

Characterization of the genetic relationship between the domains of sleep and circadian-related behaviors with substance use phenotypes

Alexander S. Hatoum * ¹

Evan A. Winiger * ²

Claire L. Morrison ^{3,4}

Emma C. Johnson ¹

Arpana Agrawal ¹

Affiliations:

1. Washington University School of Medicine, Department of Psychiatry, Saint Louis, USA
2. Department of Psychiatry, University of Colorado School of Medicine, Denver, USA
3. Institute for Behavioral Genetics, University of Colorado, Boulder, Colorado
4. Department of Psychology and Neuroscience, University of Colorado, Boulder, Colorado

* Authors contributed equally to this manuscript.

Acknowledgements and Funding: ASH receives support from DA007261-17. EAW receives support from T32 MH015442 and DA017637. CLM receives support from MH016880. ECJ receives support from K01DA051759.

Corresponding Author: Dr. Alexander S. Hatoum, ashatoum@wustl.edu

Abstract

Sleep problems and substance use frequently cooccur. While substance use can often manifest as specific sleep deficits, genetic pleiotropy could also explain part of the relationship between sleep and substance use. Here we assess the genetic overlap between substance use behaviors and both sleep and circadian-related activity measures by deriving genetic clusters between these domains and testing processes of causality vs. horizontal pleiotropy using the largest publicly available genome-wide summary statistics of substance use behaviors (N= 79,729 - 632,802) and sleep/activity phenotypes/endophenotypes to date (N=85,502 - 449,734). We found 31 genetic correlations between substance use and sleep/activity measures after Bonferroni correction. Two specific genetic clusters explained our patterns of overlap. Genes associated with tobacco use severity (age of first regular tobacco use and smoking cessation) share overlap with elements of sleep health (sleep duration, sleep efficiency, and chronotype). Substance consumption (drinks per day and cigarettes per day) and problematic substance use behaviors (cannabis use disorder, opioid use disorder, and problematic alcohol use) clustered strongly with problematic measures of sleep (insomnia, self-reported short sleep duration, increased number of sleep episodes, increased sleep duration variability, diurnal inactivity) as well as measures of circadian-related activity (L5, M10, and sleep midpoint). Latent causal variable analyses determined that horizontal pleiotropy (rather than causality) underlies a majority of the associations between substance use and sleep/circadian related measures, except one plausible genetically causal relationship for opioid use disorder on self-reported long sleep duration. Results indeed show significant genetic overlap between substance use and sleep/circadian-related activity measures.

Introduction

Sleep disturbances are one of the most common complaints in substance use treatment,^{1,2} and there is substantial comorbidity between substance use disorders (SUD) and sleep disorders^{3,4}. Attempting to quit using substances and the consequent cravings and withdrawal are often associated with sleep disturbances and these sleep deficits can in turn be reciprocally linked to substance use relapse⁵. For instance, sleep difficulties and insomnia are common clinical features of withdrawal from alcohol, nicotine, and cannabis⁶. Further complicating this relationship, common substances, such as alcohol and cannabis, are often used to self-medicate sleep issues, with evidence that cannabis⁷⁻⁹ and alcohol⁸⁻¹¹ are frequently considered sleep aids despite strong cross-sectional evidence demonstrating positive correlations between increased use of many common substances and sleep issues⁵. Relatedly, tobacco is often used as a stress and tension reliever^{12,13} but tobacco use before sleep is also associated with sleep disturbances¹⁴.

Research suggests that circadian mechanisms contribute to the association between sleep issues and substance use/abuse, and that substance use may also impact circadian rhythms, thus further disturbing sleep¹⁵. While circadian rhythm is a multi-faceted biological construct, it is often operationally defined as chronotype, a preference between either evening or morning activity, wake-up, and bedtime¹⁶. An evening chronotype has been associated with higher alcohol¹⁷⁻¹⁹, tobacco^{18,20,21}, and cannabis use^{19,20,22}. In addition to the standard self-report measure of chronotype, activity-related proxies (measured via actigraphy or accelerometer) can function as objective circadian-related measures²³⁻²⁶. Research on the relationship between circadian-related activity measures and substance use is scarce²⁷ and is yet to be fully understood.

Sleep and substance use behaviors seem to interact dynamically. There is evidence of a bi-directional relationship between the domains of substance use and sleep deficits particularly during development there is indication that early sleep problems and an evening chronotype predict later substance use^{19,28–35} and both early substance use and problematic substance use predicts later sleep issues^{28,29,36–42}. Reconciling these cross-sectional associations and bidirectional patterns, it is likely that a third variable, such as genetic pleiotropy, is driving trait-like manifestations of sleep and substance use comorbidity.

Here, we focus on trait like (rather than state-like) overlap in sleep and substance use by utilizing statistical genetics methods. Prior genetic studies implicate shared genetic influences on sleep and substance use and misuse as a likely contributor to their comorbidity. Twin studies focused on sleep/circadian-related outcomes and substance use are scarce but a few have found genetic correlations between regular cannabis use and both short sleep duration and insomnia^{41,42} as well as genetic correlations between an evening chronotype with alcohol quantity and increased binge drinking⁴³. Modern genome-wide association studies (GWASs) studies have found genetic correlations amongst subjective sleep deficits and common substance use behaviors including tobacco behaviors such as smoking initiation, smoking cessation, and cigarettes per day^{44,45}, alcohol behaviors via scores on The Alcohol Use Disorders Identification Test (AUDIT)⁴⁶, Opioid Use Disorder⁴⁷, and cannabis behaviors such as lifetime cannabis use and Cannabis Use Disorder⁴⁸. GWASs focused on chronotype/circadian-related activities and substance use are sparse. Genetic correlations have been found between an evening chronotype and lifetime cannabis use⁴⁸, but the relationship between the multitude of available substance use and both chronotype and circadian-related activity measures has yet to be explored. Still, the current limited genetically informed results imply that the genes that could be contributing to

sleep deficits and circadian-related factors such as chronotype might contribute to substance use and misuse as well or vice versa.

Genetic studies can also be used to inform causality. A prior study explored the pairwise associations between sleep and substance use disorders using Mendelian randomization, finding that insomnia had a potential positive causal influence on smoking, alcohol dependence, and cannabis initiation, while smoking initiation may be causal for insomnia⁴⁹. However, the effects of insomnia on both alcohol dependence and cannabis initiation were partially driven by pleiotropic SNPs (rather than direct causality), suggesting caution in inferring these results as causal and the possibility of a shared genetic liability. Thus, a perspective incorporating pleiotropy is needed.

Finally, there has yet to be a comprehensive genomic study to include multitude of alcohol, tobacco, cannabis, and opioid use behaviors as well as both a collection of subjective and objective/endophenotype sleep/circadian-related measures. Sleep itself is a highly heterogeneous set of behaviors, encompassing many subdomains that relate differentially. Substance use and use disorder may interfere with sleep generally, or with sleep efficiency, duration, circadian rhythm, or the sleep cycle to produce sleep problems. Further, different substances may show unique relationships with any one domain of sleep specifically. Therefore, it is important to explore subdomains of substance use, misuse, and sleep behaviors in order to localize the causes of such deficits.

No study has comprehensively looked across the relationships of substance use behaviors and sleep/activity measures with the goal of defining genetic clusters of sleep/circadian-related behaviors and substance use traits. Therefore, the goals of the current study were to:

1) Investigate the shared genetics between a multitude of substance use and sleep/activity measures. 2) Determine if genetic causality or genetic pleiotropy are responsible for the genetic relationships discovered. 3) Analyze genetic clusters amongst the domains of substance use and sleep/activity.

Methods

Measures

Subjective and Objective Sleep and Circadian-related Measures

Our analyses used summary statistics from several large scale GWAS focused on self-report sleep phenotypes and objective accelerometer derived sleep/activity phenotypes/endophenotypes. Self-report sleep phenotypes included insomnia, chronotype, self-report sleep duration, self-report short sleep duration, and self-report long sleep duration⁵⁰⁻⁵³. Objective accelerometer derived sleep/circadian-related activity phenotypes/endophenotypes included sleep duration, standard deviation of sleep duration (a measure of sleep variability), sleep efficiency, number of sleep episodes, diurnal inactivity (inactive states such as napping and wakeful rest), sleep midpoint (a proxy of chronotype^{54,55}), least active 5 hours of the day (L5 timing, indication of a preference for going to bed earlier or later in the day²³), and most active 10 hours of the day (M10 timing, indication of if a person is most active earlier or later in the day²³)⁵⁶. Diurnal inactivity can be conceptualized as a sleep problem due to increased naps/rest/inactivity that are a compensatory behavior for non restful sleep⁵⁷. L5, M10, and sleep midpoint will be conceptualized as activity measures that are circadian-related proxy measures. **Table 1** details sleep/activity trait in terms of type of measure, measure construct, coding, and discovery sample.

Substance Use Measures

Our analysis used summary statistics from several of the largest GWAS of substance use behaviors. Alcohol behaviors included alcoholic drinks per week (DPW)⁵⁸ and problematic alcohol use (PAU)⁴³. Tobacco behaviors included lifetime tobacco use, age of first becoming a regular smoker, smoking cessation, and cigarettes per day (CPD)⁵⁸. Cannabis behavior included lifetime cannabis use⁵⁹ and Cannabis Use Disorder (CUD)⁶⁰. Lastly, we included Opioid Use Disorder (OUD)⁴⁷. All trait discovery cohorts were of European ancestry. **Table 1** details each substance use trait in terms of type of measure, measure construct, coding, and discovery sample.

