



OPEN

Machine learning and expression analyses reveal circadian clock features predictive of anxiety

Aziz Zafar^{1,2}, Rebecca Overton^{1,2}, Ziad Attia¹, Ahmet Ay¹✉ & Krista Ingram¹✉

Mood disorders, including generalized anxiety disorder, are associated with disruptions in circadian rhythms and are linked to polymorphisms in circadian clock genes. Molecular mechanisms underlying these connections may be direct—via transcriptional activity of clock genes on downstream mood pathways in the brain, or indirect—via clock gene influences on the phase and amplitude of circadian rhythms which, in turn, modulate physiological processes influencing mood. Employing machine learning combined with statistical approaches, we explored clock genotype combinations that predict risk for anxiety symptoms in a deeply phenotyped population. We identified multiple novel circadian genotypes predictive of anxiety, with the *PER3*(rs17031614)-*AG/CRY1*(rs2287161)-*CG* genotype being the strongest predictor of anxiety risk, particularly in males. Molecular chronotyping, using clock gene expression oscillations, revealed that advanced circadian phase and robust circadian amplitudes are associated with high levels of anxiety symptoms. Further analyses revealed that individuals with advanced phases and pronounced circadian misalignment were at higher risk for severe anxiety symptoms. Our results support both direct and indirect influences of clock gene variants on mood: while sex-specific clock genotype combinations predictive of anxiety symptoms suggest direct effects on mood pathways, the mediation of *PER3* effects on anxiety via diurnal preference measures and the association of circadian phase with anxiety symptoms provide evidence for indirect effects of the molecular clockwork on mood. Unraveling the complex molecular mechanisms underlying the links between circadian physiology and mood is essential to identifying the core clock genes to target in future functional studies, thereby advancing the development of non-invasive treatments for anxiety-related disorders.

Mood disorders, including depression and anxiety, are becoming more prevalent globally, affecting nearly one-fifth of the adult population¹. These disorders negatively impact productivity, social relationships, and overall quality of life in individuals. The search for genetic and environmental factors contributing to the epidemic of mental health has uncovered numerous links between circadian rhythm disruptions and mood disorders including major depressive disorder (MDD), schizophrenia, bipolar disorder (BD), and generalized anxiety disorder^{2–7}. Clinical studies have also demonstrated that circadian rhythms and circadian clock genes can modulate mood and psychiatric disorders but few of these studies have explicitly focused on anxiety⁸.

Circadian rhythms regulate a sleep–wake cycle that is reset every 24 h based on exposure to natural or artificial light–dark cycles. The molecular clock driving these rhythms is created by feedback loops in core clock genes and their associated transcription factors that control physiological cycles in the body via the regulation of over a third of all transcribed genes^{9,10}. The core feedback loop includes the transcription factor CLOCK, which regulates the transcription of genes in the Period (*per1*, *per2*, and *per3*) and Cryptochrome (*cry1* and *cry2*) gene families. The CLOCK protein forms a heterodimer with BMAL1, which activates additional core clock genes. PER/CRY heterodimers, in turn, inhibit the activity of the BMAL1–CLOCK complex. This cycle of transcription activation and repression is vital for the 24-h circadian cycle. Mutations in these core clock genes may affect mood via direct transcriptional activity of downstream physiological pathways that influence mood. Alternatively, clock gene mutations may modulate mood pathways indirectly—through disruptions in the phase and amplitude of circadian rhythms¹¹. Individual core clock genes may, in fact, be involved in both direct and indirect mechanisms modulating the relationship between mood and circadian rhythm.

Evidence supporting indirect impacts of the core circadian oscillator on mood pathways is derived from studies examining the association of diurnal preference, differences in timing of activity levels, or chronotype,

¹Department of Biology, Colgate University, 13 Oak Drive, Hamilton, NY 13346, USA. ²These authors contributed equally: Aziz Zafar and Rebecca Overton. ✉email: aay@colgate.edu; k Ingram@colgate.edu

differences in sleep–wake timing, with mood^{7,12–16}. Individuals that are morning types tend to have earlier circadian phases and sleep–wake rhythms than the intermediate and evening types. These behavioral patterns parallel physiological changes, such as changes in body temperature and melatonin profiles¹⁷ and clock gene expression oscillations (molecular chronotypes) associated with diurnal preference^{5,13,15,18–21}. Studies on diurnal preference show that evening types are more likely to experience depression or anxiety at one point in their lifetime^{17,22–29}. Mood disorders have also been linked to diurnal preference and circadian misalignment—a mismatch between an individual’s physiological circadian rhythms and their behavioral cycles^{10,16}. Previous genome-wide association studies (GWAS) and candidate gene studies have also found associations between multiple circadian genes and diurnal preference^{7,14} as well as depression or other psychological disorders¹³. These studies provide strong evidence that diurnal variation in circadian rhythms plays a role in modulating the physiology of mood disorders. However, attempts to determine the molecular mechanisms underlying these indirect circadian influences on mood and anxiety, in particular, are in a nascent stage.

Core circadian genes also function as transcriptional regulators and can influence neurotransmitter signaling in well-known mood pathways, including serotonin, dopamine, and glucocorticoid pathways³⁰. Recent studies demonstrate that components of the circadian clock can directly modulate mood disorders. In mice, knock down of the *Neuronal PAS Domain Protein 2* (*NPAS2*), *period* (*PER*), or *cryptochrome* (*CRY*) genes results in altered anxiety levels^{9,31,32}. In humans, genome- and phenotype-wide association studies (GWAS and PheWAS) have not yielded strong evidence of associations between core clock genes and mood until Ho et al.³³ identified a clock-related gene associated with seasonal affective disorder. However, population-level candidate gene studies have identified multiple links between clock genes and depressive disorders, including anxiety^{3,7,32}. Mathematical models have also predicted links between circadian clock disruptions and anxiety¹⁴. Most interestingly, a recent clinical study demonstrated that melatonin treatment used to correct circadian misalignment in anxious patients helped to mitigate anxiety³⁴. The lack of clarity in large GWAS and PheWAS studies with complex phenotypes like mood disorders suggests that directed candidate gene studies using deep phenotyping are needed to identify the influence of specific clock genes and/or synergistic clock gene interactions on mood pathways.

In the current study, we performed a deep phenotypic analysis of anxiety (State-Trait Anxiety Inventory (STAI))³⁵ diurnal preference (Morningness–Eveningness Questionnaire (MEQ))³⁶, and molecular chronotype (via gene expression analyses based on clock gene expression phase and amplitude). We explore the synergistic effects of multiple genotypes on phenotypes using an array of machine learning algorithms, including feature selection and association rule learning, as well as statistical approaches. Unlike PheWAS studies, machine learning techniques are not constrained by and can be robust to the smaller sample sizes typical of deeply phenotyped datasets. In addition, feature selection can be used to incorporate both clinical and genotypic features to reduce the data’s dimensionality and identify the most predictive disease risk factors. Here, we employ deep-phenotyping and machine learning to identify direct and indirect mechanisms of circadian influences on anxiety.

Methods

Experimental data collection. Study participants were recruited from Colgate University and the surrounding community in Hamilton, NY, USA (n = 982; males = 318, females = 664, ages 17–79; median = 19). Participants were predominantly Caucasians of European descent. All participants gave written informed consent, and all procedures followed the principles of the Declaration of Helsinki. The Institutional Review Board at Colgate University authorized all consent forms and procedures (#FR-F13-07, #ER-F14-12, #F15-13, and #ER-F16-19).

