

1 Untargeted Metabolome Atlas for Sleep Phenotypes in the Hispanic 2 Community Health Study/Study of Latinos 3

4 Ying Zhang¹, Brian W Spitzer², Yu Zhang², Danielle A Wallace^{1,2,3}, Bing Yu⁴, Qibin Qi⁵, Maria
5 Argos^{6,7}, M Larissa Avilés-Santa⁸, Eric Boerwinkle⁴, Martha L Daviglus⁹, Robert Kaplan^{5,10},
6 Jianwen Cai¹¹, Susan Redline^{1,3}, Tamar Sofer^{1,2,3,12,*}

7 ¹Division of Sleep Medicine and Circadian Disorders, Department of Medicine, Brigham and Women's
8 Hospital, Boston, MA, USA

9 ²CardioVascular Institute, Beth Israel Deaconess Medical Center, Boston, MA, USA

10 ³Department of Medicine, Harvard Medical School, Boston, MA, USA

11 ⁴Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at
12 Houston, Houston, TX, USA

13 ⁵Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New
14 York, USA

15 ⁶Department of Epidemiology and Biostatistics, School of Public Health, University of Illinois Chicago,
16 Chicago, IL, USA

17 ⁷Department of Environmental Health, School of Public Health, Boston University, Boston, MA, USA

18 ⁸Division of Clinical and Health Services Research, National Institute on Minority Health and Health
19 Disparities, National Institutes of Health, Bethesda, MD, USA

20 ⁹Institute for Minority Health Research, University of Illinois at Chicago, Chicago, IL, USA

21 ¹⁰Division of Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, WA, USA

22 ¹¹Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at
23 Chapel Hill, Chapel Hill, NC 27599, USA

24 ¹²Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
25

26
27 *Correspondence:

28 Tamar Sofer
29 Center for Life Sciences, CLS-934
30 3 Blackfan St
31 Boston, MA, 02116
32 tsofer@bidmc.harvard.edu
33

34
35
36
37
38
39

40

41 Abstract

42 Sleep is essential to maintaining health and wellbeing of individuals, influencing a variety of outcomes
43 from mental health to cardiometabolic disease. This study aims to assess the relationships between
44 various sleep phenotypes and blood metabolites. Utilizing data from the Hispanic Community Health
45 Study/Study of Latinos, we performed association analyses between 40 sleep phenotypes, grouped in
46 several domains (i.e., sleep disordered breathing (SDB), sleep duration, timing, insomnia symptoms, and
47 heart rate during sleep), and 768 metabolites measured via untargeted metabolomics profiling. Network
48 analysis was employed to visualize and interpret the associations between sleep phenotypes and
49 metabolites. The patterns of statistically significant associations between sleep phenotypes and
50 metabolites differed by superpathways, and highlighted subpathways of interest for future studies. For
51 example, some xenobiotic metabolites were associated with sleep duration and heart rate phenotypes
52 (e.g. 1H-indole-7-acetic acid, 4-allylphenol sulfate), while ketone bodies and fatty acid metabolism
53 metabolites were associated with sleep timing measures (e.g. 3-hydroxybutyrate (BHBA), 3-
54 hydroxyhexanoylcarnitine (1)). Heart rate phenotypes had the overall largest number of detected
55 metabolite associations. Many of these associations were shared with both SDB and with sleep timing
56 phenotypes, while SDB phenotypes shared relatively few metabolite associations with sleep duration
57 measures. A number of metabolites were associated with multiple sleep phenotypes, from a few
58 domains. The amino acids vanillylmandelate (VMA) and 1-carboxyethylisoleucine were associated with
59 the greatest number of sleep phenotypes, from all domains other than insomnia. This atlas of sleep-
60 metabolite associations will facilitate hypothesis generation and further study of the metabolic
61 underpinnings of sleep health.

62

63 Introduction

64 Sleep plays an important role in the health and wellbeing of individuals. Insufficient quality, timing, and
65 duration of sleep have a major public health impact, and are associated with daytime sleepiness, poor
66 mental health, impaired cognitive function, and increased risk of cardiovascular morbidity and mortality
67 (1–3). Sleep is increasingly recognized as a crucial factor in cardiovascular health, evident by the addition
68 of sleep duration to the “life’s essential 8” metric developed by the American Heart Association(4). In
69 addition to sleep duration, measures of suboptimal sleep, such sleep disturbances and quality (or
70 insomnia symptoms), irregularity of sleep timing, sleep fragmentation, and sleep disordered breathing
71 (SDB), are also associated with poor health outcomes (5). In fact, there is growing recognition of the
72 importance in measuring and characterizing multi-dimensional sleep health— a framework that
73 concurrently considers these varied aspects of sleep (6–8).

74

75 Despite the strong epidemiological evidence observed in many cohort and clinical studies for the
76 connection between suboptimal sleep health and increased risks for poor health outcomes, the biology
77 and physiology behind these links are not fully understood. While many sleep behaviors and outcomes
78 share some underlying genetic and physiological pathways (9–11), or have, potentially bidirectional,
79 causal relationships (12), there may also be distinct mechanisms that underlie specific sleep
80 disturbances or sleep subtypes (13–15). Untangling these shared and distinct mechanisms underlying
81 sleep phenotypes has the potential to inform sleep health intervention efforts.

82

83 Biological sampling to measure molecular markers of health, such as in metabolomics, can be used to
84 investigate the mechanisms and pathways underpinning sleep phenotypes. Metabolites are small
85 molecules produced in the formation and/or breakdown of endogenous or exogenous substances and
86 are oriented at the closest layer to phenotypes compared to other underlying biochemical layers (e.g.,
87 genome, transcriptome and proteome). The increasing availability of large datasets with untargeted
88 metabolomics profiling has unveiled metabolic outcomes and correlates of numerous health
89 phenotypes, including sleep measures (16–31). A large-scale study of metabolites in relation to sleep
90 phenotypes may shed light on how underlying biological processes may converge and differ among
91 common sleep phenotypes, the complex interplay between sleep and the metabolic environment, and,
92 ultimately, potential interactions among sleep disorders and progression of cardiometabolic and other
93 health conditions. Here, we study the associations between a range of sleep phenotypes and the
94 metabolic environment in a large population-based study using a high-dimensional set of measured
95 metabolites. We create an “atlas” – a resource for the sleep research community that will facilitate
96 hypotheses formulation and accelerate studies on sleep and its association with other health outcomes.

97

98 Our study has taken a comprehensive approach, covering key sleep phenotypes including sleep
99 disordered breathing (SDB), sleep duration, sleep timing, insomnia symptoms, and heart rate (HR)
100 during sleep. Each category of sleep phenotypes provides a different perspective on sleep, while
101 together may highlight some shared biological processes within this complex physiological
102 phenomenon. We also conducted network analysis to better understand the interconnectedness
103 between multiple sleep phenotypes and metabolites –representing significant associations as links in a
104 bipartite network allows for simultaneous visualization of many associations, enabling researchers to
105 perceive connectivity patterns that might otherwise be obscured when looking at the individual
106 relationships. By reporting a large number of associations between metabolites and sleep phenotypes,

107 this resource may provide researchers with a starting point for more targeted inquiries into the
108 metabolic environment changes induced by sleep disorders, and facilitate hypothesis generation for
109 future metabolomic sleep research. This may ultimately contribute to our understanding of the
110 pathogenesis of sleep disorders and pave the way for developing more effective diagnostic and
111 therapeutic strategies.

112

113 Methods

114 The Hispanic Community Health Study/Study of Latinos

115 The Hispanic Community Health Study / Study of Latinos (HCHS/SOL) is a prospective community-based
116 cohort study of 16,415 Hispanic/Latino individuals aged 18–74 years at the baseline examination (2008-
117 2011). Study participants were selected using a multi-stage stratified random sampling from four
118 geographic regions: Bronx NY, Chicago IL, Miami FL, and San Diego CA (32,33). Description of major
119 ancillary studies and findings in the context of cardiovascular health is provided in a prior publication
120 (34). Fasting blood samples were collected at the baseline examination, and within the subsequent
121 week, 14,440 of these participants underwent an evaluation for SDB using a validated Type 3 home
122 sleep apnea test (ARES Unicorder 5.2; B-Alert, Carlsbad, CA) that measured nasal airflow, position,
123 snoring, heart rate and oxyhemoglobin saturation with measures of SDB scored by a central reading
124 center as detailed previously (35). All sleep phenotypes used and their definitions are provided in
125 Supplementary Note 1.

126

127 Metabolomics profiling

128 From those who attended the HCHS/SOL baseline assessment and also underwent genotyping (12,803
129 of the study individuals (36)), 4,002 individuals were selected at random for metabolomics profiling of
130 fasting serum samples collected at baseline (metabolomics batch 1, processed in 2017). In 2021,
131 additional 2,368 serum samples from 2,330 participants, also collected at baseline, were profiled in a
132 second metabolomics batch 2. Serum samples were stored at -70°C at the HCHS/SOL Core Laboratory at
133 the University of Minnesota until analysis by Metabolon, Inc. (Durham, NC) in 2017 (batch 1) and 2021
134 (batch 2). Serum samples were then extracted and prepared using Metabolon's standard solvent
135 extraction method. Extracts were split into a five fractions to use in four liquid chromatography-mass
136 spectrometry (LC-MS)-based metabolomic quantification platforms (two reverse phase methods with
137 positive ion mode electrospray ionization (EI), one reverse phase method with negative ion mode EI, and
138 one hydrophilic interaction liquid chromatography with negative ion mode EI), with the fifth fraction
139 reserved for backup. Instrument variability was assessed by calculating the median relative standard
140 deviation (SD) for the internal standards added to each sample prior to injection into the mass
141 spectrometers. Overall process variability was determined by calculating the median relative SD for all
142 endogenous metabolites (i.e., non-instrument standards) present in 100% of the technical replicate
143 samples.

144

145 Metabolomic data pre-processing

146 Preprocessing of the metabolomic data is described in Supplementary Figure S1. First, we removed
147 batch 2 individuals who overlapped with batch 1 and replicate samples from the same individuals,
148 resulting in 2,178 remaining batch 2 observations. We then computed percentages of missing values of
149 each metabolite in each batch separately. We excluded metabolites with missing values in more than

150 75% of the individuals in either batch. For xenobiotic metabolites (metabolite annotation was provide by
151 Metabolon), we assumed that missing values were due to concentrations below the minimum detection
152 limits, thus imputed the missing values for each metabolite with half of the lowest non-missing value of
153 that metabolite across the sample within the batch. For non-xenobiotic metabolites, we applied multiple
154 imputation using the futuremice function from the R mice package (version 3.15.0) that implements
155 fully conditional imputation in a computationally efficient manner (using parallelization). Each variable
156 in the imputation dataset is imputed (if it has missing values) using a model that predicts potential
157 values based on other variables in the dataset. The dataset was imputed 5 times to generate 5
158 completed datasets. Differences between the computed datasets are due to randomness (e.g. random
159 residuals added to the predicted values of a variable). We imputed metabolites, separately in each
160 batch, together with a set of covariates that are strongly linked to the metabolic environment: age, sex,
161 BMI, waist-to-hip ratio, fasting insulin, fasting glucose, type 2 diabetes status (American Diabetes
162 Association definition: fasting glucose ≥ 126 mg/dL, or post-OGTT glucose ≥ 200 mg/dL or A1C $\geq 6.5\%$,
163 or self-report of diabetes), estimated glomerular filtration rate, and lipid measures – total cholesterol,
164 triglycerides, high- and low-density lipoprotein measures. We then rank normalized the 768 metabolite
165 measures passing quality control in each batch separately and each imputed dataset separately, and
166 finally, aggregated data from the two batches, such that the “first” imputed batch 1 dataset was
167 aggregated with the “first” imputed batch 2 dataset, and so on. Supplementary Note 2 describes the
168 analysis we performed to inform the metabolite imputation strategy.

169

170 Sleep phenotypes and modeling approach

171 We used 40 sleep phenotypes from 5 domains, described in detail in Supplementary Note 1. In brief,
172 these included: self-reported sleep duration (sleep duration during weekdays and weekend days,

173 average sleep duration during, short and long sleep during weekdays), heart rate during sleep
174 (minimum, maximum, average, and standard deviation of HR), insomnia (the women health initiative
175 insomnia rating scale (WHIIRS), its component questions, sleeping pill use, and daytime and excessive
176 daytime sleepiness, measured via the Epworth sleepiness scale), sleep disordered breathing (respiratory
177 event indices, measures of oxyhemoglobin saturation during sleep, measures of respiratory event
178 lengths, as well as self-reported snoring), and sleep timing (bed time and wake time as well as sleep
179 midpoint during weekend and weekend days, and social jetlag – the difference in sleep midpoint
180 between weekend and weekdays). In figures and tables, these domains are referred to as Duration, HR,
181 Insomnia, SDB, and Timing. Phenotypes in the sleep timing category were analyzed as “circular”
182 variables, i.e., to avoid bias due to day thresholding (i.e., defining a day by midnight) and to account for
183 the fact that 11:59PM (23:59) is adjacent to 12:01AM (00:01). Other sleep phenotypes were treated as
184 linear or binary (when dichotomized). Two nested conceptual models were used in the analyses. Model
185 1 adjusted for batch number, demographic, and baseline clinical variables, including age, sex, field
186 center, Hispanic/Latino background (Mexican, Puerto Rican, Cuban, Central American, Dominican, South
187 American, and other/multi) and body mass index (BMI). Model 2 further adjusted for lifestyle variables –
188 alcohol use (never, former, current), cigarette use (never, former, current), total physical activity (MET-
189 min/day, computed based on self-reported time spent doing physical activities), and diet (Alternative
190 Healthy Eating Index 2010, computed based on 24 hours dietary recall data) in addition to Model 1
191 covariates. Sleep phenotypes were used as exposures with metabolites as outcomes. When testing the
192 association of a circular phenotype (i.e., sleep timing) with a metabolite, we first converted the sleep
193 timing phenotype into radians (with 12am serving as the 0) and then computed the sine and cosine of
194 these radians. We used the sine and cosine terms as predictors in the regression (37), and tested their
195 association with the metabolite using the multivariate Wald test, accounting for two predictors and their
196 estimated covariance. All association analyses were performed using the *survey* R package (version 4.1)

197 to account for the HCHS/SOL sampling design and provide effect estimates relevant for the HCHS/SOL
198 target population.

