human cellular clock model[49] (**Supplementary Table 12**). Of the 22.5% of all genes tested that were enriched in the SCN[48], 28.0% of the 804 genes mapped using MAGMA and present in the enrichment analysis, were enriched in the SCN representing a significant excess (P=2x10^{-4}). As a negative control, we tested the enrichment of MAGMA-mapped genes for several unrelated GWAS phenotypes, finding no significant excess of SCN-enriched genes (all P>0.05) (**Supplementary Table 14**). Similar enrichment was found for those chronotype genes fluctuating in the SCN, but no significant excess from the RNAi knockdown study. Enriched and fluctuating genes from the fine-mapping efforts include known circadian genes such as *FBXL3* and putative genes such as *LSM7* and *VIP*. *LSM7* encodes core components of the spliceosomal U6 small nuclear ribonucleoprotein complex for which some previous studies have suggested a role in circadian timing[49,50]. *VIP* encodes a vasoactive peptide hormone that lowers arterial blood pressure and relaxes muscles of the stomach and trachea. Evidence from mouse models indicates that it has a role in generating and light-entrainment of circadian oscillations[51].

Chronotype is heritable and demonstrates strong genetic correlation with several psychiatric traits. As a strategy to prioritise traits for subsequent causal analyses, as previous studies have shown a strong correlation between genetic and phenotypic correlations[52,53], and to identify genetic overlap between chronotype and other diseases and

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

traits, we performed LD score regression analyses against a range of other diseases and traits where GWAS summary statistics are publicly available (**Supplementary Table 15**). We estimated the heritability of chronotype to be 13.7% (95% CI 13.3%, 14.0%), as calculated by BOLT-REML in the UK Biobank data alone, which is towards the lower end of previously reported figures (12-21%)[24–26]. The most genetically correlated trait was subjective well-being, which was positively correlated with being a morning person (r_G=0.17, P=6x10^{-9}). Psychiatric traits schizophrenia (r_G=-0.11, P=1x10^{-7}), depressive symptoms (r_G=-0.16; P=2x10^{-6}), major depressive disorder (r_G=-0.19; P=3x10^{-5}) and intelligence (r_G=-0.11; P=8x10^{-6}) were all negatively correlated with the morning chronotype. Metabolic traits fasting insulin (r_G=-0.09, P=0.03) and HOMA-IR (r_G=-0.12, P=0.009) were negatively correlated with being a morning person but did not reach our Bonferroni-corrected significance threshold. BMI (r_G=0.007, P=0.74) and T2D (r_G=0.02, P=0.60) were not genetically correlated with morningness.

Mendelian randomisation (MR) analyses provide evidence for a causal link between chronotype and mental health. Genetic correlations do not allow for magnitudes of causality to be determined between an exposure and an outcome. We therefore performed two-sample Mendelian Randomisation (MR) analyses against the five psychiatric traits that showed evidence of a genetic correlation to estimate causal effects. Because of the extensive literature on the link between chronotype and metabolic disease and because the well-known SNP in *FTO* (rs1558902) previously associated with higher BMI[54,55] was also associated with being a morning person (OR=1.04, P= 4.9x10^{-32}), we also performed two-sample MR against the metabolic phenotypes BMI, type 2 diabetes and fasting insulin levels. For individual instrument effects on chronotype, log odds ratios (representing liability for morningness) from the secondary morning person meta-analysis were used, as no effect sizes were obtained in the primary meta-analysis. With chronotype as an exposure, we implemented the R package TwoSampleMR[56] to report causal associations of chronotype on these eight outcomes (**Supplementary Table 16**). We saw evidence that being a morning person confers a liability to lower risk of schizophrenia and greater subjective well-being, with a genetically determined unit log odds increase in self-report morningness being associated with a liability for reduced schizophrenia (odds ratio of 0.89 (0.82, 0.96); IVW P=0.004) and higher subjective well-being (0.04SD (0.02, 0.06); IVW P=5x10^{-5}), and with good agreement amongst the different MR methods (**Figures 5-6**). There was suggestive evidence that morningness decreases the liability of depression: one-unit log odds increase in morningness was associated with an odds ratio of 0.65 (0.44, 0.95; IVW P=0.03) for major depressive disorder and 0.02 SD lower (0.002, 0.04; IVW P=0.03) for depressive symptoms (**Supplementary Figures 3-4**) but these did not reach our multiple testing threshold of

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

P_{bonf}=0.005. There was no strong statistical evidence that chronotype was causally associated with BMI, fasting insulin and risk of type 2 diabetes (IVW *P*>0.1), as previously reported[24–26].

No evidence that schizophrenia or depression influence morningness.

To assess whether our genetically-correlated phenotypes were causally influencing chronotype, we performed two-sample Mendelian Randomisation analyses with chronotype as the outcome. Owing to a limited number of genetic instruments, of the original five genetically-correlated psychiatric phenotypes we were able to test only schizophrenia and major depressive disorder, in addition to the metabolic phenotypes BMI, insulin secretion and type 2 diabetes (**Supplementary table 17**). We observed only weak evidence of liability effects of type 2 diabetes (IVW *P*=0.01), insulin secretion (IVW *P*=0.04) and BMI (IVW *P*=0.05) on chronotype. Despite strong genetic correlations with chronotype, we see no strong evidence that schizophrenia (IVW *P*=0.07) or major depressive disorder (IVW *P*=0.62) causally influence liability for morningness.

Discussion

Using data from 697,828 individuals, we have performed the largest GWAS study of chronotype and expanded the number of chronotype-associated loci from 24 to 351. Using activity monitor data from 85,760 we show that these variants are associated with measures of objective sleep timing. We confirm previously reported enrichment of circadian rhythm pathways and retina and brain expressed genes at associated loci, and demonstrate further enrichment of genes in the cAMP, cGMP, NMDA and insulin signalling pathways as well as those in pituitary gland tissue and the SCN. We fine-map the loci and provide target genes for other researchers to perform in depth functional investigation into chronobiology. We have provided more accurate genetic correlation estimates of chronotype with a range of traits and disease and provide some evidence for a causal link between chronotype and mental health.

We have found evidence that the natural variation in circadian preference amongst the human population can be ascribed to several different mechanisms. Given the prominence of genetic variants in or near multiple core circadian rhythm genes (*PER1*, *PER2*, *PER3*, *CRY1*, *FBXL3*, *ARNTL*), we infer that some of the variation is attributed to subtle differences in the biochemical feedback mechanism of the circadian clock. This is supported by evidence of the chronotype-associated loci being enriched in the SCN, suggesting that variants that also subtly affect the modification and regulation of the circadian clock contribute to the population variation of chronotype. Entrainment of circadian rhythms

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

through external stimuli such as light and temperature is well-known but through this study and previous GWAS efforts, we find that an individual's chronotype is also influenced by variants in genes important in the correct formation and functioning of retinal ganglion cells (*RGS16* and *INADL*), highlighting that some natural variation could be explained by better detection and communication (to the SCN) of external light signals. Variants in genes with known roles in appetite regulation (*FTO*), insulin secretion (*MADD*) and even nicotine and caffeine metabolism (*CYP2A6*) point to other processes that impact an individual's chronotype, though it is unclear whether the effect of these on chronotype are mediated through the modulation of the circadian clock or by other means, such as through sleep-wake homeostasis.

Reported observational associations of chronotype with metabolic diseases are particularly strong[57,58], but we found no evidence for a causal effect of morningness on type 2 diabetes, BMI or insulin levels and could exclude the observational association effect sizes. One possibility which future studies should investigate is whether circadian misalignment, rather than chronotype itself, is more strongly associated with disease outcomes. For example, are individuals who are genetically evening people but have to wake early because of work commitments particularly susceptible to obesity and diabetes?

There are clear epidemiological associations reported in the literature between mental health traits and chronotype, with mental health disorders typically being overrepresented in evening types[59–61], and in this study we show that morningness is negatively genetically correlated with both depression and schizophrenia, and positively correlated with wellbeing. Previous studies have found a link between schizophrenia and circadian dysregulation and misalignment[62,63] with schizophrenics displaying greater variation in sleep and activity timing and misaligned melatonin and sleep cycles, but no evidence exists for the effect of chronotype on schizophrenia risk. Our Mendelian Randomisation analyses support a causal role of eveningness on increased risk of schizophrenia, though the statistical significance is not overwhelming strong. We do not find evidence of schizophrenia causally influencing chronotype. However, several of the mapped genes at the chronotype associated loci are well known schizophrenia loci such as NRXN1 (as well at NRXN2 and NRXN3) and RELN[39,40] and subsequent studies will be necessary to understand the shared biological mechanisms between chronotype and schizophrenia risk.