Self-report surveys. Participants completed computer-based surveys, which included the trait version of the Spielberger’s State-Trait Anxiety Scale (STAI)³⁷, Beck Depression Inventory (BDI-II)³⁸, and the short form of the Patient-Reported Outcomes Measurement Information System (PROMIS™)³⁹ Sleep Disturbance. The STAI was used to indicate the anxiety scores of individuals ranging from 20 to 80, with scores of 20–37 indicating “no or low anxiety”, 38–44 indicating “moderate anxiety”, and 45–80 indicating “high anxiety.” The Horne–Östberg Morningness-Eveningness Questionnaire (MEQ)³⁶ survey was administered to measure diurnal preference.

Genotyping. DNA was extracted from 10 to 20 hair follicles from each participant. The hair samples were digested with Proteinase K at 56 °C for 24 h, and purified using the Qiagen DNeasy Micro Kit. Genotyping for single nucleotide polymorphisms (SNPs) was performed using a TaqMan SNP Genotyping assay (Applied Biosystems, Foster City, CA) on an ABI 3700HT real-time qPCR instrument. Participants were identified as homozygous or heterozygous for the major and minor alleles (Suppl. Table 1).

A fragment length analysis of the *PER3* VNTR length polymorphism repeat region was conducted using PCR fluorescent primers on GeneScan software with an ABI 3100 sequencer. The forward primer fluorescently labeled with 6-FAM was used with the following PCR primers: forward, 5'-CAAAATTTTA TGACACTACCAGAAT GGCTGAC-3', and reverse, 5'-AACC TTGTACTTCCACATCAGTGCCTGG-3'⁴⁰. The PCR was performed in a 25- μ l volume using Qiagen PCR Mastermix. The PCR cycling conditions were 3 min at 94 °C, followed by 35 cycles of 45 s at 94 °C, 45 s at 58 °C, and 45 s at 72 °C, with a final step at 72 °C for 3 min. Capillary electrophoresis was then used to separate *PER3* alleles on an ABI 3700 sequencer and sized using ABI ROX standards. The genotype of each participant was identified as *PER3* 4/4, *PER3* 4/5, or *PER3* 5/5.

Circadian gene expression analysis (molecular chronotyping). Ten to twenty hair follicles were collected in RNAlater solution at four different time points during the day: 8 a.m., 4 p.m., 5 p.m., and 8 p.m.¹⁶. All hair samples were stored at –80 °C prior to analysis. RNA was extracted and purified from hair follicles using the RNeasy Micro purification kit according to the protocol provided by Qiagen. The purified RNA was converted

to cDNA using rt-PCR (TaqMan Gold rt-PCR, ABI). Nanodrop was used to quantify the cDNA. Expression levels of clock genes *PER3* and *NR1D2* were measured using quantitative PCR on an ABI 7900HT instrument (Applied Biosystems). *GUSB* and *18S* were used as control genes, and each analysis was performed in a replicate of three. Relative mRNA levels were determined using the standard curve method as described in the ABI User Bulletin #2, and then converted into z-scores per individual. A standard curve was created based on the average data points from all the subjects⁴¹. The trained curve was fitted to the four data points of each subject using the parameter estimation method, Stochastic Ranking Evolutionary Strategy (SRES).² For phase shift estimation, the training curve was obtained from known intermediate types ($n = 20$; individuals not included in this study). The phase difference between the curve obtained from each subject's four RNA data points and the training curve gave the phase shift. Amplitudes and phases were then converted into z-scores for statistical analysis.

Feature generation and selection. *Genotypic and clinical features.* We used seven genotypic features: *CLOCK3111* (rs1801260), *CRY1* (rs228716), *CRY2* (rs10838524), *PER2* (rs10838524), *PER3A* (rs228697), *PER3B* (rs17031614), and *PER3 VNTR* (rs57875989), and four behavioral/clinical features: diurnal preference scores, age (≤ 22 or > 22), gender, and socioeconomic status (poor, lower-middle-class, upper-middle-class, affluent). For genotypic features, individuals can be homozygous dominant, homozygous recessive, or heterozygous. To reduce multicollinearity, we performed one-hot encoding for non-binary features, removing the most frequent variant (as the baseline condition). We created 2-way combinations using the seven genotypic features and their respective variants, giving us $7C_2 * 9$ more features. We removed the most frequent class for each group of 9 combinations and treated it as a reference category. After pre-processing, the data set for analysis contained 174 total features.

Feature selection. We used four feature selection methods (each method ten times with ten-fold cross-validation) to find epistatic combinations predictive of mood disorders. InfoGain (IG) and ReliefF (ReF) are ranking-based feature selection methods that rank features based on their correlation with the class^{42,43}. Minimum Redundancy Maximum Relevance (MRMR) and Joint Mutual Information (JMI) are subset-based feature selection methods that use information theory-based criteria to find possible subsets from the feature space^{44,45}. A feature was considered robust if it appeared in 95% of the runs for a certain feature selection method and in at least three out of the four feature selection methods. This subset of the features was used for in-depth statistical analysis.

Classifiers. We modeled the relationship between risk factors and anxiety using three classifiers: tree-based methods like Random Forests (RF) and XGBoost (XGB) and a linear method, Support Vector Machines (SVM)^{46–48}. We evaluated the performance of our classifiers using accuracy scores and the area under the receiver operating characteristic (AUROC) curves. Classifiers employ a variety of hyperparameters that must be tailored for each dataset. As a result, we used preliminary testing to determine an appropriate range of hyperparameters for each classifier, followed by grid searching to determine the optimal combination of hyperparameters for maximizing accuracy.

Cross-validation. We employed stratified tenfold cross-validation to determine each model's generalizability. We divided the data set into tenfolds(subsets) for each combination of feature selection method and classifier, maintaining a consistent distribution of our outcome class for each fold. We performed the k-nearest neighbors' imputation to fill in missing values for each fold⁴⁹. We repeated the cross-validation procedure ten times to ensure robust results, each time using a different random number generator seed.

SMOTE. When a dataset is unbalanced, the feature selection and classification models frequently overestimate the likelihood of the majority outcome. As a result, the model may be inaccurate. We employed Synthetic Minority Oversampling (SMOTE) technique to increase model accuracy by balancing our unbalanced dataset. SMOTE accomplishes this by identifying the k-nearest neighbors (we chose $k = 5$ based on empirical evidence) and randomly generating new data along the line connecting two neighbors of the same class⁵⁰. To ensure our dataset was balanced, we used SMOTE to oversample the number of cases for the less frequent outcome. We performed our analysis with and without SMOTE to determine whether it improved them and then reported the balanced dataset results.

Statistical analyses. *Logistic regression.* All statistical analyses were performed using R⁵¹. We performed logistic regression analysis on each feature individually, keeping one-hot encoded features grouped together for the regression. After this univariate analysis, we performed multivariate logistic regression on all the features. Due to the high dimensionality of the dataset, we observed overfitting of the initial model. As a result, we performed multivariate logistic regressions using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), employing a sequential replacement method to identify subsets of features with low multicollinearity and strong association with the target variable⁵². We conducted these analyses using the *RcmdrMisc* library in R (*RcmdrMisc*). Along with AIC and BIC, we used the results of machine learning based feature selection algorithms to identify robust features for subsequent multivariate analysis as described above. Due to unplanned pairwise comparisons between features, p-values from the regression analysis were adjusted using the Benjamini–Hochberg method⁵³.