199

200 Estimating associations between metabolites and sleep phenotypes

201 We separately assessed the association between each metabolite's concentration level, as an outcome,
202 and each sleep phenotype, as a predictor, using Model 1 and Model 2 covariates as described above, in
203 a single metabolite association analysis. Xenobiotic metabolites were imputed once, while missing
204 values in other metabolites values were imputed 5 times. Thus, we estimated the associations of non-
205 xenobiotic metabolites with sleep phenotypes 5 times (using each of the completed datasets), and then
206 combined the resulting estimated associations. Here we treated the linearly-modeled sleep phenotypes
207 differently from the circular ones. The estimated metabolite associations with linearly-modeled sleep
208 phenotypes were combined using Rubin's rule (38). Sleep timing phenotypes cannot be combined in the
209 same manner, because there is no method to combine the covariance between the sine and cosine
210 terms across several completed datasets. Instead, we focused on testing and aggregated the p-value of
211 the multivariate Wald test using the aggregated Cauchy association test (ACAT; (39)). The ACAT test was
212 developed in the context of genetic association analyses, but it is appropriate for our settings, because it
213 allows for the aggregated tests to be based on correlated data. Finally, after combining association
214 results so we had one p-value per sleep phenotype-metabolite association (per model), we
215 implemented the Benjamini-Hochberg method to control the false discovery rate (FDR) for multiple
216 testing across all metabolites in all models for each sleep phenotype (40). Any association that resulted
217 in an FDR-adjusted p-value < 0.05 in Model 1 was considered statistically significant. In secondary
218 analysis, we also performed sex-stratified analyses using the same analytic approach. For descriptive
219 purposes, we computed the number of sleep traits that each metabolite was associated with at the

220 FDR<0.05 Model 1 threshold, and identified the top 10% metabolites based on number of sleep trait
221 associations.

222

223 We summarized the associations between sleep phenotypes, individually and categorized by domains,
224 and metabolites, individually and categorized by pathways. Similarities of these associations between a
225 pair of phenotypes or domains were estimated with the Dice Similarity Coefficient (DSC) (41) defined as

226
$$DSC = \frac{2|X \cap Y|}{|X| + |Y|}$$

227 Where X and Y represent the set of metabolites with statistically significant associations (FDR<0.05) with
228 two sleep phenotypes or domains. DSC takes values between 0 and 1 where 0 indicates no similarity
229 while 1 indicates full overlap between the two sets.

230

231 Bipartite network analysis

232 Bipartite network is a type of network in which there are two groups of nodes (here: sleep phenotypes
233 and metabolites), and links, or edges, can only exist between the two types of nodes but not within a
234 group of nodes. We constructed a bipartite network using the sleep phenotypes and metabolites, where
235 an edge was added between a sleep phenotype-metabolite pair if their association was statistically
236 significant (FDR-adjusted p -value<0.05) in Model 1 association analysis. The network is built based on an
237 incidence matrix, with rows corresponding to sleep phenotypes, columns represent metabolites, and the
238 i, j cell in this matrix has a value of 1 if the i th sleep phenotype has a statistically significant association
239 with the j th metabolite, and 0 otherwise. We then aggregated metabolites by sub- and super-pathway,
240 and sleep phenotypes by domain, which results in a consolidated incidence matrix in which a value of 1
241 in cell(m, n) indicating any statistically significant association between the m th sleep phenotype domain

242 and the n th metabolite sub- and super-pathway and a weight matrix documenting the number of total
243 statistically significant associations between the m th sleep phenotype domain and the n th metabolite
244 sub- and super-pathway. Several network property metrics were computed to offer insights into the
245 structure and associations of the sleep domain – metabolite sub- and super- pathway network.

246

247 For visualization, we converted the bipartite network into a univariate network and visualized the
248 network using the Fruchterman-Reingold force-directed algorithm, which optimizes the placement of
249 nodes based on connectivity similarity between nodes, where similar connectivity is reflected as
250 proximity of nodes (42), providing an intuitive spatial representation of the network structure.

251

252 Statistical software

253

254 All analyses were done in R version 4.2.3. The function `svyglm` from the `survey` package was used for
255 survey-weighted generalized linear regression models. The `car` package (version 3.1) was used for
256 multivariate Wald test. The `bipartite` package (version 2.18) was used for bipartite network analysis, and
257 `igraph` (version 1.5) and `ggnetwork` (version 0.5) packages were used for visualizing network graph.

258

259 Results

260

261 Study sample characteristics

262 Table 1 characterizes the HCHS/SOL target population (using means and percentages weighted to
263 account for study participation), with more comprehensive data provided in Supplementary Table S1.

264 The analytic sample, batch 1 and batch 2 metabolomic dataset combined, included 6,180 participants

265 with a mean age of 44.31 years (SD=15.27), of whom 40.1% were males, 20.7% were non-alcohol users,
266 58.5% never smoked, and 11.7% had moderate-to-severe OSA (REI3>=15). The average sleep duration
267 was 7.98 hours (SD=1.46), while 22.7% reported restless or very restless sleep on a typical night in the
268 last month, and 15.8% reported excessive sleepiness (ESS>=10). Batch 2 participants were older
269 compared to batch 1 – the mean age is 41.54 years (SD) in batch 1 and 51.11 years (SD) in batch 2. The
270 baseline rates of diabetes and hypertension were higher in batch 2 (diabetes: 29.4%; hypertensions:
271 44.1%) compared to batch 1 (diabetes: 20.1%; hypertensions: 31.5%), consistent with the age difference
272 between the two batches.

273

Table 1: HCHS/SOL target population characteristics by sex, batch, and overall					
	Female	Male	Batch 1	Batch 2	Overall
n	3704	2476	4002	2178	6180
Age (mean (SD))	44.98 (15.17)	43.56 (15.35)	41.54 (15.18)	51.11 (13.21)	44.31 (15.27)
Gender = Male (%)	0 (0.0)	2476 (100.0)	1705 (42.6)	771 (35.4)	2476 (40.1)
Hispanic/Latino background (%)					
BMI (mean (SD))	30.15 (6.69)	28.78 (5.37)	29.42 (6.27)	29.72 (5.80)	29.50 (6.14)
Current alcohol drinking (%)	1401 (37.9)	1490 (60.2)	1973 (49.3)	918 (42.3)	2891 (46.8)
Current smoking (%)	596 (16.1)	668 (27.1)	863 (21.6)	401 (18.5)	1264 (20.5)
Physical activity (MET-min/day) (mean (SD))	424.88 (714.71)	879.70 (1174.88)	697.66 (1020.41)	496.61 (880.66)	639.60 (986.25)
The Alternate Healthy Eating Index (2010) (mean (SD))	46.79 (7.42)	48.74 (7.47)	47.28 (7.52)	48.75 (7.36)	47.71 (7.50)
OSA status = OSA (%)	260 (7.9)	383 (17.4)	388 (10.9)	255 (13.0)	643 (11.7)
Baseline Diabetes status (ADA) = Yes (%)	863 (23.3)	579 (23.4)	803 (20.1)	639 (29.4)	1442 (23.3)
Baseline Hypertension status = Yes (%)	1357 (36.6)	865 (34.9)	1261 (31.5)	961 (44.1)	2222 (36.0)
Sleep Duration: Work/School Days (in hours) (mean (SD))	7.89 (1.66)	7.75 (1.56)	7.82 (1.62)	7.83 (1.60)	7.82 (1.62)
Womens's Health Initiative Insomnia Rating Scale (WHIIRS) total score (mean (SD))	7.64 (5.58)	6.16 (5.13)	6.78 (5.37)	7.34 (5.54)	6.95 (5.42)
Epworth Sleepiness Scale (ESS) total score (mean (SD))	5.44 (4.71)	5.87 (4.87)	5.62 (4.71)	5.70 (4.99)	5.64 (4.79)
OSA was defined as respiratory event index >=15.					
Means and percentages have been weighted to provide values representative of the HCHS/SOL target population.					

275 Results from sleep phenotype-metabolite association analysis

276 The single metabolite association analysis was conducted in pair-wise fashion between 40 sleep
277 phenotypes from five domains (i.e., sleep duration, HR during sleep, insomnia, SDB, and sleep timing)
278 and 768 metabolites, including 113 unknown metabolites and 77 xenobiotic metabolites. Of the
279 phenotypes, 35 sleep phenotypes had statistically significant associations with at least one metabolite
280 (Table 2). When limited to Model 1, the median number of significant associations for each sleep
281 phenotype is 16.5 (range: 0 - 304), corresponding to 2.15% (range: 0% – 39.58%) of all tested
282 metabolites. The number of statistically significant associations is much lower among dichotomized
283 sleep phenotypes (median: 5; range: 0 – 73) compared to non-dichotomized sleep phenotypes (median:
284 37; range: 0 – 304), likely corresponding to loss of power due to dichotomization. Phenotypes from the
285 heart rate during sleep domain had the highest number of statistically significant associations with
286 metabolites (median: 174, range: 2 - 304), followed by the sleep timing domain (median: 117, range: 64
287 - 197). The SDB domain had the lowest number of statistically significant associations with tested
288 metabolites (median: 8, range: 0 – 68) (Supplementary Table S2). Figure 2 visualizes the strength of
289 associations (i.e., negative logarithm of the FDR-adjusted p) between sleep phenotypes and individual
290 metabolites grouped by superpathway. Compared to the sex-stratified analysis, sex-combined analysis
291 tends to identify more statistically significant associations between sleep phenotypes and metabolites
292 (Figure 3), in accordance with the higher power due to larger sample size of the sex-combined analysis.

293

294

Table 2: Statistically significant metabolite associations by sleep phenotype

Sleep phenotype	Number of statistically significant associations	Percentage of statistically significant associations (%)	Sleep phenotype domain	Sleep phenotype dichotomized or continuous
weekday long sleep	2	0.26	Duration	Dichotomized
weekday short sleep	0	0.00	Duration	Dichotomized
sleep duration	107	13.93	Duration	Continuous
weekday sleep duration	101	13.15	Duration	Continuous
social jetlag	0	0.00	Duration	Continuous
weekend sleep Duration	3	0.39	Duration	Continuous
min HR during sleep	225	29.30	Heart rate (HR)	Continuous
max HR during sleep	123	16.02	Heart rate (HR)	Continuous
avg HR during sleep	304	39.58	Heart rate (HR)	Continuous
std HR during sleep	2	0.26	Heart rate (HR)	Continuous
difficulty back to sleep	0	0.00	Insomnia	Dichotomized
early wake	16	2.08	Insomnia	Dichotomized
ESS	34	4.43	Insomnia	Continuous
ESS>10 (EDS)	8	1.04	Insomnia	Dichotomized
difficulty fall asleep	17	2.21	Insomnia	Dichotomized
frequent wake	0	0.00	Insomnia	Dichotomized
taking sleep pill	73	9.51	Insomnia	Dichotomized
restless sleep	6	0.78	Insomnia	Dichotomized
WHIIRS	35	4.56	Insomnia	Continuous
avg event Duration	0	0.00	SDB	Continuous
hypoxic burden	7	0.91	SDB	Continuous
per90	1	0.13	SDB	Dichotomized
rei0	11	1.43	SDB	Continuous
rei0>15	4	0.52	SDB	Dichotomized
rei0>5	27	3.52	SDB	Dichotomized
total event count	12	1.56	SDB	Continuous
total event Duration	7	0.91	SDB	Continuous
rei3	68	8.85	SDB	Continuous
rei3>15	8	1.04	SDB	Dichotomized
rei3>5	43	5.60	SDB	Dichotomized
min o2	38	4.95	SDB	Continuous
avg o2	36	4.69	SDB	Continuous
perlt90	8	1.04	SDB	Continuous
snore	4	0.52	SDB	Dichotomized
weekday bed time	158	20.57	Timing	Continuous
weekday midpoint time	197	25.65	Timing	Continuous
weekday wake time	153	19.92	Timing	Continuous

weekend bed time	80	10.42	Timing	Continuous
weekend midpoint time	81	10.55	Timing	Continuous
weekend wake time	64	8.33	Timing	Continuous

295

296 Among the 768 included metabolites, carbohydrates had the highest average number of statistically
297 significant associations with sleep phenotypes (mean: 4.23, range: 1 – 8 per metabolite), followed by
298 cofactors and vitamins (mean: 3.92, range 0 – 12 per metabolite) and lipids (mean: 3.04, range: 0 – 11
299 per metabolite). Partially characterized molecules (mean: 1.78, range: 0 – 8) and xenobiotics (mean:
300 1.78, range: 0 – 9) had the lowest number of statistically significant associations with sleep phenotypes
301 per metabolite (Supplementary Table S3).

