Machine Learning and In-Vivo Expression Analyses Reveal Circadian Clock Features Predictive of Anxiety in Humans

Aziz Zafar  
Colgate University

Rebeccah Overton  
Colgate University

Ziad Attia  
Colgate University

Ahmet Ay  
Colgate University

Krista Ingram (kingram@colgate.edu)  
Colgate University

Research Article

Keywords: anxiety, disruptions, polymorphisms, phenotyped population, chronotyping

Posted Date: December 29th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1190238/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Mood disorders, including anxiety, 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 PER3B-AG/CRY1-CG genotype being the strongest predictor of anxiety risk 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 PER3B 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.

Introduction

Mood disorders, including depression and anxiety, are becoming more prevalent globally, affecting nearly one-fifth of the adult population (Steel et al., 2014). 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 anxiety (Roybal et al., 2007; Lavebratt et al., 2010a,b; Shi et al., 2016; Partonen, 2012; Benedetti et al., 2003, 2008; Li et al., 2013; Desan et al., 2000; Kishi et al., 2008; Liberman et al., 2017, 2018; McCarthy & Welsh, 2012; Ketchesin et al., 2020). 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 anxiety (Karatsoreos, 2014; Landgraf et al., 2014).

Circadian rhythms regulate a sleep-wake cycle that is reset every 24 hours 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 (Evans et al., 2012a; Walker et al., 2020). The core feedback loop includes the transcription factor CLOCK, which positively regulates the transcription of genes in the Period (PER1, PER2, and PER3) and Cryptochrome (CRY1, and CRY2) complexes. The CLOCK
protein forms a heterodimer with BMAL1, which activates additional core clock genes. PER/CRY heterodimers, in turn, inhibit the transcriptional activity of the BMAL1-CLOCK complex. This cycle of transcription activation and repression is vital for the 24-hour circadian cycle. Mutations in these core clock genes may affect mood via direct transcriptional activity of downstream physiological pathways or indirectly through disruptions in the phase and strength of circadian rhythms (Albrecht, 2017). 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, differences in sleep-wake timing, with mood (Nguyen et al., 2019; Archer and Oster, 2015; Lewy et al., 2007; Liberman et al., 2017, 2018; Kim et al., 2015; Soria et al., 2010). 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 profiles (Robillard et al., 2018) and clock gene expression oscillations (molecular chronotypes) associated with diurnal preference (Zhang et al., 2016; Lavebratt et al., 2010; Hida et al., 2014; Archer et al., 2003; C. Lavebratt et al., 2010; Partonen et al., 2007; Soria et al., 2010; Kim et al., 2015). Studies on diurnal preference show that evening types are more likely to experience depression or anxiety at one point in their lifetime (Hidalgo et al., 2009; Kitamura et al., 2010; Levandovski et al., 2011; Adan et al., 2012; Merikanto et al., 2013; Fares et al., 2015; Antypa et al., 2016; Au and Reece, 2017, Buoli et al., 2018). 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 (Nguyen et al., 2019a; Walker et al., 2020b). Previous GWAS and candidate gene studies have also found associations between multiple circadian genes and diurnal preference (Hida et al., 2014; Jones et al., 2016; Liberman et al., 2017a, 2018) as well as depression or other psychological disorder (Soria et al., 2010). 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 neurotransmitters and can directly influence well-known mood pathways, including serotonin, dopamine, and glucocorticoid pathways (McClung, 2013). Recent studies demonstrate that components of the circadian clock can directly modulate mood disorders. In mice, knock down of the NPAS2, period, or cryptochrome genes results in altered anxiety levels (Evans et al., 2012; Savalli et al., 2015; Ozburn et al., 2017a). 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 and colleagues (2018) 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 (Lavebratt et al., 2010b; Liberman et al., 2017; Ozburn et al., 2017b; Partonen, 2012; Silva et al., 2020). Mathematical models have also predicted links between circadian clock disruptions and anxiety.
(Liberman et al., 2018). Most interestingly, a recent clinical study demonstrated that melatonin treatment used to correct circadian misalignment in anxious patients helped to mitigate anxiety (Satyanarayanan et al., 2020). 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)) (Spielberger et al., 1983), diurnal preference (Morningness-Eveningness Questionnaire (MEQ) (Horne and Ostberg, 1976), and molecular chronotype (via in-vivo 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), Beck Depression Inventory (BDI-II), and the short form of the Patient-Reported Outcomes Measurement Information System (PROMIS™) Sleep Disturbance. The STAI was used to indicate the anxiety scores of individuals ranging from 20-80, 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) survey was administered to measure diurnal preference.

