SOF Analysis Plan Submission Form

Date: 12/21/2010

Investigator’s Name: Daniel S. Evans

Clinical Center: SFCC

Sponsor (if not a SOF investigator): Greg Tranah

Telephone: 415-600-7435 e-mail: devans@sfcc-cpmc.net

Other investigators who will be working on this analysis: Neeta Parimi, Katie Stone, Terri Blackwell, Sonia Ancoli-Israel, Steve Cummings, Greg Tranah, and others expressing interest.

Analysis Plan Title: Association between candidate circadian and melatonin gene variants and actigraphic measures of sleep and activity rhythms.

Data sets to be used: Visit 8 actigraphy data, SNP genotypes from circadian and melatonin gene data available from Greg Tranah’s ancillary study.

Primary variables to be used in the analysis: SNP genotypes from circadian and melatonin gene dataset, sleep latency, wake after sleep onset, sleep efficiency, total sleep time, bed time, wake time, acrophase, amplitude, mesor, F-statistic, age, clinic site, and multidimensional scaling estimates of population stratification.

Do you plan to submit an abstract based on these results? ☐ YES ☒ NO
If YES, when is the abstract due?

Who will perform the analyses?

☒ Coordinating Center
☐ Other local analyst, please specify:

Please attach a 1-2 page description of your analysis plan. Please include the following:

1) Short background/rationale for addressing the research question
2) Brief description of statistical methods
3) Mock tables

E-mail this completed form (as an attachment) to Dana Kriesel (dkriesel@sfcc-cpmc.net).
SOF Analysis Plan #674

**SOF Analysis Proposal**

Association between candidate circadian and melatonin gene variants and actigraphic measures of sleep and activity rhythms.

Dan Evans  
Greg Tranah  
Katie Stone  
Terri Blackwell  
Sonia Ancoli-Israel  
Steve Cummings  
Others…

**Introduction**

Sleep patterns change with age. Total sleep time typically decreases and is more fragmented in the elderly with less time spent in slow wave sleep \(^1\) as compared to younger adults. Increased difficulty initiating and maintaining sleep leads to increased daytime sleepiness and more frequent daytime naps. In one large epidemiologic study of elderly subjects, over half of the subjects reported at least one sleep disturbance, including trouble falling asleep, frequent arousals from sleep, early morning awakenings, daytime naps and fatigue \(^2\). The deterioration of solid sleep at night and sustained wake in the day is thought to be due in part to weakening of the circadian system, possibly related to loss of suprachiasmatic nucleus (SCN) neurons or decreases in nocturnal melatonin secretion. Specific sleep disorders, especially sleep apnea and periodic leg movements, are also more common in the elderly \(^3,4\). There is increasing evidence that sleep disturbances may actually accelerate the aging process, increasing the risk for a variety of diseases and conditions such as cognitive impairments \(^5-11\), diabetes \(^12-15\), cardiovascular disease \(^16-21\) and premature death \(^19, 22-28\).

The role of circadian genes in the study of sleep-wake regulation is currently being elucidated through several animal models; however, little is known about the effect of circadian gene variants on human sleep function, especially in the older age groups where sleep problems are most severe. In addition, a variety of age-related diseases and conditions have been associated with poor sleep. It is uncertain whether predisposing genetic factors that are associated with sleep characteristics may also influence longevity and risk for associated age-related conditions in the elderly. Genetic methods have the potential to uncover some of the causes of sleep disorders and lead to specific treatments. Furthermore, increasing evidence suggests that aging and longevity are regulated at least in part by some components of the circadian clock-signaling pathway.

We propose to analyze common genetic variants in the >20 circadian rhythm (*Clock*, *Npas2*, *Bmal1*, *Per1*, *Per2*, *Per3*, *Cry1*, *Cry2*, *Dec1*, *Dec2*, *CK1ε*, *CK1δ*, *Rev-ERBa*, *RORα*, *RXRa*, *RARα*, *TIM*, *DBP*, *GSK3β*, *MTNR1a* and *MTNR1b*) and 6 melatonin-related genes (*TPH1*, *AADC*, *AA-NAT*, *HIOMT/ASMT*) for associations with sleep and activity rhythm characteristics measured by actigraphy. Data for ~620 single nucleotide polymorphisms (SNPs) were generated using the Illumina BeadArray platform as part of Greg Tranah’s ancillary study.
Research question:  
Among older women, what is the relationship between candidate circadian and melatonin genes and actigraphic measures of sleep performance and activity rhythms?

Data:  
Visit 8 actigraphy data: Total sleep time, Sleep Efficiency, Sleep Latency, Wake after Sleep Onset, Bed time, Wake time, Amplitude, Mesor, pseudo-F statistic, and Acrophase.  
Circadian and melatonin gene data.