302

303 We aggregated metabolites by subpathway and sleep phenotypes by domain, then identified and
304 visualized top metabolomic subpathways with over 25% statistically significant associations (defined as
305 FDR-adjusted $p < 0.05$) among all tested associations (Figure 4) by sleep phenotype domain. Primary bile
306 acid metabolism showed the highest cumulative percentage of statistically significant associations
307 across all five sleep phenotype domains, although no significant association was identified for SDB
308 phenotypes (Supplementary Table S4). Both lipids and cofactors/vitamins had subpathways (e.g., ketone
309 bodies, acyl glutamine, pantothenate and CoA metabolism, nicotinate and nicotinamide metabolism)
310 that had high percentage of statistically significant associations with sleep timing phenotypes. SDB
311 phenotypes had the highest percentage of statistically significant associations among lipids, especially
312 progestin steroids, phosphatidylethanolamine, diacylglycerols, pregnenolone steroids and
313 sphingomyelins, which is consistent with our previous research findings (43). As for the sleep duration
314 domain, the top subpathways with the highest percentage of significant associations were
315 bacterial/fungal, amino sugar metabolism, nicotinate and nicotinamide metabolism, primary bile acid

316 metabolism, corticosteroids, and polyamine metabolism. Insomnia domain generally had fewer
317 statistically significant associations, among which pregnenolone steroids and vitamin A metabolism
318 showed the highest percentage of statistically significant associations.

319
320 Comparing men and women, potential sex differences can be observed (Supplementary Figure S2), with
321 the limitation that the sample size of male participants was lower. Relatively more subpathways had
322 statistically significant associations with SDB phenotypes among females compared to males (these
323 include oxidative phosphorylation, lactosylceramides fructose, mannose and galactose metabolism). The
324 subpathways from which metabolites were associated with insomnia phenotypes were mostly
325 pregnenolone steroids, nicotinate and nicotinamide metabolism and carnitine metabolism among
326 females, while long chain saturated and monounsaturated fatty acid and amino sugar metabolism were
327 more associated with insomnia phenotypes among males.

328
329 We also aggregated metabolites by superpathway and visualized the percentage of statistically
330 significant associations between superpathways and sleep phenotype domains in the format of a
331 heatmap (Figure 5). The percentage of statistically significant associations were highest for HR during
332 sleep, especially among carbohydrates (40.34%), followed by cofactors and vitamins (26.92%), amino
333 acids (24.70%) and peptides (24.04%) (Supplementary Table S5). Insomnia and SDB domain showed
334 overall lower percentage of statistically significant associations with metabolites across all
335 superpathways, all of which were below 10%.

336

337 Among all metabolites associated with at least one sleep phenotype, we identified the top 10% (n=37)
338 with the highest number of total associations, regardless of domains (Figure 6). Sixteen metabolites
339 were lipids, 8 were amino acids, 4 were cofactors and vitamins, 4 were unnamed, 2 were peptides, 1
340 was carbohydrate, 1 was xenobiotic and 1 was partially characterized. Every metabolite from the top
341 10% list had associations with more than one sleep phenotype domain, among which 11 metabolites
342 were associated with four different domains of sleep phenotypes. The amino acids Vanillylmandelate
343 (VMA) and 1-carboxyethylisoleucine were associated with the greatest number of sleep phenotypes
344 from all sleep phenotype domains except for the insomnia domain.

345
346 We also compared the similarities among sleep phenotype domains by calculating the Dice Similarity
347 Coefficients (DSC) which quantifies the associated metabolites overlap between any two sleep domains.
348 As shown in Figure 7, the highest overlap was observed between HR during sleep phenotypes and SDB
349 (DSC=0.35), and with sleep timing domain (DSC=0.35), respectively. The overlap between SDB and sleep
350 timing domain was lower (DSC=0.26). High overlap was also observed between sleep duration and sleep
351 timing (DSC=0.31). The lowest levels of overlap were observed for sleep duration traits and SDB domain
352 (DSC=0.13) and sleep duration and HR during sleep phenotypes (DSC=0.19).

353
354 Bipartite network analysis based on statistically significant associations between
355 metabolites and sleep phenotypes

356 Figure 8 shows a network of nodes and edges in which each node represents either a sleep phenotype
357 domain or a metabolomic subpathway, while each edge represents one or more statistically significant
358 associations between two nodes based on the single metabolite association analysis results (FDR-
359 adjusted $p < 0.05$). The width of the edge indicates the number of statistically significant associations
360 between two nodes. Some metabolomic subpathways have exclusive connections with one sleep

361 domain, such as fatty acid and dicarboxylate subpathways and sleep timing domain. Others tend to have
362 many connections with multiple sleep domains, such as sphingomyelins with sleep timing, heart rate
363 and SDB phenotype domain. The placement of the nodes reflects their connectivity similarities – the
364 closer two nodes are, the more similar their overall connectivity patterns are. SDB and heart rate
365 phenotypes, for instance, are closely placed together which also corresponds to the high DSC values
366 between the two groups (Figure 7). Metabolomic subpathways located close to the center of the
367 network indicates their connectivity with multiple sleep domains, while metabolomic subpathways
368 located at the outskirts of the network indicates their connections with one or few closely located sleep
369 phenotype domains. For instance, the cluster of fatty acids (i.e., fatty acid, dicarboxylate, fatty acid
370 metabolism (acyl carnitine, hydroxy), fatty acid metabolism (acyl choline), long chain saturated fatty
371 acid, and fatty acid, monohydroxy, appearing in the left quadrant of Figure 8) are mostly connected to
372 sleep timing, heart rate, and insomnia sleep domains, while lipids such as lysophospholipid, secondary
373 bile acid metabolism and androgenic steroids are connected to sleep duration and SDB phenotype
374 domain.

375
376 Treated as a bipartite network model, we further characterized the connectivity between the
377 metabolites, grouped at the superpathway level, and sleep domains quantified by network properties.
378 The number of connected metabolites is the lowest for energy metabolites and the highest for lipids,
379 which corresponds to the number of connected sleep phenotypes for the two superpathways and the
380 average connection per node (i.e., links per node) (Figure 9). Similar trend can be observed for the mean
381 number of shared metabolites among sleep phenotypes – the number of overlapping metabolites for
382 sleep phenotypes is the highest among lipids and lowest among energy metabolites, which suggests that
383 more lipids are connected to the same sleep phenotype compared to metabolites from other
384 superpathways. The mean number of shared sleep phenotypes among metabolites, however, peaked

385 for carbohydrates and cofactors and vitamins, followed by lipids, suggesting metabolites from the
386 former two superpathways are more likely to be connected to similar sleep phenotypes than other
387 superpathways, despite their low number of total metabolites.

388

389 Cluster coefficient, the number of realized links divided by the number of possible links, is the highest
390 among carbohydrates and nucleotides. This pattern indicates a higher degree of overlap in the
391 neighboring nodes among carbohydrates and nucleotides, compared to unnamed metabolites and
392 when combining all metabolites together regardless of superpathways, which may suggest a lower
393 degree of diversity in terms of their connectivity patterns.

394

395 Nestedness of a graph is the property where groups of connected nodes are nested within larger groups
396 of connected nodes. Two versions of nestedness measures were calculated. Both weighted (using the
397 number of sleep domain-superpathway associations) and non-weighted nestedness measures generally
398 agree, and suggest relatively high nestedness among lipids, amino acids, unnamed metabolites as well
399 as combining all metabolites together, and relatively low nestedness among energy, carbohydrates and
400 nucleotides (Supplementary Table S6).

401

402 Modularity Q is a measure of community structure. Modularity values are higher when there are more
403 clusters of connected nodes in a graph. Modularity Q was the highest among energy (Q=0.57) and
404 xenobiotics (Q=0.57), followed by peptides (Q=0.52), and the lowest among cofactors and vitamins
405 (Q=0.39) and lipids (Q=0.4), suggesting that among the former three superpathways, there are groups of
406 metabolites in the superpathway that tend to connect to the same sleep phenotype. Among lipids, and

407 cofactors and vitamins, there are less such “communities” of metabolites with similar sleep domain
408 connections.

409

410 Discussion

411

412 In this study we performed single metabolite association analyses for a variety of sleep phenotypes
413 including SDB, insomnia, sleep duration, sleep timing, and heart rate measures during sleep, adjusted
414 for common demographic covariates and BMI, and identified metabolites with statistically significant
415 associations with each sleep phenotype. Taking a network analysis approach, we developed a bipartite
416 network based on the identified associations between metabolites and sleep phenotypes. This approach
417 characterizes the connectivity patterns of metabolites from different metabolite super- and sub-
418 pathways, and identifies “neighboring” metabolites and sleep phenotypes that are inter-connected, as
419 well as metabolomic subpathways that may be of interest for further study.

420

421 Many of the metabolites identified in our study as associated with sleep phenotypes have been
422 previously reported to be linked to various sleep phenotypes, usually concurring with our findings.
423 Among the top 10% of connected metabolites with the highest number of associations, regardless of
424 sleep phenotype domain, almost all metabolites were reported in prior publications (Table 3), of which
425 22 were related to sleep. These include some published results from our prior work in HCHS/SOL (i.e.,
426 non-independent findings). Glycine, for instance, has shown associations with multiple SDB phenotypes,
427 and was reported to be associated with sleep deprivation (44,45). In a recent study, glycine was found to
428 be depleted in blood among Hispanic/Latino individuals with severe SDB, likely due to microbiome
429 changes in an oxygen-poor environment (46). Four other metabolites – 1-oleoyl-GPE (18:1), 1-(1-enyl-

430 palmitoyl)-GPC (P-16:0)*, pregnenolone sulfate, 5alpha-pregnan-3beta,20alpha-diol monosulfate (2),
431 previously reported to be associated with novel SDB phenotype metrics after dimension reduction in our
432 prior work in HCHS/SOL (43), were here also associated with several SDB phenotypes as well as with
433 phenotypes from domains including insomnia, sleep timing, and heart rate during sleep.

434

435 Some of the metabolites identified in this study were also reported to be associated with other
436 comorbidities such as cardiovascular, metabolic, and neurodegenerative diseases, potentially
437 connecting sleep to a wide range of chronic adverse health outcomes. For instance, 1-
438 carboxyethylisoleucine and 1-carboxyethylphenylalanine, both associated with sleep duration, SDB,
439 sleep timing and heart rate during sleep, were both reported to be altered in blood among type II
440 diabetes patients (47), while 1-carboxyethylisoleucine, among several other branched-chain amino acid
441 metabolites, was also altered in serum levels among gestational diabetes patients (48). Another amino
442 acid from the top 10% connected metabolites list, 1-carboxyethylvaline, was reported to be associated
443 with hypertension (49). Hexadecanedioate (C16-DC), a dicarboxylate fatty acid, was reported to be
444 associated with blood pressure regulation (50) and neurodegenerative diseases (51). Beta-
445 cryptoxanthin, an antioxidant and pre-vitamin A carotenoid found in fruits and vegetables, was found
446 positively associated with cognition (analysis was not adjusted for sleep traits) in individuals of diverse
447 race/ethnic backgrounds (52). These metabolites suggest some shared biochemical mechanisms
448 between sleep and other chronic adverse health outcomes.

449

450

Table 3: Top sleep-connected metabolites and their relevant previously-reported associations.