Chronotype is influenced by circadian rhythms and innate sleep homeostatic mechanisms but is also dependent on societal pressures. It is also a self-report measure which means the interpretation of the phenotype and the genetic association is complicated. In this study,

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

however, we show, using objective measures derived from activity monitor data that these chronotype variants do affect objectives measures of sleep timing, but not other aspects of sleep including duration and timing, providing evidence that we are identifying biologically meaningful associations and allowing us to quantity the effect of these variants on sleep timing.

The response to UK Biobank participation was <5% and this has resulted in selection for healthier individuals, which may introduced bias into our analyses, including in GWAS and MR[64]. Here GWAS results replicated those of 23andMe, a study that may also suffer from selection bias but of a different nature to UK Biobank. Adopting two-sample MR we attempted to maximise statistical power by using publicly available aggregated data based on consortia of studies that had considerably greater response rates, and avoided winner's curse which can lead to underestimation of causal effects[65]. MR of a binary (or other broad category) exposure that is derived from an underlying continuous trait, as is the case with chronotype, may be biased by horizontal pleiotropy from within-category variation in the trait that cannot be identified by alternative MR methods, such as MR-Egger. As effect sizes for MR analyses were derived from log odds ratios in the secondary morning person meta-analysis, there may be the possibility of undetected pleiotropy and so our findings should therefore be treated with some caution.

In conclusion, we have identified 327 novel loci that regulate circadian rhythms and sleep timing in humans and provide new insights into the association of chronotype with disease.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Materials and Methods

Cohorts

The UK Biobank is described in detail elsewhere[27]. We used data on 451,454 individuals from the full UK Biobank data release that we identified as White European and that had genetic data available. To define a set of White Europeans, we performed Principal Components Analysis (PCA) in the 1000 Genomes (1KG) reference panel using a subset of variants that were of a high quality in the UK Biobank. We projected these principal components into the set of related UK Biobank participants to avoid the relatedness confounding the principal components. We then adopted a k-means clustering approach to define a European cluster, initialising the ethnic centres defined by the population-specific means of the first four 1KG principal components. This analysis was performed only within individuals self-reporting as “British”, “Irish”, “White” or “Any other white background”. Because association analyses are performed using linear mixed-model (LMM) method, we included related individuals.

We used summary statistics from a morning chronotype GWAS performed by 23andMe of 248,100 (N_{case}=120,478; $N_{control}$=127,622) participants with a minimum of 97% European ancestry. GWAS analysis was performed in a maximal set of unrelated participants, where pairs of individuals were considered related if they shared 700cM IBD of genomic segments, roughly corresponding to first cousins in an outbred population. The 23andMe cohort is described in more detail elsewhere[24].

Activity Monitor Data

A subset of the UK Biobank cohort was invited to wear a wrist-worn activity monitor for a period of one week. Individuals were mailed the device and asked to wear it continuously for seven days, including while bathing, showering and sleeping. In total, 103,720 participants returned their activity monitor devices with data covering at least three complete 24-hour periods. We downloaded the raw activity monitor data (data-field 90001) for these individuals, in the form of binary Continuous Wave Accelerometer (cwa) files. Further information, along with details of centrally-derived variables, is available elsewhere[66]. Detailed protocol information can be found online at http://biobank.ctsu.ox.ac.uk/crystal/docs/PhysicalActivityMonitor.pdf and a sample instruction letter at http://biobank.ctsu.ox.ac.uk/crystal/images/activity_invite.png (UKB Resources 131600 and 141141, respectively; both accessed January 30th 2018). We converted the .cwa files to .wav format using the open-source software “omconvert”, recommended by the activity monitor manufacturers Axivity, which is available online (see https://github.com/digitalinteraction/openmovement/tree/master/Software/AX3/omconvert).

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

To process the raw accelerometer data, we used the freely available R package "GGIR" (v1.5-12)[67,68]. The list of our GGIR settings is provided in **Supplementary File 1** and the full list of variables produced by GGIR can be found in the CRAN GGIR reference manual (see https://cran.r-project.org/web/packages/GGIR/GGIR.pdf).

Genotyping and quality control

The 23andMe cohort was genotyped on one of four custom arrays: the first two were variants of the Illumina HumanHap550+ BeadChip (N_{case}=4,966; $N_{control}$=5,564), the third a variant of the Illumina OmniExpress+ BeadChip (N_{case}=53,747; $N_{control}$=61,637) and the fourth a fully custom array (N_{case}=61,765; $N_{control}$=60,421). Successive arrays contained substantial overlap with previous chips. These genotypes were imputed to ~15.6M variants using the September 2013 release of the 1000 Genomes phase 1 reference panel. For analyses, we used ~11.9M imputed variants with imputation $r^2 \geq 0.3$, MAF ≥ 0.001 (0.1%) and that showed no sign of batch effects.

The UK Biobank cohort was genotyped on two almost identical arrays. The first ~50,000 samples were genotyped on the UK BiLEVE array and the remaining ~450,000 samples were genotyped on the UK Biobank Axiom array in two groups (interim and full release). A total of 805,426 directly-genotyped variants were made available in the full release. These variants were centrally imputed to ~93M autosomal variants using two reference panels: a combined UK10K and 1000 Genomes panel and the Haplotype Reference Consortium (HRC) panel. For all analyses, we used ~12.0M Haplotype Reference Consortium (HRC) imputed variants with an imputation $r^2 \geq 0.3$, MAF ≥ 0.001 (0.1%) and with a Hardy–Weinberg equilibrium $P>1x10^{-12}$. We excluded non-HRC imputed variants on advice from the UK Biobank imputation team. Further details on the UK Biobank genotyping, quality control and imputation procedures can be found elsewhere[28].

Self-Report Phenotypes

Morning Person (23andMe)

Responses to two identical questions were used to define the dichotomous morning person phenotype in the 23andMe cohort, with one question having a wider selection of neutral options. More details are given in Supplementary Table 2 of the 23andMe morning person GWAS[24]. Morning people were coded as 1 (cases; N=120,478) and evening people were coded as 0 (controls; N=127,622).

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Chronotype (UK Biobank)

The UK Biobank collected a single self-reported measure of Chronotype ("Morning/evening person (chronotype)"; data-field 1180). Participants were prompted to answer the question "Do you consider yourself to be?" with one of six possible answers: "Definitely a 'morning' person", "More a 'morning' than 'evening' person", "More an 'evening' than a 'morning' person", "Definitely an 'evening' person", "Do not know" or "Prefer not to answer", which we coded as 2, 1, -1, -2, 0 and missing respectively (distribution summarised in **Table 1**). Prior to association testing, we adjusted the phenotype for age, gender and study centre (categorical). Of the 451,454 white European participants with genetic data, 449,734 were included in the GWAS (had non-missing phenotype and covariates).

Morning Person (UK Biobank)

In order to provide interpretable odds ratios for our genome-wide significant variants, we also defined a binary phenotype using the same data-field as for Chronotype. Participants answering "Definitely an 'evening' person" and "More an 'evening' than a 'morning' person" were coded as 0 (controls) and those answering "Definitely a 'morning' person" and "More a 'morning' than 'evening' person" were coded as 1 (cases). Participants answering "Do not know" or "Prefer not to answer" were coded as missing. A total of 403,195 participants were included in the GWAS (252,287 cases and 150,908 controls).

Activity monitor Phenotypes

Identifying the sleep period window

The software package GGIR[68,69] produces quantitative and timing measures relating to both activity levels and sleep patterns, with a day-by-day breakdown, as well averages across the period of wear. A new algorithm, implemented in version 1.5-12 of the GGIR R package and validated using PSG in an external cohort[70], allows for detection of sleep periods without the use of a sleep diary and with minimal bias. Briefly, for each individual, median values of the absolute change in z-angle (representing the dorsal-ventral direction when the wrist is in the anatomical position) across 5-minute rolling windows were calculated across a 24-hour period, chosen to make the algorithm insensitive to activity monitor orientation. The 10th percentile was incorporated into the threshold to distinguish movement from non-movement. Bouts of inactivity lasting ≥30 minutes are recorded as inactivity bouts. Inactivity bouts that are <60 minutes apart are combined to form inactivity blocks. The start and end of longest block defines the start and end of the sleep period time-window (SPT-window).