Analyses

Linkage Disequilibrium Score Regression

We used Linkage Disequilibrium Score Regression (LDSC)⁶¹ to estimate genetic correlations between traits. GWAS summary statistic SNPs were removed if they had $MAF > 0.01$ and $INFO > 0.70$. SNPs, duplicated rs numbers, were multi-allelic variants, were strand ambiguous, had deletions/insertions or had low Ns. Alleles were merged with the Hap Map 3⁶² reference panel (major histone complex removed). Beta weights and linkage disequilibrium (LD) were pre-generated from 1000 Genomes European GWAS data included in the LDSC software download. LDSC regresses Chi-square statistics from the summary stats of GWAS on LD scores of the trait of interest⁶¹. LD scores for each SNP are calculated via the sum of the variance explained by LD of that SNP with other SNPs⁶³. Genetic correlations were estimated using overlapping SNPs from filtered summary statistic files provided from GWAS summary statistics. LDSC accounts for possible sample overlap and additional sources of confounding (e.g., population stratification). We estimated a pair-wise genetic correlation matrix that included all sleep/activity and substance use measures (figure 7). Due to the large number of genetic correlations estimated, we utilized Bonferroni correction⁶⁴ to adjust for potential false positives.

K-means Clustering of correlations

K-means clustering is an unsupervised machine learning clustering technique that uses a centroid or distance-based algorithm to assign correlations to a cluster of a predefined number^{65,66}. After assigning the K number of clusters, the algorithm shuffles the data to clusters and assigns them to initial random centroids. It then determines the sum of squares (or distance) between each data point and the initial centroids and does a series of reassignments to the centroids until the algorithm is finished with the appropriate clusters, attempting to make data in the clusters similar while making each individual cluster separate from the others. By comparing the distance (sum of squares) within clusters across different number of clusters (i.e. 1 vs. 2 and 2 vs. 3) we can estimate a silhouette coefficient that shows the best number of clusters to account for patterns in the data, i.e. the cluster solution that allows for the smallest sum of squares across clusters.

An advantage of K-means over other clustering algorithms is that clustering is done at the variable level with the correlation matrix (unlike mixture modeling or DBSCAN) in a hypothesis free format (unlike GenomicSEM). The latter is necessary for this analysis as large GWAS studies do not share individual data. For the current analyses, the genetic correlation matrix was read into the K-Means algorithm. The silhouette coefficient determined how many centers were needed to keep each substance use and sleep measure closest together, compared to other potential cluster solutions. K-means clustering was conducted in R using the R packages “cluster”⁶⁷ and “factoextra”⁶⁸.

Latent Causal Variable Analysis

To examine evidence for genetic causality between sleep and substance use/misuse phenotypes, we used latent causal variable analysis (LCV)⁶⁹ on genetic correlations that

survived Bonferroni correction. LCV allows genetic correlations between two traits to be mediated by a latent variable with a causal effect on each trait. This model was designed to account for genetic pleiotropy by partitioning the genetic correlation into pleiotropy vs partial causality. This is done using the 4th order moments from LD Score Regression⁶⁹. Using this model, causality is implied when trait one is strongly correlated with the causal latent variable compared to the second trait, suggesting that part of the genetic component of trait one is causal for the second trait in the relationship. If trait one is perfectly genetically correlated with the latent variable, it can be considered fully genetically causal. The extent to which the latent causal variable causes trait 1 versus trait 2 is expressed as a ratio, referred to as the genetic causality proportion (gcp). The gcp is an estimate of the degree to which each trait is correlated with the latent genetic variable with a score that can range from 0 (reflecting no genetic causality) to 1/-1 (signifying full genetic causality). For instance, a gcp of 0.70 would suggest 70% of SNP effect sizes are consistent with trait 1 causing trait 2. It is worth noting that LCV estimates the direction of causality in terms of the order of variables, if the gcp is positive, trait 1 is genetically causal for trait 2. If the gcp is negative, trait 2 is genetically causal for trait 1. For the current analyses, the sleep/activity measure is trait 1 and the substance use measure is trait 2.

While LCV is similar to other methods of determining genetic causality in that it utilizes SNPs to derive instrumental variables (like traditional Mendelian randomization), LCV has advantages over other methods of estimating genetic causality. First, sample overlap is accounted for by the LDSC intercept. Second, the model produces a genetic causality proportion, that is the proportion of genetic effects that are consistent with a model of genetic causality, allowing us to account for partial overlap. Third, the gcp is robust to pleiotropy, which is accounted for using 4th order moments from LDSC. Fourth, the pleiotropic effects across the entire genome are

accounted for, rather than only at a few SNPs⁶⁹. Indeed, simulation designs shows that the method outperforms Mendelian randomization⁶⁹ when the genetic correlation between traits is non-zero. Considering that many of our samples overlap and that we want to account for pleiotropy, LCV is ideal over MR in this analysis.

Results

Genetic Correlations Across Sleep and Substance Use

Figure 1 displays the genetic correlation matrix between all substance use behaviors and sleep/activity measures. We found 31 significant pair-wise correlations across domains that survived Bonferroni correction (**Table 2**), with several trends suggesting strong genetic relationships between substance use behaviors and sleep/activity measures. In all cases, greater vulnerability for substance use was associated with more sleep problems. Measures of both subjective self-reported sleep duration (long, full, and short) and objective accelerometer derived sleep duration had genetic correlations with numerous substance use traits including lifetime tobacco use, age of first becoming a regular smoker, CPD, smoking cessation, PAU, lifetime cannabis use, CUD, and OUD (absolute values of rGs between 0.11 - 0.32). Increased diurnal inactivity was genetically correlated with lifetime tobacco use, age of initiation of smoking, CPD, and DPW (Absolute values of rGs between 0.12 – 0.18), Insomnia was also genetically correlated with numerous substance use traits including lifetime tobacco use, age of initiation of regular smoking, CPD, smoking cessation, DPW, PAU, CUD, and OUD (absolute values of rGs between 0.11-0.36).

Only a few of the genetic correlations involving the activity based circadian-related endophenotype measures survived Bonferroni corrections including 1) sleep midpoint with lifetime cannabis use (rG = 0.26), and DPW (rG = 0.24) and 2) M10 with DPW (rG = 0.24).

Focusing on measures of chronotype, in addition to the circadian-related sleep midpoint relationships mentioned above, significant genetic correlations were found between self-reported chronotype and lifetime cannabis use, lifetime tobacco use, and DPW (Absolute values of rGs between 0.08 – 0.25).

K-means clustering within and between sleep and substance use domains

A silhouette coefficient determined that 2 clusters were optimal to explain the overlap between sleep/circadian-related activity and substance use dimensions (**Figure 2**). **Figure 3** displays our optimal cluster solution. We refer to the first cluster as the “tobacco use severity” cluster, which grouped the substance use behaviors of age of initiation of regular smoking and smoking cessation with elements of sleep health such as self-report long sleep duration, self-report sleep duration, accelerometer derived sleep duration, sleep efficiency, and self-report chronotype. The traits most central to this cluster included sleep efficiency (sleep duration divided by the time between the start and end of the first and last nocturnal inactivity period) and self-reported sleep duration. The second cluster reflected “substance use and use disorders” and contained common substance use behaviors (lifetime tobacco use and lifetime cannabis use), consumption behaviors (CPD and DPW) as well as problematic substance use behaviors (CUD, OUD, and PAU) with measures of sleep difficulties (insomnia, self-report short sleep duration, increased number of sleep episodes, increased standard deviation of accelerometer derived sleep duration, and diurnal inactivity) and circadian-related activity measures (L5, M10, and sleep midpoint). The traits most central to this cluster included DPW, PAU, and the standard deviation of accelerometer derived sleep duration.

Latent Causal Variable Analysis

We used latent causal variable analysis between each of the 31 substance use and

sleep/activity measure genetic correlations that survived Bonferroni correction. Of the 31 LCV models, one model survived further Bonferroni correction among the LCV models ($.05/31=.0016$), with evidence of OUD being genetically causal for self-report long sleep duration ($gcp = 0.50$, corrected $p = 0.01$). Several models reached nominally significant p -values before Bonferroni correction, including CUD with subjective short sleep duration ($gcp = -0.23$, nominal $p = 0.04$), CUD with insomnia ($gcp = -0.29$, nominal $p = 0.05$), and OUD with insomnia ($gcp = -0.35$, nominal $p = 0.05$). A majority of the models lacked significance before correction, indicating a lack of support for causality between substance use/misuse and sleep measures (**Table 3**). Thus, there was scarce evidence of genetic causality once pleiotropy, polygenicity, and sample overlap are appropriately accounted for.

Discussion

Using summary data from the largest publicly available GWAS of both sleep/circadian-related activity and substance use measures to date, we found 31 significant genetic correlations between traits in these domains that survived Bonferroni correction. Clustering analysis uncovered two principal genetic clusters: 1) the tobacco use severity cluster and 2) the substance use and use disorders cluster. Latent causal variable analyses confirmed that the associations between sleep/activity measures and substance use behaviors were driven primarily by common or shared genetic influences, with one exception, a model which implied that OUD may plausibly exert a genetically causal influence on self-reported long sleep duration. The results herein suggest a strong pleiotropic genetic relationship between the domains of substance use and sleep/circadian-related activity measures.

