Genetic basis of chronotype in humans: Insights from three landmark GWAS

David A Kalmbach PhD,[1] Logan D Schneider MD,[2] Joseph Cheung MD MS,[2] Sarah J Bertrand PhD,[3] Thiruchelvam Kariharan PhD,[4] Allan I Pack MBChB PhD,[5*] & Philip R Gehrman PhD[5*]

[1]Departments of Psychiatry and Neurology, University of Michigan Medical School, Ann Arbor, MI 48109

[2]Center for Sleep Sciences and Medicine, Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA 94063

[3]Department of Anesthesiology and Critical Care Medicine, Johns Hopkins Hospital School of Medicine, Baltimore, MD 21205

[4]Department of Molecular, Cellular, and Developmental Biology, University of Michigan, Ann Arbor 48109

[5]Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA 19104

*co-senior authors

Total number of Figures: 3
Total number of Table: 1

This work was supported by NIH grant T32 HL110952-04

Conflicts of Interest Disclosures: Dr. Gehrman has had research grant support from Merck, Inc. and has served as a consultant for Johnson & Johnson and General Sleep Corporation. The other authors reported no financial conflicts of interest.

Corresponding author:

Philip R Gehrman PhD
Department of Psychiatry
University of Pennsylvania
3535 Market Street, Suite 670
Philadelphia, PA 19104
gehrman@exchange.upenn.edu
Phone: 215-746-3578
Fax: 215-573-0759

© Sleep Research Society 2017. Published by Oxford University Press on behalf of the Sleep Research Society. All rights reserved. For permissions, please e-mail journals.permissions@oup.com.

Accepted Manuscript

Abstract

Chronotype, or diurnal preference, refers to behavioral manifestations of the endogenous circadian system that governs preferred timing of sleep and wake. As variations in circadian timing and system perturbations are linked to disease development, the fundamental biology of chronotype has received attention for its role in the regulation and dysregulation of sleep and related medical and psychiatric illnesses. Twin and family studies have shown that chronotype is a heritable trait, thus directing attention toward its genetic basis. Chronotype is influenced by a large number of genetic variants, making it a *complex phenotype*. Although discoveries from molecular studies of candidate genes have shed light onto its genetic architecture, the contribution of genetic variation to chronotype has remained unclear with few related variants identified. In the advent of large-scale genome-wide association studies (GWASs), scientists now have the ability to discover novel common genetic variants associated with complex phenotypes such as chronotype. Three recent large-scale GWASs of chronotype were conducted on subjects of European ancestry from the 23andMe cohort and the UK Biobank. Several genes previously unknown to influence chronotype were identified. The present review discusses findings of these three GWASs in the context of prior research. Replicated and non-replicated results across these GWASs are discussed along with explorations of methodological differences. Future directions in molecular science, model systems, and discovery of rare variants are discussed.

Keywords: circadian rhythms; chronotype; genome-wide association study; genetics; sleep

Accepted Manuscript

Statement of significance

An individual's tendency to be a *night owl* (late chronotype) or a *morning lark* (early chronotype) is influenced by his or her genetic makeup. Expression of chronotype (across early and late types) is normally distributed in the population, which suggests that many common genetic variants with modest effects contribute to the phenotype. In our review, we discuss three recent hallmark studies using genome-wide association analyses for the identification of genes associated with trait chronotype. Findings identified a number of genes with previously known roles in the circadian clock, as well as several novel gene loci that had not been previously identified as having a role in chronotype or circadian rhythms. Genome-wide analysis is an important tool for identifying novel common gene variants, which illuminates previously unknown genetic influences. Future research on the genetic basis of chronotype should consider using model systems to elucidate circadian function of novel loci.

Accepted Manuscript

Introduction

Circadian rhythms are cyclical changes in cellular, molecular, and biological processes that repeat approximately once every 24 hours. Driven by an internal 'master clock' located in the suprachiasmatic nucleus (SCN) and influenced by environmental stimuli/*zeitgebers* (e.g., light-dark cycles), circadian rhythms play a central role in the regulation of many aspects of physiological processes, including sleep-wake cycles. Typical endogenous circadian rhythmicity in humans tends to cycle across a period of 24.2 hours with little day-to-day variation in length.[1] The timing of the circadian system, however, varies considerably across individuals. Chronotype, or diurnal preference, refers to behavioral patterns or manifestations indicative of underlying circadian-governed biological processes.[2] In the study of human sleep-wake regulation, chronotype represents preferred timing of sleep and wake (independent of environmental factors, such as work schedules) and corresponds to the timing of the circadian system.

The timing of circadian rhythmicity falls on a continuum, at either end of which are individuals colloquially referred to as *morning larks* and *night owls*. Morning larks are considered to have *early* chronotype or *advanced sleep phase* (also referred to as *morning* type/preference). Early chronotypes prefer to awaken early in the morning and feel most active during the earlier parts of the day. Some evidence suggests that short circadian periods (e.g., < 24 hours) may play a role in extreme advanced sleep phase.[3] Night owls, or *late* chronotypes with *delayed sleep phase* (also *evening* type/preference), prefer to awaken much later in the day and typically feel most active and motivated in the late evening or at night. In direct contrast to early types, some evidence suggests that individuals with extreme delayed sleep phase may have circadian periods longer than 24.2 hours.[4] Individuals who fall between the ends of the

Accepted Manuscript

continuum are referred to as *intermediate* types. Epidemiological studies characterize chronotype as near-normally distributed with robust developmental changes in diurnal preference across the lifespan, i.e., earlier preference during childhood development, later in adolescence and early adulthood, then progressively earlier as age advances.[2,5]

Genetics of the circadian clock

There has long been an interest in the genetic basis for circadian rhythmicity. Tremendous advances have been made in identifying a series of molecules that comprise a core circadian clock that exists in every cell, rather than being exclusive to the SCN. Each cell maintains its own circadian rhythmicity, and the master clock in the SCN seems to play the role of maintaining synchronization among these cellular clocks. Primarily based on work in model systems, it is now understood that the molecular circadian clock consists of a negative feedback loop involving the *Period (PER1, PER2,* and *PER3*) and *Cryptochrome* (*CRY1* and *CRY2*) genes.[6] Other genes involved in the molecular generation of circadian rhythms include: Casein Kinase *1δ* and *1ε* (*CK1*) and transcription factors *Circadian Locomotor Output Cycles Kaput Protein* (*CLOCK*), *Brain and Muscle ARNT-like Protein* (*BMAL1* and *BMAL2*), and Neuronal Pas Domain Protein (*NPAS1* and *NPAS2*); for greater in-depth discussion of circadian clock genes, see Lowrey and Takahashi,[7] McClung,[8] and Pack et al.[9]

Genetics of circadian timing

In parallel with the identification of the molecular mechanisms of circadian rhythms, studies sought to determine whether chronotype, as a behavioral manifestation of circadian rhythms, is also determined by genetic factors by exploring its heritability. Several twin and

Accepted Manuscript

family studies have estimated the heritability of chronotype to be approximately 50% based on estimates in the US,[10] UK,[11] Scandinavia,[12-14] and Brazil,[15] whereas studies of Hutterites[16] and Amazonians[17] suggest lower heritability at 14% and 23%, respectively. Taken together, heritability estimates suggest that genetic factors explain a considerable proportion, up to 50%, of the population variability in circadian timing. Further, the normal distribution of chronotypes indicates a polygenic basis such that multiple genes contribute modest, aggregative effects to this complex phenotype.[18] Given the heritability of circadian typology and prior success in identifying the components of the molecular circadian clock, chronotype is a logical target for genetic studies.