Activity monitor exclusions and adjustments

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The UK Biobank made multiple activity monitor data-quality variables available. From our activity monitor phenotypes, we excluded 4,925 samples with a non-zero or missing value in data field 90002 ("Data problem indicator"). We then excluded any individuals with the "good wear time" flag (field 90015) set to 0 (No), "good calibration" flag (field 90016) set to 0 (No), "calibrated on own data" flag (field 90017) set to 0 (No), "data recording errors" (field 90182) > 788 (Q_3 + 1.5xIQR) or a non-zero count of "interrupted recording periods" (field 90180). Phenotypes determined using the SPT-window (all phenotypes except L5 and M10 timing) had additional exclusions based on short (<3 hours) and long (>12 hours) mean sleep duration and too low (<5) or too high (>30) mean number of sleep episodes per night (see below). These additional exclusions were to ensure that individuals with extreme (outlying), and likely incorrect, sleep characteristics were not included in any subsequent analyses. Prior to association testing, we adjusted all phenotypes for age activity monitor worn (derived from month and year of birth and date activity monitor worn), gender, season activity monitor worn (categorical; winter, spring, summer or autumn; derived from date activity monitor worn) and number of valid measurements (SPT-windows for sleep phenotypes, number of valid days for diurnal inactivity or number of L5 or M10 detections).

Sleep midpoint (UK Biobank)

Sleep midpoint was calculated as the time directly between the start and end of the SPT-window and is defined as the number of hours elapsed since midnight at the start of the calendar day on which the STP-window started (e.g. 02:30 = 26.5; 23:45 = 23.75) with a cut-off at midday (12:00 and 36:00). This takes account of participants whose sleep midpoint occurs before midnight. Our sleep midpoint phenotype represented the average of each participant over all their valid SPT-windows. After exclusions and adjustments, 84,810 participants had valid sleep midpoint, covariates and genetic data.

L5 and M10 timing (UK Biobank)

L5 and M10 refer to the least-active five and the most-active ten hours of each day, and are commonly studied measures relating to circadian activity and sleep. L5 (M10) defines a five-hour (ten-hour) daily period of minimum (maximum) activity, as calculated by means of a moving average with a five-hour (ten-hour) window. As with sleep midpoint, we defined our L5 (M10) timing phenotype as the number of hours elapsed from the previous midnight to the L5 (M10) midpoint, averaged over all valid wear days. Of the 103,711 participants with activity monitor data, there were 85,205 and 85,670 with valid L5 and M10 timing measures respectively, covariates and genetic data. Basic summaries of these and other raw activity monitor phenotypes are given in **Supplementary Table 3**.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Sleep duration (UK Biobank)

Sleep episodes within the SPT-window were defined as periods of at least 5 minutes with no change larger than 5° associated with the z-axis of the activity monitor, as described previously[68]. The summed duration of all sleep episodes provided the sleep duration for a given SPT-window. We took both the mean and standard deviation of sleep duration across all valid SPT-windows to provide a measure of average sleep quantity and a measure of variability. After exclusions and adjustments, we had 85,449 (84,441) participants with valid sleep duration mean (SD), covariates and genetic data.

Sleep efficiency (UK Biobank)

This was calculated as a ratio of sleep duration (defined above) to SPT-window duration. The phenotype represented the mean across all valid SPT-windows and after exclusions and adjustments, left us with 84,810 participants with valid sleep efficiency, covariates and genetic data.

Number of sleep episodes (UK Biobank)

This is defined as the number of sleep episodes of at least 5 minutes separated by at least 5 seconds of wakefulness within the SPT-window. The phenotype represents the mean across all SPT-windows and can be interpreted as a measure of sleep disturbance or fragmentation. After exclusions and adjustments, we had 84,810 participants with valid sleep efficiency, covariates and genetic data.

Diurnal inactivity duration (UK Biobank)

The total daily duration of estimated bouts of inactivity that fall outside of the SPT-window. This comprises the total length of periods of sustained inactivity (>5 minutes) and captures sleep (naps) but does not include other inactivity such as sitting and reading or watching television, which typically involve a detectable level of movement. This variable likely captures some non-sleep rest as it is impossible to separate these without detailed activity diaries. The phenotype is calculated as the mean across all valid days and, after exclusions and adjustments, we were left with 84,757 participants with a valid measure, covariates and genetic data.

Genome-wide association analysis

We performed all association test using BOLT-LMM[71] v2.3, which applies a linear mixed model (LMM) to adjust for the effects of population structure and individual relatedness, and

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

allowed us to include all related individuals in our white European subset, boosting our power to detect associations. This meant a sample size of up to 449,734 individuals, as opposed to the set of 379,768 unrelated individuals. BOLT-LMM approximates relatedness within a cohort by using LD blocks and avoids the requirement of building a genetic-relationship matrix (GRM), with which calculations are intractable in cohorts of this size. From the ~805,000 directly-genotyped (non-imputed) variants available, we identified 524,307 "good-quality" variants (bi-allelic SNPs; MAF≥1%; HWE $P>1x10^{-6}$; non-missing in all genotype batches, total missingness<1.5% and not in a region of long-range LD[72]) that BOLT-LMM used to build its relatedness model. For LD structure information, we used the default 1000 Genomes LD-Score table provided with the software. We forced BOLT-LMM to apply a non-infinitesimal model, which provides better effect size estimates for variants with moderate to large effect sizes, in exchange for increased computing time. Prior to association testing, continuous phenotypes were first adjusted for relevant covariates, as indicated above, and at runtime we included "release" (categorical; UKBiLEVE array, UKB Axiom array interim release and UKB Axiom array full release) as a further covariate. The binary morning person phenotype was adjusted at runtime for age, gender, study centre and release.

In the 23andMe morning person GWAS, the summary statistics were generated through logistic regression (using an additive model) of the phenotype against the genotype, adjusting for age, gender, the first four principal components and a categorical variable representing genotyping platform. Genotyping batches in which particular variants failed to meet minimum quality control were not included in association testing for those variants, resulting in a range of sample sizes over the whole set of results. A λ_{GC} of 1.325 was reported for this GWAS. Lead variants for the 23andMe only morning person GWAS are provided in **Supplementary Table 18**.

Sensitivity Analysis

To avoid issues with stratification, we performed a sensitivity GWAS, in UK Biobank alone, to assess whether any of the associations were driven by a subset of the cohort with specific conditions. We excluded those reporting shift or night shift work at baseline, those taking medication for sleep or psychiatric disorders and those with either with a HES ICD10 or self-reported diagnosis of depression, schizophrenia, bipolar disorder, anxiety disorders or mood disorder (see **Supplementary Methods** for further details). Results for the 341 lead chronotype variants available in the UK Biobank are provided in **Supplementary Table 1** alongside the main meta-analysis results.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Meta-analysis of GWAS results

Meta-analysis was performed using the software package METAL[73]. To obtain the largest possible sample size, and thus maximising statistical power, we performed a sample-size meta-analysis, using the results from the UK Biobank chronotype GWAS and the 23andMe morning person GWAS. Genomic control was not performed on each set of summary statistics prior to meta-analysis but instead the meta-analysis chi-squared statistics were corrected using the LD score intercept (I_{LDSC} = 1.0829), calculated by the software LDSC, as using λ_{GC} is considered overly conservative and the LD score intercept better captures inflation due to population stratification[74]. For interpretable results, we reported the odds ratio from a secondary effect size meta-analysis between our dichotomous UK Biobank morning person GWAS and the 23andMe morning person GWAS. The primary chronotype sample-size meta-analysis produced results for 15,880,941 variants in up to 697,828 individuals, with the secondary effect-size morning person meta-analysis producing results for 15,880,664 variants in up to 651,295 individuals (372,765 cases and 278,530 controls).

Post-GWAS analyses

Pathway analysis and tissue-enrichment

We used MAGENTA[37], DEPICT[38], PASCAL[36] and MAGMA[34] to perform pathway and tissue enrichment. For MAGENTA and DEPICT, we included all variants from the meta-analysis, whereas for PASCAL, we included only those with an RSID as the software assigns variants to genes using their RSID. For the MAGENTA analysis, we used upstream and downstream limits of 110Kb and 40Kb to assign variants to genes by position, we excluded the HLA region from the analysis and set the number of permutations for gene-set enrichment analysis (GSEA) to 10,000. For DEPICT, we used the default settings and the annotation and mapping files provided with the software. As each of the four pieces of software adopts a different gene prioritisation method or relies on different databases, we included results from all four to cover all bases and to allow for better comparison with other studies, where only a single method may have been used. Briefly PASCAL corrects for the effect of LD blocks by accounting for the LD structure between associated variants, MAGENTA uses distance-based mapping but allows the user to set the upstream and downstream distances for inclusion, DEPICT makes use of large-scale data on gene co-regulation to prioritise genes before calculating enrichment in its own reconstituted gene sets and MAGMA, the most recent method (and implemented in the FUMA GWAS[35] platform), claims greater statistical power to detect enriched gene sets than methods such as MAGENTA and PASCAL, without affecting the type 1 error rate. By using multiple methods and looking for

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

consistency, we provide more compelling evidence of enrichment in specific pathways and tissues.