We found significant genetic correlations that align with prior findings, including genetic relationships between subjective sleep measures and substance use behaviors involving cannabis

⁴⁸, tobacco^{44,45}, alcohol⁴⁶, and opioids⁴⁷. We also found novel relationships between subjective measures of sleep and substance use, such as OUD with subjective long sleep duration. Of primary interest are the genetic correlations involving substance use and objective measures of sleep and circadian-related activity, most of which have not been reported previously. For example, genetic associations between lifetime cannabis use and objectively measured later sleep mid-point and decreased accelerometer derived sleep duration are the first reports of a genetic relationship between cannabis use and objective measures in this domain of research. Later sleep midpoint is an objective proxy for an evening chronotype^{26,54,55}, and these results corroborate prior subjective findings regarding a genetic relationship between lifetime cannabis use and self-report evening chronotype⁴⁸. These findings suggest a shared genetic relationship of lifetime cannabis use with circadian-related measures, implying that people who have more evening or nighttime activity are more likely to have used cannabis due to a shared genetic vulnerability.

While a prior result of objective sleep duration variation and lifetime tobacco use were not replicated⁵⁶, this is the first report of significant genetic correlations between numerous tobacco behaviors and subjective sleep measures. These results imply shared genetic underpinnings between a range of increased tobacco behaviors (lifetime use, earlier age of first becoming a regular smoker, increased cigarettes per day, and smoking cessation) with both problematic objective sleep duration (short and long sleep duration) and increased diurnal inactivity during the day. These relationships suggest the genetics responsible for increased tobacco use are also responsible for improper sleep duration and the potential sleep compensation/rest that can accompany. Genetic correlations between drinks consumed (DPW) and sleep midpoint, diurnal inactivity, and M10 are the first report of a genetic relationship between alcohol use and objective measures of sleep and circadian-related activity. Interestingly,

increased DPW being genetically correlated with less diurnal inactivity, later midpoint (proxy for evening chronotype), and increased M10 (more active later in the day) imply that genetic predisposition to increased drinking might also be associated with increased activity throughout the day and being more active later in the day or evening. Thus, these differential associations may be of interest in developing sleep-assisted interventions, such that adjusting chronotype could be a novel intervention for issues related to alcohol use.

These genetic correlations formed two distinguishable and well-delineated clusters. The first cluster suggests that genes associated with aspects of tobacco use severity (a younger age of regular smoking initiation and smoking cessation) share genetic overlap with elements of sleep health such as sleep duration (self-report long sleep duration, shorter self-report sleep duration, and shorter accelerometer derived sleep duration), sleep efficiency, and self-report chronotype. Sleep health is an established research term that provides a frame of reference for the field⁷⁰ and includes sleep elements such as sleep duration, efficiency, and chronotype. These sleep health elements clustering together imply a genetic relationship between indicators of heavy tobacco use and sleep health. Tobacco use is highly comorbid with insufficient sleep⁷¹ and it is speculated that the relationship could be centered on potential nightly withdrawals, the strong stimulant effects of nicotine, and an increased prevalence of sleep disordered breathing in comparison to nonsmokers⁷². Exploration of sleep deficits in tobacco withdrawal may benefit from understanding shared genetic vulnerabilities to poor sleep health in particular.

The second cluster found that substance use behaviors, such as consumption (lifetime tobacco use, lifetime cannabis use, DPW, and CPD) and problematic use (CUD, OUD, and PAU) clustered strongly with measures of problematic sleep (insomnia, self-report short sleep duration, increased number of sleep episodes, increased standard deviation of accelerometer

derived sleep duration, and diurnal inactivity) as well as measures of circadian-related activity (L5, M10, and sleep midpoint). Insomnia deficits are thought to be part of the “dark side ” of substance use ⁷³ that reflects negative reinforcement-related drug intake, and seems to manifest specifically with diagnosis (and likely heavy use) based on this cluster. Interestingly, recent drug targets for these negative affect/withdrawal-related symptomology of addiction have been shown to improve insomnia symptoms as well. For instance, Acamprosate that is used to treat alcohol use disorder targets withdrawal symptoms and thus improves sleep symptoms during treatment ⁷⁴ .

Though some previous work argues the association between insomnia and substance use is causal (via Mendelian randomization) ⁴⁹, based on results herein, these findings are likely confounded by pleiotropy. Our approach accounted for pleiotropy and did not confirm the previous findings, although we did find one novel association. Specifically, our LCV analyses provided novel evidence that liability to OUD may be genetically causal for self-report long sleep duration, even after accounting for pleiotropy between these traits. Over 80% of individuals with OUD report poor sleep quality and sleep problems which likely impedes opioid maintenance and other pharmacological treatments ^{75,76}. Opioid therapy for pain and related conditions has been well-documented to disrupt sleep and due to its respiratory depressant effects, exacerbate risk for sleep-disordered breathing ⁷⁷. While this is the first finding of a significant relationship between OUD and long sleep duration, long sleep duration has been found to be phenotypically correlated with depression, antidepressant use, benzodiazepine use, and heavy drinking ⁷⁸, as well as genetically correlated with depression ⁵². Thus, the causal genetic relationship demonstrated could reflect more of an underpinning of the genetics associated with a general maladaptive behavioral profile in comparison to specific opiate use

genetics. While three other models were nominally significant before correcting for multiple tests, the lack of significance in these models ultimately implies pleiotropic inferences as the more supportive explanation of the associations. Thus, it is worth noting that pleiotropy may underlie sleep problems in substance use, and that populations entering substance use treatment are also likely to be at greater vulnerability for sleep deficits.

Limitations

There are several limitations of this study that would influence LDSC and LCV analyses and the generalizability of our findings. First, GWAS results are comprised of mostly common variants; therefore, no rare variation would be included in our analyses, but rare variants may be an additional source of genetic overlap or provide stronger instruments for causality. Second, the GWAS summary statistics used in our analyses assume that the genomic liability to a trait is a good instrument for manifesting the phenotype (for example, genomic liability for insomnia is a good predictor of having insomnia), but this may not necessarily be true for all traits. Third, LCV and LDSC are best when used in one ancestral population, and all GWAS traits were of European ancestry. There is a need to include more racially diverse and multi-racial samples in genetics research. Finally, the objectively measured sleep/activity phenotypes/endophenotypes (the accelerometer derived traits) had much smaller samples than the self-report sleep measures; this difference in statistical power may mean that genetic correlations with the accelerometer traits were less likely to be statistically significant than those for self-report sleep measures, and therefore less likely to be carried forward in the causal analyses.

Conclusions

While substances such as cannabis and alcohol are often used as sleep aids, individuals with substance use disorders also struggle with sleep difficulties, and poor sleep complicates

pathways to sustained remission from substance use disorders. Our study documents the role of shared genetic influences on substance use disorders and both sleep and circadian-related measures indicating specific domains of overlap that may be used to discover mechanisms of intervention. For opioid use disorder in particular, mechanisms of association likely extend beyond pleiotropy into potential causality. Together, these results imply a strong shared genetic relationship between the domains of common substance use behaviors and sleep traits.

References

1. Teplin D, Raz B, Daiter J, Varenbut M, Tyrrell M. Screening for substance use patterns among patients referred for a variety of sleep complaints. *Am J Drug Alcohol Abuse*. 2006. doi:10.1080/00952990500328695
2. Hasler BP, Martin CS, Wood DS, Rosario B, Clark DB. A Longitudinal Study of Insomnia and Other Sleep Complaints in Adolescents With and Without Alcohol Use Disorders. *Alcohol Clin Exp Res*. 2014. doi:10.1111/acer.12474
3. Chakravorty S, Vandrey RG, He S, Stein MD. Sleep Management Among Patients with Substance Use Disorders. *Med Clin North Am*. 2018. doi:10.1016/j.mcna.2018.02.012
4. Diep C, Tian C, Vachhani K, et al. Recent cannabis use and nightly sleep duration in adults: a population analysis of the NHANES from 2005 to 2018. *Reg Anesth Pain Med*. December 2021:rapm-2021-103161. doi:10.1136/RAPM-2021-103161
5. Angarita GA, Emadi N, Hodges S, Morgan PT. Sleep abnormalities associated with alcohol, cannabis, cocaine, and opiate use: A comprehensive review. *Addict Sci Clin Pract*. 2016. doi:10.1186/s13722-016-0056-7
6. APA. *American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (5th Ed.)*; 2013.
7. Altman BR, Mian MN, Slavin M, Earleywine M. Cannabis Expectancies for Sleep. *J Psychoactive Drugs*. 2019. doi:10.1080/02791072.2019.1643053
8. Goodhines PA, Gellis LA, Kim J, Fucito LM, Park A. Self-Medication for Sleep in College Students: Concurrent and Prospective Associations With Sleep and Alcohol Behavior. *Behav Sleep Med*. 2019. doi:10.1080/15402002.2017.1357119