Metabolite	Super pathway	Sub pathway	Prior publications
glycine	Amino Acid	Glycine, Serine and Threonine Metabolism	Sleep deprivation (44), OSA (46)
N-acetylglycine	Amino Acid	Glycine, Serine and Threonine Metabolism	Sleep restriction (53)
1-carboxyethylisoleucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism	T2DM(47) and GDM (48)
1-carboxyethylvaline	Amino Acid	Leucine, Isoleucine and Valine Metabolism	HTN (49)
lysine	Amino Acid	Lysine Metabolism	Insomnia (54), repeated sleep disruption (55)
1-carboxyethylphenylalanine	Amino Acid	Phenylalanine Metabolism	T2DM (47)
vanillylmandelate (VMA)	Amino Acid	Tyrosine Metabolism	Sleep deprivation (56)
pro-hydroxy-pro	Amino Acid	Urea cycle; Arginine and Proline Metabolism	Sleep midpoint (27)
erythronate*	Carbohydrate	Aminosugar Metabolism	Sleep midpoint and wake time (57)
1-methylnicotinamide	Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism	Sleep restriction (58)
N1-Methyl-2-pyridone-5-carboxamide	Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism	Sleep restriction (58)
nicotinamide N-oxide	Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism	Anti-inflammatory effect (59)
beta-cryptoxanthin	Cofactors and Vitamins	Vitamin A Metabolism	Sleep duration (60), sleepiness and sleep disturbance (61) cognitive function (52)
behenoyl dihydrosphingomyelin (d18:0/22:0)*	Lipid	Dihydrosphingomyelins	Physical exercise (62)
sphingomyelin (d18:0/20:0, d16:0/22:0)*	Lipid	Dihydrosphingomyelins	Blood pressure (63)
hexadecanedioate (C16-DC)	Lipid	Fatty Acid, Dicarboxylate	Blood pressure regulation (50), neurodegenerative diseases (51)
1-linoleoyl-GPE (18:2)*	Lipid	Lysophospholipid	Depression (64)
1-oleoyl-GPE (18:1)	Lipid	Lysophospholipid	SDB (43)
1-(1-enyl-palmitoyl)-GPC (P-16:0)*	Lipid	Lysoplasmalogen	SDB (43)
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Lipid	Phosphatidylethanolamine (PE)	Infant sepsis (65)
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	Lipid	Phosphatidylethanolamine (PE)	OSA (24)
pregnenolone sulfate	Lipid	Pregnenolone Steroids	SDB (43), REM sleep (66)
glycochenodeoxycholate	Lipid	Primary Bile Acid Metabolism	Poor sleep quality (67)
glycochenodeoxycholate 3-sulfate	Lipid	Primary Bile Acid Metabolism	OSA (29)

glycocholate	Lipid	Primary Bile Acid Metabolism	OSA (26)
taurochenodeoxycholate	Lipid	Primary Bile Acid Metabolism	OSA (26), sleep midpoint (27)
5alpha-pregnan-3beta,20alpha-diol monosulfate (2)	Lipid	Progestin Steroids	SDB (43)
glycoursodeoxycholate	Lipid	Secondary Bile Acid Metabolism	Bed time (68)
ursodeoxycholate	Lipid	Secondary Bile Acid Metabolism	OSA intermittent hypoxia/hypercapnia (69)
pentose acid*	Partially Characterized Molecules	Partially Characterized Molecules	REM vs wakefulness (70)
gamma-glutamylisoleucine*	Peptide	Gamma-glutamyl Amino Acid	Sleep midpoint (27), wake time and sleep midpoint (57)
gamma-glutamyltryptophan	Peptide	Gamma-glutamyl Amino Acid	Late-onset sepsis among premature infants (71)
mannonate*	Xenobiotics	Food Component/Plant	Undernutrition (72)

451

452 When working with complex data such as untargeted metabolomic profiling and multiple phenotypes,
453 effectively summarize data and develop useful insights is challenging (73). Different visualization and
454 analytical approaches have been developed to facilitate this process (74). Here, we applied a systems
455 biological approach – network analysis which has been widely used in gene expression, gene regulation,
456 gene-disease network, and drug-drug interaction studies, on statistical relations among multiple
457 phenotypes and metabolites. This data-driven network approach is different from knowledge-based
458 network construction approaches (75) built on biochemical relations such as KEGG networks (76). Here,
459 the network was built on inferred statistical relations among metabolites and phenotypes, leveraging,
460 for interpretation purposes, well-studied network properties metrics from other fields such as ecology,
461 socioeconomics, neuroscience, drug-disease networks, among others (77–80). One interesting
462 observation is the relatively high nestedness among lipids and amino acids. High nestedness indicates
463 that metabolites from these two superpathways form such a structure that metabolites with fewer
464 connections (referred to as “degree”) are more likely to “connect”, via their mutual connections with
465 sleep domains, with metabolites with higher degree of connections, rather than with other metabolites

466 with a similar low degree of connectiveness. An intuitive depiction of such network is a “core-periphery”
467 structure, in which a “core” of nodes is connected with other nodes, where a “periphery” of nodes tends
468 to only connect with the nodes in the core. Contextually, among lipids and amino acids associated with
469 sleep phenotypes, higher nestedness, compared to other superpathways, implies that these
470 superpathways are more likely to have a subset of metabolites that play a “key” role forming
471 connections with many, and the same, sleep phenotypes, rather than an evenly distributed network
472 where metabolites form connections with various sleep phenotypes from different domains in a random
473 manner. This is supported by the pattern observed in Figure 8 where subpathways from amino acids and
474 lipids superpathways tend to be placed in the center of the network graph. Metabolites with high
475 connectiveness from these two super-pathways likely have roles in shared biological processes across
476 multiple sleep phenotypes, especially considering 24 out of 37 top 10% connected metabolites are
477 either amino acids or lipids. Modularity metrics, on the other hand, measure how well a network can be
478 partitioned into clusters or compartments in which there are dense connections internally and sparser
479 connections with other clusters (81). Among energy and peptide metabolites associated with sleep
480 phenotypes, relatively higher modularity compared to other superpathways suggests the existence of
481 such subgroups with distinct relationships with sleep phenotypes.

482

483 HR domain had the highest percentage of metabolite associations of these assessed. HR during sleep
484 reflects activity of the autonomic nervous system, and is influenced by cardiac function, sleep stage (i.e.,
485 lowest in deep restorative sleep and highest and most variable in REM sleep and wakefulness during the
486 sleep period), and sleep apnea-related heart rate response (82,83). Notably, both low and elevated
487 heart rate response to SDB events are associated with biomarkers of cardiovascular disease, while
488 elevated heart rate response to SDB events is predictive of incident fatal and non-fatal CVD (84). The
489 finding that HR during sleep associated with hundreds of metabolites, from all superpathways, may

490 reflect the multiple biological mechanisms that underlie this phenotype, as well as support the sleep-
491 related HR as a marker of multiple biological processes that may be targeted for interventions. Sleep
492 timing also was associated with a relatively high number of metabolite associations, supportive of
493 growing data implicating timing-related behaviors on metabolic outcomes (85), as well as the correlation
494 of sleep timing with other sleep domains (86). In contrast, fewer associations were observed between
495 SDB phenotypes and metabolites. This may be because that SDB measured by traditional metrics such as
496 the AHI may poorly characterize disease that is influenced by multiple mechanistic pathways (87).

497
498 An analytic choice that we made that is worth discussing is missing metabolite data imputation. Here,
499 we imputed missing values of metabolites with no more than 25% missing values. For non-xenobiotic
500 metabolites, we selected the imputation method based on an empirical investigation of the proportion
501 of replicated associations between batches, and the selected method was multiple imputation that
502 included all metabolite and other measures (covariates and lab values) that have strong associations
503 with metabolite levels. It is natural to question whether this may somehow bias results: for example, is it
504 possible that, for instance, using diabetes and BMI (among the rest) to impute metabolite values will
505 lead to metabolite values that are “too reflective” or BMI and diabetes and therefore will somehow
506 generate spurious metabolite associations with sleep measures that are associated with BMI and
507 diabetes? The answer is that this is very unlikely. Generation of metabolite values that are overly similar
508 to BMI or diabetes (in this example) by the predictive mean matching function suggests overfitting to
509 the values of these covariates. Such overfitting will reduce the likelihood of replication of associations in
510 the second batch, rather than increase it. Further, the predictive mean matching has some randomness
511 due to sampling, further limiting overfitting to the variables used in imputation. Finally, it is important to
512 note that the correlation between metabolites and covariates, as with sleep, is an inherent

513 characteristic of this biological signal (i.e. it is a “feature, not a bug”), as are the metabolite associations
514 with sleep measures. Therefore, leveraging this characteristic is useful.

515

516 There are a few strengths to this study: we looked at multiple sleep phenotypes simultaneously which
517 provides a unique opportunity for pattern recognition across phenotypes. We also explored the network
518 analysis approach in summarizing the metabolite-sleep phenotypes associations which enabled us to
519 leverage well developed network property metrics from other fields to offer new insights and potentially
520 lead to more hypothesis generation. Additionally, we compared the metabolite-sleep phenotype
521 associations in combined sexes and sex-stratified study populations, recognizing more work is needed to
522 further understand the implications of potential differences. Lastly, the atlas created in this study will be
523 a useful resource for the scientific community. This study also has a few limitations. First, our study
524 population is based on the HCHS/SOL cohort, representative of the Hispanic/Latino population in the US.
525 Although it’s important to study under-represented populations such as Hispanic/Latino individuals,
526 further studies on other populations are needed to increase the generalizability of the findings. Second,
527 the network analysis conducted in this study is not to make network inference but to summarize the
528 associations results. Third, due to the large breadth of analyses we did not account for medication use in
529 this work. This can influence associations as medications can affect metabolite levels and sleep
530 phenotypes. Similarly, fourth, our primary model for which we summarize the results was not adjusted
531 for comorbidities. In all, one cannot infer mechanisms and directionality of associations from these
532 analyses. Fifth, another limitation is that the objective overnight sleep measures did not use EEG,
533 limiting SDB (e.g. we do not have measures of arousals) as well as potential sleep staging measures.
534 Finally, any comparison between the sleep domains used in this analysis is limited by the available sleep
535 phenotypes, their number, and the correlation patterns between them.

536

537 In summary, we studied the associations between multiple sleep phenotypes from multiple sleep
538 domains and the metabolomic environment in a large population-based cohort study. Using network
539 analysis, we were able to visualize the interconnectedness between multiple sleep phenotypes and
540 associated metabolites simultaneously, which provides an opportunity to glean into connectivity
541 patterns that otherwise might be obscure when presented as individual relationships. We also created a
542 resource for the sleep research community that will facilitate hypothesis generation in future
543 metabolomic studies on sleep health. As sleep is highly affected by the social and built environment, it
544 would be important, in the future, to use metabolomics to glean into the pathways by which the
545 environment impacts sleep.

546

547 Figure legends

548

549 **Figure 1. Study design diagram**

550 **Figure 2. Scatter plot of single metabolite association analysis with sleep phenotypes**

551 $-\log_{10}(\text{FDR-}p)$ is based on FDR-adjusted p in which the raw p was derived by accounting for the complex
552 sampling design-based degrees of freedom, using adjusted standard errors to compute the t-statistic in
553 single metabolite association analysis with each sleep phenotype as dependent variables, while the FDR
554 adjustment was based on the Benjamini-Hochberg method to control false discovery rate (FDR) for
555 multiple testing among all metabolites in all models for each sleep phenotype. Black horizontal line
556 indicates $\text{FDR-}p=0.05$.

557

558 **Figure 3. Number of statistically significant associations between metabolites and sleep phenotypes**
559 **by domain and sex strata**

560 Associations between metabolite and sleep phenotype are identified as statistically significant based on
561 FDR- $p < 0.05$. The raw p -value was derived by accounting for the complex sampling design-based degrees
562 of freedom, using adjusted standard errors to compute the t-statistic in single metabolite association
563 analysis with each sleep phenotype as dependent variables. The FDR adjustment was based on the
564 Benjamini-Hochberg method to control false discovery rate (FDR) for multiple testing among all
565 metabolites in all models for each sleep phenotype.

566

567 **Figure 4. Number of statistically significant associations between metabolites and sleep phenotypes**
568 **aggregated by subpathway and sleep phenotype domain**

569

570 **Figure 5. Percentage of statistically significant associations between metabolites and sleep**
571 **phenotypes aggregated by superpathway and sleep phenotype domain**

572

573 **Figure 6. Number of significant associations aggregated by sleep phenotype domain among the top**
574 **10% connected metabolites with the most statistically significant associations**

575

576 **Figure 7. Dice coefficient matrix of sleep phenotype domain based on associated metabolites**

577 Dice coefficient was calculated based on the shared associations with metabolites between any two
578 sleep phenotype domain over the total significant associations for both sleep phenotype domains

579 combined. A high value (maximum value of 1) indicates complete overlap, while a low value (minimum
580 value of 0) indicates no overlap between two sleep phenotype domains.

581 **Figure 8. Network based on associations between metabolites and sleep phenotypes aggregated by**
582 **subpathway and sleep phenotype domain using the Fruchterman-Reingold force-directed algorithm**

583 Here we present a network of nodes and edges in which each node represents either a sleep phenotype
584 domain or a metabolomic subpathway, while each edge represents one or more statistically significant
585 associations between two nodes based on the single metabolite association analysis results (FDR-
586 corrected $p < 0.05$). The width of the edge (i.e., degree) indicates the number of statistically significant
587 associations between two nodes. The placement of nodes is based on connectivity similarity between
588 nodes, where similar connectivity is reflected as proximity of nodes.

589

590 **Figure 9. Network structural property metrics summarized by superpathway**

591 Network properties metrics were calculated for the bipartite network based on the significant
592 associations between metabolites and sleep phenotypes (FDR-corrected $P < 0.05$). Cluster coefficients is
593 the number of realized links divided by the number of possible links
594 [<https://www.rdocumentation.org/packages/bipartite/versions/2.19/topics/networklevel>]. Nestedness
595 measures how the interactions of less connected nodes are a subset of the interactions of more
596 connected nodes. A value of 0 indicates high nestedness, while a value of 100 indicates “chaos”.
597 Weighted nestedness considers interaction frequencies of the network, proposed by Galeano et al.(75).
598 It ranges between 1 (perfect nestedness) and 0 (perfect chaos). Modularity Q is a measure to quantify
599 how well a network can be partitioned into different groups of nodes such that nodes that belong to the
600 same group are more likely to be connected than nodes that belong to different groups.