Several studies have used a candidate-gene approach and examined the associations between circadian genes and either chronotype or circadian rhythm sleep disorders. An association was identified and supported between the *3111C* allele of the *CLOCK* gene and eveningness,[19,20] though some studies failed to replicate this finding.[21-23] Associations have been found between extreme delayed sleep phase and length polymorphisms of the *PER3* gene[24-26] in addition to the *3111C* allele of the *CLOCK* gene.[19] Morningness has been found to be associated with polymorphisms in the clock genes *PER1* and *PER2* in other studies.[27,28] In addition, polymorphisms in the *PER3* and *ARNTL2* genes are associated with chronotype in a sample of 966 British adults.[29]

Perhaps the most studied human gene variant with respect to its role in diurnal preference is a *PER3* polymorphism with variable number of tandem repeats (VNTR) of a 54-bp motif in exon 18 that is repeated either 4 or 5 times (see Dijk and Archer[30] for review of *PER3*). This polymorphism is found only in primates, for which humans can be homozygotic (4/4 or 5/5) or heterozygotic (4/5) for the number of repeats.[31,32] Investigations of *PER3* and chronotype have

Accepted Manuscript

offered conflicting results such that the 5-repeat allele has been found to be positively associated with morningness,[26,33] but has also been shown to be elevated in individuals with extreme delayed sleep phase,[34] while other studies have shown no association between the *PER3* VNTR and diurnal preference[35-38] nor in melatonin or cortisol as objective markers of circadian rhythms.[39]

Studies of extreme chronotype expression can yield highly informative results. Research on Familial Advanced Sleep Phase Syndrome (FASPS) has identified specific genetic variants that segregate with the disorder. Linkage analysis, a statistical gene mapping approach used in affected families, has identified a rare missense mutation (i.e., replacement of a single nucleotide) in the clock component gene, *PER2,* which alters the circadian period and is associated with FASPS.[40] However, this finding was not replicated in two FASPS pedigrees in Japan,[41] which highlights the difficulty of replicating results of rare mutations and the limited generalizability of findings concerning extreme phenotype expressions.

Although results from these genetic studies are mixed, overall patterns suggest that multiple genes play important roles in influencing circadian typology. As chronotype is a complex trait, a large number of genes with modest influence likely contribute to phenotypic manifestation. The detection of multiple small genetic effects requires very large sample sizes, and genome-wide discovery approaches are necessary to identify novel genetic loci. Importantly, going beyond classically gene-centric approaches is essential to elucidate unexplained heritability that remains.

Genome-wide discovery

Accepted Manuscript

The first genome-wide analysis of circadian-related phenotypes analyzed data from 749 subjects in *The Sleep Heart Health Study*.[42] The authors employed linkage analysis, which is used to map genetic loci associated with a phenotype based on observations of related individuals. Linkage analysis for habitual bedtime revealed no significant genetic associations, though statistical power was likely limited by insufficient sample size for genome-wide analysis (as large samples are required to detect variants of even moderate effect, i.e., OR = 1.5-2.0)[43] and poor quality of genetic data. Notably, near-significant genetic associations for habitual bedtime were observed near *CLOCK*, *PROK2*, *CSNK2A2*, and *NSPR1*, all known circadian clock-related genes. However, three recent, large-scale genome-wide association studies (GWASs) of subjects of European ancestry from the 23andMe cohort[44] (23andMe, Inc. Mountain View, CA, USA) and the UK Biobank[45,46] have uncovered novel gene loci associated with diurnal chronotypic preference. Here, we will compare methodologies and significant findings of each GWAS (overlap and divergence), closing with a discussion of the findings in the context of prior knowledge in this area.

Unlike candidate gene approaches, GWAS does not require any *a priori* knowledge of genes underlying the expression of the phenotype of interest. Indeed, GWAS is a discovery approach wherein the entire genome is searched for small genetic variations that occur more commonly in individuals with a particular phenotypic trait. Specifically, a GWAS tests associations between single nucleotide polymorphisms (SNPs, i.e., the most common type of genetic variation among people, represented by a difference in a single DNA base pair) and a phenotype of interest (often a trait or disease). GWAS are typically restricted to common genetic variants (i.e., with minor alleles occurring in more than 5% of the population), which tend to have small or modest effects on complex phenotypes. Because SNPs are often correlated with

Accepted Manuscript

other nearby SNPs, a substantial proportion of the variability in the genome is captured by measuring several hundred thousand SNPs and imputing millions of non-genotyped SNPs.[47,48]

GWAS analyses involve conducting individual association models of the relationships between each SNP and the phenotype of interest (e.g., linear regression models for continuous phenotypes or logistic regression models for binary phenotypes). Due to the large number of tests resulting from all SNPs being examined individually, a multiple comparisons correction is necessary to determine statistical significance and protect against false positive associations. While there are different methods for performing this correction, a p-value less than 5×10^{-8} is typically required to claim genome-wide significance.[48-51]

The first large-scale GWAS of diurnal preference was published by Hu and colleagues[44] from a European-ancestry sample of worldwide customers of 23andMe, a direct-to-consumer genotyping service. Subjects ranged from early adulthood (<30y) to old age (>60y; exact age range not reported). Given the known effect of age on circadian preference,[2,5] this is an important issue. Subjects in this study responded to two survey items regarding natural tendency of being a night person/night owl vs. morning person/early bird. Specifically, subjects were asked twice: *Are you naturally a night person or a morning person?* with a unique set of response options for each administration: (1) *night owl*, *early bird*, or *neither*, then (2) *night person*, *morning person*, *neither*, *it depends*, or *I'm not sure*. Responses indicating intermediate type or ambiguity (viz. *neither*, *it depends*, and *I'm not sure*) were treated as missing data. Thus, those who indicated intermediate type on both items were not included in analyses, whereas those who indicated intermediate type on one item, but also early or late type on another were classified based on their early/late responses. Subjects with diametrically opposed responses, i.e., indicating early chronotype on one item and late chronotype on another, were excluded from analyses entirely,

Accepted Manuscript

resulting in 89,283 unrelated individuals (38,937 morning types, 50,346 evening types) included in the analysis. The primary chronotype phenotype in this study was a binary outcome value (early vs late chronotype) and analyses included as covariates age and sex. In addition, the authors adjusted for the top five principle components accounting for population stratification, which controls for differences in allele frequencies among populations of different ancestry. The authors identified 15 loci associated with self-rated diurnal preference (Table 1, see Figure 1 for Manhattan plot, which shows the genome-wide significant SNPs and nearby loci). Although Hu et al. conducted a follow-up GWAS with chronotype scored on a continuum including intermediate types, SNP association results from this analysis were only provided for two newly significant SNPs (designated by * in Table 1). Notably, because chronotype is a continuous trait, phenotyping chronotype on a continuum yields greater statistical power to detect genetic associations than operationally defining chronotype as a binary trait with arbitrary cutoffs.

Lane et al[45] and Jones et al[46] each conducted genome-wide association analyses for chronotype on data from white British, Caucasian subjects from the UK Biobank, a prospective study of >500,000 people in the United Kingdom. Individuals in the UK Biobank range in age from 37 to 73 years, though Lane et al[45] analyzed data only from subjects between the ages of 40 and 69 to increase sample homogeneity. In these studies, GWASs were conducted for chronotype defined as both a binary and a continuous trait, which was self-reported in response to the single item from the *Morningness-Eveningness Questionnaire* (MEQ) that explains the highest fraction of variance in preferences of sleep-wake timing:[52,53] *Do you consider yourself to be…* responses to which included self-description as *definitely a morning person, more a morning than evening person, more an evening person than a morning person, definitely an*

Accepted Manuscript

evening person, *do not know*, and *prefer not to answer* (subjects endorsing *prefer not to answer* were excluded from analyses in both studies).

From their GWAS analyses, Lane et al[45] excluded shift workers, subjects taking sleep medications, and those who responded *do not know* to the chronotype item, resulting in a sample size of n=100,420. The authors conducted two parallel GWASs: a continuous chronotype outcome (1-4, *definitely a morning person* to *definitely an evening person*) and a binary comparison of the two extreme phenotypes (26,948 definite morning types vs 8,724 definite evening types). Genetic association analysis was performed using a SNPTEST while controlling for age, sex, and 10 principal components of ancestry and genotyping array (two different, though highly similar, Affymetrix arrays were used for different subjects in the UK Biobank). The authors identified 12 significant loci, and 1 suggestive secondary signal, associated with chronotype (Table 1).