9. Goodhines PA, Gellis LA, Ansell EB, Park A. Cannabis and Alcohol Use for Sleep Aid: A Daily Diary Investigation. *Heal Psychol*. 2019. doi:10.1037/hea0000765
10. Taylor DJ, Bramoweth AD. Patterns and Consequences of Inadequate Sleep in College Students: Substance Use and Motor Vehicle Accidents. *J Adolesc Heal*. 2010. doi:10.1016/j.jadohealth.2009.12.010
11. Roehrs T, Papineau K, Rosenthal L, Roth T. Ethanol as a hypnotic in insomniacs: Self administration and effects on sleep and mood. *Neuropsychopharmacology*. 1999. doi:10.1016/S0893-133X(98)00068-2
12. Warbut DM, Revel AD, Thomp DH. Smokers of the future. *Br J Addict*. 1991. doi:10.1111/j.1360-0443.1991.tb01819.x
13. Creamer MLR, Delk J, Case K, Perry CL, Harrell MB. Positive Outcome Expectations and Tobacco Product Use Behaviors in Youth. *Subst Use Misuse*. 2018. doi:10.1080/10826084.2017.1404104
14. Spadola CE, Guo N, Johnson DA, et al. Evening intake of alcohol, caffeine, and nicotine: Night-to-night associations with sleep duration and continuity among African Americans in the Jackson Heart Sleep Study. *Sleep*. 2019. doi:10.1093/sleep/zsz136
15. Hasler BP, Smith LJ, Cousins JC, Bootzin RR. Circadian rhythms, sleep, and substance abuse. *Sleep Med Rev*. 2012. doi:10.1016/j.smr.2011.03.004
16. Kerkhof GA. Inter-individual differences in the human circadian system: A review. *Biol Psychol*. 1985. doi:10.1016/0301-0511(85)90019-5
17. A. A. Chronotype and personality factors in the daily consumption of alcohol and psychostimulants. *Addiction*. 1994.
18. Fernández-mendoza J, Ilioudi C, Montes MI, et al. Circadian preference, nighttime sleep

- and daytime functioning in young adulthood. *Sleep Biol Rhythms*. 2010.
doi:10.1111/j.1479-8425.2010.00430.x
19. Hasler BP, Franzen PL, de Zambotti M, et al. Eveningness and Later Sleep Timing Are Associated with Greater Risk for Alcohol and Marijuana Use in Adolescence: Initial Findings from the National Consortium on Alcohol and Neurodevelopment in Adolescence Study. *Alcohol Clin Exp Res*. 2017. doi:10.1111/acer.13401
 20. Prat G, Adan A. Influence of circadian typology on drug consumption, hazardous alcohol use, and hangover symptoms. *Chronobiol Int*. 2011. doi:10.3109/07420528.2011.553018
 21. Patterson F, Malone SK, Lozano A, Grandner MA, Hanlon AL. Smoking, Screen-Based Sedentary Behavior, and Diet Associated with Habitual Sleep Duration and Chronotype: Data from the UK Biobank. *Ann Behav Med*. 2016. doi:10.1007/s12160-016-9797-5
 22. Kervran C, Fatséas M, Serre F, et al. Association between morningness/eveningness, addiction severity and psychiatric disorders among individuals with addictions. *Psychiatry Res*. 2015. doi:10.1016/j.psychres.2015.05.026
 23. Mitchell JA, Quante M, Godbole S, et al. Variation in actigraphy-estimated rest-activity patterns by demographic factors. *Chronobiol Int*. 2017.
doi:10.1080/07420528.2017.1337032
 24. Saito Y, Kume Y, Kodama A, Sato K, Yasuba M. The association between circadian rest-activity patterns and the behavioral and psychological symptoms depending on the cognitive status in Japanese nursing-home residents. *Chronobiol Int*. 2018.
doi:10.1080/07420528.2018.1505752
 25. Tsanas A, Woodward E, Ehlers A. Objective characterization of activity, sleep, and circadian rhythm patterns using a wrist-worn actigraphy sensor: Insights into

- posttraumatic stress disorder. *JMIR mHealth uHealth*. 2020. doi:10.2196/14306
26. Santisteban JA, Brown TG, Gruber R. Association between the Munich Chronotype Questionnaire and Wrist Actigraphy. *Sleep Disord*. 2018. doi:10.1155/2018/5646848
 27. Gross G, Maruani J, Vorspan F, et al. Association between coffee, tobacco, and alcohol daily consumption and sleep/wake cycle: an actigraphy study in euthymic patients with bipolar disorders. *Chronobiol Int*. 2020. doi:10.1080/07420528.2020.1725542
 28. Pieters S, Burk WJ, Van der Vorst H, Dahl RE, Wiers RW, Engels RCME. Prospective Relationships Between Sleep Problems and Substance Use, Internalizing and Externalizing Problems. *J Youth Adolesc*. 2014. doi:10.1007/s10964-014-0213-9
 29. Pasch KE, Latimer LA, Cance JD, Moe SG, Lytle LA. Longitudinal Bi-directional Relationships Between Sleep and Youth Substance Use. *J Youth Adolesc*. 2012. doi:10.1007/s10964-012-9784-5
 30. Wong MM, Brower KJ, Nigg JT, Zucker RA. Childhood sleep problems, response inhibition, and alcohol and drug outcomes in adolescence and young adulthood. *Alcohol Clin Exp Res*. 2010. doi:10.1111/j.1530-0277.2010.01178.x
 31. Wong MM, Brower KJ, Zucker RA. Childhood sleep problems, early onset of substance use and behavioral problems in adolescence. *Sleep Med*. 2009. doi:10.1016/j.sleep.2008.06.015
 32. Roane BM, Taylor DJ. Adolescent insomnia as a risk factor for early adult depression and substance abuse. *Sleep*. 2008. doi:10.5665/sleep/31.10.1351
 33. Wong MM, Brower KJ, Fitzgerald HE, Zucker RA. Sleep Problems in Early Childhood and Early Onset of Alcohol and Other Drug Use in Adolescence. *Alcohol Clin Exp Res*. 2004. doi:10.1097/01.ALC.0000121651.75952.39

34. Miller MB, Janssen T, Jackson KM. The Prospective Association Between Sleep and Initiation of Substance Use in Young Adolescents. *J Adolesc Heal*. 2017.
doi:10.1016/j.jadohealth.2016.08.019
35. Nguyen-Louie TT, Brumback T, Worley MJ, et al. Effects of sleep on substance use in adolescents: a longitudinal perspective. *Addict Biol*. 2018. doi:10.1111/adb.12519
36. Brook JS, Zhang C, Rubenstone E, Brook DW. Insomnia in adults: the impact of earlier cigarette smoking from adolescence to adulthood. *J Addict Med*. 2015;9(1):40-45.
doi:10.1097/ADM.0000000000000083
37. Rognmo K, Bergvik S, Rosenvinge JH, Bratlid KL, Friborg O. Gender differences in the bidirectional relationship between alcohol consumption and sleeplessness: The Tromsø study. *BMC Public Health*. 2019. doi:10.1186/s12889-019-6801-6
38. Haynie DL, Lewin D, Luk JW, et al. Beyond sleep duration: Bidirectional associations among chronotype, social jetlag, and drinking behaviors in a longitudinal sample of US high school students. *Sleep*. 2018. doi:10.1093/sleep/zsx202
39. Ogeil RP, Phillips JG, Rajaratnam SMW, Broadbear JH. Risky drug use and effects on sleep quality and daytime sleepiness. *Hum Psychopharmacol*. 2015. doi:10.1002/hup.2483
40. Haario P, Rahkonen O, Laaksonen M, Lahelma E, Lallukka T. Bidirectional associations between insomnia symptoms and unhealthy behaviours. *J Sleep Res*. 2013.
doi:10.1111/j.1365-2869.2012.01043.x
41. Winiger EA, Huggett SB, Hatoum AS, Stallings MC, Hewitt JK. Onset of regular cannabis use and adult sleep duration: Genetic variation and the implications of a predictive relationship. *Drug Alcohol Depend*. 2019.
doi:10.1016/j.drugalcdep.2019.06.019

42. Winiger EA, Huggett SB, Hatoum AS, et al. Onset of regular cannabis use and young adult insomnia: an analysis of shared genetic liability. *Sleep*. 2019.
doi:10.1093/sleep/zsz293
43. Watson NF, Buchwald D, Harden KP. A twin study of genetic influences on diurnal preference and risk for alcohol use outcomes. *J Clin Sleep Med*. 2013.
doi:10.5664/jcsm.3282
44. Hammerschlag AR, Stringer S, de Leeuw CA, et al. Genome-wide association analysis of insomnia complaints identifies risk genes and genetic overlap with psychiatric and metabolic traits. *Nat Genet*. 2017;49(11):1584-1592. doi:10.1038/ng.3888
45. Gibson M, Munafò MR, Taylor AE, Treur JL. Evidence for genetic correlations and bidirectional, causal effects between smoking and sleep behaviors. *Nicotine Tob Res*. 2018. doi:10.1093/ntr/nty230
46. Kranzler HR, Zhou H, Kember RL, et al. Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat Commun*. 2019;10(1). doi:10.1038/s41467-019-09480-8
47. Zhou H, Rentsch CT, Cheng Z, et al. Association of OPRM1 Functional Coding Variant with Opioid Use Disorder: A Genome-Wide Association Study. *JAMA Psychiatry*. 2020.
doi:10.1001/jamapsychiatry.2020.1206
48. Winiger EA, Ellingson JM, Morrison CL, et al. Sleep deficits and cannabis use behaviors: an analysis of shared genetics using linkage disequilibrium score regression and polygenic risk prediction. *Sleep*. 2020. doi:10.1093/sleep/zsaa188
49. Pasma JA, Smit DJA, Kingma L, Vink JM, Treur JL, Verweij KJH. Causal relationships between substance use and insomnia. *Drug Alcohol Depend*. 2020.

