

# Untargeted metabolome atlas for sleep-related phenotypes in the Hispanic community health study/study of Latinos



Ying Zhang,<sup>a</sup> Brian W. Spitzer,<sup>b</sup> Yu Zhang,<sup>b</sup> Danielle A. Wallace,<sup>a,b,c</sup> Bing Yu,<sup>d</sup> Qibin Qi,<sup>e</sup> Maria Argos,<sup>f,g</sup> M Larissa Avilés-Santa,<sup>h</sup> Eric Boerwinkle,<sup>d</sup> Martha L. Daviglus,<sup>i</sup> Robert Kaplan,<sup>e,j</sup> Jianwen Cai,<sup>k</sup> Susan Redline,<sup>a,c</sup> and Tamar Sofer<sup>a,b,c,l,\*</sup>



<sup>a</sup>Division of Sleep Medicine and Circadian Disorders, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

<sup>b</sup>CardioVascular Institute, Beth Israel Deaconess Medical Center, Boston, MA, USA

<sup>c</sup>Department of Medicine, Harvard Medical School, Boston, MA, USA

<sup>d</sup>Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

<sup>e</sup>Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA

<sup>f</sup>Department of Epidemiology and Biostatistics, School of Public Health, University of Illinois Chicago, Chicago, IL, USA

<sup>g</sup>Department of Environmental Health, School of Public Health, Boston University, Boston, MA, USA

<sup>h</sup>Division of Clinical and Health Services Research, National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD, USA

<sup>i</sup>Institute for Minority Health Research, University of Illinois at Chicago, Chicago, IL, USA

<sup>j</sup>Division of Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, WA, USA

<sup>k</sup>Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

<sup>l</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, 02115, USA

## Summary

**Background** Sleep is essential to maintaining health and wellbeing of individuals, influencing a variety of outcomes from mental health to cardiometabolic disease. This study aims to assess the relationships between various sleep-related phenotypes and blood metabolites.

**Methods** Utilising data from the Hispanic Community Health Study/Study of Latinos, we performed association analyses between 40 sleep-related phenotypes, grouped in several domains (sleep disordered breathing (SDB), sleep duration, sleep timing, self-reported insomnia symptoms, excessive daytime sleepiness (EDS), and heart rate during sleep), and 768 metabolites measured via untargeted metabolomics profiling. Network analysis was employed to visualise and interpret the associations between sleep phenotypes and metabolites.

**Findings** The patterns of statistically significant associations between sleep phenotypes and metabolites differed by superpathways, and highlighted subpathways of interest for future studies. For example, primary bile acid metabolism showed the highest cumulative percentage of statistically significant associations across all sleep phenotype domains except for SDB and EDS phenotypes. Several metabolites were associated with multiple sleep phenotypes, from a few domains. Glycochenodeoxycholate, vanillyl mandelate (VMA) and 1-stearoyl-2-oleoyl-GPE (18:0/18:1) were associated with the highest number of sleep phenotypes, while pregnenolone sulfate was associated with all sleep phenotype domains except for sleep duration. N-lactoyl amino acids such as N-lactoyl phenylalanine (lac-Phe), were associated with sleep duration, SDB, sleep timing and heart rate during sleep.

**Interpretation** This atlas of sleep–metabolite associations will facilitate hypothesis generation and further study of the metabolic underpinnings of sleep health.

**Funding** R01HL161012, R35HL135818, R01AG80598.

**Copyright** © 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

**Keywords:** Metabolomics; Sleep; Atlas; Bipartite network; Hispanic or latino

eBioMedicine  
2025;111: 105507  
Published Online xxx  
<https://doi.org/10.1016/j.ebiom.2024.105507>

\*Corresponding author. Center for Life Sciences, CLS-934, 3 Blackfan St, Boston, MA, 02115, USA.  
E-mail address: [tsofer@bidmc.harvard.edu](mailto:tsofer@bidmc.harvard.edu) (T. Sofer).

### Research in context

#### Evidence before this study

The authors searched standard databases such as PubMed and Google Scholar to identify studies reporting metabolomics associations with sleep-related phenotypes. Studies linking metabolites to sleep-related phenotypes are increasingly reported.

#### Added value of this study

This study used a relatively large dataset, in terms of both the sample size ( $n \approx 6000$ ) and the number of metabolites (over 700) to estimate the associations between metabolites and a range of sleep-related phenotypes from multiple domains. The authors used the resulting atlas to identify patterns of associations across sleep-related phenotype domains, and

metabolite pathways. Metabolites that were linked to multiple sleep phenotypes were studied.

#### Implications of all the available evidence

The patterns of statistically significant associations between sleep phenotypes and metabolites differed by superpathways, highlighting subpathways of interest for future studies. Several metabolites, such as glycochenodeoxycholate, vanillyl mandelate (VMA) and 1-stearoyl-2-oleoyl-GPE (18:0/18:1), and N-lactoyl amino acids, were associated with multiple sleep phenotypes, from a few domains. The authors expect that the generated atlas of associations will be used to develop in-depth studies of sleep and circadian measures and contextualise them within other aspects of health and disease.