Jones and colleagues[46] also conducted parallel GWASs for continuous and binary operationalizations of chronotype. However, their continuous chronotype was coded differently such that the response of *do not know* served as a proxy for intermediate type, rather than excluded as missing data. Thus, their continuous variable included five groups (i.e., *definite morning*, *moderate morning*, *intermediate*, *moderate evening*, *definite evening types*) rather than four as in the Lane et al study. Chronotypes were then normalized for analysis through adjustment of the raw phenotype for age, gender, and study center, followed by inverse normalization of the resulting residuals. Unlike Lane et al who compared just extreme phenotypes in their binary GWAS, Jones et al collapsed across *definite* and *moderate* groups within morning and evening types, resulting in a sample size of 114,765 for the case-control GWAS (compared to n=35,672 in the case-control GWAS of extreme phenotypes in Lane et al).

Accepted Manuscript

In addition, exclusionary criteria differed from the prior two investigations such that Jones et al excluded diabetics who reported insulin use within the first year of diagnosis, diabetics diagnosed younger than the age of 35 years or with no known age of diagnosis, and diabetics diagnosed within a year of the study. These exclusions were due to the authors focus on the relationship between chronotype and diabetes. The authors used linear mixed models to conduct GWAS analyses to account for population structure and subject relatedness, while controlling for age, sex, study center, and genotyping array. Jones et al[46] identified 16 significant loci associated with chronotype in the UK Biobank (Table 1).

All three GWASs independently supported associations with chronotype for genes *PER2*, *RGS16*, *FBXL13*, and *AK5* based on nearby significant SNPs. Of these identified genes, *PER2* and *RGS16* are both known for their roles in circadian regulation. The degradation of *PER2*, the speed of which influences circadian period length, is governed by Casein kinase 1's phosphorylation of *DEC2* (*BHLHE41*).[54] A regulator of G-protein signaling 16, *RGS16* has been shown to regulate $G_{\alpha i/o}$-mediated cyclic adenosine monophosphate (cAMP) synthesis in the SCN. *RGS16* deficiency/ablation leads to delayed dorsomedial *PER1* expression resulting in circadian period lengthening in mice, highlighting the association between chronotype and *RGS16*.[55] The F-Box protein, *FBXL13*, has a suspected circadian influence,[56-58] whereas *AK5* had not previously been linked to chronotype or circadian processes.[59] Regarding *FBXL13*, the SCF^{FBXL3} complex regulates expression of *CRY* and *PER* through ubiquitination and degradation of CRY proteins via the SCF^{FBXl3} ubiquitin ligase complex, and mutations in *FBXL3* are associated with lengthened circadian periods in mice.[56-58] Adenylate kinase 5, i.e., *AK5*, has been shown to regulate the adenine nucleotides within cells,[59] but has not been tied to chronotype or the circadian pacemaker prior to these studies.

Accepted Manuscript

Five additional genes were found to be associated with chronotype in two of the three studies (Table 1). Hypocretin (orexin) Receptor Type 2 (*HCRTR2*) is a gene that regulates the expression of the G-protein coupled receptor hypocretin receptor type 2. Deficient hypocretin systems are thought to be associated with a reduction of wakefulness and circadian alertness signals in narcolepsy.[60] 5-Hydroxytryptamine (serotonin) Receptor 6 (*HTR6*) has not been previously linked to chronotype, but is known for its connection to sleep regulation in rats such that 5-HTR_6 agonists lead to reductions in sleep.[61,62] Trinucleotide Repeat Containing 6B (*TNRC6B*) is a protein-coding gene that belongs to the *GW182* family of proteins, which are necessary for micro-RNA gene silencing in animal cells. These proteins have been linked to circadian behavior in *Drosophila*,[63] and bound to circadian transcription factors in mouse liver,[64] although these GWASs offer the first evidence linking this gene to human circadian typology. Both Hu et al and Lane et al identified the association of *APH1A* and chronotype. *APH1A* encodes a subunit of the γ-secretase complex which cleaves the amyloid-β precursor protein.[65] Hu et al highlighted that amyloid-β levels are higher in wakefulness than sleep and are affected by manipulations of the hypocretin (orexin) system,[66,67] in addition to clearance by the glymphatic system of the brain during sleep.[68] The last gene to be implicated by two independent GWASs was *ERC2* (ELKS/RAB6-Interacting/CAST Family Member 2), a protein-coding gene belonging to the Rab3-interacting molecule (RIM)-binding protein family involved in the regulation of neurotransmitter release through involvement in the organization of the cytomatrix at the nerve terminals active zone. This gene has no previously known association with sleep or circadian rhythmicity.

Highlighting the genetic complexity of chronotype as well as the limitations of genome-wide association analysis, 24 SNPs across these three GWASs were genome-wide significant in

Accepted Manuscript

just one of the three studies (see Table 1 for full list). Unlike the replicated genes discussed above, these unreplicated genes' relationships with chronotype were largely unclear, e.g., *ALG10B*, *LINC01128*, and *UNC5D*. Even so, five genes with well-documented circadian and/or sleep regulatory roles—*PER3*, Vasoactive Intestinal Peptide (*VIP*),[69] *FBXL3*, Methionine Sulfoxide Reductase A (*MSRA*), and RAS Dexamethasone-Induced 1 (*RASD1*)—were significant in one of the three GWASs. Given the lack of replication across studies, these associations, particularly for those with no known circadian roles, should be interpreted with great caution.

Importantly, of the nine genes identified in multiple studies (4 found in all three studies, and 5 found in two studies), only three SNPs were shared: rs372229746 in *FBXL13*, rs2050122 in *HTR6*, and rs4821940 in *TNRC6B*. Effects of these shared SNPs on chronotype were in the same direction and exhibited similar magnitudes. Only the Lane and Jones studies shared SNPs in common, given the usage of the same genotype data from the UK Biobank. Both Lane et al and Jones et al attempted to replicate the findings from the 23andMe SNPs reported in the Hu study. In Lane et al.'s GWAS of extreme chronotype, 8 of the 15 significant SNPs identified in Hu et al were found to have genome-wide significance, and all 15 signals had the same direction of effect in both studies. In a follow-up meta-analysis of both samples, Lane et al identified three additional SNPs near *PER3*, *VIP*, and *TOX3*. In the Jones study, the authors conducted a meta-analysis on their UK Biobank sample and the 23andMe cohort from the Hu study. Findings supported 13 of Jones et al.'s initially identified 16 loci to be associated with chronotype, though two signals were in the opposite direction.

Accepted Manuscript

Discussion:

The results of these three GWAS reports highlight the potential of discovery-based approaches for the identification of novel candidate genes involved in chronotype, circadian rhythms, and circadian rhythm sleep disorders, while also pointing to the need for large sample sizes in order to find common genetic variants with small effects for complex phenotypes. In particular, GWAS have demonstrated promise in revealing unexpected genetic effects that have gone undetected in more traditional candidate-gene studies.[70,71] Despite some differences in phenotyping and analytic approach, there was a good deal of consistency across studies with nine genes identified in at least two of the three main studies reviewed. On the other hand, given that two studies both used data from the UK Biobank, one might expect the results of these analyses to be more similar. There are several important lessons that can be drawn from these studies.