doi:10.1016/j.drugalcdep.2020.108151

50. P.R. J, K. W, S. S, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet.* 2019.
51. Jones SE, Lane JM, Wood AR, et al. Genome-wide association analyses of chronotype in 697,828 individuals provides insights into circadian rhythms. *Nat Commun.* 2019.
doi:10.1038/s41467-018-08259-7
52. Dashti HS, Jones SE, Wood AR, et al. Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat Commun.* 2019. doi:10.1038/s41467-019-08917-4
53. Jansen PR, Watanabe K, Stringer S, et al. Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways. *Nat Genet.* 2019.
doi:10.1038/s41588-018-0333-3
54. Kaufmann CN, Gershon A, Depp CA, Miller S, Zeitzer JM, Ketter TA. Daytime midpoint as a digital biomarker for chronotype in bipolar disorder. *J Affect Disord.* 2018.
doi:10.1016/j.jad.2018.08.032
55. Fekedulegn D, Andrew ME, Shi M, Violanti JM, Knox S, Innes KE. Actigraphy-based assessment of sleep parameters. *Ann Work Expo Heal.* 2020.
doi:10.1093/ANNWEH/WXAA007
56. Jones SE, van Hees VT, Mazzotti DR, et al. Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat Commun.* 2019.
doi:10.1038/s41467-019-09576-1
57. Häusler N, Marques-Vidal P, Haba-Rubio J, Heinzer R. Does sleep predict next-day napping or does napping influence same-day nocturnal sleep? Results of a population-

- based ecological momentary assessment study. *Sleep Med.* 2019.
doi:10.1016/j.sleep.2019.04.014
58. Liu M, Jiang Y, Wedow R, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet.* 2019;51(2):237-244. doi:10.1038/s41588-018-0307-5
59. Pasmán JA, Verweij KJH, Gerring Z, et al. GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat Neurosci.* 2018;21(9):1161-1170. doi:10.1038/s41593-018-0206-1
60. Johnson EC, Demontis D, Thorgeirsson TE, et al. A large-scale genome-wide association study meta-analysis of cannabis use disorder. *The Lancet Psychiatry.* 2020.
doi:10.1016/S2215-0366(20)30339-4
61. Bulik-Sullivan B, Loh PR, Finucane HK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet.* 2015.
doi:10.1038/ng.3211
62. Duan S, Zhang W, Cox NJ, Dolan ME. FstSNP-HapMap3: a database of SNPs with high population differentiation for HapMap3. *Bioinformatics.* 2008.
doi:10.6026/97320630003139
63. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015. doi:10.1038/ng.3406
64. Bonferroni CE. Il Calcolo delle Assicurazioni su Gruppi di Teste. In: *Studi in Onore Del Professore Salvatore Ortu Carboni.* ; 1935.
65. Wagstaff K, Cardie C, Rogers S, Schrödl S. Constrained K-means Clustering with Background Knowledge. In: *International Conference on Machine Learning ICML.* ;

- 2001.
66. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithms: Analysis and implementation. *IEEE Trans Pattern Anal Mach Intell*. 2002. doi:10.1109/TPAMI.2002.1017616
 67. Maechler, M., Rosseeuw, P., Struyf, Al, Hubert, M., Hornik K. cluster: Cluster Analysis Basics and Extensions. *R Packag Version 212*. 2021.
 68. Kassambara, Alboukadel, Mundt F. factoextra: Extract and Vizualize the Results of Multivariate Data Analyses. *R Packag version 107*. 2020.
 69. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet*. 2018;50(12):1728-1734. doi:10.1038/s41588-018-0255-0
 70. Buysse DJ. Sleep health: can we define It? does it matter? *Sleep*. 2014. doi:10.5665/sleep.3298
 71. Sabanayagam C, Shankar A. The association between active smoking, smokeless tobacco, second-hand smoke exposure and insufficient sleep. *Sleep Med*. 2011. doi:10.1016/j.sleep.2010.09.002
 72. Wetter DW, Young TB. The relation between cigarette smoking and sleep disturbance. *Prev Med (Baltim)*. 1994;23(3):328-334. doi:10.1006/pmed.1994.1046
 73. Koob GF. The dark side of addiction the horsley gantt to joseph brady connection. *J Nerv Ment Dis*. 2017;205(4):270-272. doi:10.1097/NMD.0000000000000551
 74. Luc S, Peter B, Thierry D, et al. Effects of Acamprosate on Sleep During Alcohol Withdrawal: A Double-Blind Placebo-Controlled Polysomnographic Study in Alcohol-Dependent Subjects. *Alcohol Clin Exp Res*. 2006;30(9):1492-1499. doi:10.1111/J.1530-

0277.2006.00180.X

75. Dunn KE, Finan PH, Andrew Tompkins D, Strain EC. Frequency and correlates of sleep disturbance in methadone and buprenorphine-maintained patients. *Addict Behav.* 2018. doi:10.1016/j.addbeh.2017.07.016
76. Wilkerson AK, McRae-Clark AL. A review of sleep disturbance in adults prescribed medications for opioid use disorder: potential treatment targets for a highly prevalent, chronic problem. *Sleep Med.* 2021. doi:10.1016/j.sleep.2021.05.021
77. Rosen IM, Aurora RN, Kirsch DB, et al. Chronic opioid therapy and sleep: An American academy of sleep medicine position statement. *J Clin Sleep Med.* 2019. doi:10.5664/jcsm.8062
78. Patel SR, Malhotra A, Gottlieb DJ, White DP, Hu FB. Correlates of long sleep duration. *Sleep.* 2006. doi:10.1093/sleep/29.7.881

Table 1. Trait, type of measure, measure construct, coding, and discovery sample for all substance use and sleep and activity phenotypes/endophenotypes.

Trait	Type of measure	Discovery Sample	Measure construct	Coding
Insomnia (Jansen et al. 2019)	Self-report	UK Biobank	Insomnia complaints were assessed by asking: “Do you have trouble falling asleep at night or do you wake up in the middle of the night?” Insomnia cases were defined as participants who answered this question with “usually”, while participants answering “never/rarely” or “sometimes” were defined as controls.	$n = 386,533$
Chronotype (Jones et al. 2019)	Self-report	UK Biobank	Self-reported measure phrased as “Do you consider yourself to be?” with one of six possible answers: “Definitely a ‘morning’ person”, “More a ‘morning’ than ‘evening’ person”, “More an ‘evening’ than a ‘morning’ person”, “Definitely an ‘evening’ person”, “Do not know” or “Prefer not to answer”.	$n = 449,734$ Definitely morning, 2, $n = 107,555$ More morning than evening, 1, $n = 144,731$ Don’t know, 0, $n = 46,538$ More evening than morning, -1, $n = 115,090$ Definitely evening, -2, $n = 35,818$
Self-report sleep duration (Dashti et al. 2019)	Self-report	UK Biobank	A continuous variable categorized as short (6h or less), normal (7–8h), or long (9h or more) sleep duration. Responses of less than 3h or more than 18h excluded.	$n = 446,118$
Self-report short sleep duration (Dashti et al. 2019)	Self-report	UK Biobank	Short sleep duration (<7h) compared to 7–8h sleep duration.	$n = 446,118$; short (<7h; $n = 106,192$ cases) compared to 7–8h sleep duration ($n = 305,742$ controls)
Self-report long sleep duration (Dashti et al. (2019)	Self-report	UK Biobank	Long sleep duration (≥ 9 h) compared to 7–8h sleep duration.	$n = 446,118$; long (≥ 9 h; $n = 34,184$ cases) compared to 7–8h sleep duration ($n = 305,742$ controls)
L5 timing (Jones et al. 2019)	Accelerometer	UK Biobank	The midpoint of the least active 5h of each day. This measure implies whether someone prefers to go to bed earlier or later in the day.	$n = 85,830$

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/) .

M10 timing (Jones et al. 2019)	Accelerometer	UK Biobank	The most active 10 h of each day. This measure implies whether a person is most active earlier or later in the day.	$n = 85,723$
Sleep duration in hours (Jones et al. 2019)	Accelerometer	UK Biobank	Summed duration of all sleep episodes.	$n = 85,068$
Sleep duration in hours (standard deviation) (Jones et al. 2019)	Accelerometer	UK Biobank	Standard deviation of the summed duration of all sleep episodes.	$n = 85,068$
Sleep midpoint (Jones et al. 2019)	Accelerometer	UK Biobank	The midpoint between the start of the first detected sleep episode and the end of the last sleep episode.	$n = 85,502$
Sleep efficiency (%) (Jones et al. 2019)	Accelerometer	UK Biobank	sleep duration divided by the time between the start and end of the first and last nocturnal inactivity period.	$n = 85,502$
Number of sleep episodes (Jones et al. 2019)	Accelerometer	UK Biobank	Periods of at least 5 min with no change larger than 5° associated with the z-axis of the activity-monitor.	$n = 85,502$
Diurnal inactivity (Jones et al. 2019)	Accelerometer	UK Biobank	The total daily duration of estimated bouts of inactivity that fell outside of the sleep period time window. This measure captures very inactive states such as napping and wakeful rest but not inactivity such as sitting and reading or watching television.	$n = 85,502$
Alcoholic drinks per week (DPW) (Liu et al., 2019)	Self-report	Meta-analysis performed from, Add Health, Avon Longitudinal Study of Parents and Children, Atherosclerosis Risk in Communities, Biobank Japan Project, The Barrett's and Esophageal Adenocarcinoma Genetic Susceptibility Study, Brisbane Longitudinal Twin Study, Center on Antisocial Drug Dependence, Collaborative Genetic Study of Nicotine Dependence, Genetics of Chronic Obstructive Pulmonary Disease, deCODE Genetics/AMGEN, Inc., Estonian Genome Center, Electronic Medical Records and Genomics, Finnish Twin	Average number of drinks a participant reported drinking each week.	$n = 537,349$