Analysis:  
For each SNP, we will use model-free genotypic coding to estimate the effect of heterozygote and homozygote genotypes. The significance of both parameters for each SNP will be assessed using a 2 degree of freedom (2df) anova. Among SNPs with a 2df $p$-value less than 0.05, the two separate parameters of the genotypic coding will be examined to determine the appropriate inheritance model. Possible inheritance models include additive, dominant and recessive. All SNPs will be tested for deviation from Hardy Weinberg equilibrium. Ordinary least-squares linear regression models will be used for analyses involving genotype associations with continuous outcome measures. Given the complexity of multi-locus haplotype data, we plan to use 3 different methods to correct for multiple hypothesis testing. First, we propose to use a Bonferroni correction that has been adapted for linkage among SNPs. Bonferroni correction is appropriate for independent tests, but can be overly conservative in haplotype-based analyses and result in a loss of power since alleles at neighboring loci will be associated via LD. Thus, we will estimate the effective number of independent tests corrected for the LD structure, and use this estimate for a standard Bonferroni correction. Second, we will calculate the expected false discovery rate (FDR), defined as the fraction of false rejections among those hypotheses rejected. The FDR procedure controls the increased error rate from multiple testing while maintaining the ability to detect real differences. For assessing associations between SNPs in ~50 circadian and melatonin gene SNPs we will assume a type 1 error rate of 0.05 prior to multiple comparison correction. Third, we will perform a permutation procedure to estimate the empirical $p$-value for each SNP. This will be done by randomly shuffling the phenotype variable while keeping the SNP data intact. For each random shuffle, the minimum $p$-value among all SNPs will be stored, and this process will be repeated at least 10,000 times, forming an empirical $p$-value distribution under the null hypothesis of no association. The observed $p$-values will be evaluated on this distribution by calculating the percentage of $p$-values under the null that are more extreme than the observed $p$-value. This has the effect of preserving the LD structure within the SNP data so that multiple testing is corrected and the LD structure of the SNP data is taken into account.

Assessment of stratification has been carried out using the program Structure, which is a model-based clustering program that parses the participants into sub-populations and assesses if there are distinct populations or admixed individuals. To account for any population sub-structure, a principal components method of analysis that corrects for population substructure will be employed. All analyses will adjust for age and study site in addition to components of population sub-structure.
A meta-analysis will be performed with results from the SOF and MrOS study using a fixed effect model with inverse variance weighting. Heterogeneity between studies will be assessed using the $I^2$ statistic and the $p$-value from the Q-test.

**MOCK TABLE 1.**

**ADDITIVE REGRESSION MODEL FOR PIM:MEAN SLEEP EFFICIENCY IN BED**

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP RS number</th>
<th>Regression P-value</th>
<th>False Discovery Rate</th>
<th>Beta estimate and 95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLOCK</td>
<td>rs1019731</td>
<td>0.0185</td>
<td>0.6151</td>
<td>2.400(0.404,4.395)</td>
</tr>
<tr>
<td>BMAL1</td>
<td>rs1137101</td>
<td>0.0058</td>
<td>0.4317</td>
<td>2.066(0.602,3.530)</td>
</tr>
<tr>
<td>CRY</td>
<td>rs1152003</td>
<td>0.0032</td>
<td>0.3986</td>
<td>-2.25(-3.737,-0.757)</td>
</tr>
<tr>
<td>PER</td>
<td>rs11575194</td>
<td>0.0274</td>
<td>0.6151</td>
<td>-3.49(-6.584,-0.389)</td>
</tr>
<tr>
<td>AANAT</td>
<td>rs11804091</td>
<td>0.0218</td>
<td>0.6151</td>
<td>-2.33(-4.325,-0.341)</td>
</tr>
<tr>
<td>TPH1</td>
<td>rs1319868</td>
<td>0.0026</td>
<td>0.3986</td>
<td>-3.60(-5.945,-1.262)</td>
</tr>
<tr>
<td>CLOCK</td>
<td>rs1801282</td>
<td>0.0298</td>
<td>0.6151</td>
<td>2.294(0.226,4.362)</td>
</tr>
<tr>
<td>BMAL1</td>
<td>rs2197423</td>
<td>0.0228</td>
<td>0.6151</td>
<td>2.393(0.334,4.452)</td>
</tr>
<tr>
<td>CRY</td>
<td>rs2684792</td>
<td>0.0447</td>
<td>0.6969</td>
<td>1.563(0.037,3.090)</td>
</tr>
<tr>
<td>PER</td>
<td>rs2946385</td>
<td>0.0447</td>
<td>0.6969</td>
<td>1.632(0.039,3.224)</td>
</tr>
<tr>
<td>AANAT</td>
<td>rs3020368</td>
<td>0.0279</td>
<td>0.6151</td>
<td>-2.55(-4.831,-0.279)</td>
</tr>
</tbody>
</table>

**REFERENCES:**