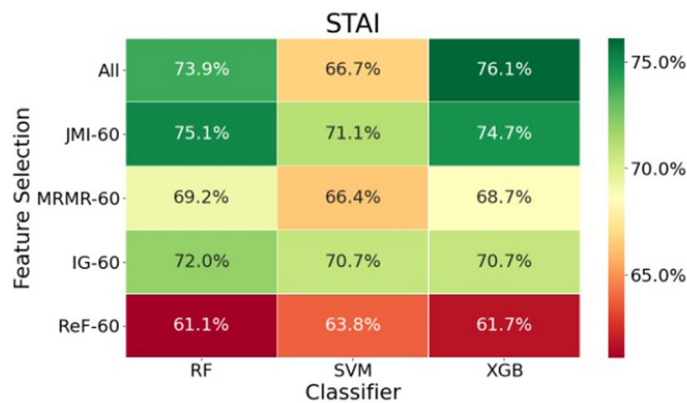


Figure 1. Heat map of prediction accuracy for feature selection and classifier methods. Our analyses yielded up to 26% higher prediction accuracy than baseline (50%) on a balanced data set.

Mediation analyses. We used mediation analysis to determine whether MEQ scores were statistically significant mediators between genotypic and clinical factors and mood disorders. The mediation's significance was found using the R package *mediate*, which employs a nonparametric bootstrapping method to compute a confidence interval for the mediatory effects⁵⁴.

Fisher's exact tests. Fisher's exact tests were used to identify features with a strong association with human anxiety individually. We created heatmaps illustrating patterns of association between molecular chronotype and human anxiety using different phase, amplitude, and STAI cutoffs.

Analysis of variance. To identify sex-specific differences in the average STAI scores for different two-way gene combinations, we used the *car* library in R to conduct a Type-3 Sum of Squares two-way ANOVA⁵⁵. Tukey's follow-up tests on significant factors were performed using the *emmeans* library. The normality of data was assessed by visual inspection and Shapiro–Wilk's test in R.

Association rule learning analyses. We performed association analysis using the *arules* package⁵⁶. The probability that an association occurs in the dataset is called its support. The lift of a rule is defined as the ratio of the observed support to that expected if the left-hand side and right-hand side of the relationship were independent. Since we had eleven variables, it was computationally intractable to find rules of length up to eleven with suitable support and lift. We limited ourselves to rules of size at most six, with at least 90% confidence. For each sex, we found rules that code for both categories of the target variable and then sorted them by their respective lift values. We visualized these relationships in sex-specific network plots, using the *igraph* library⁵⁷.

Gene networks using mutual information. We employed the Algorithm for the Reconstruction of Gene Regulatory Networks (ARACNE) to find the direct and indirect interactions between genes, clinical features, MEQ, and mood disorders⁵⁸. ARACNE constructs a network of relationships between nodes using a distance metric such as mutual information and correlation; for each triplet of edges, drops the edge with the lowest value. We used the *minet* package in R to create the network and then plotted it using *Rgraphviz*⁵⁹. We used bootstrapping to determine the frequency (i.e., confidence level) with which each link in the network appears.

Results

Synergistic, two-way genotype combinations are predictive of human anxiety. RF, SVM, and XGB classifiers predicted anxiety symptoms with an accuracy of 61–76% using all or a subset of genotypic and clinical factors chosen by feature selection methods (Fig. 1). The XGB method achieves the highest accuracy (76%) when all features are used. However, if sixty features selected using the JMI method are used, a similar accuracy level (75%) can also be obtained using XGB and RF classifiers. These accuracy levels are 25%–26% more accurate than random chance in our balanced dataset, which has a baseline accuracy of 50%.

Multivariate logistic regression analysis of the top features revealed that two-genotype combinations predicted a more robust risk of anxiety symptoms relative to single gene variants (Table 1). In the overall dataset, the combination of *PER3B*-AG and *CRY1*-CG was most strongly associated with the risk of having anxiety symptoms (Fig. 2; OR 15.3, $p = 0.026$). Average anxiety scores for individuals with *PER3B*-AG and *CRY1*-CG genotypes were higher (males: 54.8 ± 4.1 ; females: 46.7 ± 2.5) than for individuals of other genotypes (males: 41.3 ± 0.8 , females: 44.8 ± 0.6 ; Fig. 2A,B).

The combination of *CLOCK3111*-TC and *CRY2*-AG also significantly increased the odds of anxiety symptoms (Fig. 2; OR 2.5 (1.3–5.8), $p = 0.026$). Average anxiety scores for individuals with *CLOCK3111*-TC and *CRY2*-AG were higher (males: 45.5 ± 1.7 ; females: 45.4 ± 1.5) than for individuals of other genotypes (males: 40.8 ± 0.9 , females: 44.7 ± 0.6 ; Fig. 2C,D).

Variable	Odds ratio	95% CI	Adj P
PER3B_AG/CRY1_CG	15.262	2.352–319.019	0.026
PER3B_AG/CRY2_AA	0.17	0.034–0.694	0.026
AGE	3.851	1.234–14.006	0.031
CLOCK3111_TC/CRY2_AG	2.524	1.226–5.761	0.026
CRY1_CC/PER3_VNTR_4,4	2.432	1.029–6.781	0.061
GENDER	1.826	1.121–2.972	0.026
MEQ	0.958	0.935–0.982	0.005

Table 1. Risk factors for anxiety symptoms. Results from a multivariate logistic regression model based on top-ranked selected features identify two-way gene combinations, most notably *PER3B*-AG/*CRY1*-CG, as well as demographic features that are strongly associated with anxiety symptoms. Age is coded as 1 for 18–22 years old and 0 for >22 years old. Gender is coded as 1 for females, and 0 for males.