		<p>Cohort, Framingham Heart Study, Genetic Epidemiology Research in Adult Health and Aging, Genes for Good, The Nord-Trøndelag Health Study, Health and Retirement Study, Minnesota Center for Twin and Family Research, Multi-Ethnic Study of Atherosclerosis, Metabolic Syndrome in Men, Netherlands Study on Cognition, Environment and Genes, Nurses' Health Study, Nurses' Health Study II, and Health Professionals' Follow-up Study, The National Institute of Neurological Disorders and Stroke Genetics Network, Netherlands Twin Register, Australian Twin-Family Studies on Nicotine and Alcohol Genetics, SardiNIA project), UK Biobank, and Women's Health Initiative</p>		
<p>Problematic alcohol use (PAU) (Zhou et al., 2020)</p>	<p>Self-report</p>	<p>Million Veteran Program, UK Biobank, Psychiatric Genomics Consortium Substance Use Disorders working group AUD and AUDIT-P, European-ancestry individuals</p>	<p>DSM-IV alcohol dependence (AD), AUD via the International Classification of Diseases (ICD) codes, Alcohol Use Disorders Identification Test-Problems (AUDIT-P).</p>	<p>$n = 435,563$</p>
<p>Lifetime tobacco use (Liu et al., 2019)</p>	<p>Self-report</p>	<p>Meta-analysis performed from 23&me, Add Health, Avon Longitudinal Study of Parents and Children, Atherosclerosis Risk in Communities, Biobank Japan Project, The Barrett's and Esophageal Adenocarcinoma Genetic Susceptibility Study, Brisbane Longitudinal Twin Study, Center on Antisocial Drug Dependence, Collaborative Genetic Study of Nicotine Dependence, Genetics of Chronic Obstructive Pulmonary Disease, deCODE Genetics/AMGEN, Inc., Estonian Genome Center, Electronic Medical Records and Genomics, Finnish Twin Cohort, Framingham Heart Study, Genetic Epidemiology Research in Adult Health and Aging, Genes for Good, The Nord-Trøndelag Health Study, Health and Retirement Study, Minnesota Center for Twin and Family Research, Multi-Ethnic</p>	<p>Reporting ever being a regular smoker in their life (current or former) coded as "2" and any participant who reported never being a regular smoker in their life coded as "1".</p>	<p>$n = 632,802$</p>

		Study of Atherosclerosis, Metabolic Syndrome in Men, Netherlands Study on Cognition, Environment and Genes, Nurses' Health Study, Nurses' Health Study II, and Health Professionals' Follow-up Study, The National Institute of Neurological Disorders and Stroke Genetics Network, Netherlands Twin Register, Australian Twin-Family Studies on Nicotine and Alcohol Genetics, SardiNIA project), UK Biobank, and Women's Health Initiative		
Age of first becoming a regular smoker (Liu et al., 2019)	Self-report	Meta-analysis performed from, Add Health, Avon Longitudinal Study of Parents and Children, Atherosclerosis Risk in Communities, Biobank Japan Project, The Barrett's and Esophageal Adenocarcinoma Genetic Susceptibility Study, Brisbane Longitudinal Twin Study, Center on Antisocial Drug Dependence, Collaborative Genetic Study of Nicotine Dependence, Genetics of Chronic Obstructive Pulmonary Disease, deCODE Genetics/AMGEN, Inc., Estonian Genome Center, Electronic Medical Records and Genomics, Finnish Twin Cohort, Framingham Heart Study, Genetic Epidemiology Research in Adult Health and Aging, Genes for Good, The Nord-Trøndelag Health Study, Health and Retirement Study, Minnesota Center for Twin and Family Research, Multi-Ethnic Study of Atherosclerosis, Metabolic Syndrome in Men, Netherlands Study on Cognition, Environment and Genes, Nurses' Health Study, Nurses' Health Study II, and Health Professionals' Follow-up Study, The National Institute of Neurological Disorders and Stroke Genetics Network, Netherlands Twin Register, Australian Twin-Family Studies on Nicotine and Alcohol Genetics, SardiNIA project), UK Biobank, and Women's Health Initiative	Age at which an individual started smoking cigarette regularly measured by several ways including "At what age did you begin smoking regularly?" and "How long have you smoked?" combined with "What is your current age?".	<i>n</i> = 262,990

It is made available under a [CC-BY-NC 4.0 International license](https://creativecommons.org/licenses/by-nc/4.0/) .

Smoking cessation (Liu et al., 2019)	Self-report	Meta-analysis performed from Add Health, Avon Longitudinal Study of Parents and Children, Atherosclerosis Risk in Communities, Biobank Japan Project, The Barrett's and Esophageal Adenocarcinoma. Genetic Susceptibility Study, Brisbane Longitudinal Twin Study, Center on Antisocial Drug Dependence, Collaborative Genetic Study of Nicotine Dependence, Genetics of Chronic Obstructive Pulmonary Disease, deCODE Genetics/AMGEN, Inc., Estonian Genome Center, Electronic Medical Records and Genomics, Finnish Twin Cohort, Framingham Heart Study, Genetic Epidemiology Research in Adult Health and Aging, Genes for Good, The Nord-Trøndelag Health Study, Health and Retirement Study, Minnesota Center for Twin and Family Research, Multi-Ethnic Study of Atherosclerosis, Metabolic Syndrome in Men, Netherlands Study on Cognition, Environment and Genes, Nurses' Health Study, Nurses' Health Study II, and Health Professionals' Follow-up Study, The National Institute of Neurological Disorders and Stroke Genetics Network, Netherlands Twin Register, Australian Twin-Family Studies on Nicotine and Alcohol Genetics, SardiNIA project), UK Biobank, and Women's Health Initiative	Current smokers coded as "2" and former smokers coded as "1", and never smokers are coded as missing.	<i>n</i> = 312,821
Cigarettes per day (CPD) (Liu et al., 2019)	Self-report	Meta-analysis performed from Add Health, Avon Longitudinal Study of Parents and Children, Atherosclerosis Risk in Communities, Biobank Japan Project, The Barrett's and Esophageal Adenocarcinoma. Genetic Susceptibility Study, Brisbane Longitudinal Twin Study, Center on Antisocial Drug Dependence, Collaborative Genetic Study of Nicotine Dependence, Genetics of Chronic Obstructive Pulmonary Disease, deCODE Genetics/AMGEN, Inc., Estonian Genome Center, Electronic Medical Records	Cigarettes per day responses were binned as a. 1=1-5 b. 2 = 6-15 c. 3 = 16-25 d. 4 = 26-35 e. 5=36+	<i>n</i> = 263,954

		and Genomics, Finnish Twin Cohort, Framingham Heart Study, Genetic Epidemiology Research in Adult Health and Aging, Genes for Good, The Nord-Trøndelag Health Study, Health and Retirement Study, Minnesota Center for Twin and Family Research, Multi-Ethnic Study of Atherosclerosis, Metabolic Syndrome in Men, Netherlands Study on Cognition, Environment and Genes, Nurses' Health Study, Nurses' Health Study II, and Health Professionals' Follow-up Study, The National Institute of Neurological Disorders and Stroke Genetics Network, Netherlands Twin Register, Australian Twin-Family Studies on Nicotine and Alcohol Genetics, SardiNIA project), UK Biobank, and Women's Health Initiative		
Lifetime cannabis use (Pasman et al., 2018)	Self-report	UK Biobank and International Cannabis Consortium	Defined as any use of cannabis during lifetime	$n = 162,082$
Cannabis Use Disorder (Johnson et al., 2020)	Clinician ratings or semi-structured interviews	Psychiatric Genomics Consortium Substance Use Disorders working group, The Integrative Psychiatric Research, and deCODE	Met criteria for a lifetime diagnosis of DSM-IV (or DSM-III-R) cannabis abuse or dependence, DSM-5 cannabis use disorder and ICD-10 codes of F12.1 (cannabis abuse) or F12.2 (cannabis dependence).	$n = 384,032$
Opioid Use Disorder (Zhou et al., 2020)	Clinician ratings	Million Veteran Program, Yale-Penn, and Study of Addiction: Genetics and Environment	International Classification of Diseases, Ninth Revision-diagnosed OUD or International Statistical Classification of Diseases and Related Health Problems, Tenth Revision-diagnosed OUD (Million Veteran Program), and DSM-IV-defined opioid dependence (Yale-Penn and Study of Addiction: Genetics and Environment).	$n = 114,759$

Table 2. Significant pair-wise correlations that survived Bonferroni multiple corrections between substance use and sleep/activity traits.