601 **Supplementary Figure S1: Study sample selection and pre-processing for the metabolomic analysis.**

602 **Supplementary Figure S2: Number of statistically significant associations between metabolites and**
603 **sleep phenotypes aggregated by subpathway and sleep phenotype domain stratified by sex.**

604 Data availability

605
606 HCHS/SOL data are available through application to the data base of genotypes and phenotypes (dbGaP)
607 accession phs000810. HCHS/SOL metabolomics data are available via data use agreement with the
608 HCHS/SOL Data Coordinating Center (DCC) at the University of North Carolina at Chapel Hill, see
609 collaborators website: <https://sites.csc.unc.edu/hchs/>. The metabolite association data generated in
610 this study are provided in a shiny app https://bws-bidmc.shinyapps.io/20240410_for_shinyappsio/.

611

612 Code availability

613 The code used in this work has been deposited in the public repository
614 https://github.com/yzhang104/HCHS_SOL_Sleep_Metabolomics_Atlas.git.

615

616 Ethics statement

617 The HCHS/SOL was approved by the institutional review boards (IRBs) at each field center, where all
618 participants gave written informed consent, and by the Non-Biomedical IRB at the University of North
619 Carolina at Chapel Hill, to the HCHS/SOL Data Coordinating Center. All IRBs approving the HCHS/SOL
620 study are: Non-Biomedical IRB at the University of North Carolina at Chapel Hill. Chapel Hill, NC; Einstein
621 IRB at the Albert Einstein College of Medicine of Yeshiva University. Bronx, NY; IRB at Office for the
622 Protection of Research Subjects (OPRS), University of Illinois at Chicago. Chicago, IL; Human Subject
623 Research Office, University of Miami. Miami, FL; Institutional Review Board of San Diego State

624 University, San Diego, CA. All methods and analyses of HCHS/ SOL participants' materials and data were
625 carried out in accordance with human subject research guidelines and regulations. This work was
626 approved by the Mass General Brigham IRB and by the Beth Israel Deaconess Medical Center Committee
627 on Clinical Investigations.

628

629 Declaration of interests

630 Dr. Redline discloses consulting relationships with Eli Lilly Inc. Additionally, Dr. Redline serves as an
631 unpaid member of the Apnimed Scientific Advisory Board, as an unpaid board member for the Alliance
632 for Sleep Apnoea Partners, and has received loaned equipment for a multi-site study: oxygen
633 concentrators from Philips Respironics and polysomnography equipment from Nox Medical.

634

635

636 Acknowledgements

637 The authors thank the staff and participants of HCHS/SOL for their important contributions.
638 Investigators website - <http://www.csc.unc.edu/hchs/>. This work was supported by National Heart
639 Lung and Blood Institute (NHLBI) grants R01HL161012 to TS, R35HL135818 to SR, and National Institute
640 on Aging grant R01AG80598 to TS. Support for metabolomics data was graciously provided by the JLH
641 Foundation (Houston, Texas) and by NHLBI grant R01HL141824. The Hispanic Community Health
642 Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung,
643 and Blood Institute (NHLBI) to the University of North Carolina (HHSN268201300001I / N01-HC-65233),
644 University of Miami (HHSN268201300004I / N01-HC-65234), Albert Einstein College of Medicine
645 (HHSN268201300002I / N01-HC-65235), University of Illinois at Chicago (HHSN268201300003I / N01-
646 HC-65236 Northwestern Univ), and San Diego State University (HHSN268201300005I / N01-HC-65237).

647 The following Institutes/Centers/Offices have contributed to the HCHS/SOL through a transfer of funds
648 to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on
649 Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research,
650 National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological
651 Disorders and Stroke, NIH Institution-Office of Dietary Supplements.

652

653 Contributions of the authors

654 Ying Z performed association analyses, network analyses, and data visualization. BWS developed the
655 R/shiny app for visualization of results from single metabolite association analysis. Yu Z performed data
656 imputation. DAW provided code for modeling circular sleep phenotypes. BY, QQ, and EB designed and
657 established the metabolomics ancillary studies. MA, LA-S, MD, RK, and JC participated in HCHS/SOL
658 study design, recruitment, and operations. Ying Z and TS drafted the manuscript. BWS, Yu Z, DAW, BY,
659 QQ, MA, LA-S, EB, MD, RK, JC, and SR critically reviewed and approved the manuscript. TS supervised the
660 work.

661 References

662

663

664

- 665 1. Chattu VK, Manzar MD, Kumary S, Burman D, Spence DW, Pandi-Perumal SR. The global problem
666 of insufficient sleep and its serious public health implications. *Healthcare (Basel)*. 2018 Dec
667 20;7(1).
- 668 2. Worley SL. The extraordinary importance of sleep: the detrimental effects of inadequate sleep on
669 health and public safety drive an explosion of sleep research. *P T*. 2018 Dec;43(12):758–63.
- 670 3. Ramar K, Malhotra RK, Carden KA, Martin JL, Abbasi-Feinberg F, Aurora RN, et al. Sleep is
671 essential to health: an American Academy of Sleep Medicine position statement. *J Clin Sleep*
672 *Med*. 2021 Oct 1;17(10):2115–9.

- 673 4. Ioachimescu OC. From seven sweethearts to life begins at eight thirty: A journey from life's
674 simple 7 to life's essential 8 and beyond. *J Am Heart Assoc*. 2022 Nov;11(21):e027658.
- 675 5. Buysse DJ. Sleep health: can we define it? Does it matter? *Sleep*. 2014 Jan 1;37(1):9–17.
- 676 6. Chung J, Goodman M, Huang T, Wallace ML, Lutsey PL, Chen JT, et al. Multi-dimensional sleep
677 and mortality: The Multi-Ethnic Study of Atherosclerosis. *Sleep*. 2023 Sep 8;46(9).
- 678 7. Lee S, Mu CX, Wallace ML, Andel R, Almeida DM, Buxton OM, et al. Multidimensional sleep
679 health problems across middle and older adulthood predict early mortality. *J Gerontol A Biol Sci*
680 *Med Sci*. 2024 Mar 1;79(3).
- 681 8. Lee S, Mu CX, Wallace ML, Andel R, Almeida DM, Buxton OM, et al. Sleep health composites are
682 associated with the risk of heart disease across sex and race. *Sci Rep*. 2022 Feb 7;12(1):2023.
- 683 9. Sofer T, Kurniansyah N, Murray M, Ho Y-L, Abner E, Esko T, et al. Genome-wide association study
684 of obstructive sleep apnoea in the Million Veteran Program uncovers genetic heterogeneity by
685 sex. *EBioMedicine*. 2023 Apr;90:104536.
- 686 10. Madrid-Valero JJ, Sánchez-Romera JF, Martínez-Selva JM, Ordoñana JR. Phenotypic, genetic and
687 environmental architecture of the components of sleep quality. *Behav Genet*. 2022 Sep;52(4–
688 5):236–45.
- 689 11. Wang H, Lane JM, Jones SE, Dashti HS, Ollila HM, Wood AR, et al. Genome-wide association
690 analysis of self-reported daytime sleepiness identifies 42 loci that suggest biological subtypes.
691 *Nat Commun*. 2019 Aug 13;10(1):3503.
- 692 12. Sweetman A, Lack L, McEvoy RD, Smith S, Eckert DJ, Osman A, et al. Bi-directional relationships
693 between co-morbid insomnia and sleep apnea (COMISA). *Sleep Med Rev*. 2021 Dec;60:101519.
- 694 13. Amatory J, Azarbarzin A, Younes M, Jordan AS, Wellman A, Eckert DJ. Arousal intensity is a
695 distinct pathophysiological trait in obstructive sleep apnea. *Sleep*. 2016 Dec 1;39(12):2091–100.
- 696 14. Keenan BT, Kim J, Singh B, Bittencourt L, Chen N-H, Cistulli PA, et al. Recognizable clinical
697 subtypes of obstructive sleep apnea across international sleep centers: a cluster analysis. *Sleep*.
698 2018 Mar 1;41(3).
- 699 15. Zinchuk AV, Jeon S, Koo BB, Yan X, Bravata DM, Qin L, et al. Polysomnographic phenotypes and
700 their cardiovascular implications in obstructive sleep apnoea. *Thorax*. 2018 May;73(5):472–80.
- 701 16. Zhang L, Wei T-T, Li Y, Li J, Fan Y, Huang F-Q, et al. Functional Metabolomics Characterizes a Key
702 Role for N-Acetylneuraminic Acid in Coronary Artery Diseases. *Circulation*. 2018 Mar
703 27;137(13):1374–90.
- 704 17. Buerger T, Steinfeldt J, Ruyoga G, Pietzner M, Bizzarri D, Vojinovic D, et al. Metabolomic profiles
705 predict individual multidisease outcomes. *Nat Med*. 2022 Nov;28(11):2309–20.
- 706 18. Nikolic SB, Sharman JE, Adams MJ, Edwards LM. Metabolomics in hypertension. *J Hypertens*.
707 2014 Jun;32(6):1159–69.

- 708 19. Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. *Nat Rev*
709 *Cardiol.* 2011 Sep 20;8(11):630–43.
- 710 20. Yu B, Flexeder C, McGarrah RW, Wyss A, Morrison AC, North KE, et al. Metabolomics identifies
711 novel blood biomarkers of pulmonary function and COPD in the general population. *Metabolites.*
712 2019 Apr 1;9(4).
- 713 21. Depner CM, Cogswell DT, Bisesi PJ, Markwald RR, Cruickshank-Quinn C, Quinn K, et al. Developing
714 preliminary blood metabolomics-based biomarkers of insufficient sleep in humans. *Sleep.* 2020
715 Jul 13;43(7).
- 716 22. He S, Granot-Herskovitz E, Zhang Y, Bressler J, Tarraf W, Yu B, et al. Blood metabolites predicting
717 mild cognitive impairment in the study of Latinos-investigation of neurocognitive aging
718 (HCHS/SOL). *Alzheimers Dement (Amst).* 2022 Feb 23;14(1):e12259.
- 719 23. Fritz J, Huang T, Depner CM, Zeleznik OA, Cespedes Feliciano EM, Li W, et al. Sleep duration,
720 plasma metabolites, and obesity and diabetes: a metabolome-wide association study in US
721 women. *Sleep.* 2023 Jan 11;46(1).
- 722 24. Zhang Y, Ngo D, Yu B, Shah NA, Chen H, Ramos AR, et al. Development and validation of a
723 metabolite index for obstructive sleep apnea across race/ethnicities. *Sci Rep.* 2022 Dec
724 16;12(1):21805.
- 725 25. Lin BM, Zhang Y, Yu B, Boerwinkle E, Thygarajan B, Yunes M, et al. Metabolome-wide association
726 study of estimated glomerular filtration rates in Hispanics. *Kidney Int.* 2022 Jan;101(1):144–51.
- 727 26. Pinilla L, Benítez ID, Santamaria-Martos F, Targa A, Moncusí-Moix A, Dalmases M, et al. Plasma
728 profiling reveals a blood-based metabolic fingerprint of obstructive sleep apnea. *Biomed*
729 *Pharmacother.* 2022 Jan;145:112425.
- 730 27. Xiao Q, Derkach A, Moore SC, Zheng W, Shu X-O, Gu F, et al. Habitual Sleep and human plasma
731 metabolomics. *Metabolomics.* 2017 May;13(5).
- 732 28. Humer E, Pieh C, Brandmayr G. Metabolomics in sleep, insomnia and sleep apnea. *Int J Mol Sci.*
733 2020 Sep 30;21(19).
- 734 29. Xu H, Zheng X, Qian Y, Guan J, Yi H, Zou J, et al. Metabolomics profiling for obstructive sleep
735 apnea and simple snorers. *Sci Rep.* 2016 Aug 2;6:30958.
- 736 30. Lebkuchen A, Carvalho VM, Venturini G, Salgueiro JS, Freitas LS, Dellavance A, et al. Metabolomic
737 and lipidomic profile in men with obstructive sleep apnoea: implications for diagnosis and
738 biomarkers of cardiovascular risk. *Sci Rep.* 2018 Jul 26;8(1):11270.
- 739 31. Mohit, Tomar MS, Sharma D, Nandan S, Pateriya A, Shrivastava A, et al. Emerging role of
740 metabolomics for biomarker discovery in obstructive sleep apnea. *Sleep Breath.* 2023
741 Aug;27(4):1247–54.
- 742 32. Lavange LM, Kalsbeek WD, Sorlie PD, Avilés-Santa LM, Kaplan RC, Barnhart J, et al. Sample design
743 and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol.*
744 2010 Aug;20(8):642–9.