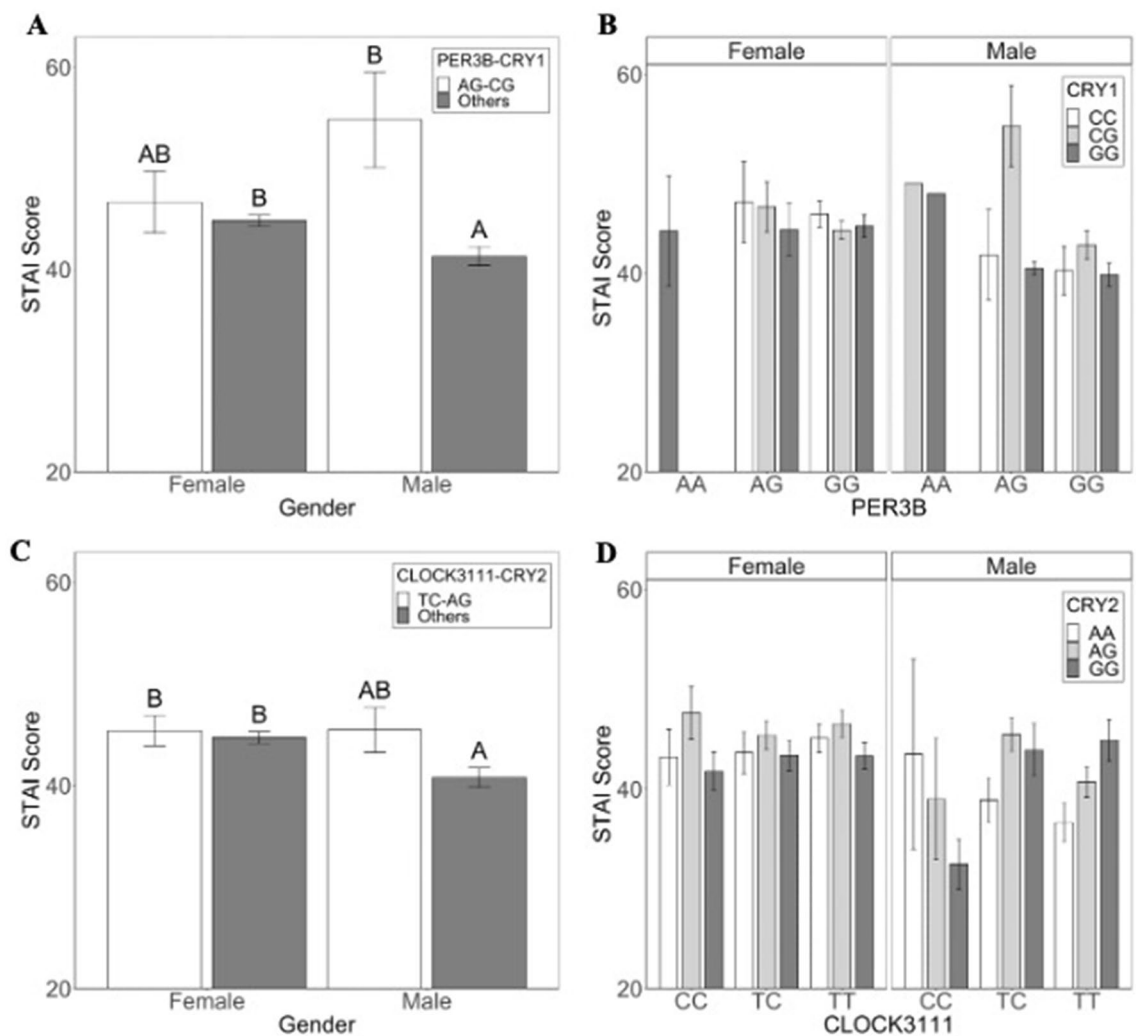


Figure 2. Genotype combinations predictive of anxiety symptoms in males. (A,B) Average anxiety scores for males with a combination of *PER3B*-AG and *CRY1*-CG are higher than average anxiety scores of individuals with other genotype combinations (Gender: $F_{1,479} = 0.661$, $p = 0.417$; Genotype: $F_{1,479} = 7.174$, $p = 0.008$; Gender x Genotype: $F_{1,479} = 4.141$, $p = 0.042$). Anxiety scores are measured using the self-reported State-Trait Anxiety Index (± 1 SE). Tukey's posthoc tests showed that males with AC-CG combination were significantly different from males with other genotypes and from females. (C,D) Average anxiety scores of males with a combination of *CLOCK3111*-TC and *CRY2*-AG also tend to be higher than average anxiety scores of individuals with other genotype combinations, but this is not significant at $p < 0.05$. (Gender: $F_{1,517} = 1.773$; $p = 0.184$, Genotype: $F_{1,517} = 3.417$; $p = 0.065$; Gender x Genotype: $F_{1,517} = 1.912$, $p = 0.167$).

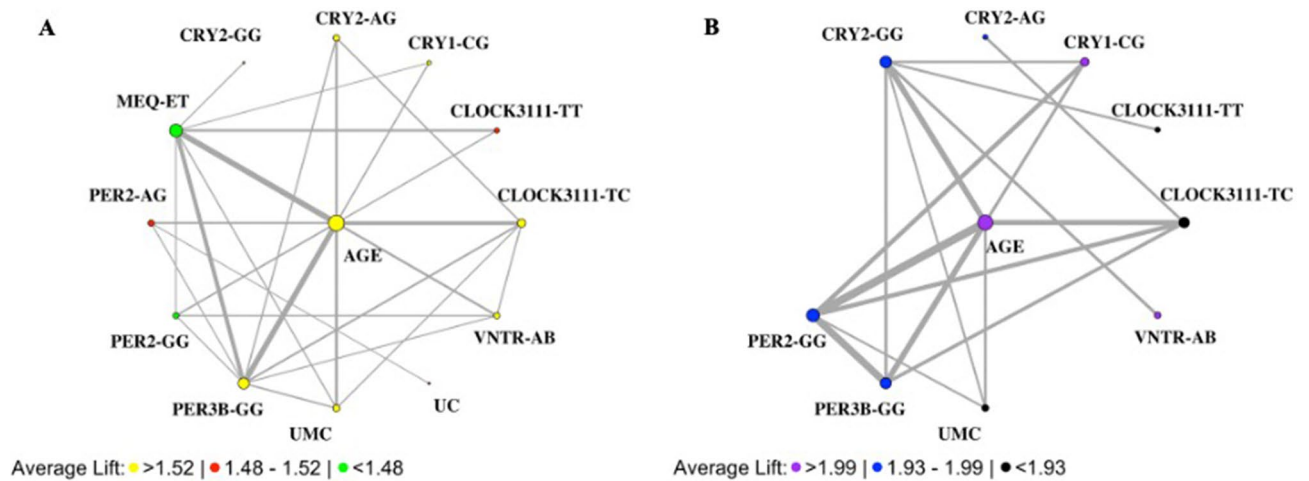


Figure 3. Association rules networks for anxiety symptoms. (A) In females, diurnal preference (MEQ), age, and *PER3BGG* co-occurred most frequently and had the highest average lift in the analysis. (B) In males, age, *PER2-GG*, and *PER3B-GG* co-occurred most frequently, but combinations with *CLOCK3111TC* had the highest average lift.

Additionally, the combination of *PER3B-AG* and *CRY2-AA* was protective against anxiety (Table 1; OR 0.17 (0.04–0.69), $p = 0.026$). Clinical features are also predictive of anxiety symptoms; females and young adults had a significantly higher risk of reporting anxiety symptoms (Table 1; age: OR 3.9 (1.2–14.0), $p = 0.026$; gender: OR 1.8 (1.1–3.0), $p = 0.026$). Preference for morningness had a slight protective effect on the odds of reporting anxiety symptoms (OR 0.96, $p = 0.005$).

The association rule learning results identified additional multi-way genotype and clinical feature combinations that were strong predictors of human anxiety (Suppl. Table 2); risk genotypes differed for males and females (Fig. 3). For females, the most frequently appearing SNP variants in the top predictors of anxiety were *CLOCK3111-TC* and *PER3B-GG*, with age and MEQ as important clinical factors (Fig. 3A). For males, the most frequently occurring variants were *PER2-GG*, *PER3B-GG*, *CRY2-GG*, and *CLOCK3111-TC*, with age, but not MEQ, significantly predicting the risk of anxiety (Fig. 3B).

Genotypic associations with anxiety symptoms can be direct or mediated by diurnal preference. Our network analysis using ARACNE on the interactions between genotypes, clinical features, and anxiety symptoms show that only the *PER3B* SNP variant (rs17031614) shares mutual information with anxiety symptoms via diurnal preference scores, which acts as a mediator (Suppl. Fig. 1). All other variants (*PER2*, *PER3 VNTR*, *PER3A*, *CLOCK3111*, *CRY1*, and *CRY2* SNPs) were directly associated with anxiety symptoms following bootstrap analysis. The most robust links to anxiety symptoms are diurnal preference scores and depressive symptoms; the latter is likely due to the well-known co-morbidity of anxiety and depressive symptoms.