Sleep and Activity Trait	Substance Use Trait	rG	P-Values before correction	Bonferroni-corrected P-value
Subjective Sleep Duration	Lifetime Tobacco Use	-0.11	4.61E-07	0.00010603
Short Sleep	Lifetime Tobacco Use	0.2304	2.77E-24	6.37E-22
Long Sleep	Lifetime Tobacco Use	0.1318	6.59E-06	0.0015157
Insomnia	Lifetime Tobacco Use	0.2872	6.44E-20	1.48E-17
Diurnal	Lifetime Tobacco Use	0.1453	9.68E-07	0.00022264
Chronotype	Lifetime Tobacco Use	-0.0769	2.00E-04	0.046
Short Sleep	Age of first becoming a regular smoker	-0.3188	4.41E-22	1.01E-19
Long Sleep	Age of first becoming a regular smoker	-0.273	7.63E-12	1.75E-09
Insomnia	Age of first becoming a regular smoker	-0.3568	1.25E-24	2.88E-22
Diurnal	Age of first becoming a regular smoker	-0.1557	5.61E-05	0.012903
Short Sleep	CPD	0.2277	8.86E-13	2.04E-10
Long Sleep	CPD	0.2754	2.44E-12	5.61E-10
Insomnia	CPD	0.2872	4.34E-15	9.98E-13
Diurnal	CPD	0.185	4.96E-09	1.14E-06
Short Sleep	Smoking Cessation	-0.254	6.51E-13	1.50E-10
Long Sleep	Smoking Cessation	-0.2585	5.95E-09	1.37E-06

Insomnia	Smoking Cessation	-0.2763	2.72E-13	6.26E-11
Sleep Midpoint	DPW	0.2364	9.14E-09	2.10E-06
Insomnia	DPW	0.1077	3.07E-05	0.007061
Diurnal	DPW	-0.1167	2.00E-04	0.046
M10	DPW	0.2388	6.70E-09	1.54E-06
Chronotype	DPW	-0.1246	4.26E-10	9.80E-08
Long Sleep	PAU	0.1652	1.20E-06	0.000276
Insomnia	PAU	0.2097	2.43E-13	5.59E-11
Sleep Midpoint	Lifetime Cannabis Use	0.263	4.03E-06	0.0009269
Accelerometer Sleep duration	Lifetime Cannabis Use	-0.168	3.72E-05	0.008556
Chronotype	Lifetime Cannabis Use	-0.2505	3.90E-21	8.97E-19
Short Sleep	CUD	0.2746	2.89E-14	6.65E-12
Insomnia	CUD	0.3121	5.21E-12	1.20E-09
Long Sleep	ODU	0.2863	2.00E-04	0.046
Insomnia	ODU	0.2509	1.35E-05	0.003105

Numerous significant genetic associations between substance use and sleep/activity traits survived Bonferroni correction, demonstrating shared genetic liability between these domains.

Table 3. Results from latent causal variable analysis to test causality between associated sleep/activity and substance use dimensions.

Sleep and Activity Trait	Substance Use Trait	GCP (standard Error)	P-values before correction	Bonferroni-corrected P-values
Self-report Short Sleep Duration	Lifetime Tobacco Use	-0.14(0.10)	0.10	1
Insomnia	Lifetime Tobacco Use	-0.08(0.10)	0.32	1
Self-report Full Sleep Duration	Lifetime Tobacco Use	-0.20(0.28)	0.16	1
Diurnal inactivity	Lifetime Tobacco Use	0.12(0.29)	0.39	1
Self-report Long Sleep Duration	Lifetime Tobacco Use	-0.37(0.37)	0.39	1
Chronotype	Lifetime Tobacco Use	-0.26(0.41)	0.25	1
Insomnia	Age of first becoming a regular smoker	-0.10(0.11)	0.39	1
Self-report Short Sleep Duration	Age of first becoming a regular smoker	-0.03(0.08)	0.79	1
Self-report Long Sleep Duration	Age of first becoming a regular smoker	-0.33(0.19)	0.16	1
Diurnal inactivity	Age of first becoming a regular smoker	0.13(0.26)	0.50	1
Insomnia	CPD	0.31(0.41)	0.63	1
Self-report Short Sleep Duration	CPD	-0.21(0.20)	0.32	1
Self-report Long Sleep Duration	CPD	-0.06(0.27)	0.63	1
Diurnal inactivity	CPD	-0.41(0.35)	0.32	1
Insomnia	Smoking Cessation	-0.43(0.27)	0.13	1

Self-report Short Sleep Duration	Smoking Cessation	0.01(0.12)	0.79	1
Self-report Long Sleep Duration	Smoking Cessation	-0.10(0.19)	0.63	1
Chronotype	DPW	-0.05(0.21)	0.63	1
M10	DPW	-0.32(0.21)	0.25	1
Sleep Midpoint	DPW	-0.45(0.32)	0.32	1
Insomnia	DPW	-0.21(0.47)	0.50	1
Diurnal inactivity	DPW	0.04(0.53)	0.63	1
Insomnia	PAU	-0.08(0.12)	0.50	1
Self-report Long Sleep Duration	PAU	0.03(0.28)	1	1
Chronotype	Lifetime Cannabis Use	0.15(0.16)	0.39	1
Sleep Midpoint	Lifetime Cannabis Use	0.21(0.38)	0.63	1
Accelerometer derived sleep duration	Lifetime Cannabis Use	-0.05(0.46)	0.16	1
Self-report Short Sleep Duration	CUD	-0.23(0.11)	0.04	1
Insomnia	CUD	-0.29(0.14)	0.05	1
Insomnia	OUD	-0.35(0.17)	0.05	1
Self-report Long Sleep Duration	OUD	-0.50(0.25)	0.00036	0.01116

Significant models that survived Bonferroni correction in bold

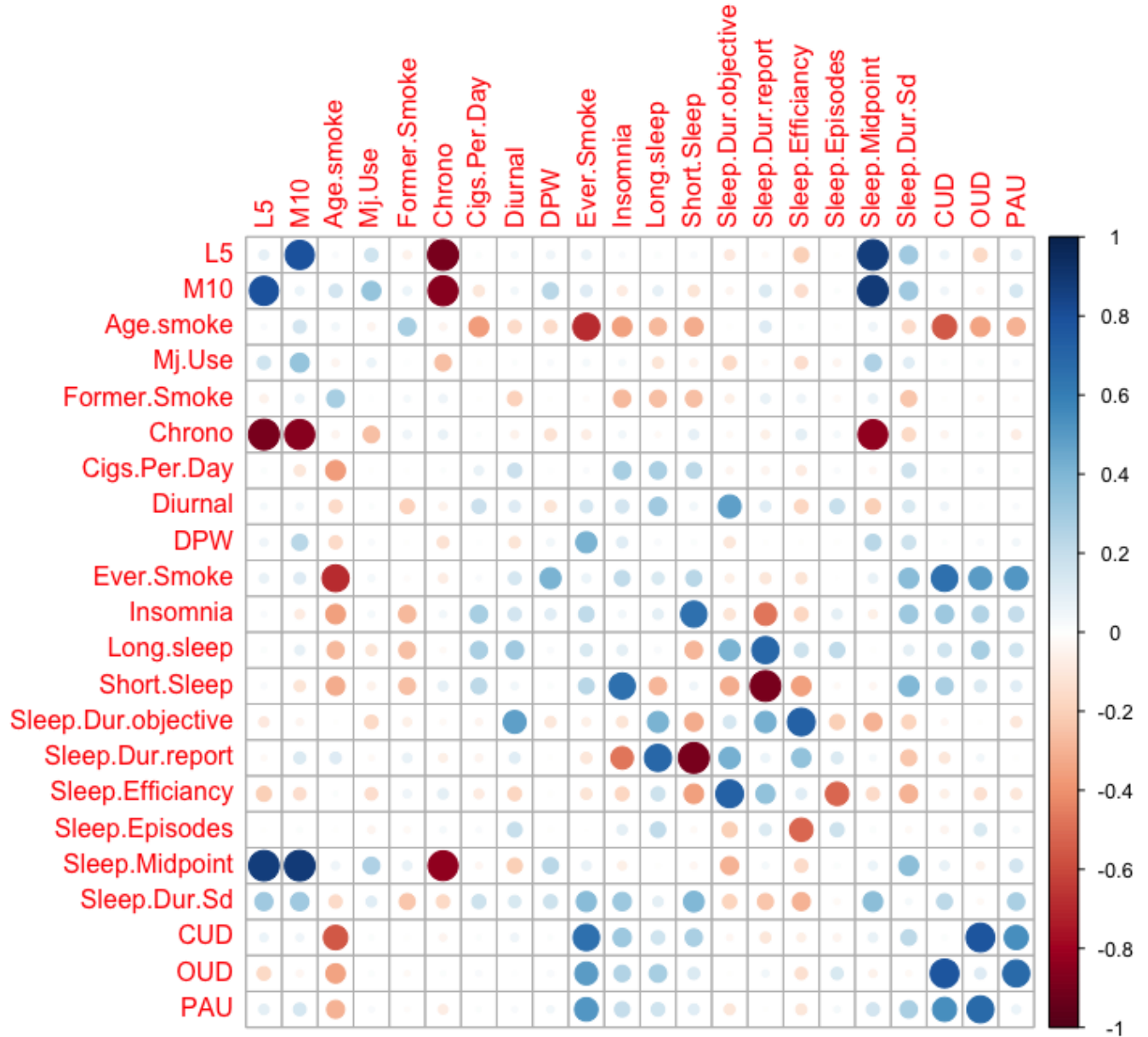


Figure 1. The genetic (rG) correlation matrix between all substance use behaviors and sleep/activity measures. Positive rGs shown in blue and negative in red. Correlations are plotted on X and Y axis based of centroid clustering. Age. Smoke = Age of first becoming a regular smoker, CPD = Cigarettes per week, Former.Smoke = current smoker versus formal smoker, DPW= Drinks per week, OUD= Opiate Use Disorder, CUD = Cannabis Use Disorder, PAU = Problematic Alcohol Use, Ever.Smoke = lifetime tobacco use, Long sleep = Subjective self-report short sleep duration, Short Sleep = Subjective self-report short sleep duration, Sleep.Dur.Sd = Accelerometer Sleep Duration (Standard Deviation), Sleep.Dur.report = subjective self-report sleep duration.