### Introduction

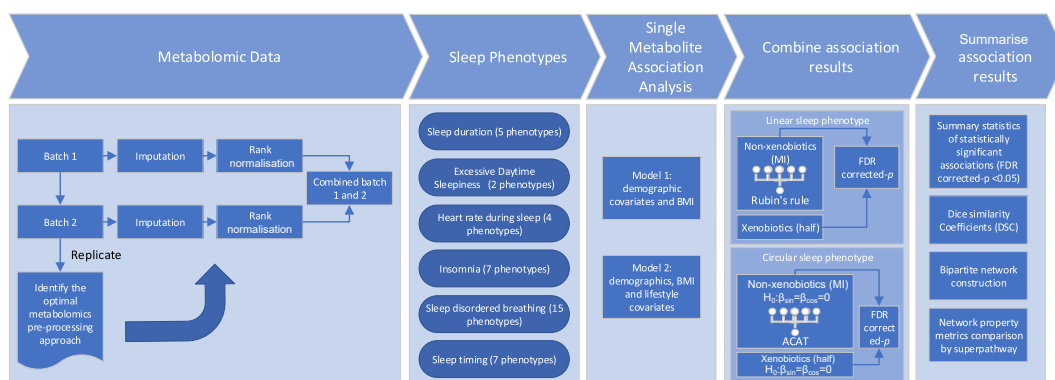
Sleep plays an important role in the health and well-being of individuals. Insufficient quality, timing, and duration of sleep have a major public health impact, and are associated with daytime sleepiness, poor mental health, impaired cognitive function, and increased risk of cardiovascular morbidity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

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

Biological sampling to measure molecular markers of health, such as in metabolomics, can be used to investigate the mechanisms and pathways underpinning sleep phenotypes. Metabolites are small molecules produced in the formation and/or breakdown of endogenous or exogenous substances and are oriented

at the closest layer to phenotypes compared to other underlying biochemical layers (e.g., genome, transcriptome and proteome). The increasing availability of large datasets with untargeted metabolomics profiling has unveiled metabolic outcomes and correlates of numerous health phenotypes, including sleep measures.<sup>16–31</sup> A large-scale study of metabolites in relation to sleep phenotypes may shed light on how underlying biological processes may converge and differ among common sleep phenotypes, the complex interplay between sleep and the metabolic environment, and, ultimately, potential interactions among sleep disorders and progression of cardiometabolic and other health conditions. Here, we study the associations between a range of sleep phenotypes and the metabolic environment in a large population-based study using a high-dimensional set of measured metabolites. We create an “atlas”—a resource for the sleep research community that will facilitate hypotheses formulation and accelerate studies on sleep and its association with other health outcomes.

Our study has taken a comprehensive approach (Fig. 1), covering key sleep-related phenotypes including sleep disordered breathing (SDB), sleep duration, sleep timing, self-reported insomnia symptoms, excessive daytime sleepiness (EDS), and heart rate (HR) during sleep. Each domain of sleep-related phenotypes provides a different perspective on sleep, while together may highlight some shared biological processes within this complex physiological phenomenon. We also conducted network analysis to better understand the interconnectedness between multiple sleep phenotypes and metabolites—representing significant associations as links in a bipartite network allows for simultaneous visualisation of many associations, enabling researchers to observe connectivity patterns that might otherwise be obscured when looking at the individual relationships. By reporting a large number of associations between



**Fig. 1: Study design diagram.** MI: multiple imputation; half: imputation of the missing values for each metabolite with half of the lowest non-missing value of that metabolite across the sample within the batch; ACAT: the aggregated Cauchy association test; FDR: false discovery rate.

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

## Methods

### The Hispanic community health study/study of Latinos

The Hispanic Community Health Study/Study of Latinos (HCHS/SOL) is a prospective community-based cohort study of 16,415 Hispanic/Latino individuals aged 18–74 years at the baseline examination (2008–2011). Study participants were selected using a multi-stage stratified random sampling from four geographic regions: Bronx NY, Chicago IL, Miami FL, and San Diego CA.<sup>32,33</sup> Sex was ascribed by the interviewer as male or female, and was not verified by other means. Description of major ancillary studies and findings in the context of cardiovascular health is provided in a prior publication.<sup>34</sup> Fasting blood samples were collected at the baseline examination, and within the subsequent week, 14,440 of these participants underwent an evaluation for SDB using a validated Type 3 home sleep apnoea test (ARES Unicorder 5.2; B-Alert, Carlsbad, CA) that measured nasal airflow, position, snoring, heart rate and oxyhaemoglobin saturation with measures of SDB scored by a central reading centre as detailed previously.<sup>35</sup> All sleep phenotypes used, and their definitions are provided in [Supplementary Note 1](#).

### Metabolomics profiling

From those who attended the HCHS/SOL baseline assessment and also underwent genotyping (12,803 of the study individuals<sup>36</sup>), 4002 individuals were selected

at random for metabolomics profiling of fasting serum samples collected at baseline (metabolomics batch 1, processed in 2017). In 2021, additional 2368 serum samples from 2330 participants, also collected at baseline, were profiled in a second metabolomics batch 2. This batch included repeated measures within the batch for quality control purposes. Serum samples were stored at  $-70^{\circ}\text{C}$  at the HCHS/SOL Core Laboratory at the University of Minnesota until analysis by Metabolon, Inc (Durham, NC) in 2017 (batch 1) and 2021 (batch 2). Serum samples were then extracted and prepared using Metabolon's standard solvent extraction method. Extracts were split into five fractions to use in four liquid chromatography-mass spectrometry (LC-MS)-based metabolomic quantification platforms (two reverse phase methods with positive ion mode electrospray ionisation (EI), one reverse phase method with negative ion mode EI, and one hydrophilic interaction liquid chromatography with negative ion mode EI), with the fifth fraction reserved for backup. Instrument variability was assessed by calculating the median relative standard deviation (SD) for the internal standards added to each sample prior to injection into the mass spectrometers. Overall process variability was determined by calculating the median relative SD for all endogenous metabolites (i.e., non-instrument standards) present in 100% of the technical replicate samples. Following a recent publication,<sup>37</sup> we re-labelled a few metabolites, previously labelled as 1-carboxyethyl molecules, to N-lactoyl molecules.