- 745 33. Sorlie PD, Avilés-Santa LM, Wassertheil-Smoller S, Kaplan RC, Daviglius ML, Giachello AL, et al.
746 Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann*
747 *Epidemiol.* 2010 Aug;20(8):629–41.
- 748 34. Pirzada A, Cai J, Heiss G, Sotres-Alvarez D, Gallo LC, Youngblood ME, et al. Evolving science on
749 cardiovascular disease among hispanic/latino adults: JACC international. *J Am Coll Cardiol.* 2023
750 Apr 18;81(15):1505–20.
- 751 35. Redline S, Sotres-Alvarez D, Loredó J, Hall M, Patel SR, Ramos A, et al. Sleep-disordered breathing
752 in Hispanic/Latino individuals of diverse backgrounds. *The Hispanic Community Health*
753 *Study/Study of Latinos. Am J Respir Crit Care Med.* 2014 Feb 1;189(3):335–44.
- 754 36. Gonzalez S, Strizich G, Isasi CR, Hua S, Comas B, Sofer T, et al. Consent for Use of Genetic Data
755 among US Hispanics/Latinos: Results from the Hispanic Community Health Study/ Study of
756 Latinos. *Ethn Dis.* 2021 Oct 21;31(4):547–58.
- 757 37. Kim S, SenGupta A. Regressions involving circular variables: an overview. In: Chattopadhyay AK,
758 Chattopadhyay G, editors. *Statistics and its applications: platinum jubilee conference, kolkata,*
759 *india, december 2016.* Singapore: Springer Singapore; 2018. p. 25–33.
- 760 38. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc.* 1996 Jun;91(434):473.
- 761 39. Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. ACAT: A Fast and Powerful p Value
762 Combination Method for Rare-Variant Analysis in Sequencing Studies. *Am J Hum Genet.* 2019
763 Mar 7;104(3):410–21.
- 764 40. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach
765 to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological).* 1995
766 Jan;57(1):289–300.
- 767 41. Sørensen T, Sørensen T, Biering-Sørensen T, Sørensen T, Sorensen JT. A method of establishing
768 group of equal amplitude in plant sociobiology based on similarity of species content and its
769 application to analyses of the vegetation on Danish commons. 1948.
- 770 42. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Softw: Pract Exper.*
771 1991 Nov;21(11):1129–64.
- 772 43. Zhang Y, Yu B, Qi Q, Azarbarzin A, Chen H, Shah NA, et al. Metabolomic profiles of sleep-
773 disordered breathing are associated with hypertension and diabetes mellitus development. *Nat*
774 *Commun.* 2024 Feb 28;15(1):1845.
- 775 44. Gou X-J, Cen F, Fan Z-Q, Xu Y, Shen H-Y, Zhou M-M. Serum and brain metabolomic variations
776 reveal perturbation of sleep deprivation on rats and ameliorate effect of total ginsenoside
777 treatment. *Int J Genomics.* 2017 Aug 16;2017:5179271.
- 778 45. Ma W-N, Zhou M-M, Gou X-J, Zhao L, Cen F, Xu Y, et al. Urinary Metabolomic Study of
779 Chlorogenic Acid in a Rat Model of Chronic Sleep Deprivation Using Gas Chromatography-Mass
780 Spectrometry. *Int J Genomics.* 2018 Feb 11;2018:1361402.

- 781 46. Elgart M, Zhang Y, Zhang Y, Yu B, Kim Y, Zee PC, et al. Anaerobic pathogens associated with OSA
782 may contribute to pathophysiology via amino-acid depletion. *EBioMedicine*. 2023
783 Dec;98:104891.
- 784 47. Liu Y, Gan L, Zhao B, Yu K, Wang Y, Männistö S, et al. Untargeted metabolomic profiling identifies
785 serum metabolites associated with type 2 diabetes in a cross-sectional study of the Alpha-
786 Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study. *Am J Physiol Endocrinol Metab*.
787 2023 Feb 1;324(2):E167–75.
- 788 48. Harville EW, Bazzano L, Qi L, He J, Dorans K, Perng W, et al. Branched-chain amino acids, history
789 of gestational diabetes, and breastfeeding: The Bogalusa Heart Study. *Nutr Metab Cardiovasc Dis*.
790 2020 Oct 30;30(11):2077–84.
- 791 49. Shi M, He J, Li C, Lu X, He WJ, Cao J, et al. Metabolomics study of blood pressure salt-sensitivity
792 and hypertension. *Nutr Metab Cardiovasc Dis*. 2022 Jul;32(7):1681–92.
- 793 50. Menni C, Graham D, Kastenmüller G, Alharbi NHJ, Alsanosi SM, McBride M, et al. Metabolomic
794 identification of a novel pathway of blood pressure regulation involving hexadecanedioate.
795 *Hypertension*. 2015 Aug;66(2):422–9.
- 796 51. Chen H, Qiao J, Wang T, Shao Z, Huang S, Zeng P. Assessing causal relationship between human
797 blood metabolites and five neurodegenerative diseases with GWAS summary statistics. *Front*
798 *Neurosci*. 2021 Dec 9;15:680104.
- 799 52. Granot-Herskovitz E, He S, Bressler J, Yu B, Tarraf W, Rebholz CM, et al. Plasma metabolites
800 associated with cognitive function across race/ethnicities affirming the importance of healthy
801 nutrition. *Alzheimers Dement*. 2023 Apr;19(4):1331–42.
- 802 53. Bell LN, Kilkus JM, Booth JN, Bromley LE, Imperial JG, Penev PD. Effects of sleep restriction on the
803 human plasma metabolome. *Physiol Behav*. 2013 Oct 2;122:25–31.
- 804 54. Gehrman P, Sengupta A, Harders E, Ubeydullah E, Pack AI, Weljie A. Altered diurnal states in
805 insomnia reflect peripheral hyperarousal and metabolic desynchrony: a preliminary study. *Sleep*.
806 2018 May 1;41(5).
- 807 55. Bowers SJ, Vargas F, González A, He S, Jiang P, Dorrestein PC, et al. Repeated sleep disruption in
808 mice leads to persistent shifts in the fecal microbiome and metabolome. *PLoS ONE*. 2020 Feb
809 20;15(2):e0229001.
- 810 56. Müller HU, Riemann D, Berger M, Müller WE. The influence of total sleep deprivation on urinary
811 excretion of catecholamine metabolites in major depression. *Acta Psychiatr Scand*. 1993
812 Jul;88(1):16–20.
- 813 57. Gordon-Dseagu VLZ, Derkach A, Xiao Q, Williams I, Sampson J, Stolzenberg-Solomon RZ. The
814 association of sleep with metabolic pathways and metabolites: evidence from the Dietary
815 Approaches to Stop Hypertension (DASH)-sodium feeding study. *Metabolomics*. 2019 Mar
816 16;15(4):48.

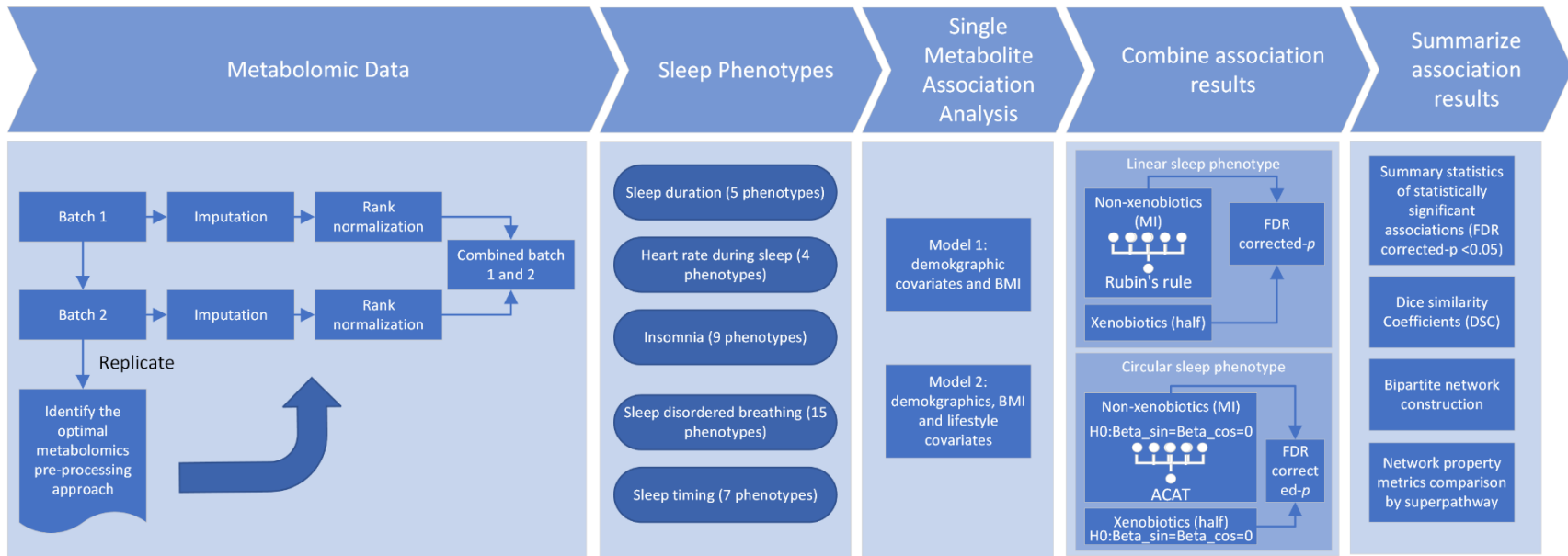
- 817 58. Sengupta A, Rhoades SD, Kim EJ, Nayak S, Grant GR, Meerlo P, et al. Sleep restriction induced
818 energy, methylation and lipogenesis metabolic switches in rat liver. *Int J Biochem Cell Biol.* 2017
819 Dec;93:129–35.
- 820 59. Song X, Cao W, Wang Z, Li F, Xiao J, Zeng Q, et al. Nicotinamide n-Oxide Attenuates HSV-1-
821 Induced Microglial Inflammation through Sirtuin-1/NF- κ B Signaling. *Int J Mol Sci.* 2022 Dec
822 16;23(24).
- 823 60. Deng M-G, Liu F, Wang K, Liang Y, Nie J-Q, Liu J. Relationship between dietary carotenoid intake
824 and sleep duration in American adults: a population-based study. *Nutr J.* 2023 Dec 8;22(1):68.
- 825 61. Beydoun MA, Gamaldo AA, Canas JA, Beydoun HA, Shah MT, McNeely JM, et al. Serum
826 nutritional biomarkers and their associations with sleep among US adults in recent national
827 surveys. *PLoS ONE.* 2014 Aug 19;9(8):e103490.
- 828 62. Gaitán JM, Moon HY, Stremlau M, Dubal DB, Cook DB, Okonkwo OC, et al. Effects of Aerobic
829 Exercise Training on Systemic Biomarkers and Cognition in Late Middle-Aged Adults at Risk for
830 Alzheimer’s Disease. *Front Endocrinol (Lausanne).* 2021 May 20;12:660181.
- 831 63. He WJ, Li C, Mi X, Shi M, Gu X, Bazzano LA, et al. An untargeted metabolomics study of blood
832 pressure: findings from the Bogalusa Heart Study. *J Hypertens.* 2020 Jul;38(7):1302–11.
- 833 64. Jansen R, Milaneschi Y, Schraner D, Kastenmuller G, Arnold M, Han X, et al. The Metabolome-
834 Wide Signature of Major Depressive Disorder. *Res Sq.* 2023 Sep 21;
- 835 65. Wang L, Cha X, Zhang Z, Qian J. Discrimination of serum metabolomics profiles in infants with
836 sepsis, based on liquid chromatography-mass spectrometer. *BMC Infect Dis.* 2023 Jan
837 23;23(1):46.
- 838 66. Darnaudéry M, Bouyer JJ, Pallarés M, Le Moal M, Mayo W. The promnesic neurosteroid
839 pregnenolone sulfate increases paradoxical sleep in rats. *Brain Res.* 1999 Feb 13;818(2):492–8.
- 840 67. Huang T, Zeleznik OA, Poole EM, Clish CB, Deik AA, Scott JM, et al. Habitual sleep quality, plasma
841 metabolites and risk of coronary heart disease in post-menopausal women. *Int J Epidemiol.* 2019
842 Aug 1;48(4):1262–74.
- 843 68. Xiang X, Chen J, Zhu M, Gao H, Liu X, Wang Q. Multiomics Revealed the Multi-Dimensional Effects
844 of Late Sleep on Gut Microbiota and Metabolites in Children in Northwest China. *Nutrients.* 2023
845 Oct 10;15(20).
- 846 69. Xue J, Allaband C, Zhou D, Poulsen O, Martino C, Jiang L, et al. Influence of intermittent
847 hypoxia/hypercapnia on atherosclerosis, gut microbiome, and metabolome. *Front Physiol.* 2021
848 Apr 8;12:663950.
- 849 70. Nowak N, Gaisl T, Miladinovic D, Marcinkevics R, Osswald M, Bauer S, et al. Rapid and reversible
850 control of human metabolism by individual sleep states. *Cell Rep.* 2021 Oct 26;37(4):109903.
- 851 71. Liu J, Zhang L, Li D, Yu X, Gao Y, Zhou Y. Intestinal metabolomics in premature infants with late-
852 onset sepsis. *Sci Rep.* 2024 Feb 26;14(1):4659.