Importance of rigorous and consistent phenotyping. Multiple methods exist to capture the timing of the circadian system, including assessments of core body temperature, dim light melatonin onset, self-reported diurnal preference, and actigraphy- or diary-based midpoint of sleep.[72] For large-sample research endeavors, non-invasive, cost-effective, and expeditious assessments are needed to capture the phenotypes of the chronotype. At its most rudimentary level, chronotype can be assessed with a single question such as those used by 23andMe and the UK Biobank.[1] While these questions have face validity, sole reliance on a single, subjective item to define the chronotype phenotype may lead to difficulties regarding individual differences in beliefs about what constitutes *morning* or *evening* types, and fails to capture much of the variability in chronotype within the population.[2] Measurement error from non-rigorous phenotyping of complex traits decreases the signal-to-noise ratio and attenuates that measure's

Accepted Manuscript

ability to demonstrate association with genetic variants, thus reducing power to detect significant genome-wide associations. Further, although circadian rhythms play important roles in its expression, chronotype is influenced by several other genetic and environmental factors, which can add further noise to the phenotype and reduce statistical power. Thus, non-rigorous phenotyping can yield unreliable significant findings that may be less likely to be replicated in subsequent studies.[73]

In the three large GWASs reviewed above, chronotype was determined by one or two self-report questions. Specifically, Lane[45] and Jones[46] used a single item of the *Morningness-Eveningness Questionnaire*, whereas Hu[44] measured chronotype using two identically worded items (*Are you naturally a night person or a morning person?*) with two similar sets of response options. Although the differences between these items are slight, they likely contributed to disparate findings between results from the 23andMe and UK Biobank samples, highlighting the impact that small variations in phenotype definition can have on genome-wide associations. Moreover, even when the measure for phenotyping was consistent across analyses (both within studies conducting parallel GWASs, and across the two separate studies analyzing UK Biobank data), differences in response coding (e.g., binary vs continuous, treatment of ambiguous answers) affect measurement error and yield disparate findings. These are the same issues that plague other epidemiological research and produce varying rates of disease prevalence and incidence. Future endeavors should prioritize more rigorous and consistent phenotyping strategies including the use of standardized self-report instruments, clinical interviews, or objective measures of sleep-wake periodicity (e.g. actigraphy), as GWASs will have greater statistical power and produce more reliable results when phenotypic information is grounded on a solid base of defining criteria. There is also a critical need to validate commonly used self-

Accepted Manuscript

reported items by comparing them to biological markers of circadian rhythms to determine which question(s) most closely relate to endogenous processes.

Analytic approaches yield differing GWAS results. Differences in the analysis plan, including decisions defining the sample and selections regarding covariates, statistical tests, and methods to address confounders may also produce disparate findings across GWASs. The first critical analytic decisions regard inclusion and exclusion criteria. Notably, all three GWASs analyzed data from subjects with European Ancestry to increase sample homogeneity, and accounted for residual population structure in their models. Where Hu et al[44] and Lane et al[45] excluded individuals based on genetic relatedness, Jones et al[46] statistically adjusted for relatedness in their analyses. In addition, Lane et al excluded shift workers and subjects taking sleep medications, as shift work and hypnotics may potentially influence self-reported chronotype. The Jones study, however, excluded individuals based on diabetic and autoimmune considerations (see above).

Where Hu et al and Lane et al used regression for their genome-wide analyses, Jones et al employed linear mixed modeling, which allowed for adjusting for between-subjects genetic relatedness. Further, the Lane study broke from convention by using a less robust "expected" genotype count (also known as the posterior mean[74] or allele dosage[75]) in SNPTEST, rather than more traditionally employed frequentist test statistics (e.g. chi-square statistics for case-control studies, the score test for quantitative trait analyses). Lane et al.'s choice of model reflects efforts to account for uncertainty in imputed SNP quality; in particular, difficult-to-impute SNPs, i.e., those with low information scores following imputation. While both models asymptotically approximate each other, frequentist test statistics (using likelihood estimation) tend to result in

Accepted Manuscript

larger effect sizes for difficult-to-impute SNPs.[74] Thus, frequentist test statistics may overestimate genetic effects, which may have influenced why the Lane study chose a different, more conservative, statistical model. Although all studies adjusted for age, sex, and population structure, differences of methods of statistical adjustment and in covariate selection existed across studies. It is unclear whether different estimation techniques (e.g., regression vs linear mixed modeling) and covariate selection contributed to the disparate results across studies. On the other hand, the use of multiple approaches can be beneficial by allowing for sensitivity analysis such that genetic associations that are significant across different strategies are more likely to be true findings.

The complementary nature of GWAS and molecular studies. Findings from the GWAS analyses confirmed the roles of known circadian genes for influencing chronotype, a behavioral phenotype. Even so, perhaps the most exciting findings from these GWASs center on the significant SNPs near genes with little to no prior known history of circadian influence, such as *AK5*, *APH1A*, and *ERC2* among many others. Indeed, though many classical clock genes were significantly associated with chronotype in these GWASs, a critical strength of these exploratory techniques is the ability to identify novel genetic etiology of these complex phenotypes. Owing to discovery of the novel loci and newly implicated genes, efforts should be directed toward replicating these findings and elucidating their roles in circadian regulation and chronotype. Comparison of phenotype-associated variants to expression quantitative trait loci (eQTL) variants can help prioritize a gene and mechanism for functional follow-up.[76,77] As genome-wide discovery attempts to establish the genetic architecture and functional genomic model of circadian rhythms and chronotype, studies using model systems can reveal the molecular

Accepted Manuscript

mechanisms modulating circadian processes. Molecular studies can be used to support associations identified through genome-wide analysis by examining these identified regions in finer detail and finding a potential mechanism by which they act on chronotype. Indeed, genetic model systems for studying sleep and sleep-wake disorders are needed to reveal the molecular basis of chronotype and circadian typology. Genes with no previously known circadian role should now be investigated in model systems to determine whether these genes do have a hitherto unknown role in circadian regulation.[78] Importantly, failure to replicate GWAS findings in molecular investigations may reveal erroneous associations from GWAS results.

Limitations and future directions

Common vs. rare variants: While genome-wide discovery has greatly advanced our knowledge of the genetic basis of sleep and circadian regulation, genetic variants obtained through genotyping chips used in GWAS are restricted to common SNPs with minor allele frequencies greater than 5%. Importantly, common variants identified in GWAS typically result in modest influences on the manifestation of complex phenotypes (Figure 2). Despite very large samples in the three GWASs we reviewed, differences in findings across studies are likely influenced by effect allele frequency (EAF) and the effect size that each allele contributes. The graphs in Figure 3 show that the more common an effect allele is in the population (i.e., higher EAF), the greater the phenotypic effect must be to detect a genetic association. This demonstrates that achieving sufficient statistical power to detect effects of common alleles requires a greater phenotypic effect at that allele. The complicating matter, however, is that the frequency of effect alleles and their magnitudes of effect are typically inversely related; that is, the more common an effect allele is, the more modest its effect is likely to be (Figure 2).[79,80]

Accepted Manuscript

Over the range of effect sizes (OR=1.04-2.34) noted in Hu, Lane, and Jones GWASs, it is not surprising that there was divergence between significant alleles, given the sample sizes and case/control proportions in each of the studies. Next-step research on common alleles contributing to chronotype may utilize polygenic risk scores, which aggregate the multiple small effects of common variants to effectively explain much of the heritability of complex/polygenic traits,[81] such as prior work on schizophrenia[82] and height.[83] Derivation of polygenic risk scores would allow for examination of whether genetic load for late chronotype is associated with phenotypic variations in mood, sleep, or other circadian-linked health outcomes.[84]

In contrast, rare gene variants (i.e., SNPs with minor allele frequencies less than 5%) often exert larger effects on phenotype expression, but are not typically identified using genome-wide discovery techniques due to their low frequency (see Pack et al.[9] for review). Given that chronotype is relatively normally distributed in the population, a large number of common variants with small effects contribute to its expression. Even so, rare variants, quite possibly in the same genes as known common variants, likely result in greater magnitude of influence on chronotype when present. The influence of rare variants is particularly germane to extreme expressions of chronotype and circadian pathology, such as in FASPS.[3,85] Given that rare variants may disproportionately influence a trait at the same loci as the common variants identified, fine-mapping of these regions may help elucidate the roles that genes in the region play in chronotype expression, beyond simple linkage disequilibrium (i.e., associations between nearby genetic loci) with plausible sleep-related genes.[86]

Demonstrating the significant impact of rare variants, He et al[87] studied two genetically related subjects with markedly advanced circadian phase and short sleep compared to other members in their family and the general population. Despite habitual sleep periods of

Accepted Manuscript

approximately 22:00 to 04:00, these individuals denied sleep-related daytime impairment. After sequencing all clock and clock-related genes, He et al identified a rare point mutation in *DEC2* as complicit in their unique and extreme sleep phenotypes. The authors then replicated the effect of this mutation in transgenic mice and drosophila. Pellegrino and colleagues extended this line of research by demonstrating that genes with rare variants with large effects may harbor other such mutations. Specifically, Pellegrino et al[88] sequenced *DEC2* in two human samples—one from a twin study[89] and another from a study on sleep deprivation[90]—and found two new mutations in *DEC2* associated with short sleep. Taken together, these findings highlight the importance of advancements in genetic sequencing and targeted efforts to identify and characterize rare variants associated with sleep and circadian phenotypes.