### Metabolomic data pre-processing

Preprocessing of the metabolomic data is described in [Supplementary Figure S1](#). First, we removed batch 2 individuals who overlapped with batch 1 and replicate samples from the same individuals, resulting in 2178 remaining batch 2 observations. We then computed percentages of missing values of each metabolite in each batch separately. We excluded metabolites with missing values in more than 75% of the individuals in either batch. For xenobiotic metabolites (metabolite annotation

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

### Sleep phenotypes and modelling approach

We used 40 sleep phenotypes from 6 domains, described in detail in [Supplementary Note 1](#). In brief, these included: self-reported sleep duration (average sleep duration on weekdays, weekend days, and over all days, short and long sleep on weekdays), heart rate during sleep (minimum, maximum, average, and standard deviation of HR), self-reported insomnia symptoms (the women health initiative insomnia rating scale (WHIIRS), its component questions, and self-reported sleeping pill use), daytime and excessive daytime sleepiness (EDS), measured via the Epworth sleepiness scale, sleep disordered breathing (SDB, including measures of breathing disturbances during sleep, mostly measured using a home sleep apnoea testing device: respiratory event indices (REIs), measures of oxyhaemoglobin saturation during sleep, measures of respiratory event lengths; and self-reported snoring, because SDB and snoring often co-occur and share causal determinants such as upper airway collapsibility), and sleep timing (bed time and wake time as well as sleep midpoint during weekend and weekend days, and

social jetlag—the difference in sleep midpoint between weekend and weekdays). In figures and tables, these domains are referred to as Duration, HR, Insomnia, EDS, SDB, and Timing. Phenotypes in the sleep timing category were analysed as “circular” variables, i.e., to avoid bias due to day thresholding (i.e., defining a day by midnight) and to account for the fact that 11:59 PM (23:59) is adjacent to 12:01 AM (00:01). Other sleep phenotypes were treated as linear or binary (when dichotomised). Because EDS is a cardinal symptom of severe SDB, but may be caused by other factors such as insufficient sleep duration and quality independent of SDB, we used a data-driven approach to select a domain for the EDS phenotypes. As described later, we computed the correlations between metabolite associations of each of the non-circular sleep-related phenotypes (i.e., the correlation between the estimated association effect sizes between the phenotype and all considered metabolites), and visualised the results via a heatmap. EDS as a separate domain was selected based on this visualisation.

Two nested conceptual models were used in the analyses. Model 1 adjusted for batch number, demographic, and baseline clinical variables, including age, sex, field centre, self-reported Hispanic/Latino background (Mexican, Puerto Rican, Cuban, Central American, Dominican, South American, and other/multi) and body mass index (BMI). Model 2 further adjusted for lifestyle variables—alcohol use (never, former, current), cigarette use (never, former, current), total physical activity (MET-min/day, computed based on self-reported time spent doing physical activities), and diet (Alternative Healthy Eating Index 2010, computed based on 24 h dietary recall data) in addition to Model 1 covariates. To appropriately model covariates, we evaluated potential nonlinearity of continuous confounders, including age, BMI, physical activity, and diet, in relation to metabolite levels. For each metabolite, we used all individuals with available metabolite levels (i.e., excluded individuals with missing values), and fit five association models, each adjusted for Model 1 covariates only and the continuous covariate of interest (if not already included in model 1). These association models considered 5 possible modelling of the continuous covariate of interest: linear, and four cubic spline models (with 1, 2, 3, and 4 knots). We computed the Akaike Information Criterion (AIC) for each model, and the model with the lowest AIC value was selected for the main analysis of this metabolite and covariate. Within this model selection rule, we prioritised a simpler model compared to a more complex model using the standard rule of thumb where if the AIC of a complex model was lower by less than 2 points compared to the simpler model, we chose the simpler one.

Sleep phenotypes were used as exposures with metabolites as outcomes. When testing the association of a circular phenotype (i.e., sleep timing) with a metabolite,

we first converted the sleep timing phenotype into radians (with 12am serving as the 0) and then computed the sine and cosine of these radians. We used the sine and cosine terms as predictors in the regression,<sup>38</sup> and tested their association with the metabolite using the multivariate Wald test, accounting for two predictors and their estimated covariance. All association analyses were performed using the *survey* R package (version 4.1) to account for the HCHS/SOL sampling design and provide effect estimates relevant for the HCHS/SOL target population.

### Estimating associations between metabolites and sleep phenotypes

We separately assessed the association between each metabolite's concentration level, as an outcome, and each sleep phenotype, as a predictor, using Model 1 and Model 2 covariates as described above, in a single metabolite association analysis. As described in Fig. 1, xenobiotic metabolites were imputed once, while missing values in other metabolites values were imputed 5 times. Thus, we estimated the associations of non-xenobiotic metabolites with sleep phenotypes 5 times (using each of the completed datasets), and then combined the resulting estimated associations. Here we treated the linearly-modelled sleep phenotypes differently from the circular ones. The estimated metabolite associations with linearly-modelled sleep phenotypes were combined using Rubin's rule.<sup>39</sup> Sleep timing phenotypes cannot be combined in the same manner, because there is no method to combine the covariance between the sine and cosine terms across several completed datasets. Instead, we focused on testing and aggregated the *p*-value of the multivariate Wald test using the aggregated Cauchy association test (ACAT<sup>40</sup>). The ACAT test was developed in the context of genetic association analyses, but it is appropriate for our settings, because it allows for the aggregated tests to be based on correlated data. Finally, after combining association results so we had one *p*-value per sleep phenotype-metabolite association (per model), we implemented the Benjamini-Hochberg method to control the false discovery rate (FDR) for multiple testing across all metabolites within each model and sex stratum for each sleep phenotype.<sup>41</sup> Any association that resulted in an FDR-corrected  $p < 0.05$  in Model 1 was considered statistically significant. To assess whether some associations are potentially driven by confounding effects from other lifestyle measures, we examined if Model 2 effect estimates fall within the 95% CI of Model 1 effect estimates from each regression model for non-timing sleep phenotypes (as the latter phenotypes do not have effect estimates). In a secondary analysis, we also performed sex-stratified analyses using the same analytic approach. For descriptive purposes, we computed the number of sleep traits that each metabolite was associated with at the FDR-corrected  $p < 0.05$  Model 1

threshold, and identified the top 10% metabolites based on number of sleep trait associations. A Spearman correlation matrix was computed using the squared effect estimates for each pair of metabolites and sleep phenotypes for sex-combined and sex-stratified analysis.