Genetic correlation and heritability analyses

We used the LD Score Regression (LDSC) software, available at https://github.com/bulik/ldsc/, to quantify the genetic overlap between the trait of interest and 222 traits with publicly available GWA data. Details of methodology are available elsewhere[74]. We considered any correlation as statistically significant if it had a Bonferroni corrected $P<0.05$.

Fine-mapping association signals

Fine-mapping analyses were performed using FINEMAP v1.1[41] using a shotgun stochastic approach, allowing up to 20 causal SNPs at each locus and by focussing on a 1Mb (±500Kb) region around each index variant. As FINEMAP assumes a fixed sample size for all variants, we excluded variants not present in both the UK Biobank and 23andMe data, and to make the LD calculations more tractable we excluded variants with P>0.01 to limit the total number of variants at each locus. We constructed an LD matrix for each locus by calculating the Pearson correlation coefficient for all pairs of variants using dosages derived from the unrelated European-ancestry subset of the UK Biobank imputed genotype probabilities (N=379,769). A variant was considered to be causal if its $\log_{10}$ Bayes factor was 2 or larger, a limit recommended by the FINEMAP documentation (http://www.christianbenner.com/index_v1.1.html).

Alamut annotation, eQTL mapping and circadian enrichment analyses

We annotated variants identified by FINEMAP as likely to be causal using Alamut Batch v1.8 (Interactive Biosoftware, Rouen, France) with genome assembly GRCh37 and all options set to default. We retained only the canonical (longest) transcript for each variant and reported the variant location and coding effect (if applicable) in this transcript. To identify whether variants were cis-eQTLs for nearby genes, we performed a lookup of our variants in the GTEx single-tissue cis-eQTL dataset (v7), accessed at the GTEx portal (https://www.gtexportal.org/home/datasets) on 13/07/18, for significant associations. A variant was reported as an eQTL for a gene if the variant-gene association was significant (q-value ≤ 0.05) for one or more brain or non-brain tissues.

With the aim of highlighting genes that have a role in regulating the internal circadian clock, we cross-referenced the genes identified by eQTL mapping, in addition to the two nearest

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genes (within 1Mb), with catalogues from three gene expression studies. Firstly, we used data from an RNAi screen of circadian clock modifiers[49], in which a genome-wide scan was performed on the effects of single-gene knockouts on the amplitude and period of the circadian expression. Secondly, we used data from a study of gene expression in SCN tissue over a 24-hour light/dark cycle[47] to identify whether our genes exhibit fluctuating expression in SCN tissue and whether the genes show enriched expression in the SCN compared to other tissues. Finally, we used data from a meta-analysis of gene expression in the SCN[48] to investigate whether the genes were preferentially expressed in the SCN when compared to other brain tissues.

Mendelian Randomisation analyses

We undertook MR analyses to explore both the effect of chronotype on different outcomes and the effect of different exposures on chronotype as an outcome. These two-sample MR analyses can be summarised by:

1. Chronotype exposure using the 351 variants and effect sizes discovered in this meta-analysis against the five significant psychiatric outcomes from the genetic correlation analyses and three metabolic outcomes, using summary data from published GWAS (**Supplementary Table 16**).
2. Two of the five significant psychiatric exposures from the genetic correlation analyses and four metabolic exposures, all using variants from published GWAS, against chronotype as an outcome, using summary data from this meta-analysis (**Supplementary Table 17**).

In both analyses, we tested four MR methods:

a. Inverse-variance weighting (IVW)[75]
b. MR-Egger[75]
c. Weighted median (WM)[76]
d. Penalised weighted median (PWM)[76]

Analysis 1 (chronotype exposure) was performed using the R package TwoSampleMR using aggregated summary statistics available through the MR-Base platform[56]. We implemented the four MR methods listed above and also included the MR-Egger bootstrap to provide better estimates of the effect sizes and standard errors as compared to the MR-Egger method. We used data from published GWAS to test the effect of chronotype on the following exposures: schizophrenia[77], major depressive disorder[78], depressive symptoms[79], subjective wellbeing[79], PGC cross-disorder traits[80], fasting insulin[81], BMI[55,82] and T2D[83,84]. To

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

provide meaningful effect sizes for MR analyses, we used betas from the secondary effect size meta-analysis of the dichotomous UK Biobank and 23andMe morning person GWAS.

For analysis 2 (chronotype outcome) we applied the four MR methods listed above, utilising a custom pipeline. Using data from published GWAS, we tested whether chronotype is influenced by the following exposures: schizophrenia[77], major depressive disorder[78], insulin secretion[85], favourable adiposity[86], BMI[55] and T2D[87]. As with analysis 1, chronotype effect sizes represented morningness liability and were taken from the secondary morning person meta-analysis.

We used the inverse variance weighted approach as our main analysis method and MR-Egger, weighted median estimation and penalised weighted median estimation as sensitivity analyses in the event of unidentified pleiotropy of our genetic instruments. MR results may be biased by horizontal pleiotropy, i.e. where the genetic variants that are robustly related to the exposure of interest (here chronotype) independently influence the outcome, through association with another risk factor for the outcome. IVW assumes that there is either no horizontal pleiotropy (under a fixed effect model) or, if implemented under a random effects model after detecting heterogeneity amongst the causal estimates, that:

i. The strength of association of the genetic instruments with the risk factor is not correlated with the magnitude of the pleiotropic effects
ii. The pleiotropic effects have an average value of zero

MR-Egger provides unbiased causal estimates if just the first condition above holds, by estimating and adjusting for non-zero mean pleiotropy. However, MR-Egger requires that the InSIDE (Instrument Strength Independent of Direct Effect) assumption[88] holds, in that it needs the pleiotropy of the genetic instruments to be uncorrelated with the instruments' effect on the exposure. The weighted median approach is valid if less than 50% of the weight in the analysis stems from variants that are pleiotropic (i.e. no single SNP that contributes 50% of the weight or a number of SNPs that together contribute 50% should be invalid because of horizontal pleiotropy). Given these different assumptions, if all methods are broadly consistent this strengthens our causal inference. Additional care should be taken interpreting results from binary exposures or outcomes, as these MR methods assume that horizontal pleiotropy due to within-category variation of dichotomous or categorical traits is negligible.

Data availability

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Summary statistics for the top 10,000 chronotype meta-analysis variants are provided in **Supplementary Table 13** and the full set of UK Biobank-only chronotype and morning person GWAS summary statistics can be found at http://www.t2diabetesgenes.org/data/. Full meta-analysis summary statistics can be requested directly from 23andMe Inc. (see https://research.23andme.com/collaborate/#publication).

Acknowledgements

This research has been conducted using the UK Biobank Resource (application 9072). We would like to thank the research participants and employees of 23andMe for making this work possible. We also wish to acknowledge the Genotype-Tissue Expression (GTEx) Project, supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS, for providing access to the data we required.

Contributor list for the 23andMe Research Team: Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson.

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. The funders had no influence on study design, data collection and analysis, decision to publish, or preparation of the manuscript. JB and DAL work in a Unit that receives support from the University of Bristol and UK Medical Research Council (MC_UU_00011/3 and MC_UU_00011/6, respectively). DAL's contribution to this work is also supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement (Grant number 669545; DevelopObese). EMB is

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

supported by grants from the National Health and Medical Research Council of Australia 1145645, 1078901 and 1087889

Duality of Interest. DAL receives support from Roche Diagnostics and Medtronic for research that is unrelated to this study. MKR reports receiving honoraria and consulting fees from Novo Nordisk, Ascensia, Cell Catapult and Roche Diabetes Care.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Chronotype category	Phenotype coding	N	Sex (% male)	Age (SD)	TDI	BMI (SD)
Definitely morning	2	107,555	43.6	57.7 (7.7)	-1.4	27.5 (4.8)
More morning than evening	1	144,731	43.9	57.0 (7.9)	-1.7	27.1 (4.6)
Don't know	0	46,538	57.1	56.8 (8.0)	-1.43	27.3 (4.7)
More evening than morning	-1	115,090	45	56.1 (8.2)	-1.41	27.4 (4.8)
Definitely evening	-2	35,818	46.8	55.3 (8.3)	-1.05	27.9 (5.2)
All		449,732	45.7	56.8 (8.0)	-1.47	27.4 (4.8)

Table 1. Summary of sex, age, Townsend Deprivation Index (TDI) and BMI by chronotype categories in the UK Biobank study.