**Genotyping**

DNA was extracted from ten to twenty hair follicles from each participant. The hair samples were digested with Proteinase K at 56°C for 24-hours and purified using the Qiagen DNAeasy 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'-CAAAATTTTAGACACTACCAGATGGCTGAC-3'; and reverse, 5'-AACC TTGTACTTCCACATCAGTGCCTGG-3' (Ebisawa et al., 2001). The PCR was performed in a 25-µl 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. (Nguyen et al., 2019). 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 subjects (Ingram et al., 2016). The trained curve was fitted to the four data points of each subject using the parameter estimation method, Stochastic Ranking Evolutionary Strategy (SRES) (Liberman et al., 2017). 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 \(7C2 \times 9\) 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 (Quinlan & Quinlan, 1986; Robnik-Sikonja & Kononenko, 1997). 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 (Ding & Peng, 2003; Yang & Moody, 1999). 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) (Boser et al., 1992; Chen & Guestrin, 2016; Statistics & Breiman, 2001). 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 ten folds (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 fold (Fukunaga, 1981). 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 class (Chawla et al., 2002). 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 R (R Core Team, 2020). 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 variable (Bozdogan, 1987). 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 method (Benjamini & Hochberg, 1995).

**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 effects (Imai et al., 2010).

**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 ANOVA (Fox & Weisberg, 2019). 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* package (Hahsler et al., 2011) 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* library (Csardi & Nepusz, 2006).

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 disorders (Margolin et al., 2006). 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* (Meyer et al., 2008). 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 (Figure 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 (Figure 2; OR = 15.3, p = 0.026). Average anxiety scores for individuals with *PER3B*-AG and *CRY1*-CG 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; Figure 2a). The highest average STAI score occurs in males with *PER3B*-AG and *CRY1*-CG genotypes (Figure 2b).
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.

| 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  |

The combination of *CLOCK3111-TC* and *CRY2-AG* also significantly increased the odds of anxiety symptoms (Figure 4; 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; Figure 2c). The highest average anxiety score of 47.6 ± 2.6 was found for females with *CLOCK3111-CC* and *CRY2-AG* (Figure 2d).

Additionally, the combination of *PER3B-AG* and *CRY2-AA* was protective against anxiety (Table 1; OR = 0.17 (0.04-0.07), 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 (Figure 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 (Figure 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 (Figure 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. Figure 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\textsubscript{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\textsubscript{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 (Figure 4).

Using in-vivo 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 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 (Kitamura et al., 2010; Levandovski et al., 2011; Antypa et al., 2015; Fares et al., 2015, Liberman et al., 2017b). Archer et al. (2003) 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. (2018) 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 loop (Dardente et al., 2007), and these genes also operate in the retina (Hsu et al., 1996). 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 (Soria et al, 2010; Hua et al, 2014; Buoli et al., 2018).

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 (Benedetti et al., 2003; Kim et
al., 2015). The CLOCK3111 SNP (rs1801260) is also linked to diurnal preference; individuals homozygous for the C allele have a stronger evening preference (Katzenberg et al., 1998; Mishima et al., 2005). In non-human models, Roybal et al. (2007) 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 stability (Ozburn et al., 2016). In humans, Lavebratt et al. (2010b) 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, are more likely to develop if the additive effects exceed a critical threshold that disrupts circadian rhythms (Albrecht, 2017; Partonen, 2012). 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 I). 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 (Shi et al., 2016; Viena et al., 2016). Previously, Shi et al. (2016) identified significant sex-dependent associations between major depressive disorder (MDD) and common variants of the circadian clock genes CLOCK, PER3, and NPAS2. 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 oscillator. (Shi et al., 2016). One possible
mechanism is glucocorticoid regulation (Shi et al., 2016), given that males and females have different cortisol levels (Halbreich et al, 1993). Sex-specific, glucocorticoid-mediated stress responses may represent a mechanism by which clock genes affect anxiety and other mood disorders (Landgraf et al., 2014). 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 population (Nyugen et al 2019), 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. 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.