### Assessment of potential metabolite biomarkers for sleep-related phenotypes

We evaluated whether there is any specific metabolite that could potentially be used as a biomarker for one of the phenotypes. Based on the metabolite-sleep phenotype results, we selected associations with  $p$ -value  $< 10^{-5}$ . Among those, for each phenotype that had more than one metabolite meeting this criterion, we picked the top metabolite with respect to its *p*-value. Finally, we assessed how well the metabolite explains the corresponding sleep-related phenotypes, using adjusted R squared. We fit linear or logistic regression models (survey weighted) with the sleep phenotype as the outcome, and the metabolite as an exposure. One model used only the metabolite, and did not adjust for anything else. This model assessed the potential of using the metabolite individually as a biomarker. Another model used age, sex, study centre, Hispanic background, and BMI as covariates (i.e., Model 1 covariates), and did not include the metabolite. A third model adjusted for Model 1 covariates and included the metabolite. This model assessed the improvement in variance explained, over a basic model, due to the metabolite. We computed the adjusted R squared in the three models.

### Sensitivity analysis for the aggregation of two metabolomics batches

Compared to our primary approach of aggregating the two metabolomics batches, we considered whether a more appropriate approach would be to perform sleep-metabolite association analysis in each batch separately, and then combine the results via a meta-analysis. For this comparison, we used 7 sleep-related phenotypes: hypoxia, long sleep during weekdays, WHIIRS, average sleep duration, REI3%, minimum SpO2 during sleep, and mean HR during sleep. We removed individuals from batch 2 who live in the same household as individuals in batch 1, to prevent inappropriate inference of standard errors (batch combined analysis accounts for this structure using robust standard errors). In both analyses (batch aggregated and separated), metabolites were rank-normalised in each batch separately. Next, for each metabolite, we performed linear regression analysis using individuals with observed values for this metabolite, during both the primary (batch combined) and sensitivity analysis (separate batch followed by meta-analysis) approaches. For this comparison we used linear regression, and not survey-based analysis, based survey-based analysis is not recommended when there are less than 2000 individuals in a dataset, as was true for batch 2. Finally, for each compared phenotype, we computed the Pearson



and Spearman correlations between the estimated association effect sizes across all metabolites to determine whether the results are different or similar between the two modelling approaches.

### Similarity of associations between sleep-related phenotypes and domains

We summarised the associations between sleep phenotypes, individually and categorised by domains, and metabolites, individually and categorised by pathways, where superpathways and subpathways were provided by Metabolon in a metabolite annotation file. Similarities of these associations between a pair of phenotypes or domains were estimated with the Dice Similarity Coefficient (DSC)<sup>42</sup> defined as

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

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

### Bipartite network analysis

Bipartite network is a type of network in which there are two groups of nodes (here: sleep phenotypes and metabolites), and links, or edges, can only exist between the two types of nodes but not within a group of nodes. We constructed a bipartite network using the sleep phenotypes and metabolites, where an edge was added between a sleep phenotype-metabolite pair if their association was statistically significant (FDR-corrected  $p < 0.05$ ) in Model 1 association analysis. The network is built based on an incidence matrix, with rows corresponding to sleep phenotypes, columns represent metabolites, and the  $i, j$  cell in this matrix has a value of 1 if the  $i$ th sleep phenotype has a statistically significant association with the  $j$ th metabolite, and 0 otherwise. We then aggregated metabolites by sub- and super-pathway, and sleep phenotypes by domain, which results in a consolidated incidence matrix in which a value of 1 in cell  $(m, n)$  indicating any statistically significant association between the  $m$ th sleep phenotype domain and the  $n$ th metabolite sub- and super-pathway and a weight matrix documenting the number of total statistically significant associations between the  $m$ th sleep phenotype domain and the  $n$ th metabolite sub- and super-pathway. Several network property metrics were computed to offer insights into the structure and associations of the sleep domain—metabolite sub- and super-pathway network.

For visualisation, we converted the bipartite network into a univariate network and visualised the network

using the Fruchterman-Reingold force-directed algorithm, which optimises the placement of nodes based on connectivity similarity between nodes, where similar connectivity is reflected as proximity of nodes,<sup>43</sup> providing an intuitive spatial representation of the network structure.

### Statistics

Associations of metabolite levels with sleep-related phenotypes were based on the Wald test for non-circular variables, i.e., when a single effect estimate was available for a phenotype. The Wald tests were computed after effect estimates and their SEs were combined across multiply-imputed datasets using Rubin's rule. For circular variables for which two coefficients were estimated corresponding to sine and cosine terms, we first applied the multivariate Wald test, and next combined  $p$ -values across the multiply-imputed datasets using the ACAT method, because there is no existing method to combine effect estimates and covariance matrices from multiple correlated variables in the multiple imputation settings.

All analyses were done in R version 4.2.3. The function `svyglm` from the *survey* package was used for survey-weighted generalised linear regression models. Cubic splines for modelling of continuous covariates were generated using the `ns` function from the *splines* package. The *car* package (version 3.1) was used for multivariate Wald test. Adjusted R square was computed using the `rsq` function from the *rsq* package. The *bipartite* package (version 2.18) was used for bipartite network analysis, and *igraph* (version 1.5) and *ggnetwork* (version 0.5) packages were used for visualising network graph.