- 853 72. Preidis GA, Keaton MA, Campeau PM, Bessard BC, Conner ME, Hotez PJ. The undernourished
854 neonatal mouse metabolome reveals evidence of liver and biliary dysfunction, inflammation, and
855 oxidative stress. *J Nutr*. 2014 Mar;144(3):273–81.
- 856 73. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted Metabolomics
857 Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectrom*. 2016 Dec;27(12):1897–
858 905.
- 859 74. Alonso A, Marsal S, Julià A. Analytical methods in untargeted metabolomics: state of the art in
860 2015. *Front Bioeng Biotechnol*. 2015 Mar 5;3:23.
- 861 75. Galeano J, Pastor JM, Iriando JM. Weighted-Interaction Nestedness Estimator (WINE): A new
862 estimator to calculate over frequency matrices. *Environmental Modelling & Software*. 2009
863 Nov;24(11):1342–6.
- 864 76. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000
865 Jan 1;28(1):27–30.
- 866 77. Ulrich W, Almeida-Neto M, Gotelli NJ. A consumer's guide to nestedness analysis. *Oikos*. 2009
867 Jan;118(1):3–17.
- 868 78. König MD, Tessone CJ, Zenou Y. Nestedness in networks: A theoretical model and some
869 applications. *Theoretical Economics*. 2014 Sep;9(3):695–752.
- 870 79. Centeno EGZ, Moreni G, Vriend C, Douw L, Santos FAN. A hands-on tutorial on network and
871 topological neuroscience. *Brain Struct Funct*. 2022 Apr;227(3):741–62.
- 872 80. Rintala TJ, Ghosh A, Fortino V. Network approaches for modeling the effect of drugs and
873 diseases. *Brief Bioinformatics*. 2022 Jul 18;23(4).
- 874 81. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E
875 Stat Nonlin Soft Matter Phys*. 2004 Feb 26;69(2 Pt 2):026113.
- 876 82. Aldredge JL, Welch AJ. Variations of heart rate during sleep as a function of the sleep cycle.
877 *Electroencephalogr Clin Neurophysiol*. 1973 Aug;35(2):193–8.
- 878 83. Yang C-K, Jordan AS, White DP, Winkelman JW. Heart rate response to respiratory events with or
879 without leg movements. *Sleep*. 2006 Apr;29(4):553–6.
- 880 84. Azarbarzin A, Sands SA, Younes M, Taranto-Montemurro L, Sofer T, Vena D, et al. The Sleep
881 Apnea-Specific Pulse-Rate Response Predicts Cardiovascular Morbidity and Mortality. *Am J Respir
882 Crit Care Med*. 2021 Jun 15;203(12):1546–55.
- 883 85. Duffy JF, Abbott SM, Burgess HJ, Crowley SJ, Emens JS, Epstein LJ, et al. Workshop report.
884 Circadian rhythm sleep-wake disorders: gaps and opportunities. *Sleep*. 2021 May 14;44(5).
- 885 86. Chung J, Goodman M, Huang T, Bertisch S, Redline S. Multidimensional sleep health in a diverse,
886 aging adult cohort: Concepts, advances, and implications for research and intervention. *Sleep
887 Health*. 2021 Dec;7(6):699–707.

888 87. Redline S, Azarbarzin A, Peker Y. Obstructive sleep apnoea heterogeneity and cardiovascular
889 disease. *Nat Rev Cardiol.* 2023 Aug;20(8):560–73.

890

Figure 1. Study design diagram



1 Figure 2. Scatter plot of single metabolite association analysis with sleep phenotypes



Figure 3. Number of statistically significant associations between metabolites and sleep phenotypes by domain and sex strata

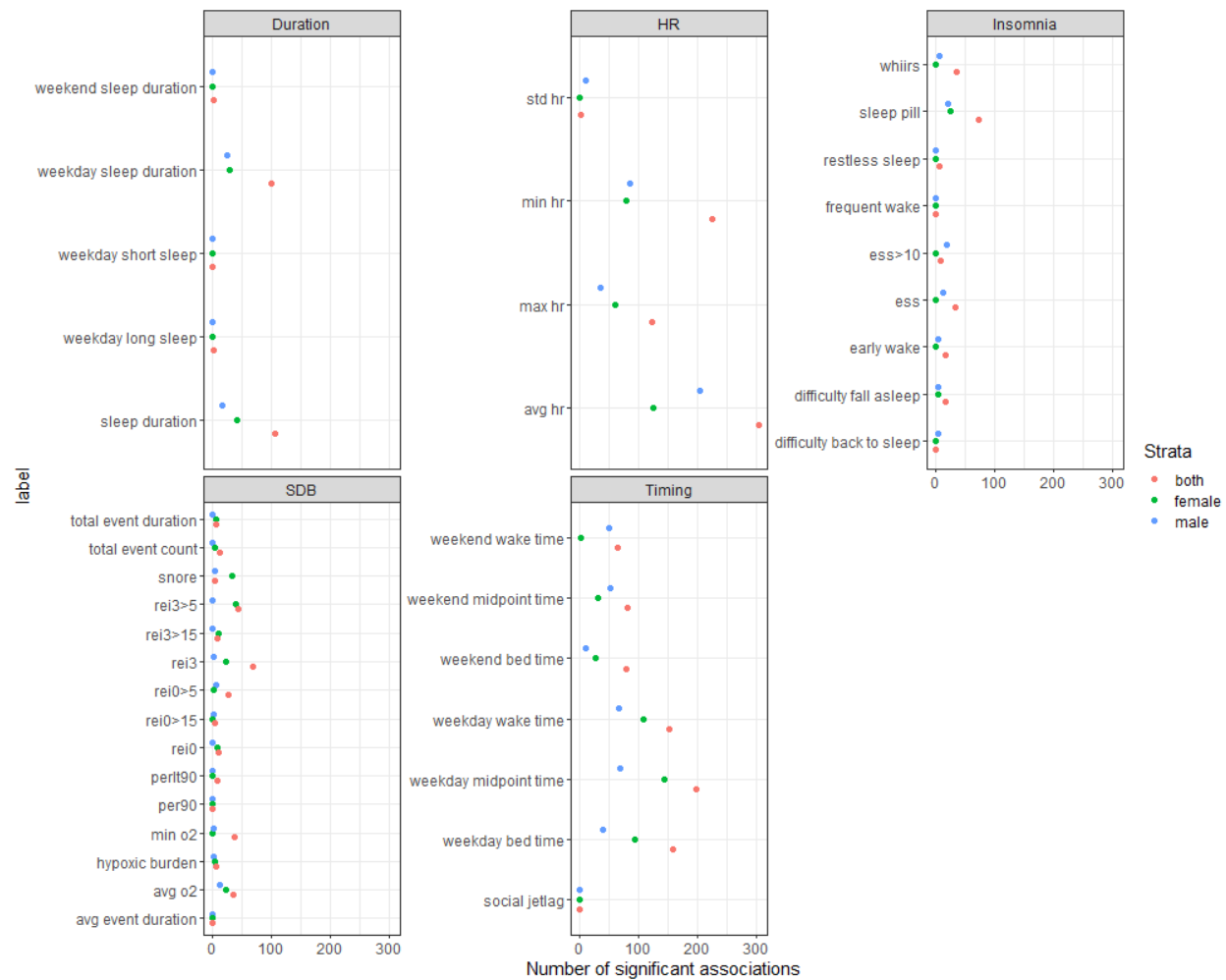


Figure 4. Number of statistically significant associations between metabolites and sleep phenotypes aggregated by subpathway and sleep phenotype domain

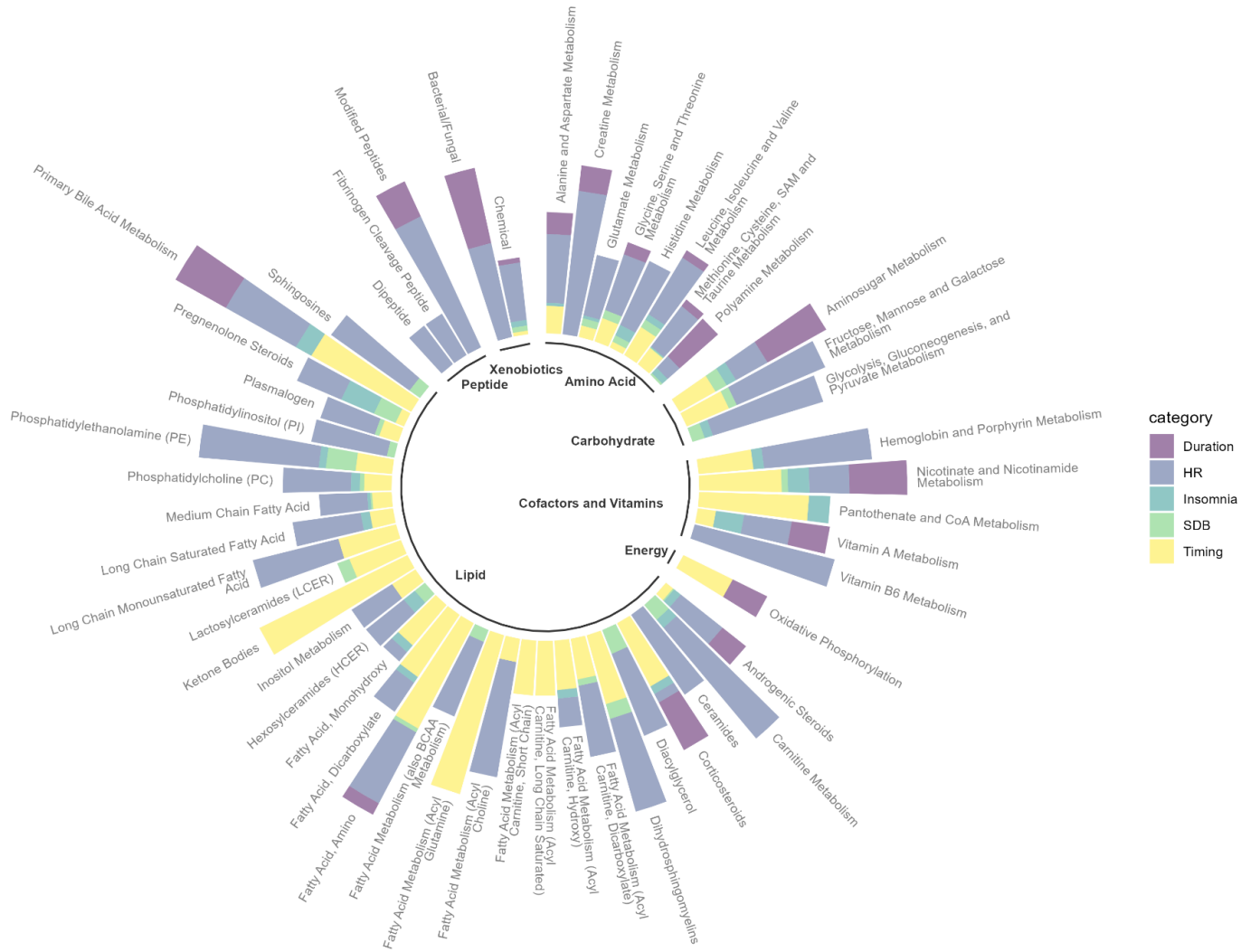


Figure 5. Percentage of statistically significant associations between metabolites and sleep phenotypes aggregated by superpathway and sleep phenotype domain

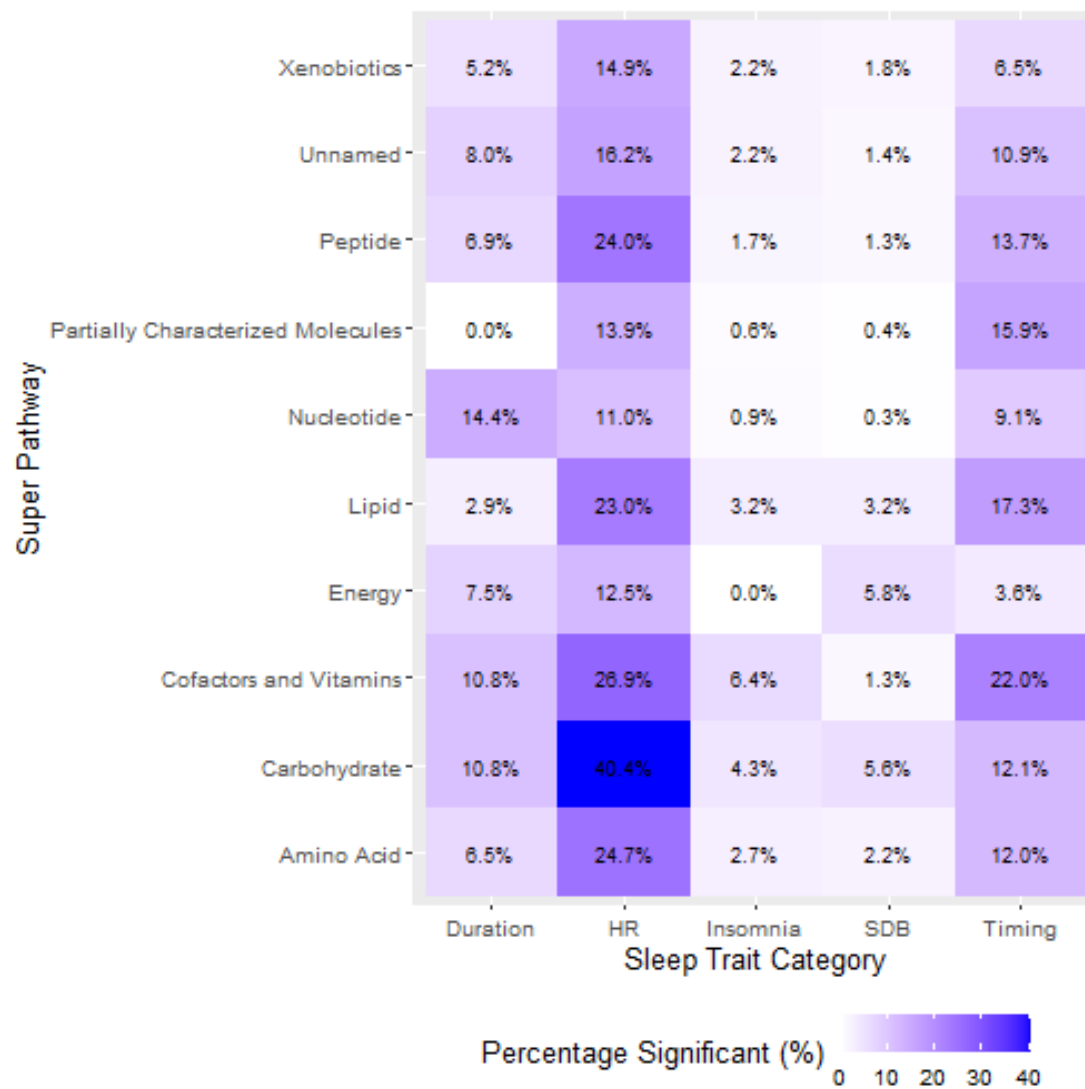


Figure 6. Number of significant associations aggregated by sleep phenotype domain among the top 10% connected metabolites with the most statistically significant associations