To further investigate the effects of MEQ on anxiety symptoms, we performed a mediation analysis of the top features. We found that the association between being a college-aged student and anxiety scores was mediated by diurnal preference scores. We also found that *PER2-GG* was strongly associated with an increase in anxiety scores (coefficient = 2.17, $t_{418} = 2.09$, $p = 0.038$) and the combination of *PER3B-AG* and *CRY2-AG* was weakly associated with anxiety scores (coefficient = 4.44, $t_{418} = 1.85$, $p = 0.065$); interestingly, these effects were completely mediated by diurnal preference score.

Circadian phase, amplitude, and misalignment are associated with anxiety. Using Fisher's exact tests to test for significant associations across circadian phase and anxiety symptom scores, we found that advanced circadian phase values measured using a *PER3* or *NR1D2* gene markers (> 2.3 and > 1.7 standard deviations above the mean, respectively) were strongly associated with high anxiety scores. Similarly, high circadian amplitudes (> 1.7 and > 2.2 standard deviations above the mean, respectively) were also strongly associated with high anxiety scores. The detailed patterns of association are shown via heatmaps (Fig. 4).

Using gene expression data to measure the degree of mismatch in phase and amplitude with self-reported chronotype, we estimated the risk of anxiety symptoms with circadian misalignment. We found that individuals with advanced circadian phase and evening preferences (low MEQ scores) are six times more likely to report anxiety (*PER3*: OR 5.88 (1.22–28.40), $p = 0.027$; *Nr1d2*: OR 6.50 (0.77–58.48), $p = 0.085$).

Discussion

A growing body of evidence indicates that alterations in circadian rhythms and clock gene mutations influence mood disorders, including anxiety. The search for molecular mechanisms underlying these connections has focused on GWAS and PheWAS studies. Still, these efforts are limited by the difficulty of predicting complex disorders with weak phenotyping and the inability to detect synergistic effects among genotypes. Further, identifying genotypes significantly associated with anxiety provides insufficient information to discern whether

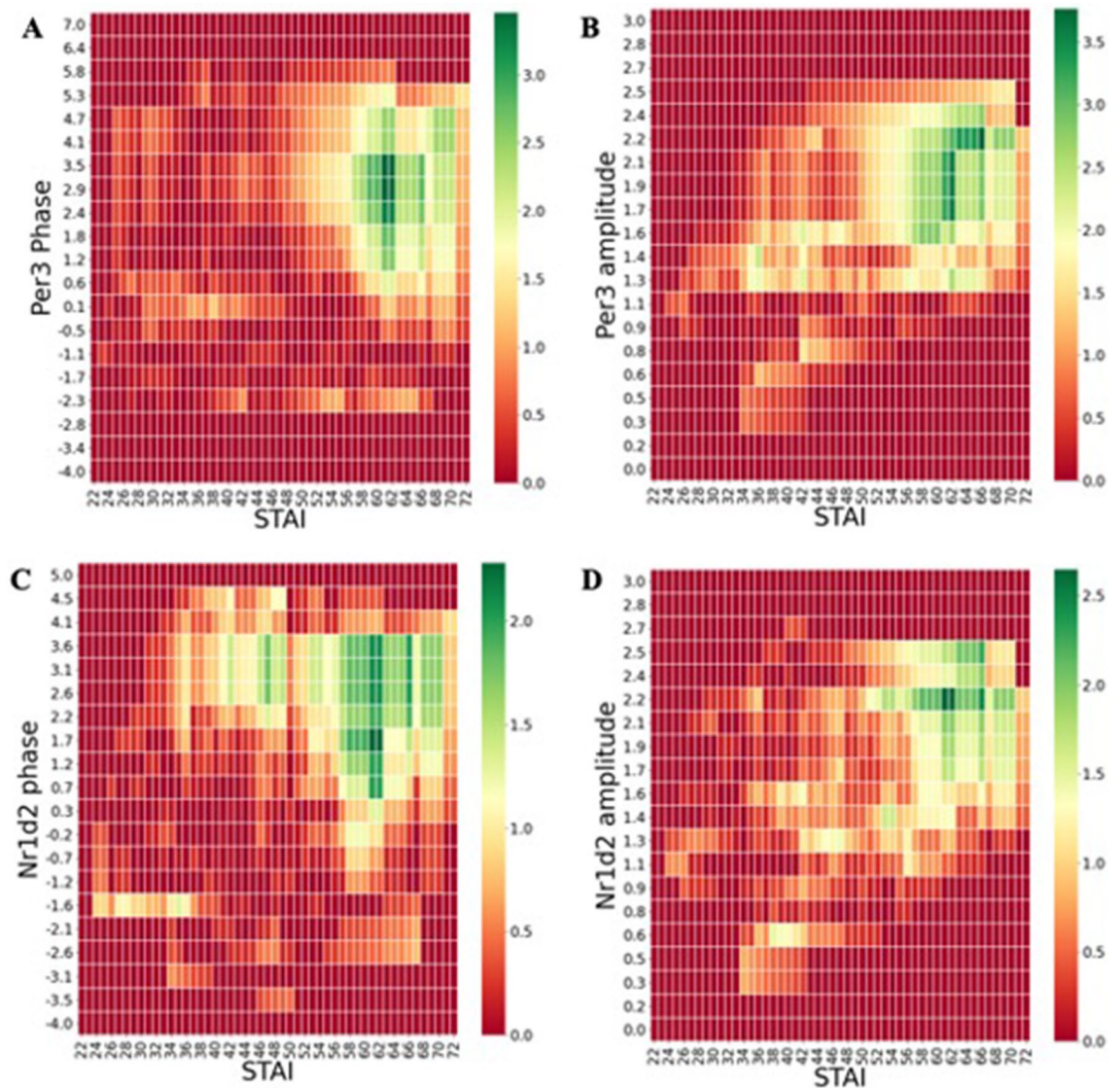


Figure 4. Heat maps of anxiety scores and circadian phenotypes. (A,B) Higher anxiety scores indicative of severe anxiety are strongly associated with positive *PER3* phase (advanced circadian phase or morning-types) and high circadian amplitude. (C,D) Similar patterns are seen with circadian phenotypes measured using the NR1D2 phase and amplitude. The heatmaps were made using Fisher's Exact Test at varying cutoff levels for both dependent and independent variables. The p-values obtained from the analysis were log-transformed (base 10).

gene variants influence symptoms directly or indirectly, impeding the utility of this information for therapeutic interventions. This study uses a novel approach based on machine learning and statistical analysis to explore associations between circadian genes and anxiety symptoms in a deeply-phenotyped population sample. We report three findings: (1) synergistic interactions between variants in *PER3B*, *CLOCK3111*, and cryptochrome genes, *CRY1* and *CRY2*, show robust associations with anxiety symptoms, (2) clock variants predictive of anxiety symptoms tend to have sex-specific effects, and (3) molecular chronotype (circadian phase) and circadian misalignment—particularly individuals with advanced phases and evening-type sleep–wake cycles—are strong predictors of anxiety symptoms. Our results suggest that circadian clock gene variants have both direct (sex-specific) and indirect (clock-mediated) effects on anxiety symptoms.