### Ethics

The HCHS/SOL was approved by the institutional review boards (IRBs) at each field centre, where all participants gave written informed consent, and by the Non-Biomedical IRB at the University of North Carolina at Chapel Hill, to the HCHS/SOL Data Coordinating Center. All IRBs approving the HCHS/SOL study are: Non-Biomedical IRB at the University of North Carolina at Chapel Hill, Chapel Hill, NC (study #07-1003); Einstein IRB at the Albert Einstein College of Medicine of Yeshiva University, Bronx, NY (reference #007788); IRB at Office for the Protection of Research Subjects (OPRS), University of Illinois at Chicago, Chicago, IL (protocol #2013-1261); Human Subject Research Office, University of Miami, Miami, FL (study number 20070461); Institutional Review Board of San Diego State University, San Diego, CA (protocol number 1586091). All methods and analyses of HCHS/SOL participants' materials and data were carried out in accordance with human subject research guidelines and regulations. This work was approved by the Mass General Brigham IRB (protocol #2022P001237) and by the

Beth Israel Deaconess Medical Center Committee on Clinical Investigations (protocol #2023P000277).

### Role of funders

As a study that used previously collected data, the funding source did not have any role in the study design, analysis, interpretation, or decision to submit this paper for publication.

## Results

### Study sample characteristics

**Table 1** characterises the HCHS/SOL target population (using means and percentages weighted to account for study participation), with more comprehensive data, including summary measures of all sleep-related phenotypes, and of common medications, provided in **Supplementary Table S1**. The analytic sample, batch 1 and batch 2 metabolomic dataset combined, included 6180 participants with a mean age of 44.31 years (SD = 15.27), of whom 40.1% were males, 20.7% did not report alcohol use, 58.5% never smoked, and 11.7% had moderate-to-severe OSA (REI3% $\geq$ 15). The average sleep duration was 7.98 h (SD = 1.46), while 22.7% reported restless or very restless sleep on a typical night in the last month, and 15.8% reported excessive sleepiness (ESS $>$ 10). Batch 2 participants were older compared to batch 1—the mean age is 41.54 years (SD) in batch 1 and 51.11 years (SD) in batch 2. The baseline rates of diabetes and hypertension were higher in batch 2 (diabetes: 29.4%; hypertensions: 44.1%) compared to batch 1 (diabetes: 20.1%; hypertensions: 31.5%), consistent with the age difference of participants between the two batches.

### Modelling of continuous covariates

**Supplementary Table S2** compares AIC from models treating continuous covariates (age, BMI, diet, and physical activity) as linear versus nonlinear terms (i.e., cubic splines with different knot numbers) and reported the selected optimal modelling for each continuous covariate. For age, only 27% of metabolites had a linear association, while 45%, 68%, and 69% of metabolites had linear associations with BMI, physical activity, and diet measures, respectively. The selected linear or spline models were used in subsequent analyses of the corresponding metabolite.

### Results from sleep phenotype-metabolite association analysis

The single metabolite association analysis was conducted in pair-wise fashion between 40 sleep phenotypes from six domains (i.e., sleep duration, HR, insomnia, EDS, SDB, and sleep timing) and 768 metabolites, including 113 unknown metabolites and 77 xenobiotic metabolites.

34 sleep phenotypes had statistically significant associations with at least one metabolite (**Table 2**). When limited to Model 1, the median number of significant associations for each sleep phenotype is 14 (range: 0–307), corresponding to 1.82% (range: 0%–39.97%) of all tested metabolites. The number of statistically significant associations is much lower among dichotomised sleep phenotypes (median: 7; range: 0–74) compared to non-dichotomised sleep phenotypes (median: 27; range: 0–307), likely corresponding to loss of power due to dichotomisation. Phenotypes representing heart rate during sleep domain had the highest number

Mean (SD) <sup>a</sup>	Female	Male	Batch 1	Batch 2	Overall
N	3704	2476	4002	2178	6180
Age	44.98 (15.17)	43.56 (15.35)	41.54 (15.18)	51.11 (13.21)	44.31 (15.27)
Sex = Male (%)	0 (0.0)	2476 (100.0)	1705 (42.6)	771 (35.4)	2476 (40.1)
BMI	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)	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)	46.79 (7.42)	48.74 (7.47)	47.28 (7.52)	48.75 (7.36)	47.71 (7.50)
OSA status = OSA (%) <sup>b</sup>	260 (7.9)	383 (17.4)	388 (10.9)	255 (13.0)	643 (11.7)
Baseline Diabetes status (ADA) = Yes (%) <sup>c</sup>	863 (23.3)	579 (23.4)	803 (20.1)	639 (29.4)	1442 (23.3)
Baseline Hypertension status = Yes (%) <sup>d</sup>	1357 (36.6)	865 (34.9)	1261 (31.5)	961 (44.1)	2222 (36.0)
Sleep Duration (hours)	8.04 (1.49)	7.90 (1.42)	7.97 (1.45)	7.98 (1.49)	7.98 (1.46)
Women's Health Initiative Insomnia Rating Scale (WHIIRS) total score	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	5.44 (4.71)	5.87 (4.87)	5.62 (4.71)	5.70 (4.99)	5.64 (4.79)

<sup>a</sup>Means and percentages have been weighted to provide values representative of the HCHS/SOL target population. <sup>b</sup>OSA was defined as respiratory event index  $\geq$ 15, with events counted based on 3% oxyhaemoglobin desaturation. <sup>c</sup>Baseline diabetes are based on American Diabetes Association definition (Diabetes Care 2010; 33:562–69), defined as fasting glucose  $\geq$ 126 mg/dL, or post-OGTT glucose  $\geq$ 200 mg/dL or A1C  $\geq$  6.5%, or self-report of diabetes. <sup>d</sup>Baseline hypertension is defined as systolic or diastolic blood pressure greater than or equal to 140/90 respectively, or current use of antihypertensive medications.