Further elucidation of both the common and rare variants constituting the genetic basis of chronotype will provide critical insights into the molecular underpinnings of the circadian system. This knowledge will enhance the understanding of the etiology of circadian rhythm and other illnesses with known circadian abnormalities, including major depression and cardiometabolic disorders. These discoveries may lead to potential novel pharmacologic and non-pharmacologic treatment options that target the circadian system. For instance, the creation of personalized genetic risk profiles for circadian abnormalities may be used to guide the design and assignment of shift work schedules to reduce risk for shift work disorder in the future.

Accepted Manuscript

Figure legends

Figure 1: Manhattan plot of GWAS of being a morning person from Hu et al. 2016.[44]

Presented in its originally published format, the grey line corresponds to threshold for genome-wide significance at $P = 5 \times 10^{-8}$, with significant results shown above this threshold in red. Gene labels are annotated as the nearby genes to the significant SNPs.

Figure 2: Relationship between allele frequency and effect on phenotype expression.

Common variants with minor allele frequencies >5% typically boast weak or modest effects on expression of complex phenotypes. The effects of common variants are often detected using GWAS. In contrast, rarer variants typically wield stronger effects and contribute to more extreme expressions of complex phenotypes. Rare variants, however, are not often detected using genome-wide analysis.

Figure 3: Comparison of statistical power in the three GWASs. Comparisons of the three different study designs (number of cases:number of controls), and their power to detect alleles with varying genetic relative risk, over the range of effect allele frequencies (EAF). The differences in the power over the range of genotype relative risks clearly indicate the impact of sample size (Jones et al. > Hu et al. > Lane et al.), which figures most prominently at higher allele frequencies. Power estimates were performed using the CaTS power calculator (http://csg.sph.umich.edu//abecasis/cats/)[91] for all studies, using a significance threshold of 5×10^{-8}, prevalence of morningness phenotype of 0.652 (estimated from total UK Biobank cohort), and EAF minimum (0.01) and minimum (0.998) values used to reflect the range across studies.

Accepted Manuscript

REFERENCES

1. Czeisler CA, Gooley J. Sleep and circadian rhythms in humans. Paper presented at: Cold Spring Harbor symposia on quantitative biology2007.
2. Roenneberg T, Kuehnle T, Juda M, et al. Epidemiology of the human circadian clock. *Sleep medicine reviews.* 2007;11(6):429-438.
3. Jones CR, Campbell SS, Zone SE, et al. Familial advanced sleep-phase syndrome: A short-period circadian rhythm variant in humans. *Nature medicine.* 1999;5(9):1062-1065.
4. Sack R, Auckley D, Auger R, Carskadon M, Wright K, Vitiello M. Zhdanova IV. Circadian rhythm sleep disorders: Part II, advanced sleep phase disorder, delayed sleep phase disorder, free-running disorder, and irregular sleep-wake rhythm. *Sleep.* 2007;30(11):1484-1501.
5. Walch OJ, Cochran A, Forger DB. A global quantification of "normal" sleep schedules using smartphone data. *Science Advances.* 2016;2(5):e1501705.
6. Lowrey PL, Takahashi JS. Genetics of circadian rhythms in Mammalian model organisms. *Advances in Genetics.* 2011;74:175-230.
7. Lowrey PL, Takahashi JS. Genetics of circadian rhythms in Mammalian model organisms. *Advances in genetics.* 2011;74:175.
8. McClung CA. Circadian genes, rhythms and the biology of mood disorders. *Pharmacology & therapeutics.* 2007;114(2):222-232.
9. Pack AI, Keenan BT, Byrne EM, Gehrman PR. Genetics and Genomic Basis of Sleep Disorders in Humans. In: Kryger MH, Roth T, Dement WC, eds. *Principles and Practice of Sleep Medicine 6th ed.* St. Louis, MO: Elsevier; 2016.
10. Hur Y-M, Bouchard Jr TJ, Lykken DT. Genetic and environmental influence on morningness–eveningness. *Personality and Individual differences.* 1998;25(5):917-925.
11. Barclay NL, Eley TC, Buysse DJ, Archer SN, Gregory AM. Diurnal preference and sleep quality: same genes? A study of young adult twins. *Chronobiology international.* 2010;27(2):278-296.
12. Vink JM, Groot AS, Kerkhof GA, Boomsma DI. Genetic analysis of morningness and eveningness. *Chronobiol. Int.* Sep 2001;18(5):809-822.
13. Koskenvuo M, Hublin C, Partinen M, HEIKKILÄ K, Kaprio J. Heritability of diurnal type: a nationwide study of 8753 adult twin pairs. *Journal of sleep research.* 2007;16(2):156-162.
14. Vink JM, Vink JM, Groot AS, Kerkhof GA, Boomsma DI. Genetic analysis of morningness and eveningness. *Chronobiology international.* 2001;18(5):809-822.
15. von Schantz M, Taporoski TP, Horimoto AR, et al. Distribution and heritability of diurnal preference (chronotype) in a rural Brazilian family-based cohort, the Baependi study. *Scientific reports.* 2015;5.
16. Klei L, Reitz P, Miller M, et al. Heritability of morningness-eveningness and self-report sleep measures in a family-based sample of 521 hutterites. *Chronobiol. Int.* 2005;22(6):1041-1054.
17. Aguiar GF, da Silva HP, Marques N. Patterns of daily allocation of sleep periods: a case study in an Amazonian riverine community. *Chronobiologia.* Jan-Mar 1991;18(1):9-19.
18. Landolt H-P, Dijk D-J, eds. *Genetic and genomic basis of sleep in healthy humans.* 6th ed. NYC: Elsevier; 2016. Kryger MH, Roth T, eds. Principles and Practice of Sleep Medicine.

Accepted Manuscript

19. Katzenberg D, Young T, Finn L, et al. A CLOCK polymorphism associated with human diurnal preference. *Sleep.* 1998;21(6):569-576.

20. Mishima K, Tozawa T, Satoh K, Saitoh H, Mishima Y. The 3111T/C polymorphism of hClock is associated with evening preference and delayed sleep timing in a Japanese population sample. *Am J Med Genet B Neuropsychiatr Genet.* Feb 5 2005;133(1):101-104.

21. Robilliard DL, Archer SN, Arendt J, et al. The 3111 Clock gene polymorphism is not associated with sleep and circadian rhythmicity in phenotypically characterized human subjects. *J. Sleep Res.* Dec 2002;11(4):305-312.

22. Iwase T, Kajimura N, Uchiyama M, et al. Mutation screening of the human Clock gene in circadian rhythm sleep disorders. *Psychiatry Res.* Mar 15 2002;109(2):121-128.

23. Pedrazzoli M, Louzada FM, Pereira DS, et al. Clock polymorphisms and circadian rhythms phenotypes in a sample of the Brazilian population. *Chronobiol. Int.* 2007;24(1):1-8.

24. Ebisawa T, Uchiyama M, Kajimura N, et al. Association of structural polymorphisms in the human period3 gene with delayed sleep phase syndrome. *EMBO Rep.* Apr 2001;2(4):342-346.

25. Pereira DS, Tufik S, Louzada FM, et al. Association of the length polymorphism in the human Per3 gene with the delayed sleep-phase syndrome: does latitude have an influence upon it? *Sleep.* 2005;28(1):29-32.

26. Archer SN, Robilliard DL, Skene DJ, et al. A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference. *Sleep.* 2003;26(4):413-415.

27. Carpen JD, Archer SN, Skene DJ, Smits M, von Schantz M. A single-nucleotide polymorphism in the 5'-untranslated region of the hPER2 gene is associated with diurnal preference. *Journal of sleep research.* 2005;14(3):293-297.

28. Carpen JD, von Schantz M, Smits M, Skene DJ, Archer SN. A silent polymorphism in the PER1 gene associates with extreme diurnal preference in humans. *J. Hum. Genet.* 2006;51(12):1122-1125.