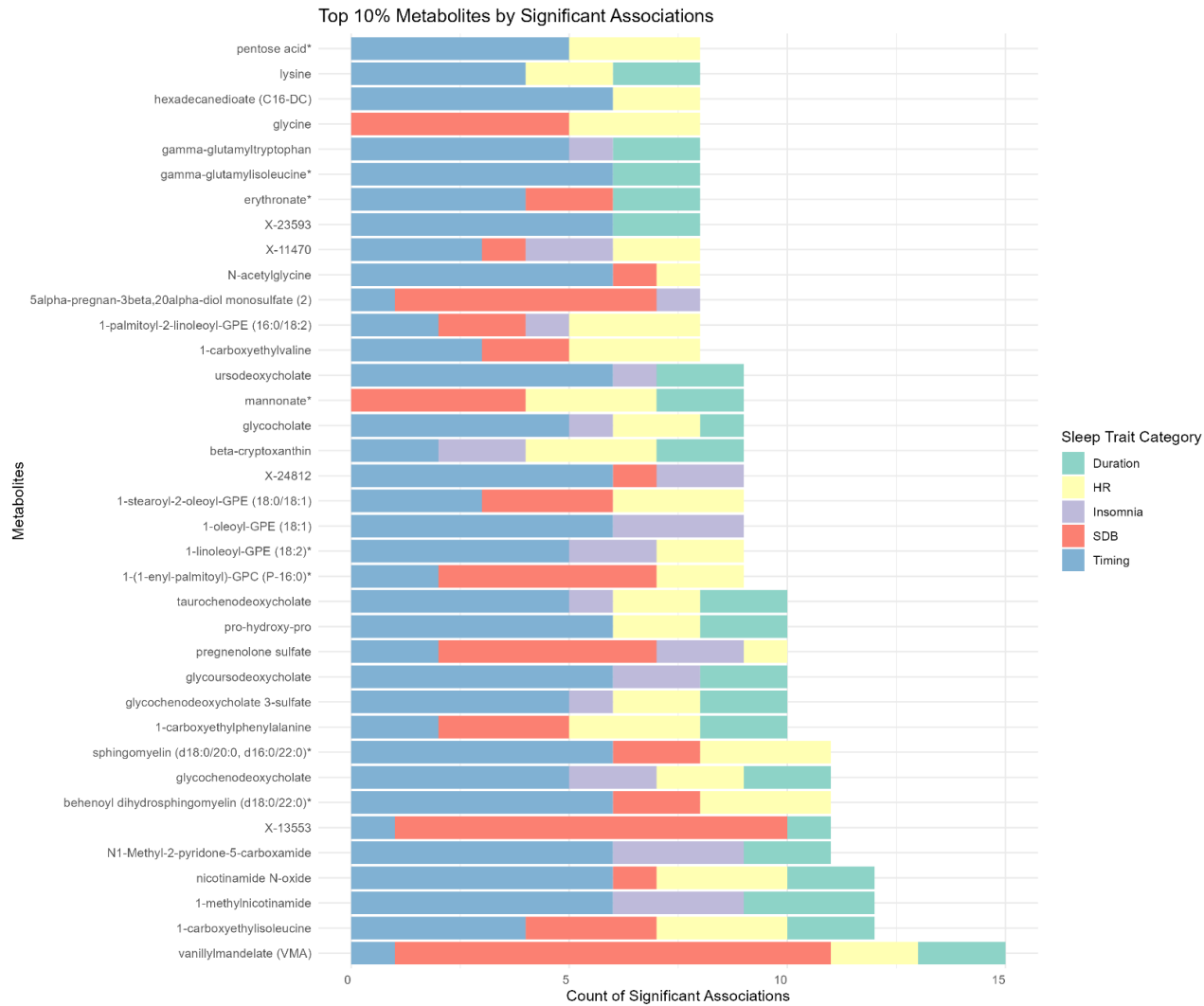


Figure 7. Dice coefficient matrix of sleep phenotype domain based on associated metabolites

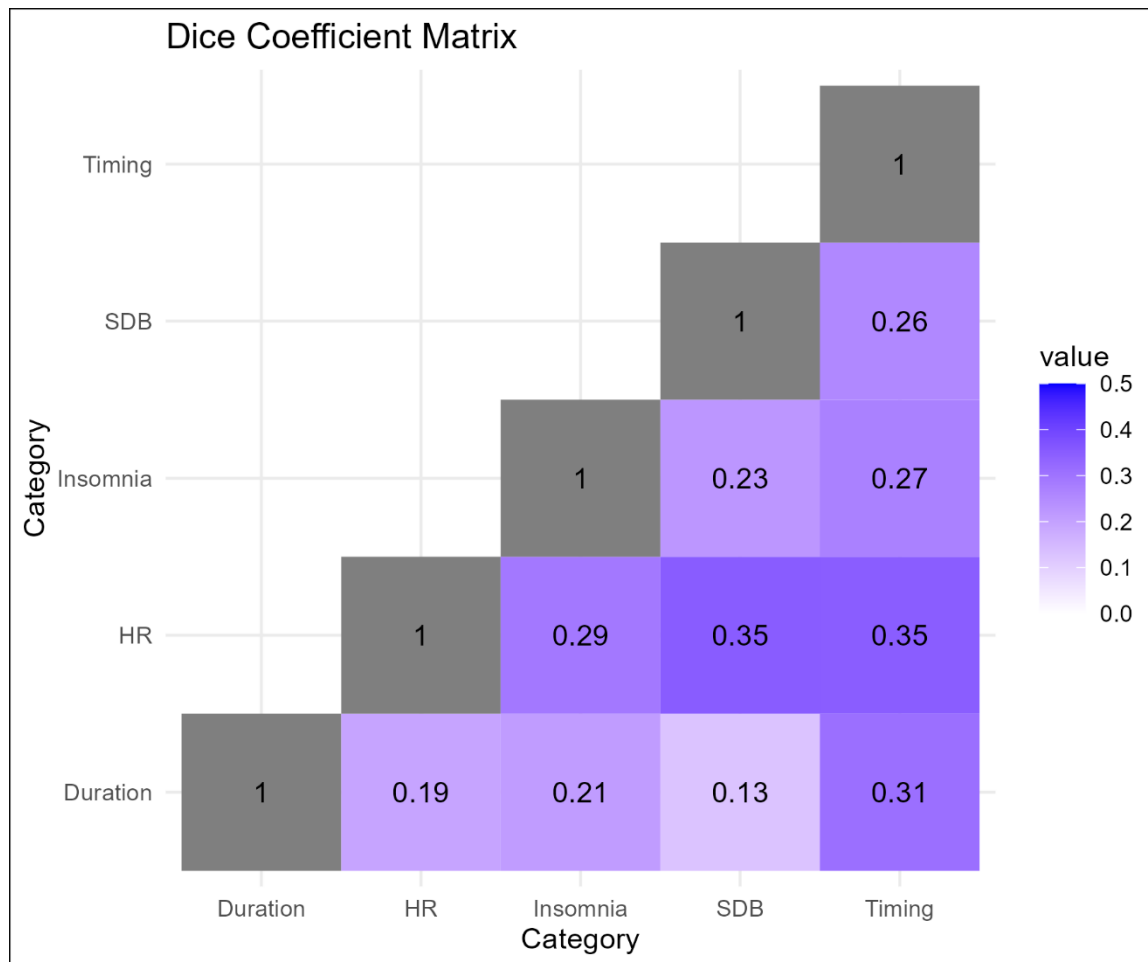
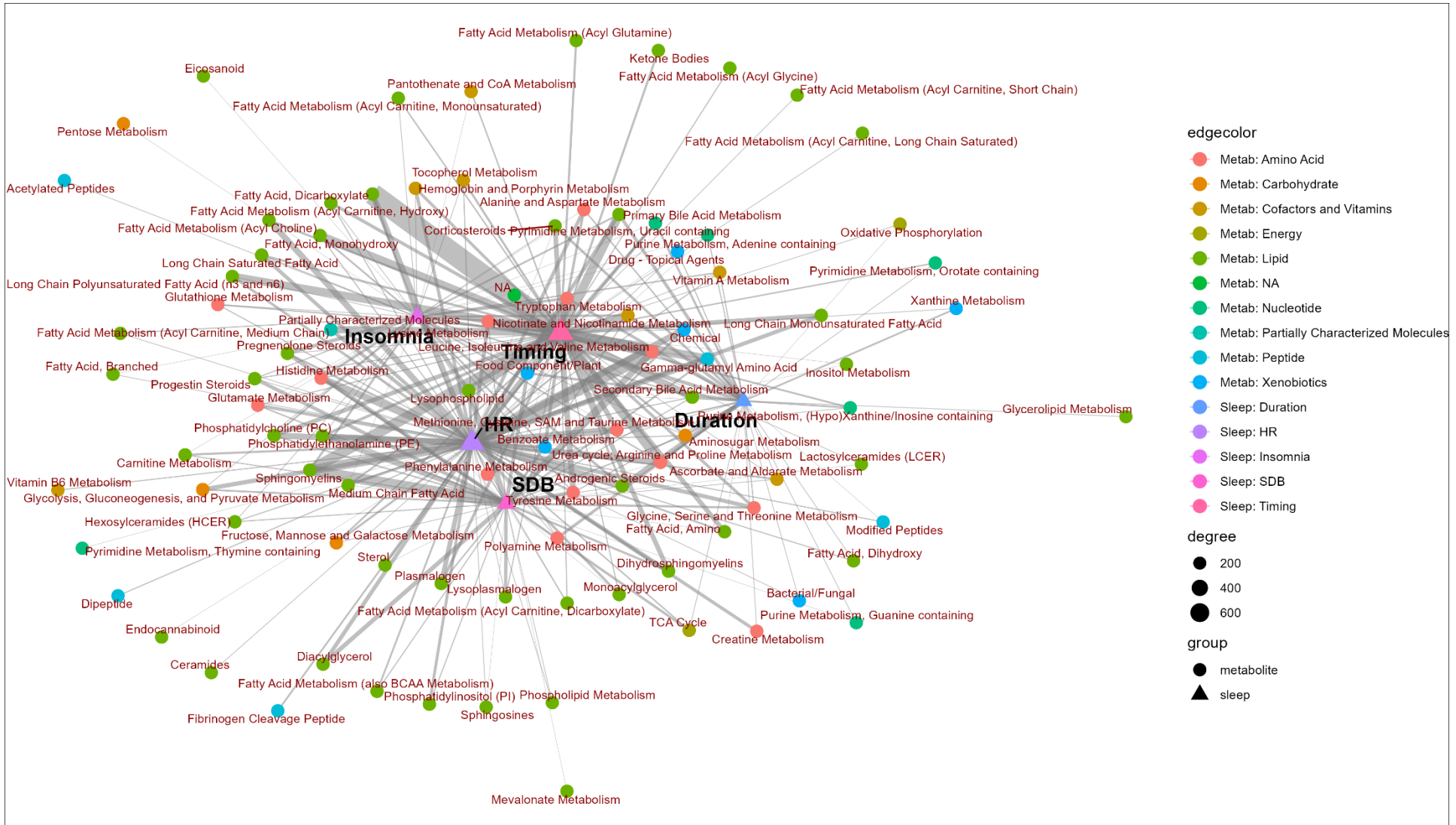


Figure 8. Network based on associations between metabolites and sleep phenotypes aggregated by subpathway and sleep phenotype domain using the Fruchterman-Reingold force-directed algorithm



1 Figure 9. Network structural property metrics summarized by superpathway



2