**Declarations**

**Acknowledgements**

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

**Contributions**

Genetic analyses were conducted by KKI and RO, statistical analyses were conducted by AZ, ZA and AA. RO and KKI wrote the manuscript; AZ and AA edited the manuscript. The study was designed and funded by KKI.
References

1. Adan, A., Archer, S. N., Hidalgo, M. P., Di Milia, L., Natale, V., & Randler, C. (2012). Circadian typology: a comprehensive review. Chronobiology international, 29(9), 1153–1175. https://doi.org/10.3109/07420528.2012.719971

2. Albrecht, U. (2017). Molecular mechanisms in mood regulation involving the circadian clock. In Frontiers in Neurology (Vol. 8, Issue FEB). Frontiers Research Foundation. https://doi.org/10.3389/fneur.2017.00030

3. Antypa, N., Vogelzangs, N., Meesters, Y., Schoevers, R., & Penninx, B. W. (2016). Chronotype associations with depression and anxiety disorders in a large cohort study. Depression and anxiety, 33(1), 75–83. https://doi.org/10.1002/da.22422

4. Archer, S. N., Robilliard, D. L., Skene, D. J., Smits, M., Williams, A., Arendt, J., & von Schantz, M. (2003). A length polymorphism in the circadian clock gene Per3 is linked to delayed sleep phase syndrome and extreme diurnal preference. Sleep, 26(4), 413–415. https://doi.org/10.1093/sleep/26.4.413

5. Archer, S. N., & Oster, H. (2015). How sleep and wakefulness influence circadian rhythmicity: effects of insufficient and mistimed sleep on the animal and human transcriptome. Journal of sleep research, 24(5), 476–493. https://doi.org/10.1111/jsr.12307

6. Au, J., & Reece, J. (2017). The relationship between chronotype and depressive symptoms: A meta-analysis. Journal of affective disorders, 218, 93–104. https://doi.org/10.1016/j.jad.2017.04.021

7. Benedetti, F., Serretti, A., Colombo, C., Barbini, B., Lorenzi, C., Campori, E., & Smeraldi, E. (2003). Influence of CLOCK gene polymorphism on circadian mood fluctuation and illness recurrence in bipolar depression. American journal of medical genetics. Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics, 123B(1), 23–26. https://doi.org/10.1002/ajmg.b.20038

8. Benedetti, F., Dallaspezia, S., Colombo, C., Pirovano, A., Marino, E., & Smeraldi, E. (2008). A length polymorphism in the circadian clock gene Per3 influences age at onset of bipolar disorder. Neuroscience letters, 445(2), 184–187. https://doi.org/10.1016/j.neulet.2008.09.002

9. Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. https://doi.org/10.1111/J.2517-6161.1995.TB02031.X

10. Boser, B. E., Guyon, I., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. COLT '92.

11. Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. Psychometrika 1987 52:3, 52(3), 345–370. https://doi.org/10.1007/BF02294361

12. Buoli, M., Serati, M., Grassi, S., Pergoli, L., Cantone, L., Altamura, A. C., & Bollati, V. (2018). The role of clock genes in the etiology of Major Depressive Disorder. Journal of affective disorders, 234, 351–357. https://doi.org/10.1016/j.jad.2017.11.015
13. Chawla, N. v, Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority over-Sampling Technique. *J. Artif. Int. Res.*, 16(1), 321–357.

14. Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. https://doi.org/10.1145/2939672.2939785

15. Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Sy*, 1695. https://igraph.org

16. Dardente, H., Fortier, E. E., Martineau, V., & Cermakian, N. (2007). Cryptochromes impair phosphorylation of transcriptional activators in the clock: A general mechanism for circadian repression. *Biochemical Journal*, 402(3), 525–536. https://doi.org/10.1042/BJ20060827

17. Desan, P. H., Oren, D. A., Malison, R., Price, L. H., Rosenbaum, J., Smoller, J., et al. (2000). Genetic polymorphism at the CLOCK gene locus and major depression. *Am. J. Med. Genet*. 96, 418–421. doi: 10.1002/1096-8628(20000612)96:3<418::aid-ajmg34>3.0.co;2-s

18. Ding, C., & Peng, H. (2003). Minimum redundancy feature selection from microarray gene expression data. *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003*, 523–528. https://doi.org/10.1109/CSB.2003.1227396

19. Ebisawa T, Uchiyama M, Kajimura N, Mishima K, Kamei Y, Katoh M, et al. (2001). Association of structural polymorphisms in the human period3 gene with delayed sleep phase syndrome. *EMBO Rep* 2(4):342e6.

20. Evans, J. A., Pan, H., Liu, A. C., & Welsh, D. K. (2012). *CRY1−/−* Circadian rhythmicity depends on SCN intercellular coupling. *Journal of Biological Rhythms*, 27(6), 443. https://doi.org/10.1177/0748730412461246

21. Fares, S., Hermens, D. F., Naismith, S. L., White, D., Hickie, I. B., & Robillard, R. (2015). Clinical correlates of chronotypes in young persons with mental disorders. *Chronobiology International*, 0528(December), 1–9. https://doi.org/10.3109/07420528.2015.1078346

22. Fox, J., & Weisberg, S. (2019). *An (R) Companion to Applied Regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

23. Fukunaga, K. (1981). The Optimal Distance Measure for Nearest Neighbor Classification. *IEEE Transactions on Information Theory*, 27(5), 622–627. https://doi.org/10.1109/TIT.1981.1056403

24. Hahsler, M., Chelluboina, S., Hornik, K., & Buchta, C. (2011). The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Datasets. *Journal of Machine Learning Research*, 12, 1977–1981. https://jmlr.csail.mit.edu/papers/v12/hahsler11a.html

25. Halbreich U, Lumley LA (1993). The multiple interactional biological processes that might lead to depression and gender differences in its appearance. *Journal of Affective Disorders* 29: 159–173.

26. Hida A, Kitamura S, Katayose Y, Kato M, Ono H, Kadotani H, et al. (2014). Screening of clock gene polymorphisms demonstrates association of a PER3 polymorphism with morningness-eveningness preference and circadian rhythm sleep disorder. *Scientific Reports* 4:6309.
27. Hidalgo, M. P., Caumo, W., Posser, M., Coccaro, S. B., Camozzato, A. L., & Chaves, M. L. (2009). Relationship between depressive mood and chronotype in healthy subjects. *Psychiatry and clinical neurosciences, 63*(3), 283–290. https://doi.org/10.1111/j.1440-1819.2009.01965.x

28. Ho, K., Han, S., Nielsen, J. V., Jancic, D., Hing, B., Fiedorowicz, J., Weissman, M. M., Levinson, D. F., & Potash, J. B. (2018). Genome-wide association study of seasonal affective disorder. *Translational psychiatry, 8*(1), 190. https://doi.org/10.1038/s41398-018-0246-z

29. Horne, J. A., & Ostberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International journal of chronobiology, 4*(2), 97–110.

30. Hsu, D. S., Zhao, X., Zhao, S., Kazantsev, A., Wang, R. P., Todo, T., Wei, Y. F., & Sancar, A. (1996). Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins. *Biochemistry, 35*(44), 13871–13877. https://doi.org/10.1021/bi962209o

31. Hua, P., Liu, W., Chen, D., Zhao, Y., Chen, L., Zhang, N., Wang, C., Guo, S., Wang, L., Xiao, H., & Kuo, S. H. (2014). *CRY1* and Tef gene polymorphisms are associated with major depressive disorder in the Chinese population. *Journal of affective disorders, 157*, 100–103. https://doi.org/10.1016/j.jad.2013.11.019

32. Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2010). Causal Mediation Analysis Using R. In H. D. Vinod (Ed.), *Advances in Social Science Research Using R*. Springer-Verlag.