Circadian genotypes most predictive of anxiety. Our results confirm previous associations of anxiety with gender and chronotype and reveal novel associations of anxiety symptoms with circadian genotypes. We found that the combination of *PER3B*-AG and *CRY1*-CG was most strongly associated with the risk of having anxiety symptoms. The *PER3* gene encodes the period circadian protein homolog 3 protein in humans and is a paralog to the *PER1* and *PER2* genes. *PER3* is not essential for maintaining the circadian rhythm but plays a vital role in sleep–wake timing and sleep homeostasis. This gene is upregulated by *CLOCK*/*BMAL* heterodimers but then is repressed in a feedback loop involving *PER*/*CRY* heterodimers via interactions with *CLOCK*/*BMAL* complex. Previous studies have linked multiple SNPs and VNTRs in *PER3* to diurnal preference^{7,17,23,24,27}. Archer

et al.¹⁸ demonstrated a strong correlation between the extreme diurnal preference and the *PER3* variable number of tandem repeat (VNTR) polymorphism (rs57875989), with the longer allele associated with morning types and the shorter allele associated with evening types and delayed sleep phase syndrome in individuals. Liberman et al.⁷ also found that the *PER3* SNP (rs228697) was significantly associated with diurnal preference and anxiety symptoms.

In mammals, the *CRY* genes act as light-independent inhibitors of the CLOCK/BMAL heterodimers, which work as activators in the main circadian core loop⁶⁰, and these genes also operate in the retina⁶¹. The *CRY1* variant (rs2287161) is an intergenic SNP located downstream from the gene's 3' polyadenylation site, suggesting that additional regulatory elements must aid in the modulation of *CRY1* gene expression. This variant has a robust association with depression in diverse populations^{13,29,62}.

By what mechanisms do variants in *PER3/CRY* genotype combinations influence mood? *PER3* is upregulated by CLOCK/BMAL heterodimers but then represses this upregulation in a feedback loop using *PER/CRY* heterodimers to interact with CLOCK/BMAL. Variants in *PER3* and *CRY* genes may affect the dimerization of the proteins and the speed or success of the binding interaction with the CLOCK/BMAL complex, thus altering the sleep-wake cycle and influencing the timing of molecular pathways regulating mood. Alternatively, changes in *PER3/CRY* binding could affect the regulation of downstream pathways that directly influence mood pathways. This highlights a potentially critical role of *CRY* protein binding in modulating mood pathways.

Combinations of CLOCK3111/*CRY2* variants may act similarly to the *PER3B/CRY1* complexes, inhibiting the phosphorylation of the BMAL1/CLOCK dimer. In humans, *CLOCK* variants have been identified as important for both seasonal affective disorder (SAD) and bipolar disorder^{8,15}. The *CLOCK3111* SNP (rs1801260) is also linked to diurnal preference; individuals homozygous for the C allele have a stronger evening preference⁶³. In non-human models, Roybal et al.⁴ show that *ClockΔ19* mice exhibit hyperactivity, reduced anxiety- and depressive-related behavior, exhibiting a strong manic-like phenotype during the daytime. Transfections of mice with the *CLOCK3111* variants revealed that the C-allele results in increased mRNA levels and stability⁸. In humans, Lavebratt et al.⁶⁴ observed a significant association between the *CRY2* haplotypes with winter depression in Northern European populations. Furthermore, depressed bipolar patients have reduced levels of *CRY2*. Results from these previous studies indicate that both *CLOCK* and *CRY2* are associated with diurnal preference and can be a risk factor for depression and/or anxiety, suggesting that variants in these genes on mood may be direct or indirect effects. Our results suggest that the *CLOCK3111* variant may directly affect anxiety symptoms, particularly in males, when found in combinations with the *CRY2-AG* variant.

Epistatic effects. Complex traits with a polygenic basis, such as anxiety, may be likely to develop if the additive effects exceed a critical threshold that disrupts circadian rhythms^{3,11}. In the current study, individuals who carry multiple circadian SNPs have an increased risk of anxiety symptoms relative to individuals who carry a single gene variant. Using feature selection methods, we can identify many two-genotype combinations that provide more significant effects on anxiety symptoms relative to the impacts of single SNP variants. Our association analysis allows for a more expansive exploration of multiple genes combinations. In these analyses, the top twelve rules with the strongest effects on anxiety symptoms included clinical features and two, three, and four-way gene combinations as strong predictors of human anxiety; only one factor represented a single gene (the *PER3* VNTR_5,5 genotype; Suppl. Table 1). The most frequently appearing SNP genotype in the association analysis was *CLOCK3111-TC*. Two of the rules with the highest lift values show the combination of *CLOCK3111-TC*, *CRY2-AG*, and *PER2-AG* as a significant predictor of anxiety. The combination of the first two SNP variants was identified as statistically significant by feature selection and logistic regression as well, but the additional additive effects of *PER2-AG* were identified by association rule learning. These findings imply that synergistic effects in the molecular clock are critical for modulating physiological pathways associated with anxiety.

Potential direct effects on anxiety: sex-specific circadian genotype effects. One clue to which pathways are influenced by disruptions in the function of *PER3B/CRY* and *CLOCK3111/CRY* complexes is the fact that the effects of these variants on anxiety symptoms may be sex-specific. Previously, Shi et al.⁶⁵ identified significant sex-dependent associations between major depressive disorder (MDD) and common variants of the circadian clock genes *CLOCK*, *PER3*, and *NPAS2*⁶. In that study, the association of *CLOCK* with MDD is also stronger in males, but the association of *PER3* and *NPAS2* with MDD is more significant in females. They propose that these SNPs have a functional effect via output transcriptional pathways that are mediated sex-dependently by the circadian system rather than the core clock oscillator⁶⁵. One possible mechanism is glucocorticoid regulation, given that males and females have different cortisol levels⁶⁶. Sex-specific, glucocorticoid-mediated stress responses may represent a mechanism by which clock genes affect anxiety and other mood disorders⁶⁷. Other targets of *CLOCK*-mediated transcription involve neuropeptides and neurotransmitters, as well as their receptors, that may act to modulate mood pathways, including serotonergic pathways.

Interestingly, our network analyses also showed clear differences in key genotypic risk factors for males and females. The top rules for females included strong predictions for anxiety with *CLOCK3111-TC*, *PER3B-GG*, age, and MEQ, as well as weaker associations for other genotypes. Average lifts were higher for all of the top rules for males, but co-occurrence was more widely distributed across the genotypes, suggesting stronger associations between multiple gene variants and anxiety. In males, age, *PER2-GG*, *PER3B-GG*, and *CRY2-GG* co-occurred most frequently, but combinations with *CLOCK3111TC* had the highest average lift. Overall, our results suggest that the sex-specific anxiety risk conferred by the genotypic combinations involving *PER3B*, *CLOCK3111*, and *CRY2* genes may be further evidence of direct effects of clock gene binding complexes on downstream mood-related physiological pathways.

Potential indirect effects on anxiety: mediation by diurnal preference and circadian misalignment. Following bootstrap analysis of the ARACNE gene network, *PER3B* was the only polymorphism in the current study with effects on anxiety that were significantly mediated by diurnal preference. In our targeted mediation analyses, associations of anxiety symptoms in *PER2* homozygotes for the G-allele were also significantly mediated by diurnal preference. This suggests that the Period family of genes may mediate mood via indirect pathways associated with circadian phenotypes and/or circadian misalignment.