**Table 1:** HCHS/SOL target population characteristics by sex, batch, and overall.

Sleep phenotype	Number of statistically significant associations	Percentage of statistically significant associations (%)	Sleep phenotype domain	Sleep phenotype dichotomised or continuous
Weekday long sleep	2	0.26	Duration	Dichotomised
Weekday short sleep	0	0	Duration	Dichotomised
Sleep duration	63	8.2	Duration	Continuous
Weekday sleep duration	68	8.85	Duration	Continuous
Weekend sleep duration	0	0	Duration	Continuous
ESS	40	5.21	EDS	Continuous
ESS>10	9	1.17	EDS	Dichotomised
Min HR during sleep	219	28.52	HR	Continuous
Max HR during sleep	135	17.58	HR	Continuous
Avg HR during sleep	307	39.97	HR	Continuous
Std HR during sleep	2	0.26	HR	Continuous
Difficulty back to sleep	0	0	Insomnia	Dichotomised
Early wake	14	1.82	Insomnia	Dichotomised
Difficulty fall asleep	14	1.82	Insomnia	Dichotomised
Frequent wake	0	0	Insomnia	Dichotomised
Sleep pill	74	9.64	Insomnia	Dichotomised
Restless sleep	7	0.91	Insomnia	Dichotomised
WHIRS	25	3.26	Insomnia	Continuous
Avg event duration	0	0	SDB	Continuous
Hypoxic burden	6	0.78	SDB	Continuous
per90	3	0.39	SDB	Dichotomised
rei0	15	1.95	SDB	Continuous
rei0 > 15	3	0.39	SDB	Dichotomised
rei0 > 5	7	0.91	SDB	Dichotomised
Total event count	12	1.56	SDB	Continuous
Total event duration	9	1.17	SDB	Continuous
rei3	60	7.81	SDB	Continuous
rei3 > 15	11	1.43	SDB	Dichotomised
rei3 > 5	32	4.17	SDB	Dichotomised
Min o2	27	3.52	SDB	Continuous
Avg o2	25	3.26	SDB	Continuous
Perlt90	2	0.26	SDB	Continuous
Snore	4	0.52	SDB	Dichotomised
Social jetlag	0	0	Timing	Continuous
Weekday bed time	132	17.19	Timing	Continuous
Weekday midpoint time	135	17.58	Timing	Continuous
Weekday wake time	98	12.76	Timing	Continuous
Weekend bed time	46	5.99	Timing	Continuous
Weekend midpoint time	47	6.12	Timing	Continuous
Weekend wake time	18	2.34	Timing	Continuous

Statistically significant associations are defined as FDR-corrected *p*-value, derived using the Benjamini-Hochberg method to control false discovery rate (FDR) for multiple testing across all metabolites within each model and sex stratum for each sleep phenotype is below 0.05; while the unadjusted two-sided *p* values were derived by accounting for the complex sampling design-based degrees of freedom, using adjusted standard errors to compute the *t*-statistic in single metabolite association analysis with each sleep phenotype as dependent variables.

**Table 2: Statistically significant metabolite associations by sleep phenotype.**

of statistically significant associations with metabolites (median: 177, range: 2–307), followed by the sleep timing domain (median: 47, range: 0–135). The SDB domain had the lowest number of statistically significant associations with tested metabolites (median: 8, range: 0–60) (Supplementary Table S3).

Supplementary Figure S2 visualises the strength of associations (i.e., negative logarithm of the FDR-corrected *p*) between sleep phenotypes and individual metabolites grouped by superpathway. Compared to the sex-stratified analysis, sex-combined analysis tended to identify more statistically significant associations between sleep phenotypes and metabolites (Supplementary Figure S3), in accordance with the higher power due to larger sample size of the sex-combined analysis.

The results from Model 2 were largely consistent with Model 1 results—less than 0.02% (6 out of 30,720) metabolite-sleep phenotype pairs resulted in effect estimates falling outside of the 95% CI of Model 1 effect estimates once controlling for additional lifestyle covariates (Supplemental Table S4). However, only two of these associations were statistically significant in the primary Model 1 analysis and none were significant in the Model 2 analysis. Both associations were with the same unnamed metabolites, and two self-reported insomnia phenotypes.

Sensitivity analysis comparing the aggregation of data across the two batches (as in the primary analysis) to separate analysis in each batch followed by meta-analysis resulted in highly correlated effect sizes across the seven compared sleep-related phenotypes (Supplementary Figure S4), suggesting that it is appropriate to aggregate data across batches while adjusting for batch as a covariate.

Results from analysis of the potential of individual metabolites to serve as biomarkers are reported in Supplementary Table S5. In essence, none of the metabolites explained a large proportion of the variance of the sleep-related phenotype, either individually, or on top of baseline covariates, as assessed using the adjusted *R* squared. The largest *R* squared observed using an individual metabolite, without accounting for other covariates, was glutamate, which had adjusted *R* squared of 0.06 in association with self-reported snoring. The next highest was isovaleryl carnitine (C5), with adjusted *R* squared of 0.04 in association with REI3%. The largest increase in adjusted *R* squared was observed for 1-(1-enyl-palmitoyl)-2-oleoyl-GPC (P-16:0/18:1), in association with average HR during sleep. It increased the adjusted *R* squared from 0.12 in the base model, to 0.14 in the model that included this metabolite.