29. Parsons MJ, Lester KJ, Barclay NL, et al. Polymorphisms in the circadian expressed genes PER3 and ARNTL2 are associated with diurnal preference and GNbeta3 with sleep measures. *J. Sleep Res.* Oct 2014;23(5):595-604.

30. Dijk DJ, Archer SN. PERIOD3, circadian phenotypes, and sleep homeostasis. *Sleep Medicine Reviews.* Jun 2010;14(3):151-160.

31. Jenkins A, Archer SN, von Schantz M. Expansion during primate radiation of a variable number tandem repeat in the coding region of the circadian clock gene period3. *J. Biol. Rhythms.* Oct 2005;20(5):470-472.

32. Sabino FC, Ribeiro AO, Tufik S, et al. Evolutionary history of the PER3 variable number of tandem repeats (VNTR): idiosyncratic aspect of primate molecular circadian clock. *PLoS ONE.* 2014;9(9):e107198.

33. Kunorozva L, Stephenson KJ, Rae DE, Roden LC. Chronotype and PERIOD3 variable number tandem repeat polymorphism in individual sports athletes. *Chronobiology international.* 2012;29(8):1004-1010.

34. Pereira DS, Tufik S, Louzada FM, et al. Association of the length polymorphism in the human Per3 gene with the delayed sleep-phase syndrome: does latitude have an influence upon it. *Sleep.* 2005;28(1):29-32.

Accepted Manuscript

35. Perea CS, Nino CL, Lopez-Leon S, et al. Study of a Functional Polymorphism in the PER3 Gene and Diurnal Preference in a Colombian Sample. *The open neurology journal.* 2014;8:7-10.

36. Osland TM, Bjorvatn BR, Steen VM, Pallesen S. Association study of a variable-number tandem repeat polymorphism in the clock gene PERIOD3 and chronotype in Norwegian university students. *Chronobiol. Int.* Nov 2011;28(9):764-770.

37. An H, Zhu Z, Zhou C, et al. Chronotype and a PERIOD3 variable number tandem repeat polymorphism in Han Chinese pilots. *International journal of clinical and experimental medicine.* 2014;7(10):3770.

38. Barclay NL, Eley TC, Mill J, et al. Sleep quality and diurnal preference in a sample of young adults: associations with 5HTTLPR, PER3, and CLOCK 3111. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics.* 2011;156(6):681-690.

39. Viola AU, Archer SN, James LM, et al. PER3 polymorphism predicts sleep structure and waking performance. *Current Biology.* 2007;17(7):613-618.

40. Toh KL, Jones CR, He Y, et al. An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science.* 2001;291(5506):1040-1043.

41. Xu Y, Padiath QS, Shapiro RE, et al. Functional consequences of a CKIdelta mutation causing familial advanced sleep phase syndrome. *Nature.* Mar 31 2005;434(7033):640-644.

42. Gottlieb DJ, T O'Connor G, Wilk JB. Genome-wide association of sleep and circadian phenotypes. *BMC medical genetics.* 2007;8(Suppl 1):S9.

43. Spencer CC, Su Z, Donnelly P, Marchini J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet.* 2009;5(5):e1000477.

44. Hu Y, Shmygelska A, Tran D, Eriksson N, Tung J, Hinds D. GWAS of 89,283 individuals identifies genetic variants associated with being a morning person. *Nature Communications.* 2016.

45. Lane JM, Vlasac I, Anderson SG, et al. Genome-wide association analysis identifies novel loci for chronotype in 100,420 individuals from the UK Biobank. *Nature communications.* 2016;7.

46. Jones SE, Tyrrell J, Wood AR, et al. Genome-wide association analyses in> 119,000 individuals identifies thirteen morningness and two sleep duration loci. *Plos Genetics.* 2016;12(8).

47. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science.* Jun 21 2002;296(5576):2225-2229.

48. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science.* Nov 7 2008;322(5903):881-888.

49. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* May 2008;9(5):356-369.

50. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet Epidemiol.* May 2008;32(4):381-385.

51. Chanock SJ, Manolio T, Boehnke M, et al. Replicating genotype-phenotype associations. *Nature.* Jun 7 2007;447(7145):655-660.

Accepted Manuscript

52. 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. *Journal of Biological Rhythms.* 2004;19(1):76-86.

53. Horne JA, Ostberg O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology.* 1976;4(2):97-110.

54. Xu Y, Toh K, Jones CR, Shin J-Y, Fu Y-H, Ptáček L. Modeling of a human circadian mutation yields insights into clock regulation by PER2. *Cell.* 2007;128(1):59-70.

55. Doi M, Ishida A, Miyake A, et al. Circadian regulation of intracellular G-protein signalling mediates intercellular synchrony and rhythmicity in the suprachiasmatic nucleus. *Nature communications.* 2011;2:327.

56. Godinho SI, Maywood ES, Shaw L, et al. The after-hours mutant reveals a role for Fbxl3 in determining mammalian circadian period. *Science.* 2007;316(5826):897-900.

57. Siepka SM, Yoo S-H, Park J, et al. Circadian mutant Overtime reveals F-box protein FBXL3 regulation of cryptochrome and period gene expression. *Cell.* 2007;129(5):1011-1023.

58. Busino L, Bassermann F, Maiolica A, et al. SCFFbxl3 controls the oscillation of the circadian clock by directing the degradation of cryptochrome proteins. *science.* 2007;316(5826):900-904.

59. Van Rompay AR, Johansson M, Karlsson A. Identification of a novel human adenylate kinase. *European Journal of Biochemistry.* 1999;261(2):509-517.

60. Taheri S, Zeitzer JM, Mignot E. The role of hypocretins (orexins) in sleep regulation and narcolepsy. *Annual review of neuroscience.* 2002;25(1):283-313.

61. Ly S, Pishdari B, Lok LL, Hajos M, Kocsis B. Activation of 5-HT6 Receptors Modulates Sleep–Wake Activity and Hippocampal Theta Oscillation. *ACS chemical neuroscience.* 2012;4(1):191-199.

62. Monti JM, Jantos H, Schechter LE. The effects of systemic and local microinjection into the central nervous system of the selective serotonin 5-HT6 receptor agonist WAY-208466 on sleep and wakefulness in the rat. *Behavioural brain research.* 2013;249:65-74.

63. Zhang Y, Emery P. GW182 controls Drosophila circadian behavior and PDF-receptor signaling. *Neuron.* 2013;78(1):152-165.

64. Koike N, Yoo S-H, Huang H-C, et al. Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science.* 2012;338(6105):349-354.

65. Francis R, McGrath G, Zhang J, et al. aph-1 and pen-2 are required for Notch pathway signaling, γ-secretase cleavage of βAPP, and presenilin protein accumulation. *Developmental cell.* 2002;3(1):85-97.

66. Kang J-E, Lim MM, Bateman RJ, et al. Amyloid-β dynamics are regulated by orexin and the sleep-wake cycle. *Science.* 2009;326(5955):1005-1007.

67. Roh JH, Huang Y, Bero AW, et al. Disruption of the sleep-wake cycle and diurnal fluctuation of β-amyloid in mice with Alzheimer's disease pathology. *Science translational medicine.* 2012;4(150):150ra122-150ra122.

68. Xie L, Kang H, Xu Q, et al. Sleep drives metabolite clearance from the adult brain. *science.* 2013;342(6156):373-377.

69. Aton SJ, Colwell CS, Harmar AJ, Waschek J, Herzog ED. Vasoactive intestinal polypeptide mediates circadian rhythmicity and synchrony in mammalian clock neurons. *Nature neuroscience.* 2005;8(4):476-483.

Accepted Manuscript

70. Klein RJ, Zeiss C, Chew EY, et al. Complement factor H polymorphism in age-related macular degeneration. *Science.* 2005;308(5720):385-389.

71. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nature genetics.* 2007;39(5):596-604.

72. Anna W. How to measure circadian rhythms in humans. *Medicographia.* 2007;29(1):84-90.

73. Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nature Reviews Genetics.* 2009;10(5):318-329.

74. Guan Y, Stephens M. Practical issues in imputation-based association mapping. *PLoS Genet.* 2008;4(12):e1000279.

75. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. *Annual review of genomics and human genetics.* 2009;10:387.