Our molecular chronotyping results provide the strongest support for the indirect role of circadian clocks on mood by linking advanced phase, robust rhythms, and circadian misalignment to high levels of anxiety symptoms. Previous studies have shown that advanced *PER3* phase is strongly associated with morning types, while delayed phases values are more commonly found in evening types. In the current study, we show that higher *PER3* phase values were strongly associated with high human anxiety scores. Similarly, a higher *PER3* amplitude was also strongly associated with high anxiety. One potential explanation for these results is that individuals with stronger, more robust circadian amplitudes tend to have more robust sleep/wake patterns. Given that a large proportion of our study population consisted of undergraduates, the altered (i.e., typically delayed) sleep/wake patterns of college life may cause significant circadian misalignment, leading to high anxiety. To test whether the pattern of advanced phase and robust rhythms with high anxiety indicated the effects of circadian misalignment on mood, we identified individuals in our dataset with advanced circadian phases but who also reported evening-type patterns of sleep–wake behavior. The risk of anxiety symptoms was six times higher for these misaligned individuals, regardless of genotype, indicating that chronic disruptions to endogenous sleep–wake patterns increase anxiety symptoms in humans. A similar result was found for depressive symptoms in the same population¹⁶, suggesting that shifts in circadian circuitry may influence parallel pathways affecting both depressive and anxious symptoms.

Limitations. This study should be viewed in the context of several limitations. Our machine learning and statistical analysis examined a large number of features for the relatively small sample size of the population. In addition, our population of primarily Caucasians of European descent limits the generalizability of our findings to diverse populations. We demonstrated that we could accurately predict anxiety using classification with a subset of features selected via feature selection. However, we are unable to quantify the classification prediction accuracy using only the top robust features since we used the entire data set (due to the small sample size) to attain these features. Our estimation of delayed or advanced circadian phase utilized an analysis of gene expression from two genes at four data points; greater accuracy in estimation may be achieved with analysis of additional genes or data points. Future studies should assess the accuracy of our anxiety risk factor predictions using an independent population sample. Finally, our statistical analyses report significant differences in anxiety risk associated with particular genotypes and circadian phenotypes; further functional and behavioral studies are needed to understand how therapeutic targets or behavioral interventions might be designed to mitigate anxiety symptoms in humans.

Conclusion

Here, we report both direct and indirect, via mediation by circadian phenotypes, effects of circadian genotypic and clinical features on anxiety symptoms. Using an approach that employs machine learning and statistical analyses to examine associations of circadian clock genes with human anxiety, our results support three conclusions. First, variants in select circadian clock genes have synergistic associations with anxiety symptoms. Second, sex-linked associations between clock gene variants and anxiety symptoms provide evidence of multiple direct pathways for clock genes to influence mood. Finally, molecular chronotype and circadian misalignment are strong predictors of anxiety symptoms, indicating that indirect effects of clock gene variants, particularly in the *PER3* gene, may also play a role in modulating anxiety symptoms. Disentangling the complex influences of clock genes on anxiety may reveal both clinical targets and non-invasive therapies that can help mitigate the causes and symptoms of anxiety.

Received: 21 December 2021; Accepted: 21 March 2022

Published online: 01 April 2022

References

1. Steel, Z. *et al.* The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *Int. J. Epidemiol.* **43**, 476–493 (2014).
2. McCarthy, M. J. & Welsh, D. K. Cellular circadian clocks in mood disorders. *J. Biol. Rhythms* **27**, 339–352 (2012).
3. Partonen, T. Clock gene variants in mood and anxiety disorders. *J. Neural Transm.* **119**, 1133–1145 (2012).
4. Roybal, K. *et al.* Mania-like behavior induced by disruption of *CLOCK*. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6406–6411 (2007).
5. Lavebratt, C., Sjöholm, L. K., Partonen, T., Schalling, M. & Forsell, Y. *PER2* variant is associated with depression vulnerability. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **153**, 570–581 (2010).
6. Shi, S. Q. *et al.* Molecular analyses of circadian gene variants reveal sex-dependent links between depression and clocks. *Transl. Psychiatry* **6**, e748 (2016).
7. Liberman, A. R. *et al.* Circadian clock model supports molecular link between *PER3* and human anxiety. *Sci. Rep.* **7**, 9893 (2017).
8. Ozburn, A. R. *et al.* Functional implications of the *CLOCK311T/C* single-nucleotide polymorphism. *Front. Psychol.* **7**, 1–8 (2016).
9. Evans, J. A., Pan, H., Liu, A. C. & Welsh, D. K. *CRY1*–/– circadian rhythmicity depends on SCN intercellular coupling. *J. Biol. Rhythms* **27**, 443 (2012).
10. Walker, W. H., Walton, J. C., DeVries, A. C. & Nelson, R. J. Circadian rhythm disruption and mental health. *Transl. Psychiatry* **10**, 28 (2020).
11. Albrecht, U. Molecular mechanisms in mood regulation involving the circadian clock. *Front. Neurol.* **8**, 30 (2017).