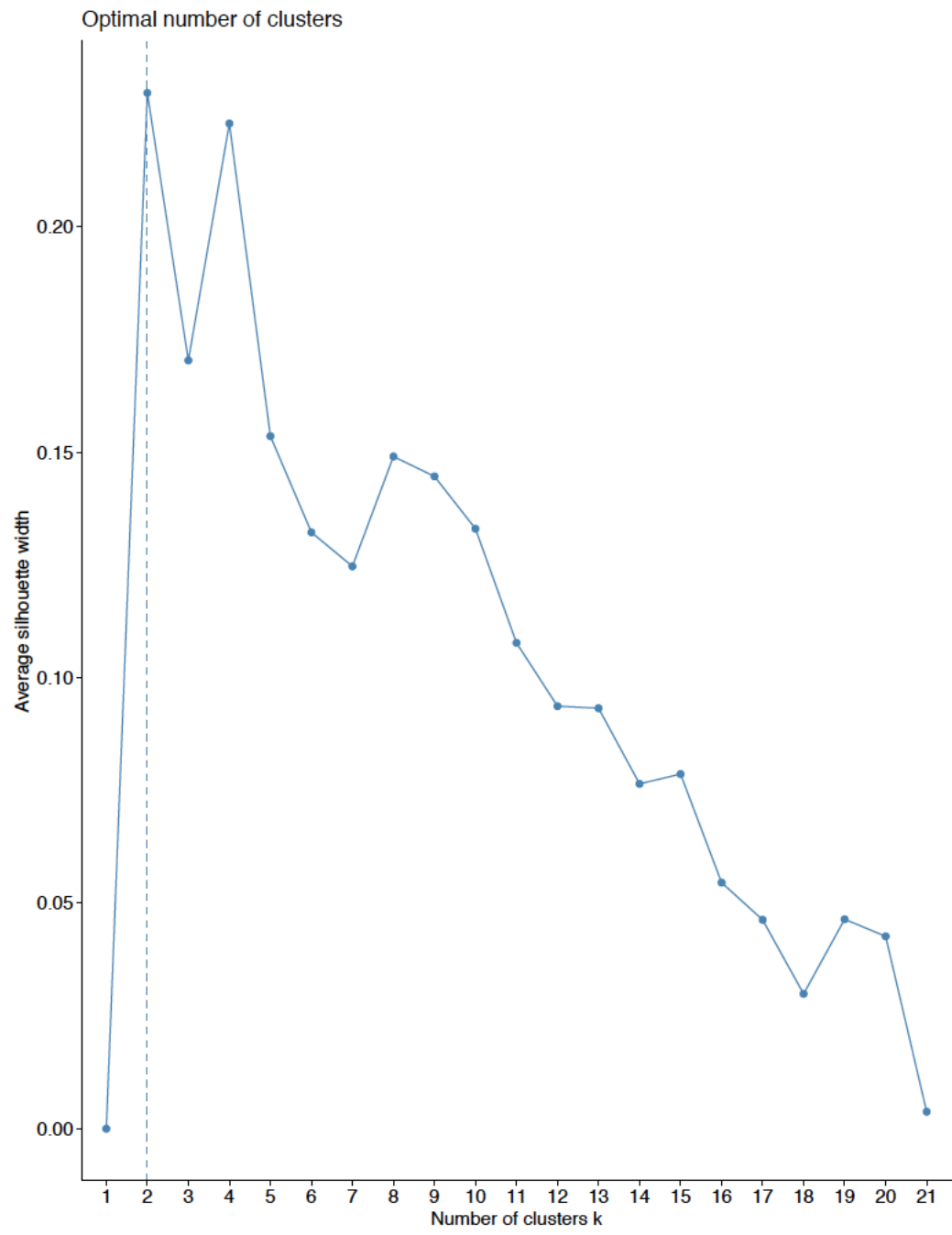


Figure 2. Silhouette Coefficients from K-Means Clustering Algorithm
A silhouette coefficient determined that a 2-cluster solution was optimal to characterize the relationships between sleep/activity and substance use dimensions.

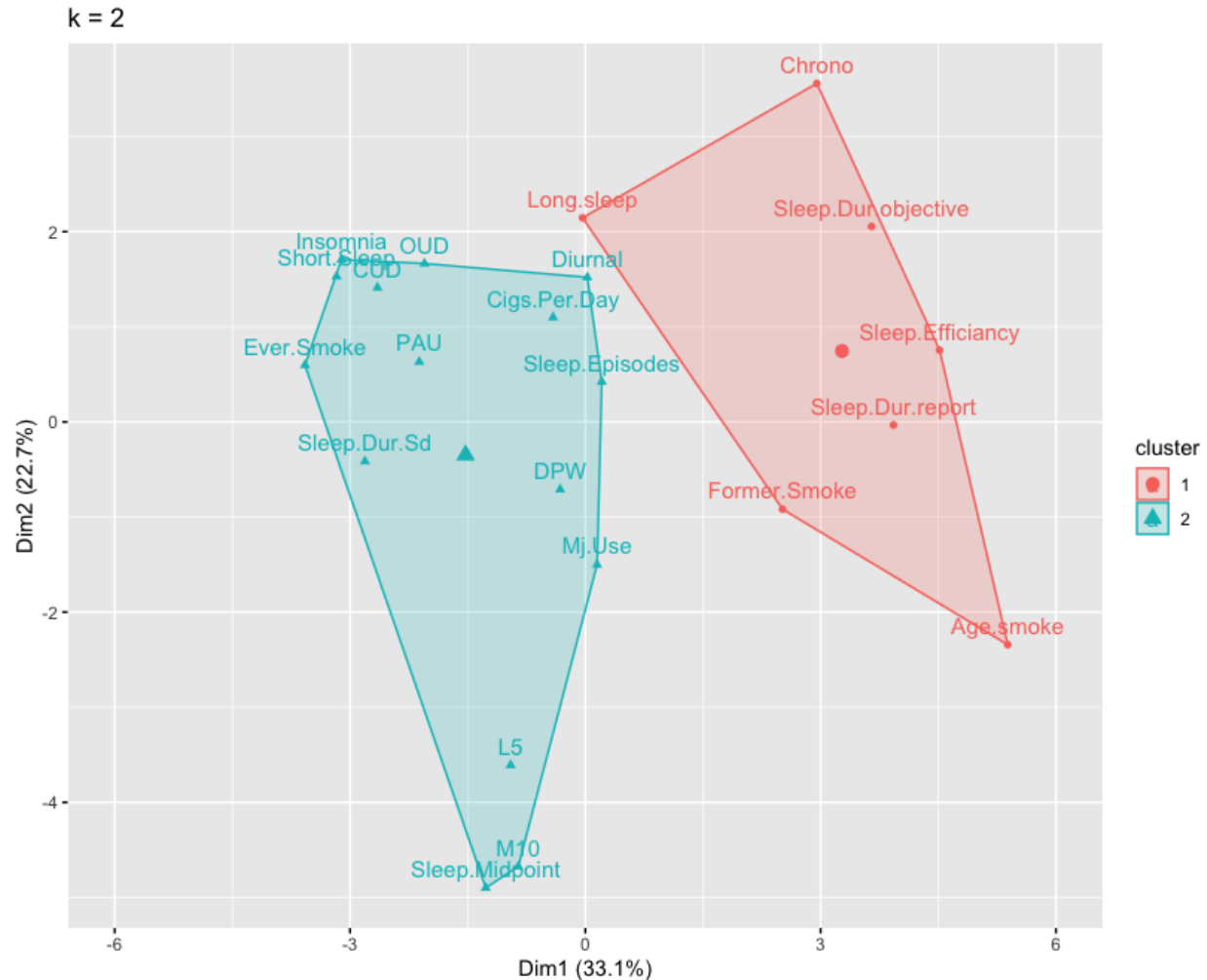


Figure 3. Optimal cluster solution demonstrated two distinct clusters, both of which were comprised of common substance use behaviors and sleep traits. Clusters are plotted by a two dimensional representation of the derived components. X and Y axis represent the first two components and their proportion of variance they explain in each trait. Clusters are plotted around their centroid. Smoke = Age of first becoming a regular smoker, CPD = Cigarettes per week, Former.Smoke = current smoker versus formal smoker, DPW= Drinks per week, OUD= Opiate Use Disorder, CUD = Cannabis Use Disorder, PAU = Problematic Alcohol Use, Ever.Smoke = lifetime tobacco use, Long sleep = Subjective self-report short sleep duration, Short Sleep = Subjective self-report short sleep duration, Sleep.Dur.Sd = Accelerometer Sleep Duration (Standard Deviation), Sleep.Dur.report = subjective self-report sleep duration.