76. Emilsson V, Thorleifsson G, Zhang B, et al. Genetics of gene expression and its effect on disease. *Nature.* 2008;452(7186):423-428.

77. Nica AC, Montgomery SB, Dimas AS, et al. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* 2010;6(4):e1000895.

78. Sehgal A, Mignot E. Genetics of sleep and sleep disorders. *Cell.* 2011;146(2):194-207.

79. Hong EP, Park JW. Sample size and statistical power calculation in genetic association studies. *Genomics & informatics.* 2012;10(2):117-122.

80. Park J-H, Gail MH, Weinberg CR, et al. Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants. *Proceedings of the National Academy of Sciences.* 2011;108(44):18026-18031.

81. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* 2013;9(3):e1003348.

82. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009;460(7256):748-752.

83. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics.* 2010;42(7):565-569.

84. Fabbian F, Zucchi B, Giorgi AD, et al. Chronotype, gender and general health. *Chronobiology International.* 2016:1-20.

85. Xu Y, Padiath QS, Shapiro RE, et al. Functional consequences of a CKIδ mutation causing familial advanced sleep phase syndrome. *Nature.* 2005;434(7033):640-644.

86. McCarthy MI. Exploring the unknown: assumptions about allelic architecture and strategies for susceptibility variant discovery. *Genome medicine.* 2009;1(7):1.

87. He Y, Jones CR, Fujiki N, et al. The transcriptional repressor DEC2 regulates sleep length in mammals. *Science.* Aug 14 2009;325(5942):866-870.

88. Pellegrino R, Kavakli IH, Goel N, et al. A novel BHLHE41 variant is associated with short sleep and resistance to sleep deprivation in humans. *Sleep.* 2014;37(8):1327-1336.

89. Kuna ST, Maislin G, Pack FM, et al. Heritability of performance deficit accumulation during acute sleep deprivation in twins. *Sleep.* Sep 2012;35(9):1223-1233.

90. Goel N, Banks S, Mignot E, Dinges DF. DQB1*0602 predicts interindividual differences in physiologic sleep, sleepiness, and fatigue. *Neurology.* Oct 26 2010;75(17):1509-1519.

Accepted manuscript

91. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nature genetics.* 2006;38(2):209-213.

Accepted Manuscript

Table 1. Gene loci associated with chronotype/diurnal preference in 23andMe and UKBiobank genome wide association analyses.

SNP	EAF	Chr:Pos	Nearest Gene(s)	β	Cont P	OR	Bin P	Study	Known CRS function
Loci identified in all three GWASs									
rs12736689	0.03	1:182,549,729	RGS16, RNASEL	--	--	0.74	7.00E-18	Hu	K
rs1144566	0.97	1:182,569,626	RGS16	-.10	2.62E-14	0.74	1.29E-08	Lane	K
rs516134	0.03	1:182,553,693	RGS16	.08	9.00E-13	1.21	3.00E-12	Jones	K
rs55694368	0.93	2:239,317,692	PER2	--	--	0.86	2.60E-09	Hu	K
rs35333999	0.043	2:239,161,957	PER2	.06	8.43E-08	1.21	9.01E-06	Lane	K
rs75804782	0.88	2:239,316,043	PER2	.03	3.00E-07	1.09	4.00E-10	Jones	K
rs3972456	0.29	7:102,436,907	FBXL13, FAM185A		--	0.92	6.00E-09	Hu	P
rs372229746	0.45	7:102,158,815	FBXL13	.03	5.18E-10	1.12	4.29E-07	Lane	P
rs372229746	0.55	7:102,158,815	FBXL13, ORAI2, RASA4	.03	4.00E-09	1.06	7.00E-07	Jones	P
rs10493596	0.24	1:77,713,434	AK5		--	1.09	8.00E-12	Hu	U
rs76681500	0.84	1:77,726,241	AK5	-.04	1.50E-12	0.86	1.77E-09	Lane	U
rs11162296	0.16	1:77,700,196	AK5, PIGK	-.04	2.00E-12	0.93	1.00E-12	Jones	U
Loci identified in two GWASs									
rs35833281	0.21	6:55,021,561	HCRTR2	--	--	0.92	2.60E-09	Hu	K
rs76899638	0.22	6:55,147,508	HCRTR2	.03	4.00E-08	1.05	2.00E-07	Jones	K
rs2050122	0.8	1:19,989,205	HTR6	.03	4.61E-08	1.12	7.94E-07	Lane	K
rs2050122	0.2	1:19,989,205	HTR6	.03	2.00E-08	1.06	3.00E-06	Jones	K
rs4821940	0.55	22:40,659,573	TNRC6B	.03	1.05E-08	1.07	8.56E-05	Lane	P
rs4821940	0.45	22:40,659,573	TNRC6B, SGSM3	.02	3.00E-08	1.05	4.00E-08	Jones	P
rs34714364	0.17	1:150,234,657	APH1A, CA14	--	--	1.12	2.00E-10	Hu	P
rs10157197	0.4	1:150,250,636	APH1A	.03	1.48E-09	1.13	1.27E-11	Lane	P
rs12635074	0.65	3:55,982,416	ERC2	-.03	3.08E-08	0.92	1.20E-06	Lane	U
rs11708779	0.32	3:55,934,939	ERC2	-.02	3.00E-08	0.96	2.00E-06	Jones	U
Loci identified in one GWAS									
rs1015197	0.6	1:150,250,636	PRPF3, TARS2	.03	1.00E-09	1.05	5.00E-07	Jones	U
rs11121022	0.42	1:7,836,659	PER3, VAMP3	--	--	1.07	2.00E-08	Hu	K
rs72720396	0.23	1:91,191,582	CALB1	-.03	1.00E-07	0.95	3.00E-08	Jones	P
rs12140153	0.9	1:62,579,891	INADL	.04	7.00E-09	1.07	4.00E-06	Jones	P
rs141175086	0.998	1:780,397	LINC01128	-.27	1.42E-04	0.46	4.38E-08	Lane	U
rs1075265	0.52	2:54,354,927	PSME4, ACYP2	-.03	2.00E-10	0.95	4.00E-08	Jones	U
rs70944707	0.23	2:24,257,444	FKBP1B	.03	3.00E-08	1.05	2.00E-05	Jones	U
rs11895698	0.14	2:239,338,495	ASB1	.04	1.15E-08	1.1	1.30E-04	Lane	U
rs1595824	0.49	2:198,874,006	PLCL1	--	--	1.08	1.20E-10	Hu	P
rs148750727	0.995	4:188,022,952	FAT1	.15	3.61E-06	2.34	1.58E-08	Lane	U
rs9479402	0.01	6:153,135,339	VIP	--	--	0.69	3.90E-11	Hu	K
*rs9357620**	*0.71*	*6:13,170,634*	*PHACTR1*	*.02*	*1.70E-08*		--	*Hu*	U
rs2948276	0.88	7:96,457,119	DLX5, SHFM1	--	--	0.92	1.10E-08	Hu	U
*rs2975734**	*0.43*	*8:10,030,097*	*MSRA*	*.02*	*2.10E-09*		--	*Hu*	K
rs192534763	0.99	8:36,202,946	UNC5D	0.1	3.00E-07	1.25	2.00E-08	Jones	U
rs17311976	0.19	8:131,637,337	ADCY8	.19	1.08E-06	1.13	3.37E-08	Lane	P
rs77641763	0.88	9:140,265,782	EXD3, GRIN1, NRARP	.04	5.00E-11	1.07	7.00E-09	Jones	U
rs6582618	0.52	12:38,726,137	ALG10B	--	--	1.07	1.50E-08	Hu	U
rs542675489	0.6	12:120,994,888	RNF10	.03	3.29E-09	1.09	1.38E-05	Lane	U