12. Archer, S. N. & Oster, H. How sleep and wakefulness influence circadian rhythmicity: Effects of insufficient and mistimed sleep on the animal and human transcriptome. *J. Sleep Res.* **24**, 476–493 (2015).
13. Soria, V. *et al.* Differential association of circadian genes with mood disorders: CRY1 and NPAS2 are associated with unipolar major depression and clock and VIP with bipolar disorder. *Neuropsychopharmacology* **35**, 1279–1289 (2010).
14. Liberman, A. R., Halitjaha, L., Ay, A. & Ingram, K. K. Modeling strengthens molecular link between circadian polymorphisms and major mood disorders. *J. Biol. Rhythms* **33**, 318–336 (2018).
15. Kim, H.-I. *et al.* Association of CLOCK, ARNTL, and NPAS2 gene polymorphisms and seasonal variations in mood and behavior. *Chronobiol. Int.* **32**, 785–791 (2015).
16. Nguyen, C., Murray, G., Anderson, S., Filipowicz, A. & Ingram, K. K. In vivo molecular chronotyping, circadian misalignment, and high rates of depression in young adults. *J. Affect. Disord.* **250**, 425 (2019).
17. Fares, S. *et al.* Clinical correlates of chronotypes in young persons with mental disorders. *Chronobiol. Int.* **32**, 1183–1191 (2015).
18. Archer, S. N. *et al.* A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference. *Sleep* **26**, 413–415 (2003).
19. Partonen, T. *et al.* Three circadian clock genes Per2, Arntl, and Npas2 contribute to winter depression. *Ann. Med.* **39**, 229–238 (2007).
20. Hida, A. *et al.* Screening of clock gene polymorphisms demonstrates association of a PER3 polymorphism with morningness-eveningness preference and circadian rhythm sleep disorder. *Sci. Rep.* **4**, 6309 (2014).
21. Zhang, L. *et al.* A PERIOD3 variant causes a circadian phenotype and is associated with a seasonal mood trait. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E1536–E1544 (2016).
22. Hidalgo, M. P. *et al.* Relationship between depressive mood and chronotype in healthy subjects. *Psychiatry Clin. Neurosci.* **63**, 283–290 (2009).
23. Kitamura, S. *et al.* Evening preference is related to the incidence of depressive states independent of sleep-wake conditions. *Chronobiol. Int.* **27**, 1797–1812 (2010).
24. Levandovski, R. *et al.* Depression scores associate with chronotype and social jetlag in a rural population. *Chronobiol. Int.* **28**, 771–778 (2011).
25. Prat, G. & Adan, A. Relationships among circadian typology, psychological symptoms, and sensation seeking. *Chronobiol. Int.* **30**, 942–949 (2013).
26. Merikanto, I. *et al.* Evening types are prone to depression. *Chronobiol. Int.* **30**, 719–725 (2013).
27. Antypa, N., Vogelzangs, N., Meesters, Y., Schoevers, R. & Penninx, B. W. J. H. Chronotype associations with depression and anxiety disorders in a large cohort study. *Depress. Anxiety* **33**, 75–83 (2016).
28. Au, J. & Reece, J. The relationship between chronotype and depressive symptoms: A meta-analysis. *J. Affect. Disord.* **218**, 93–104 (2017).
29. Buoli, M. *et al.* The role of clock genes in the etiology of major depressive disorder: special section on “translational and neuroscience studies in affective disorders”. *J. Affect. Disord.* **234**, 351–357 (2018).
30. McClung, C. A. How might circadian rhythms control mood? Let me count the ways. *Biol. Psychiatry* **1**, 119–131 (2013).
31. Savalli, G. *et al.* Anhedonic behavior in cryptochrome 2-deficient mice is paralleled by altered diurnal patterns of amygdala gene expression. *Amino Acids* **47**, 1367 (2015).
32. Ozburn, A. R. *et al.* NPAS2 regulation of anxiety-like behavior and GABAA receptors. *Front. Mol. Neurosci.* **10**, 360 (2017).
33. Ho, K. W. D. *et al.* Genome-wide association study of seasonal affective disorder. *Transl. Psychiatry* **8**, 190 (2018).
34. Satyanarayanan, S. K. *et al.* Melatonergic agonist regulates circadian clock genes and peripheral inflammatory and neuroplasticity markers in patients with depression and anxiety. *Brain Behav. Immunity* **85**, 142–151 (2020).
35. Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R. & Jacobs, G. A. State-Trait anxiety inventory for adults. In *Manual for the State-Trait Anxiety Inventory* (1983).
36. Horne, J. A. & Ostberg, O. A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *Int. J. Chronobiol.* **4**, 97–110 (1976).
37. Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R. & Jacobs, G. A. *Manual for the State-Trait Anxiety Inventory* (Consulting Psychologists Press, 1983).
38. Beck, A. T. & Beamesderfer, A. Assessment of depression: The depression inventory. In *Psychological Measurements in Psychopharmacology* Vol. 7 (eds Pichot, P. & Olivier-Martin, R.) 151–169 (Karger, 1974).
39. Yu, L. *et al.* Development of short forms from the PROMIS™ sleep disturbance and sleep-related impairment item banks. *Behav. Sleep Med.* **10**, 6–24 (2012).
40. Ebisawa, T. *et al.* Association of structural polymorphisms in the human period3 gene with delayed sleep phase syndrome. *EMBO Rep.* **2**, 342–346 (2001).
41. Ingram, K. K. *et al.* Molecular insights into chronotype and time-of-day effects on decision-making. *Sci. Rep.* **6**, 29392 (2016).
42. Robnik-Sikonja, M. & Kononenko, I. An adaptation of relief for attribute estimation in regression. In *Proc. Fourteenth International Conference on Machine Learning*, 296–304 (Morgan Kaufmann Publishers Inc., 1997).
43. Quinlan, J. R. & Quinlan, J. R. Induction of decision trees. *Mach. Learn.* **1**, 81–106 (1986).
44. Ding, C. & Peng, H. Minimum redundancy feature selection from microarray gene expression data. In *Computational Systems Bioinformatics. CSB2003. Proc. 2003 IEEE Bioinformatics Conference. CSB2003*, 523–528 (2003). <https://doi.org/10.1109/CSB.2003.1227396>.
45. Yang, H. H. & Moody, J. E. *Feature Selection Based on Joint Mutual Information* (1999).
46. Chen, T. & Guestrin, C. XGBoost: A scalable tree boosting system. In *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794 (Association for Computing Machinery, 2016). <https://doi.org/10.1145/2939672.2939785>.
47. Boser, B. E., Guyon, I. & Vapnik, V. N. A training algorithm for optimal margin classifiers. In *COLT '92* (1992).
48. Statistics, L. B. & Breiman, L. Random forests. In *Machine Learning*, 5–32 (2001).
49. Fukunaga, K. The optimal distance measure for nearest neighbor classification. *IEEE Trans. Inf. Theory* **27**, 622–627 (1981).
50. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Int. Res.* **16**, 321–357 (2002).
51. R Core Team. *R: A Language and Environment for Statistical Computing* (2020).
52. Bozdogan, H. Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika* **52**, 345–370 (1987).
53. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. Ser. B (Methodol.)* **57**, 289–300 (1995).
54. Imai, K., Keele, L., Tingley, D. & Yamamoto, T. Causal mediation analysis using R. In *Advances in Social Science Research Using R* (ed. Vinod, H. D.) (Springer, 2010).
55. Fox, J. & Weisberg, S. *An R Companion to Applied Regression* (Sage, 2019).
56. Hahsler, M., Chelluboina, S., Hornik, K. & Buchta, C. The arules R-package ecosystem: Analyzing interesting patterns from large transaction datasets. *J. Mach. Learn. Res.* **12**, 1977–1981 (2011).
57. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal* **1695**, 1–9 (2006).

58. Margolin, A. A. *et al.* ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinform.* **7**, 1–15 (2006).
59. Meyer, P. E., Lafitte, F. & Bontempi, G. MINET: An open source R/bioconductor package for mutual information based network inference. *BMC Bioinform.* **9**, 461 (2008).
60. Dardente, H., Fortier, E. E., Martineau, V. & Cermakian, N. Cryptochromes impair phosphorylation of transcriptional activators in the clock: A general mechanism for circadian repression. *Biochem. J.* **402**, 525–536 (2007).
61. Hsu, D. S. *et al.* Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins. *Biochemistry* **35**, 13871–13877 (1996).
62. Hua, P. *et al.* Cry1 and Tef gene polymorphisms are associated with major depressive disorder in the Chinese population. *J. Affect. Disord.* **157**, 100–103 (2014).
63. Katzenberg, D. *et al.* A CLOCK polymorphism associated with human diurnal preference. *Sleep* **21**, 569–576 (1998).
64. Lavebratt, C. *et al.* CRY2 is associated with depression. *PLoS ONE* **5**, e9407 (2010).
65. Shi, S. Q. *et al.* Molecular analyses of circadian gene variants reveal sex-dependent links between depression and clocks. *Transl. Psychiatry*. <https://doi.org/10.1038/tp.2016.9> (2016).
66. Halbreich, U. & Lumley, L. A. The multiple interactional biological processes that might lead to depression and gender differences in its appearance. *J. Affect. Disord.* **29**, 159–173 (1993).
67. Landgraf, D., McCarthy, M. J. & Welsh, D. K. Circadian clock and stress interactions in the molecular biology of psychiatric disorders. *Curr. Psychiatry Rep.* **16**, 483 (2014).

Acknowledgements

We would like to acknowledge K. Woods and L. Dhawka for assistance with genetic analyses.

Author contributions

Genetic analyses were conducted by K.I. and R.O., statistical analyses were conducted by A.Z., Z.A. and A.A. R.O. and K.I. wrote the manuscript; A.Z. and A.A. edited the manuscript. The study was designed by K.I. and A.A. and funded by K.I.

Funding

The funding was provided by Colgate University.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-09421-4>.

Correspondence and requests for materials should be addressed to A.A. or K.I.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022