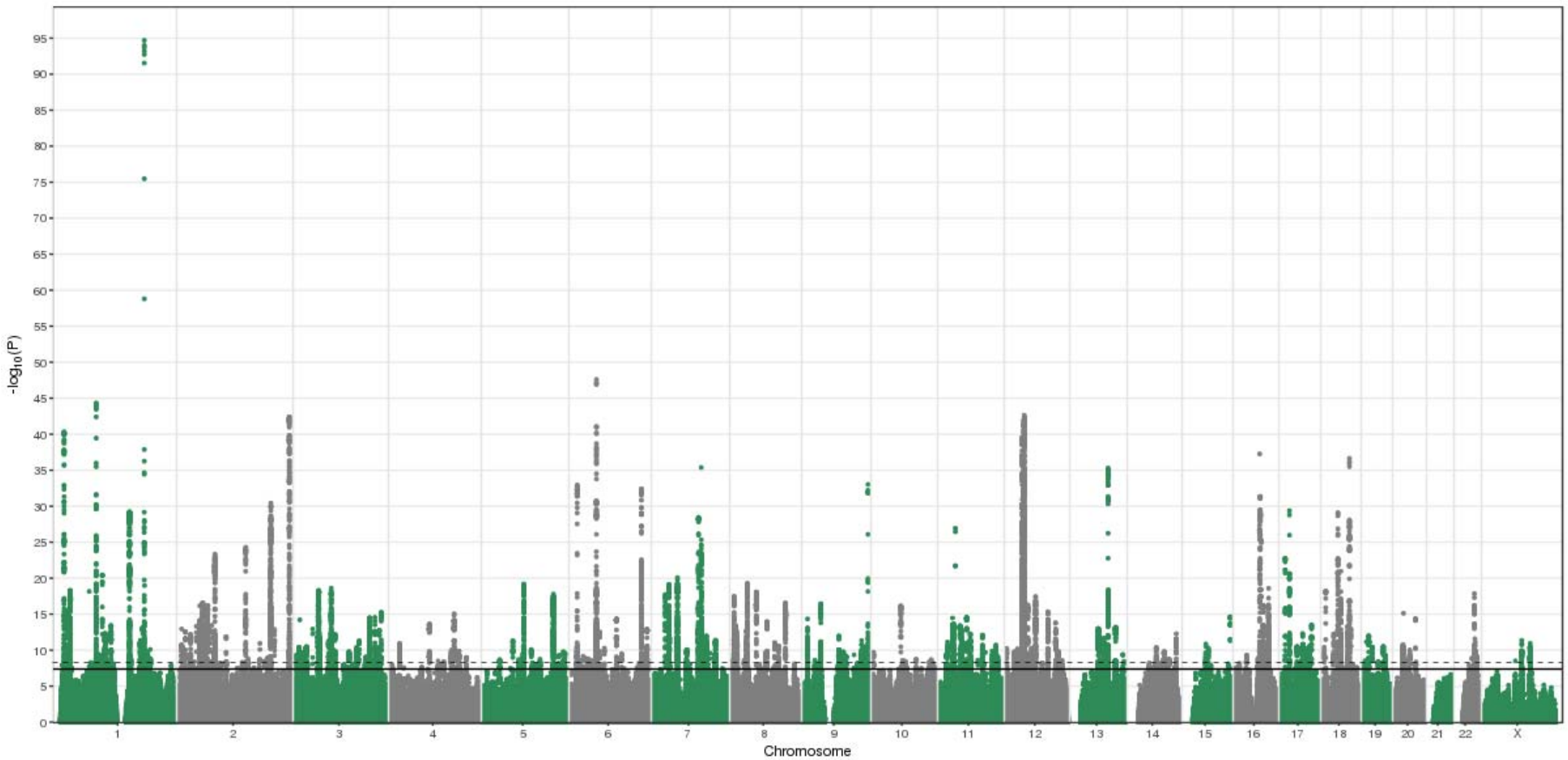

Figure 1. Manhattan plot of the chronotype meta-analysis GWAS. The solid line indicates the typical genome-wide significance threshold of P=5x10^{-8} and the dashed line marks the threshold of P=6x10^{-9} identified through permutation testing.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

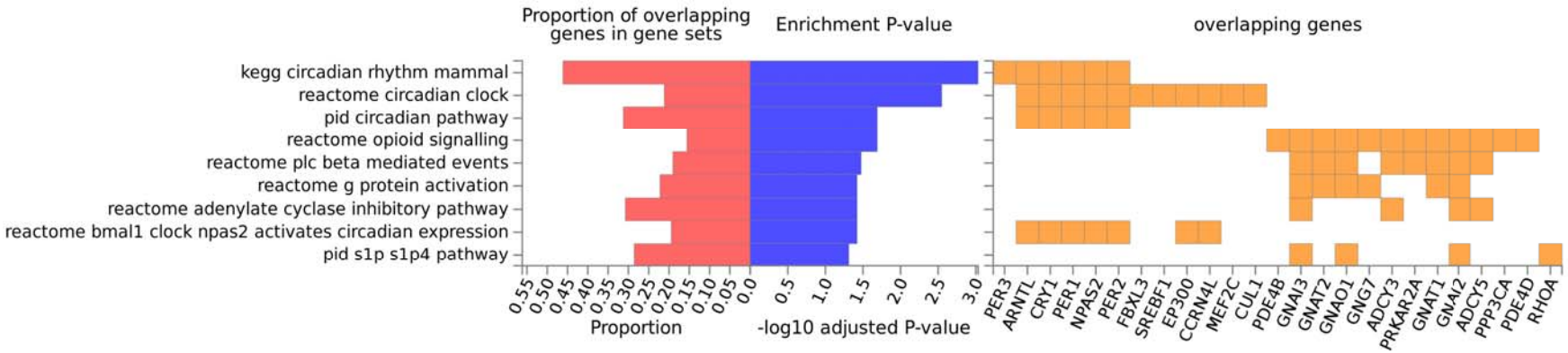

Figure 2. Reactome gene sets overlapping Chronotype genes identified using positional and eQTL mapping in FUMA's GENE2FUNC process. Note that these results may differ to those produced by MAGMA.

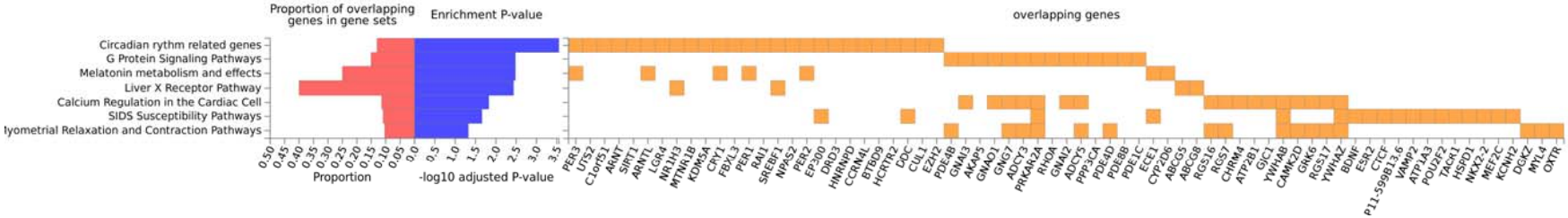

Figure 3. WikiPathways gene sets overlapping Chronotype genes identified using positional and eQTL mapping in FUMA's GENE2FUNC process. Note that these results may differ to those produced by MAGMA.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