### Aggregating association analysis results by pathway and by domain

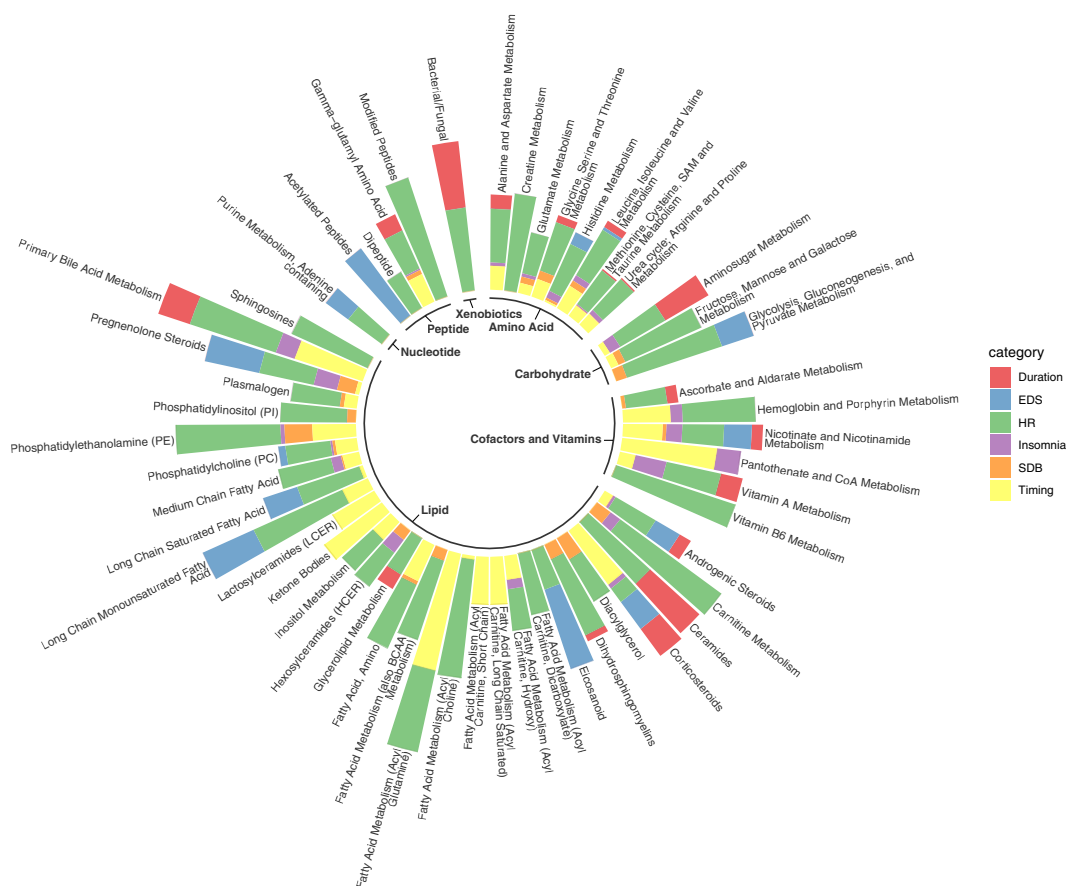
Among the 768 included metabolites, carbohydrates had the highest average number of statistically significant associations with sleep phenotypes (mean: 3.31, range: 1–6 per metabolite), followed by cofactors and vitamins (mean: 3.19, range 0–9 per metabolite) and lipids (mean: 2.53, range: 0–12 per metabolite). Partially characterised molecules (mean: 1.22, range: 0–5) and xenobiotics (mean: 1.4 range: 0–9) had the lowest number of statistically significant associations with



sleep phenotypes per metabolite (Supplementary Table S6). The grouping of sleep phenotypes into domains was supported by the correlation matrix based on the effect estimates from the single metabolite association analysis results (Supplementary Figures S5–S7 for sex combined and sex-stratified results) (Fig. 1).

We aggregated metabolites by subpathway and sleep phenotypes by domain, then identified and visualised top metabolomic subpathways with over 25% statistically significant associations (defined as FDR-corrected  $p < 0.05$ ) among all tested associations (Fig. 2) by sleep phenotype domain. Primary bile acid metabolism showed the highest cumulative percentage of statistically significant associations across four sleep phenotype domains, although no significant association was identified for SDB and EDS phenotypes (Supplementary Table S7). Subpathways from amino acid metabolism—specifically leucine, isoleucine, valine, phenylalanine, and tyrosine metabolism—as well as nicotinate and

nicotinamide metabolism (cofactors and vitamins), and androgenic steroid metabolism (lipids), showed statistically significant associations with sleep phenotypes across all six domains (Supplementary Table S7). Both lipids and cofactors/vitamins had subpathways (e.g., ketone bodies, acyl glutamine, pantothenate and CoA metabolism, nicotinate and nicotinamide metabolism) that had high percentage of statistically significant associations with sleep timing phenotypes. SDB phenotypes had the highest percentage of statistically significant associations among lipids, especially phosphatidylethanolamine (PE), diacylglycerols, progestin steroids, lysoplasmalogen, and pregnenolone steroids. As for the sleep duration domain, the top subpathways with the highest percentage of significant associations were ceramides, bacterial/fungal, amino sugar metabolism, polyamine metabolism, oxidative phosphorylation, corticosteroids, and primary bile acid metabolism. Self-reported insomnia symptoms domain generally had



**Fig. 2: Number of statistically significant associations between metabolites and sleep phenotypes aggregated by subpathway and sleep phenotype domain.** Statistically significant associations are defined as FDR-corrected  $p$ -value, derived using the Benjamini-Hochberg method to control false discovery rate (FDR) for multiple testing across all metabolites within each model and sex stratum for each sleep phenotype is below 0.05; while the unadjusted two-sided  $p$  values were derived by accounting for the complex sampling design-based degrees of freedom, using adjusted standard errors to compute the  $t$ -statistic in single metabolite association analysis with each sleep phenotype as dependent variables. The sample sizes for the association analysis of each pair of metabolite and sleep phenotype are available in figshare (doi.org/10.6084/m9.figshare.27212511).

fewer statistically significant associations, among which vitamin A metabolism showed the highest percentage of statistically significant associations. Eicosanoid, acetylated peptides, long chain monounsaturated fatty acid (LCMUFA) and pregnenolone steroids showed the highest percentage of statistically significant associations with EDS.