33. Ingram, K. K., Ay, A., Kwon, S. Bin, Woods, K., Escobar, S., Gordon, M., Smith, I. H., Bearden, N., Filipowicz, A., & Jain, K. (2016). Molecular insights into chronotype and time-of-day effects on decision-making. *Scientific Reports, 6*. https://doi.org/10.1038/srep29392

34. Jones, S. E., Tyrrell, J., Wood, A. R., Beaumont, R. N., Ruth, K. S., Tuke, M. A., Yaghoobtakar, H., Hu, Y., Teder-Laving, M., Hayward, C., Roenneberg, T., Wilson, J. F., del Greco, F., Hicks, A. A., Shin, C., Yun, C. H., Lee, S. K., Metspalu, A., Byrne, E. M., ... Weedon, M. N. (2016). Genome-Wide Association Analyses in 128,266 Individuals Identifies New Morningness and Sleep Duration Loci. *PLoS Genetics, 12*(8), 1–19. https://doi.org/10.1371/journal.pgen.1006125

35. Karatsoreos, I. N. (2014). Links between circadian rhythms and psychiatric disease. *Front. Behav. Neurosci.*, 8:162. doi: 10.3389/fnbeh.2014.00162

36. Katzenberg D, Young T, Finn L, Lin L, King DP, Takahashi JS et al. (1998) A CLOCK polymorphism associated with human diurnal preference. *Sleep* 21: 569–576.

37. Ketchesin, K. D., Becker-Krail, D., & McClung, C. A. (2020). Mood-related central and peripheral clocks. In *European Journal of Neuroscience*. https://doi.org/10.1111/ejn.14253

38. Kim, H.-I., Lee, H.-J., Cho, C.-H., Kang, S.-G., Yoon, H.-K., Park, Y.-M., Lee, S.-H., Moon, J.-H., Song, H.-M., Lee, E., & Kim, L. (2015). Association of CLOCK, ARNTL, and NPAS2 gene polymorphisms and seasonal variations in mood and behavior. *Chronobiology International, 32*(6), 785–791. https://doi.org/10.3109/07420528.2015.1049613

39. Kishi, T., Kitajima, T., Ikeda, M., Yamanouchi, Y., Kinoshita, Y., Kawashima, K., et al. (2008). Association analysis of nuclear receptor Rev-erb β gene (NR1D1) with mood disorders in the Japanese population. *Neurosci. Res.* 62, 211–215. doi: 10.1016/j.neures.2008.08.008
40. Kitamura, S., Hida, A., Watanabe, M., Enomoto, M., Aritake-Okada, S., Moriguchi, Y., Kamei, Y., & Mishima, K. (2010). Evening preference is related to the incidence of depressive states independent of sleep-wake conditions. *Chronobiology international, 27*(9-10), 1797–1812. https://doi.org/10.3109/07420528.2010.516705

41. Landgraf D, McCarthy MJ, Welsh DK. (2014). Circadian clock and stress interactions in the molecular biology of psychiatric disorders. *Current Psychiatry Reports 16*: 483

42. Lavebratt, C., Sjöholm, L. K., Partonen, T., Schalling, M., & Forsell, Y. (2010a). PER2 variation is associated with depression vulnerability. *American Journal of Medical Genetics, Part B: Neuropsychiatric Genetics, 153*(2), 570–581. https://doi.org/10.1002/ajmg.b.31021

43. Lavebratt, C., Sjöholm, L. K., Soronen, P., Paunio, T., Vawter, M. P., Bunney, W. E., Adolfsson, R., Forsell, Y., Wu, J. C., Kelsoe, J. R., Partonen, T., & Schalling, M. (2010b). CRY2 is associated with depression. *PLoS ONE, 5*(2). https://doi.org/10.1371/journal.pone.0009407

44. Levandovski, R., Dantas, G., Fernandes, L. C., Caumo, W., Torres, I., Roenneberg, T., Hidalgo, M. P., & Allebrandt, K. V. (2011). Depression scores associate with chronotype and social jetlag in a rural population. *Chronobiology international, 28*(9), 771–778. https://doi.org/10.3109/07420528.2011.602445

45. Lewy, A. J., Rough, J. N., Songer, J. B., Mishra, N., Yuhas, K., & Emens, J. S. (2007). The phase shift hypothesis for the circadian component of winter depression. *Dialogues in clinical neuroscience, 9*(3), 291–300. https://doi.org/10.31887/DCNS.2007.9.3/alewy