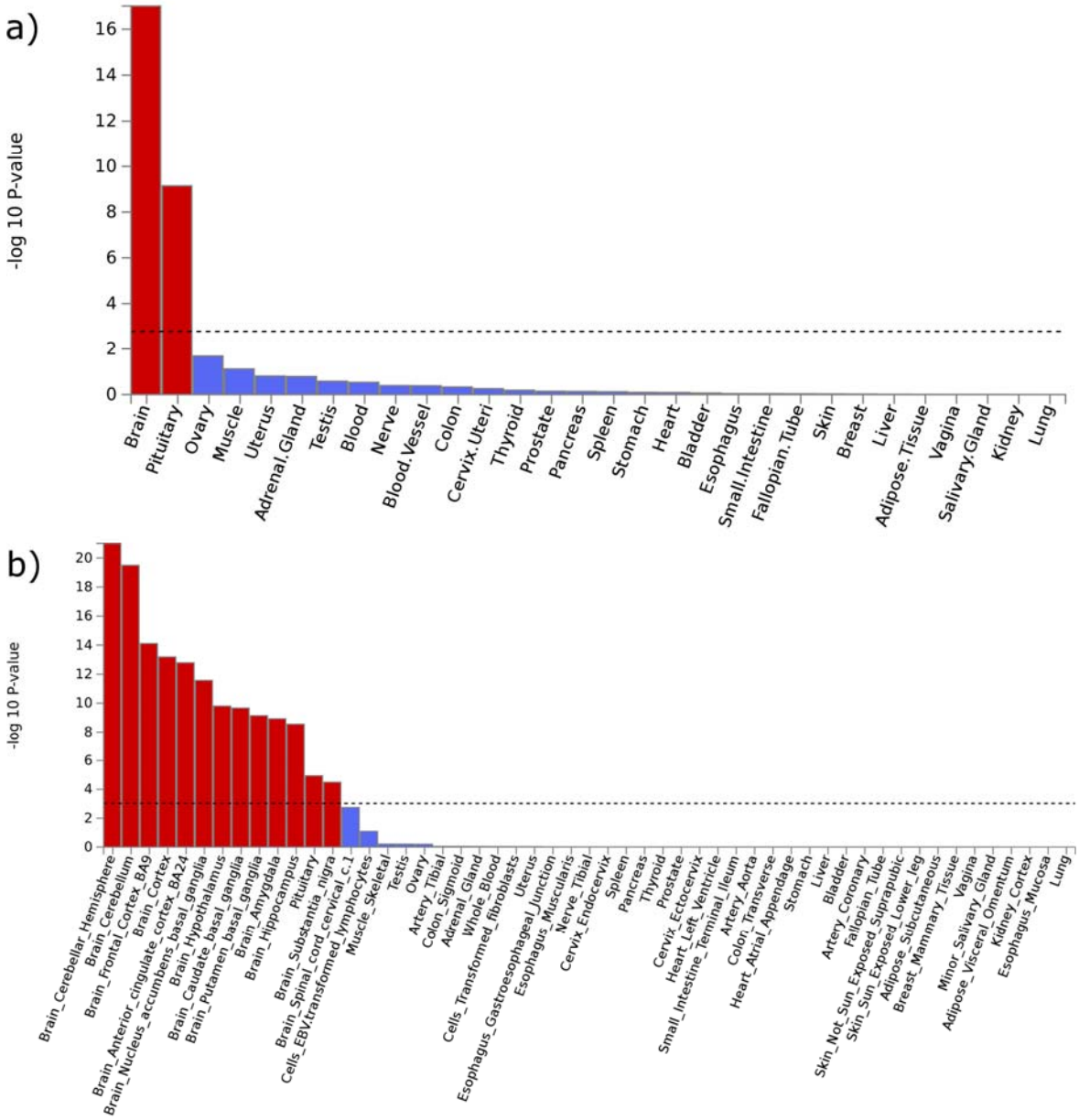

Figure 4. MAGMA tissue expression analysis using gene expression per tissue based on GTEx RNA-seq data for a) 30 general and b) 53 specific tissue types.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

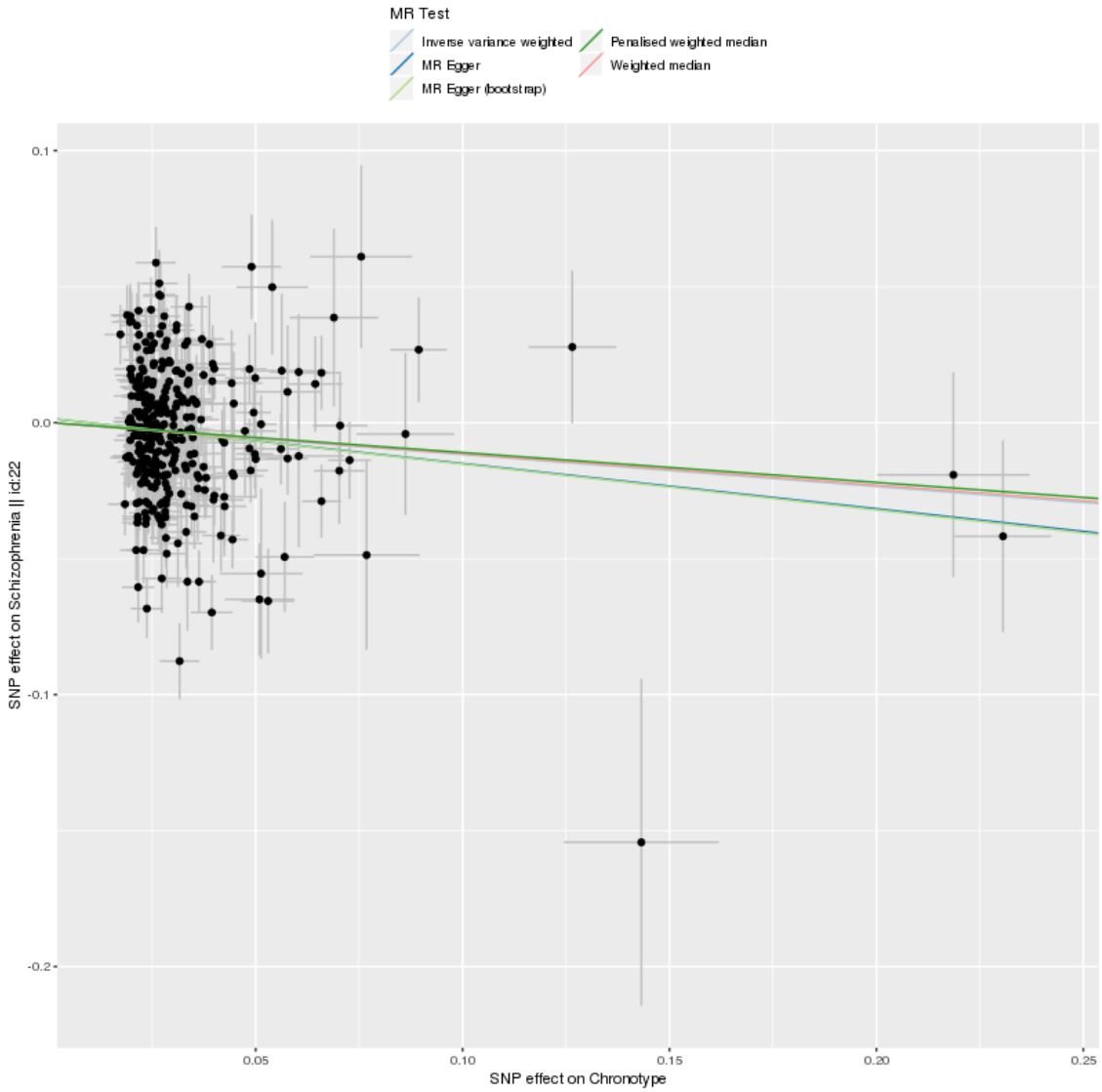

Figure 5. Scatter plot of Chronotype meta-analysis variants and their effects on schizophrenia in the PGC GWAS[77] (outcome) versus odds of being a morning person (exposure). Lines identify the slopes of the five methods tested. Log odds (and SEs) for morningness were taken from the secondary effect-size meta-analysis.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