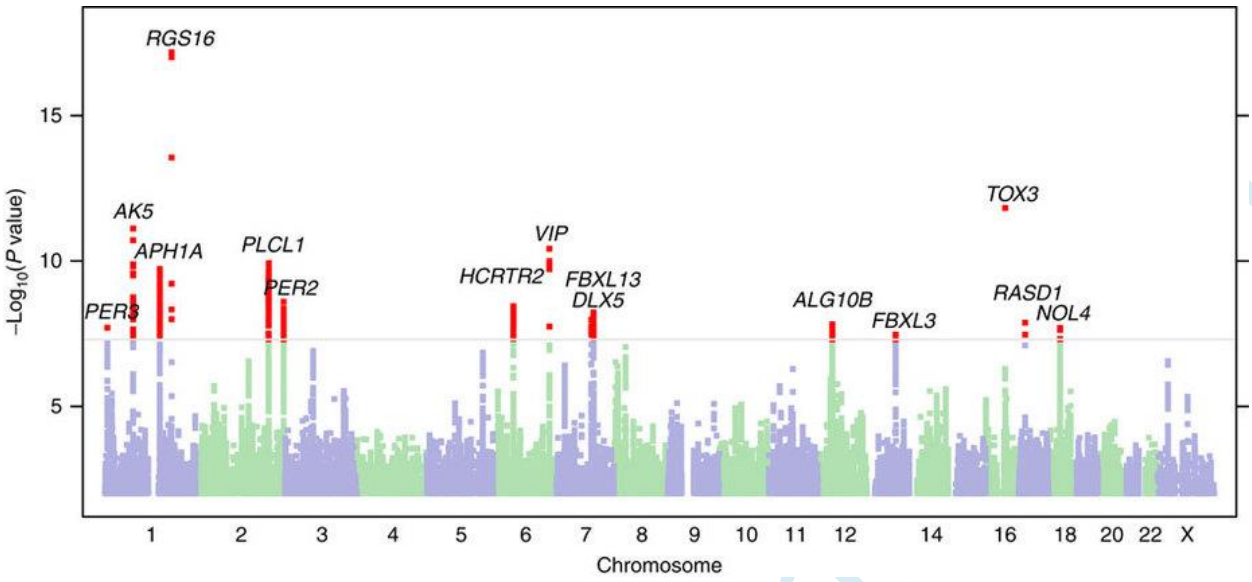
Note: SNP = single nucleotide polymorphism. EAF = effect allele frequency. Chr = Chromosome. Pos = position.

Accepted

Cont P = significance for genetic associations in GWAS on chronotype as a continuous phenotype. β = beta coefficient, i.e., effect size for genetic associations in GWAS on chronotype as a continuous phenotype. Bin P = significance for genetic associations in GWAS on chronotype as a binary (cases versus controls) phenotype. OR = odds ratio, i.e., effect size for genetic associations in GWAS on chronotype as a binary phenotype. Known CRS function = whether gene has a previously known association with circadian rhythms or sleep. K = Known. P = Probable. U = Unknown

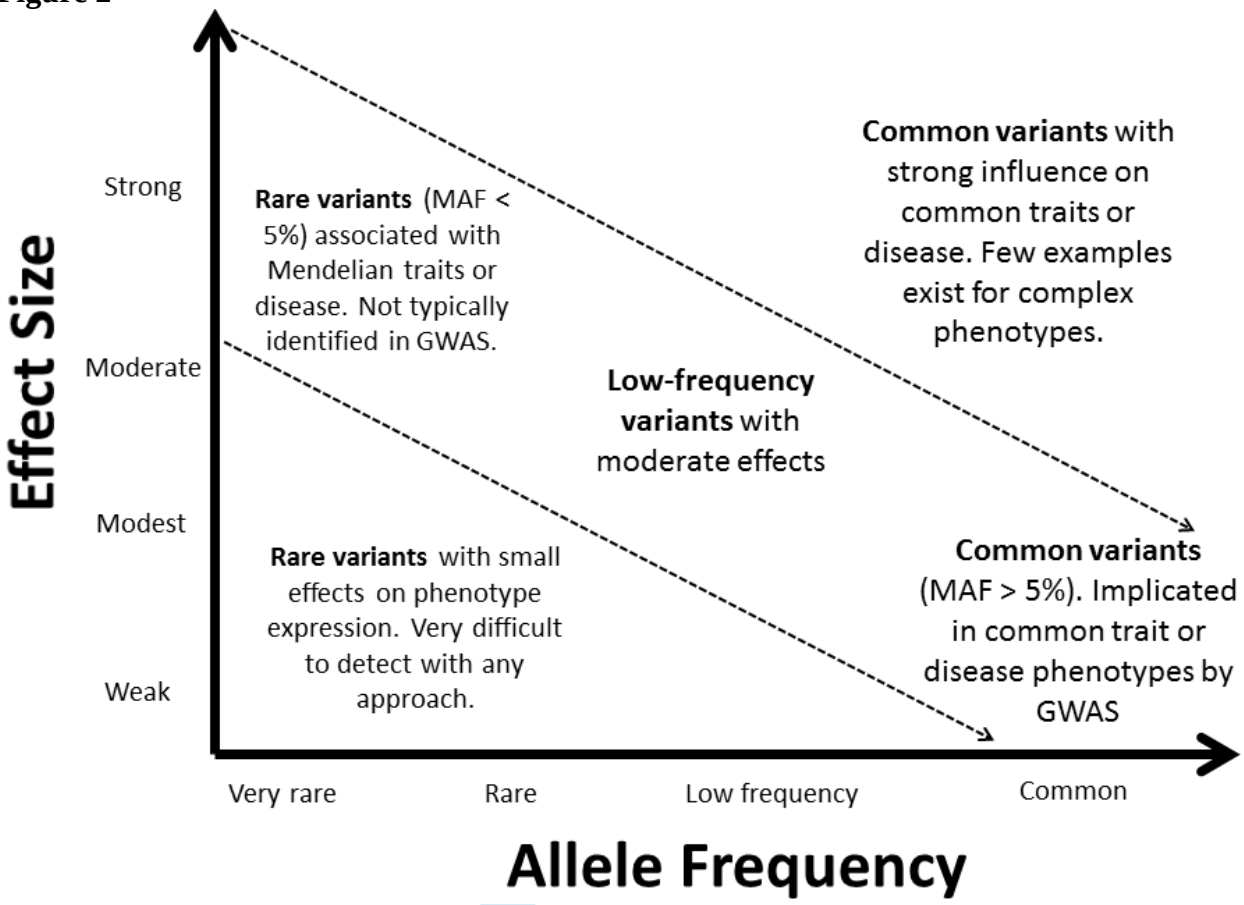
Accepted Manuscript

Figure 1

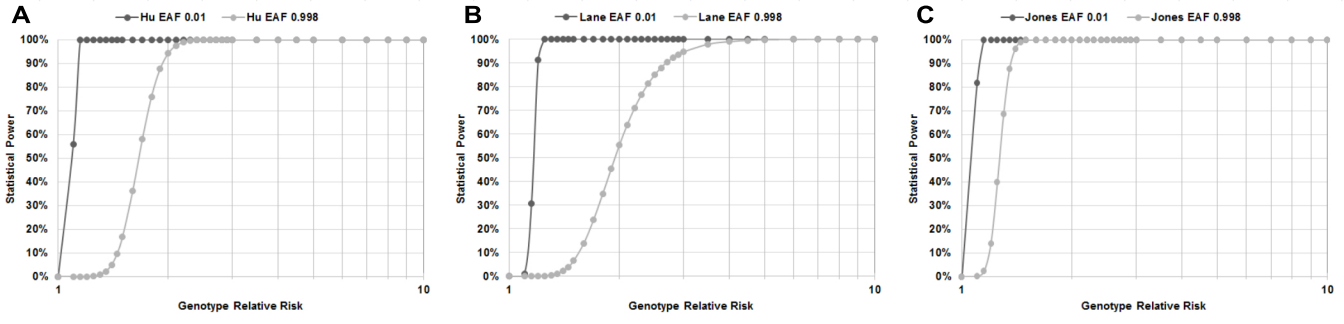


Accepted Manuscript

Figure 2



Accepted


A
Hu EAF 0.01
Hu EAF 0.998
Statistical Power
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
10
Genotype Relative Risk
B
Lane EAF 0.01
Lane EAF 0.998
Statistical Power
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
10
Genotype Relative Risk
C
Jones EAF 0.01
Jones EAF 0.998
Statistical Power
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
10
Genotype Relative Risk