46. Li, J. Z., Bunney, B. G., Meng, F., Hagenauer, M. H., Walsh, D. M., Vawter, M. P., et al. (2013). Circadian patterns of gene expression in the human brain and disruption in major depressive disorder. Proc. Natl. Acad. Sci. U S A 110, 9950–9955. doi: 10.1073/pnas.1305814110

47. Liberman, A. R., Kwon, S. bin, Vu, H. T., Filipowicz, A., Ay, A., & Ingram, K. K. (2017). Circadian Clock Model Supports Molecular Link between PER3 and Human Anxiety. *Scientific Reports, 7*(1). https://doi.org/10.1038/s41598-017-07957-4

48. Liberman, A. R., Halitjaha, L., Ay, A., & Ingram, K. K. (2018). Modeling Strengthens Molecular Link between Circadian Polymorphisms and Major Mood Disorders. *Journal of Biological Rhythms, 33*(3), 318–336. https://doi.org/10.1177/0748730418764540

49. Margolin, A. A., Nemenman, I., Bassò, K., Wiggins, C., Stolovitzky, G., Favera, R. D., & Califano, A. (2006). ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BM C Bioinformatics 2006 7:1*, 7(1), 1–15. https://doi.org/10.1186/1471-2105-7-S1-S7

50. McCarthy, M. J., & Welsh, D. K. (2012). Cellular circadian clocks in mood disorders. *Journal of Biological Rhythms, 27*(5), 339–352. https://doi.org/10.1177/0748730412456367

51. McClung C. A. (2013). How might circadian rhythms control mood? Let me count the ways... *Biological psychiatry, 74*(4), 242–249. https://doi.org/10.1016/j.biopsych.2013.02.019

52. Merikanto, I., Lahti, T., Kronholm, E., Peltonen, M., Laatikainen, T., Vartiainen, E., Salomaa, V., & Partonen, T. (2013). Evening types are prone to depression. *Chronobiology international, 30*(5), 719–
53. Meyer, P. E., Lafitte, F., & Bontempi, G. (2008). MINET: An open source R/Bioconductor Package for Mutual Information based Network Inference. *BMC Bioinformatics, 9.*
http://www.biomedcentral.com/1471-2105/9/461

54. Mishima, K., Tozawa, T., Satoh, K., Saitoh, H., & Mishima, Y. (2005). The 3111T/C polymorphism of hClock is associated with evening preference and delayed sleep timing in a Japanese population sample. *American journal of medical genetics. Part B, Neuropsychiatric genetics: the official publication of the International Society of Psychiatric Genetics, 133B*(1), 101–104.
https://doi.org/10.1002/ajmg.b.30110

55. Nguyen, C., Murray, G., Anderson, S., Filipowicz, A., & Ingram, K. K. (2019). In vivo molecular chronotyping, circadian misalignment, and high rates of depression in young adults. *Journal of Affective Disorders, 250*, 425–431. https://doi.org/10.1016/j.jad.2019.03.050

56. Ozburn, A. R., Purohit, K., Parekh, P. K., Kaplan, G. N., Falcon, E., Mukherjee, S., Cates, H. M., & McClung, C. A. (2016). Functional implications of the CLO3111T/C single-nucleotide polymorphism. *Frontiers in Psychiatry, 7*(APR), 1–8. https://doi.org/10.3389/fpsyt.2016.00067

57. Ozburn, A. R., Kern, J., Parekh, P. K., Logan, R. W., Liu, Z., Falcon, E., Becker-Krail, D., Purohit, K., Edgar, N. M., Huang, Y., & McClung, C. A. (2017). NPAS2 Regulation of Anxiety-Like Behavior and GABAA Receptors. *Frontiers in Molecular Neuroscience, 0*, 360. https://doi.org/10.3389/FNMOL.2017.00360

58. Partonen, T. (2012). Clock gene variants in mood and anxiety disorders. In *Journal of Neural Transmission* (Vol. 119, Issue 10, pp. 1133–1145). https://doi.org/10.1007/s00702-012-0810-2