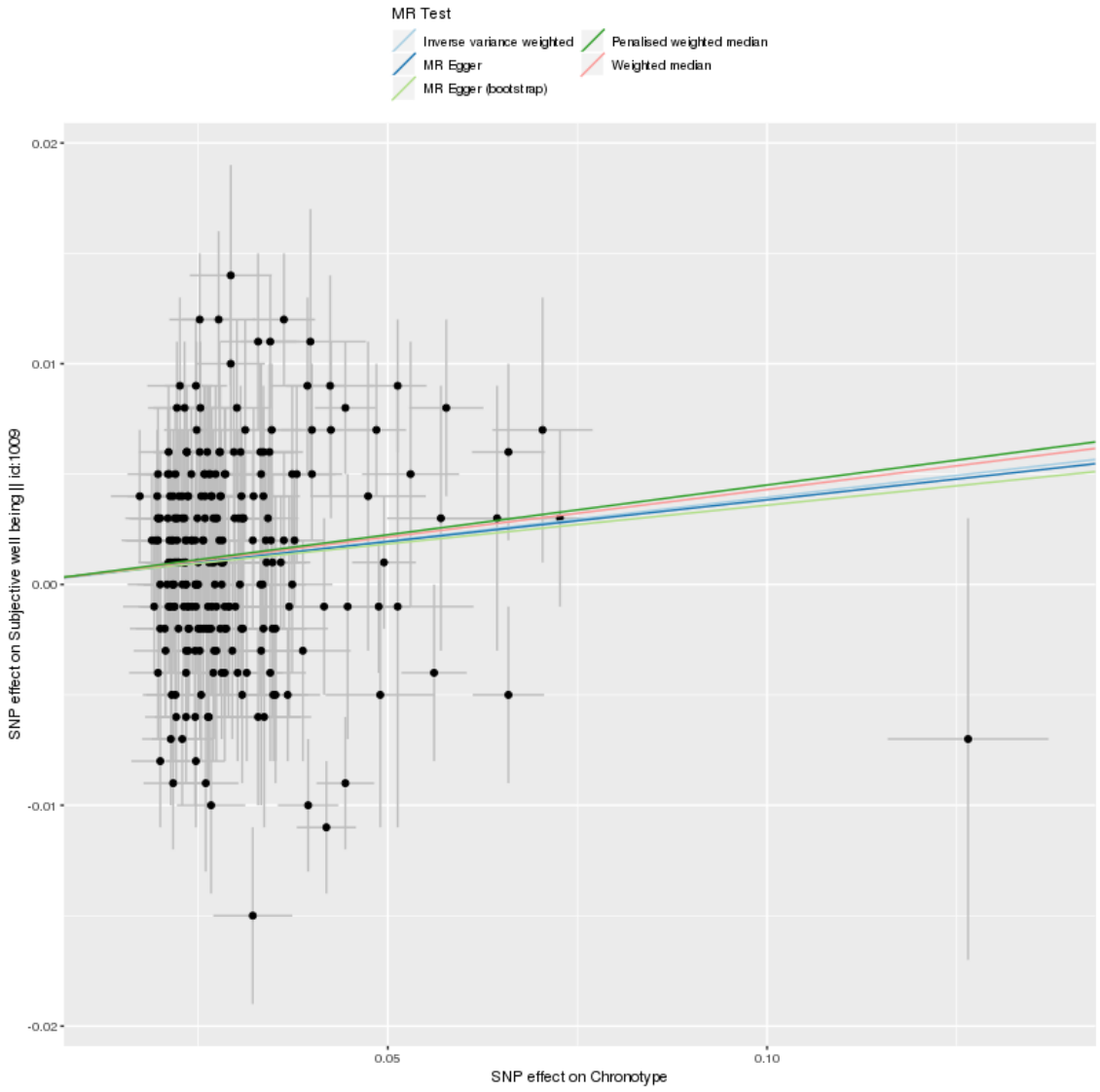

Figure 6. Scatter plot of Chronotype meta-analysis variants and their effects on subjective wellbeing in the SSGAC GWAS[79] (outcome) versus odds of being a morning person (exposure). Lines identify the slopes of the five methods tested. Log odds (and SEs) for morningness were taken from the secondary effect-size meta-analysis.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary Figure 1. Scatter plot of UK Biobank chronotype GWAS effect sizes for all 341 lead variants present in UK Biobank, against their effect sizes in the UK Biobank chronotype sensitivity GWAS. The dashed line indicates identical effect in both analyses.

Supplementary Figure 2. FUMA regional plots identifying LD patterns between chronotype-associated variants in the a) RGS16, b) PER2, c) PER3, d) PIGK/AK5, e) INADL, f) HCRTR2, g) HTR6, h) PER1, i) CRY1 and j) ARNTL loci. LD r^2 of each variant with the lead is indicated by the colour of the points, with mapped genes highlighted beneath in red.

Supplementary Figure 3. Scatter plot of Chronotype meta-analysis variants and their effects on major depressive disorder in the PGC GWAS[78] (outcome) versus odds of being a morning person (exposure). Lines identify the slopes of the five methods tested. Log odds (and SEs) for morningness were taken from the secondary effect-size meta-analysis.

Supplementary Figure 4. Scatter plot of Chronotype meta-analysis variants and their effects on depressive symptoms in the SSGAC GWAS[79] (outcome) versus odds of being a morning person (exposure). Lines identify the slopes of the five methods tested. Log odds (and SEs) for morningness were taken from the secondary effect-size meta-analysis.

Supplementary File 1. R wrapper script used to call GGIR to process the converted (.wav) UK Biobank activity monitor files. All settings, except file locations, are identical to those used in this study.

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

1. Dibner, C. & Schibler, U. Circadian timing of metabolism in animal models and humans. *J. Intern. Med.* **277,** 513–527 (2015).
2. Roenneberg, T. *et al.* Epidemiology of the human circadian clock. *Sleep Med. Rev.* **11,** 429–438 (2007).
3. Horne, J. A. & Ostberg, O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int. J. Chronobiol.* **4,** 97–110 (1976).
4. Smith, C. S., Reilly, C. & Midkiff, K. Evaluation of Three Circadian Rhythm Questionnaires With Suggestions for an Improved Measure of Morningness. *J. Appl. Psychol.* **74,** 728–738 (1989).
5. 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,** 117–120 (2002).
6. Fischer, D., Lombardi, D. A., Marucci-wellman, H. & Roenneberg, T. Chronotypes in the US – Influence of age and sex. 1–17 (2017).
7. Allebrandt, K. V *et al.* Chronotype and sleep duration: the influence of season of assessment. *Chronobiol. Int.* **31,** 731–40 (2014).
8. Leocadio-Miguel, M. A. *et al.* Latitudinal cline of chronotype. *Sci. Rep.* **7,** 2–7 (2017).
9. 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).
10. Schmid, S. M., Hallschmid, M. & Schultes, B. The metabolic burden of sleep loss. *Lancet Diabetes Endocrinol.* **3,** 52–62 (2015).
11. Kohsaka, A. *et al.* High-Fat Diet Disrupts Behavioral and Molecular Circadian Rhythms in Mice. *Cell Metab.* **6,** 414–421 (2007).
12. Marcheva, B. *et al.* Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* **466,** 627–31 (2010).
13. Turek, F. W. *et al.* Obesity and Metabolic Syndrome in Circadian Clock Mutant Mice. *Science (80-.).* **308,** 1043–1045 (2005).
14. Morris, C. J., Purvis, T. E., Mistretta, J., Hu, K. & Scheer, F. A. J. L. Circadian Misalignment Increases C-Reactive Protein and Blood Pressure in Chronic Shift Workers. *J. Biol. Rhythms* **32,** 154–164 (2017).
15. Wright, K. P. *et al.* Influence of sleep deprivation and circadian misalignment on cortisol, inflammatory markers, and cytokine balance. *Brain. Behav. Immun.* **47,** 24–34 (2015).
16. Skarke, C. *et al.* A pilot characterization of the human chronobiome. *Sci. Rep.* **7,** 1–12 (2017).
17. Patterson, F. *et al.* Interactive effects of sleep duration and morning/evening preference on cardiovascular risk factors. *Eur. J. Public Health* (2017). doi:10.1093/eurpub/ckx029
18. Merikanto, I. *et al.* Associations of Chronotype and Sleep With Cardiovascular Diseases and Type 2 Diabetes. *Chronobiol. Int.* **30,** 470–477 (2013).
19. Levandovski, R. *et al.* Depression Scores Associate With Chronotype and Social Jetlag in a Rural Population. *Chronobiol. Int.* **28,** 771–778 (2011).
20. Vetter, C. *et al.* Mismatch of sleep and work timing and risk of type 2 diabetes. *Diabetes Care* **38,** 1707–1713 (2015).
21. Smith, G. D. & Ebrahim, S. Mendelian randomization: Prospects, potentials, and limitations. *Int. J. Epidemiol.* **33,** 30–42 (2004).
22. Smith, G. D. *et al.* Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med.* **4,** 1985–1992 (2007).
23. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: Using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27,** 1133–1163 (2008).
24. Hu, Y. GWAS of 89,283 individuals identifies genetic variants associated with self-reporting of being a morning person. *Nat. Commun.* **7,** 10448 (2016).
25. Lane, J. M. Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nat. Commun.* **7,** 10889 (2016).
26. Jones, S. E. *et al.* Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci. *PLOS Genet.* **12,** e1006125 (2016).
27. Sudlow, C. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12,** e1001779 (2015).
28. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *Doi.Org* 166298 (2017). doi:10.1101/166298
29. Zhang, L. *et al.* A *PERIOD3* variant causes a circadian phenotype and is associated with a