59. Quinlan, J. R., & Quinlan, J. R. (1986). Induction of Decision Trees. *MACH. LEARN, 1*, 81–106.
http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.167.3624

60. R Core Team. (2020). *R: A Language and Environment for Statistical Computing.* https://www.r-project.org/

61. Rgraphviz: Provides plotting capabilities for R graph objects version 2.34.0 from Bioconductor. (n.d.). Retrieved November 15, 2021, from https://rdrr.io/bioc/Rgraphviz/

62. Robnik-Sikonja, M., & Kononenko, I. (1997). An Adaptation of Relief for Attribute Estimation in Regression. *Proceedings of the Fourteenth International Conference on Machine Learning*, 296–304.

63. Roybal, K., Theobold, D., Graham, A., DiNieri, J. A., Russo, S. J., Krishnan, V., Chakravarty, S., Peevey, J., Oehrlein, N., Birnbaum, S., Vitatema, M. H., Orsulak, P., Takahashi, J. S., Nestler, E. J., Carlezon, W. A., & McClung, C. A. (2007). Mania-like behavior induced by disruption of CLOCK. *Proceedings of the National Academy of Sciences of the United States of America, 104*(15), 6406–6411. https://doi.org/10.1073/pnas.0609625104

64. Satyanarayanan, S. K., Chien, Y. C., Chang, J. P., Huang, S. Y., Guu, T. W., Su, H., & Su, K. P. (2020). Melatonergic agonist regulates circadian clock genes and peripheral inflammatory and neuroplasticity markers in patients with depression and anxiety. *Brain, behavior, and immunity, 85*, 142–151. https://doi.org/10.1016/j.bbi.2019.03.003
65. Savalli, G., Diao, W., Berger, S., Ronovsky, M., Partonen, T., & Pollak, D. D. (2015). Anhedonic behavior in cryptochrome 2-deficient mice is paralleled by altered diurnal patterns of amygdala gene expression. *Amino Acids, 47*(7), 1367. https://doi.org/10.1007/S00726-015-1968-3

66. Shi, S. Q., White, M. J., Borsetti, H. M., Pendergast, J. S., Hida, A., Ciarleglio, C. M., de Verteuil, P. A., Cadar, A. G., Cala, C., McMahon, D. G., Shelton, R. C., Williams, S. M., & Johnson, C. H. (2016). Molecular analyses of circadian gene variants reveal sex-dependent links between depression and clocks. *Translational Psychiatry, 6*(December 2015). https://doi.org/10.1038/tp.2016.9

67. Silva, A., Dos Santos, M. J., Góes Gitaí, D. L., de Miranda Coelho, J., & de Andrade, T. G. (2020). Depression and anxiety symptoms correlate with diurnal preference, sleep habits, and *PER3* VNTR polymorphism (rs57875989) in a non-clinical sample. *Journal of affective disorders, 277*, 260–270.

68. Soria, V., Martínez-Amorós, È., Escaramís, G. *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. *Neuropsychopharmacol 35*, 1279–1289 (2010). https://doi.org/10.1038/npp.2009.230

69. Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press.

70. Statistics, L. B., & Breiman, L. (2001). Random Forests. *Machine Learning*, 5–32.

71. Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. *International journal of epidemiology, 43*(2), 476–493. https://doi.org/10.1093/ije/dyu038

72. Viena, T.D., Gobin, C.M., Fins, A.I., Craddock, T., Tartar, A. and Tartar, J., 2016. A *PER3* Polymorphism Interacts with Sleep Duration to Influence Transient Mood States in Women. *Journal of Circadian Rhythms, 14*(1), p.3. DOI: http://doi.org/10.5334/jcr.135

73. Walker, W. H., Walton, J. C., DeVries, A. C., & Nelson, R. J. (2020a). Circadian rhythm disruption and mental health. *Translational Psychiatry, 10*(1). https://doi.org/10.1038/S41398-020-0694-0

74. Yang, H. H., & Moody, J. E. (1999). *Feature Selection Based on Joint Mutual Information.*

**Figures**
**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.
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. **B&C)** 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.

**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).

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Suppl.Table1Anxiety.docx
- SupplTable2.docx
- SupplementaryFigure1.docx