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

seasonal mood trait. *Proc. Natl. Acad. Sci.* **113,** E1536–E1544 (2016).
30. Sabia, S. *et al.* Association between questionnaire-and accelerometer-assessed physical activity: The role of sociodemographic factors. *Am. J. Epidemiol.* **179,** 781–790 (2014).
31. Lauderdale, D. S., Knutson, K. L., Yan, L. L., Liu, K. & Rathouz, P. J. Self-Reported and Measured Sleep Duration. *Epidemiology* **19,** 838–845 (2008).
32. Slightam, C. *et al.* Assessing sleep quality using self-report and actigraphy in PTSD. *J. Sleep Res.* (2017). doi:10.1111/jsr.12632
33. Cespedes, E. M. *et al.* Comparison of Self-Reported Sleep Duration with Actigraphy: Results from the Hispanic Community Health Study/Study of Latinos Sueño Ancillary Study. *Am. J. Epidemiol.* **183,** 561–573 (2016).
34. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput. Biol.* **11,** 1–19 (2015).
35. Watanabe, K. *et al.* FUMA: Functional mapping and annotation of genetic associations. *bioRxiv* **3,** 110023 (2017).
36. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput. Biol.* **12,** 1–20 (2016).
37. Ayellet, V. S., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6,** e1001058 (2010).
38. Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6,** (2015).
39. Shifman, S. *et al.* Genome-wide association identifies a common variant in the reelin gene that increases the risk of schizophrenia only in women. *PLoS Genet.* **4,** (2008).
40. Rujescu, D. *et al.* Disruption of the neurexin 1 gene is associated with schizophrenia. *Hum. Mol. Genet.* **18,** 988–996 (2009).
41. Benner, C. *et al.* FINEMAP: Efficient variable selection using summary data from genome-wide association studies. *Bioinformatics* **32,** 1493–1501 (2016).
42. Huyghe, J. R. *et al.* Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* **45,** 197–201 (2013).
43. Yang, Y. *et al.* Regulation of behavioral circadian rhythms and clock protein PER1 by the deubiquitinating enzyme USP2. *Biol. Open* **1,** 789–801 (2012).
44. Tong, X. *et al.* CUL4-DDB1-CDT2 E3 Ligase Regulates the molecular clock activity by promoting Ubiquitination-dependent degradation of the mammalian CRY1. *PLoS One* **10,** 1–19 (2015).
45. Stojkovic, K., Wing, S. S. & Cermakian, N. A central role for ubiquitination within a circadian clock protein modification code. *Front. Mol. Neurosci.* **7,** 1–7 (2014).
46. Tian, R., Alvarez-Saavedra, M., Cheng, H.-Y. M. & Figeys, D. Uncovering the Proteome Response of the Master Circadian Clock to Light Using an AutoProteome System. *Mol. Cell. Proteomics* **10,** M110.007252 (2011).
47. Pembroke, W. G., Babbs, A., Davies, K. E., Ponting, C. P. & Oliver, P. L. Temporal transcriptomics suggest that twin-peaking genes reset the clock. *Elife* **4,** 1–15 (2015).
48. Brown, L. A. *et al.* Meta-analysis of transcriptomic datasets identifies genes enriched in the mammalian circadian pacemaker. *Nucleic Acids Res.* **45,** 9860–9873 (2017).
49. Zhang, E. E. *et al.* A Genome-wide RNAi Screen for Modifiers of the Circadian Clock in Human Cells. *Cell* **139,** 199–210 (2009).
50. Perez-Santángelo, S. *et al.* Role for *LSM* genes in the regulation of circadian rhythms. *Proc. Natl. Acad. Sci.* **111,** 15166–15171 (2014).
51. Colwell, C. S. *et al.* Disrupted circadian rhythms in VIP- and PHI-deficient mice. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **285,** R939–R949 (2003).
52. Roff, D. A. The estimation of genetic correlations from phenotypic correlations□: a test of Cheverud ' s conjecture. *Heredity (Edinb).* **74,** 481–490 (1995).
53. Sodini, S. M., Kemper, K. E., Wray, N. R. & Trzaskowski, M. Comparison of genotypic and phenotypic correlations: Cheverud's conjecture in humans. *Genetics* **209,** 941–948 (2018).
54. Speliotes, E. K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42,** 937–948 (2010).
55. Locke, A. E. *et al.* Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518,** 197–206 (2015).
56. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *Elife* **7,** e34408 (2018).

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

57. Arora, T. & Taheri, S. Associations among late chronotype, body mass index and dietary behaviors in young adolescents. *Int. J. Obes.* **39,** 39–44 (2015).
58. 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).
59. Giglio, L. M. F. *et al.* Circadian preference in bipolar disorder. *Sleep Breath.* **14,** 153–155 (2010).
60. Tonetti, L., Fabbri, M. & Natale, V. Relationship between circadian typology and big five personality domains. *Chronobiol. Int.* **26,** 337–347 (2009).
61. Merikanto, I. *et al.* Evening types are prone to depression. *Chronobiol. Int.* **30,** 719–725 (2013).
62. 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).
63. Monti, J. M. *et al.* Sleep and circadian rhythm dysregulation in schizophrenia. *Prog. Neuro-Psychopharmacology Biol. Psychiatry* **43,** 209–216 (2013).
64. Munafò, M. R., Tilling, K., Taylor, A. E., Evans, D. M. & Davey Smith, G. Collider scope: when selection bias can substantially influence observed associations. *Int. J. Epidemiol.* 226–235 (2017). doi:10.1093/ije/dyx206
65. Taylor, A. E. *et al.* Mendelian randomization in health research: Using appropriate genetic variants and avoiding biased estimates. *Econ. Hum. Biol.* **13,** 99–106 (2014).
66. Doherty, A. *et al.* Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study. *PLoS One* **12,** 1–14 (2017).
67. van Hees, V. *et al.* Auto-calibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J. Appl. Physiol.* **44,** 738–744 (2014).
68. Van Hees, V. T. *et al.* A novel, open access method to assess sleep duration using a wrist-worn accelerometer. *PLoS One* **10,** 1–13 (2015).
69. van Hees, V. T. *et al.* Separating Movement and Gravity Components in an Acceleration Signal and Implications for the Assessment of Human Daily Physical Activity. *PLoS One* **8,** 1–10 (2013).
70. van Hees, V. T. *et al.* Estimating sleep parameters using an accelerometer without sleep diary. *bioRxiv* (2018). at <http://biorxiv.org/content/early/2018/02/01/257972.abstract>
71. Loh, P.-R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47,** 284–290 (2015).
72. Price, A. L. *et al.* Long-Range LD Can Confound Genome Scans in Admixed Populations. *Am. J. Hum. Genet.* **83,** 132–135 (2008).
73. Willer, C. J., Li, Y. & Abecasis, G. R. \pkg{METAL}: Fast and Efficient Meta-analysis of Genomewide Association Scans. *Bioinformatics* **26,** 2190–2191 (2010).
74. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **advance on,** 291–295 (2015).
75. Bowden, J. *et al.* Assessing the suitability of summary data for two-sample mendelian randomization analyses using MR-Egger regression: The role of the I2statistic. *Int. J. Epidemiol.* **45,** 1961–1974 (2016).
76. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* **40,** 304–314 (2016).
77. Ripke, S. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511,** 421–427 (2014).
78. Sullivan, P. F. *et al.* A mega-Analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18,** 497–511 (2013).
79. Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48,** 624–633 (2016).
80. Smoller, J. W. *et al.* Identification of risk loci with shared effects on five major psychiatric disorders: A genome-wide analysis. *Lancet* **381,** 1371–1379 (2013).
81. Manning, A. K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44,** 659–669 (2012).
82. Randall, J. C. *et al.* Sex-stratified Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic Loci for Anthropometric Traits. *PLoS Genet.* **9,** (2013).
83. Replication, Dia. G. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the

bioRxiv preprint doi: https://doi.org/10.1101/303941; this version posted September 13, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46,** 234–244 (2014).

84. Morris, A. D. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat Genet* **44,** 981–990 (2012).
85. Dimas, A. S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63,** 2158–2171 (2014).
86. Yaghootkar, H. *et al.* Genetic evidence for a link between favorable adiposity and lower risk of type 2 diabetes, hypertension, and heart disease. *Diabetes* **65,** 2448–60 (2016).
87. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes article. *Nat. Genet.* **50,** 559–571 (2018).
88. Burgess, S., Bowden, J., Dudbridge, F. & Thompson, S. G. Robust instrumental variable methods using multiple candidate instruments with application to Mendelian randomization. 1–38 (2016). at <http://arxiv.org/abs/1606.03729>