Comparing men and women, potential sex differences can be observed (Supplementary Figure S8), with the limitation that the sample size of male participants was lower. Relatively more subpathways had statistically significant associations with SDB phenotypes among females compared to males (these include progesterin steroids, oxidative phosphorylation, lactosylceramides (LCER) and sphingomyelins). The subpathways from which metabolites were associated with self-reported insomnia symptoms were mostly progesterin steroids, amino sugar metabolism among males, as compared to pregnenolone steroids among females. 13 subpathways, mostly lipids, were significantly associated with EDS among males while no metabolite were identified for EDS in the female-only analysis.

We also aggregated metabolites by superpathway and visualised the percentage of statistically significant associations between superpathways and sleep phenotype domains in the format of a heatmap (Fig. 3). The percentage of statistically significant associations were highest for HR during sleep, especially among carbohydrates (44.2%), followed by cofactors and vitamins (28.8%), lipids (24.8%) and peptides (24.0%) (Supplementary Table S8). Self-reported insomnia and SDB domain showed overall lower percentage of statistically significant associations with metabolites across all superpathways, all of which were below 7%.

Among all metabolites associated with at least one sleep phenotype, we identified the top 10% ( $n = 37$ ) with the highest number of total associations, regardless of domains (Fig. 4). 17 metabolites were lipids, 12 were amino acids, 4 were unnamed, 3 were cofactors and vitamins, 2 were peptides, and 1 was xenobiotic. Majority of the top 10% metabolite had associations with more than one sleep phenotype domain (except for linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]\* and X-13553), among which 10 metabolites were associated with four or more domains of sleep phenotypes. Glycochenodeoxycholate (primary bile acid metabolism), vanillyl mandelate (VMA) and 1-stearoyl-2-oleoyl-GPE (18:0/18:1) were associated with the highest number of sleep phenotypes; while pregnenolone sulfate was associated with all sleep phenotype domains except for sleep duration (Fig. 4).

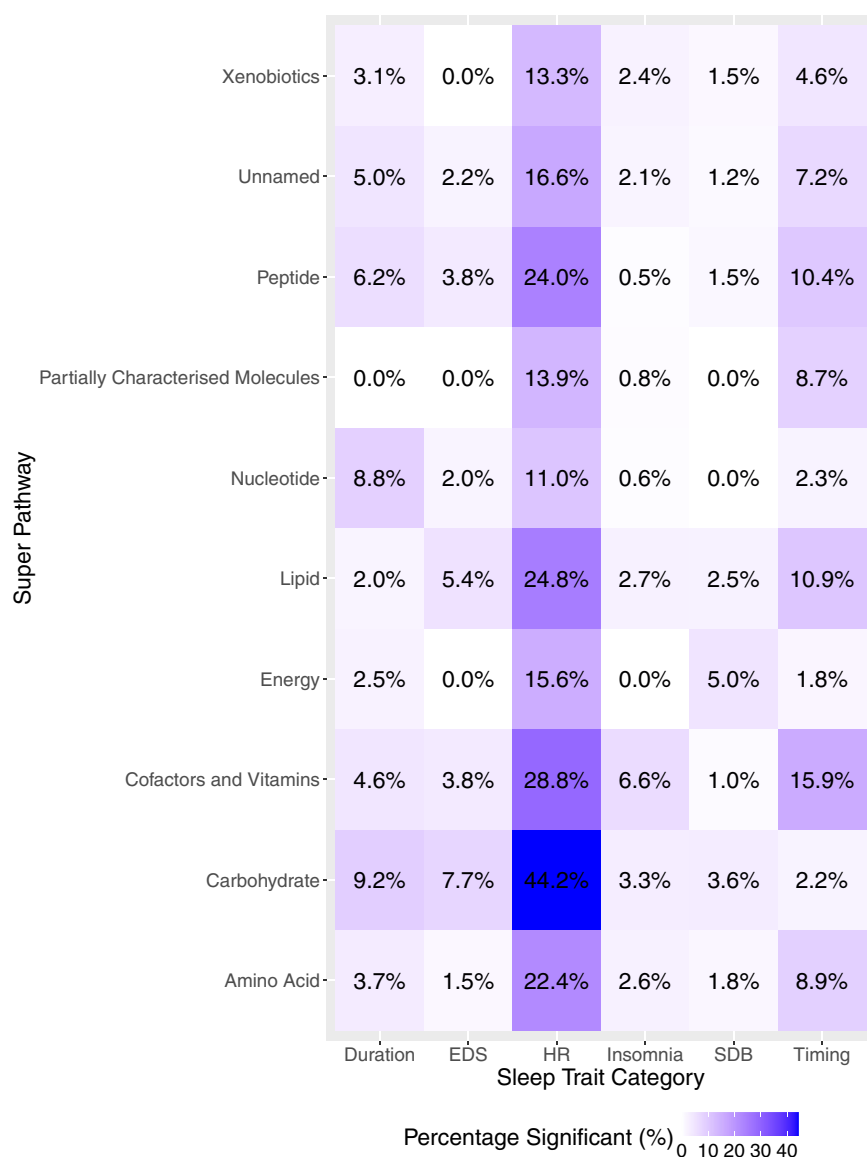
We also compared the similarities among sleep phenotype domains by calculating the Dice Similarity Coefficients (DSC) which quantifies the associated metabolites overlap between any two sleep domains. As shown in Fig. 5, the highest overlap was observed between HR during sleep phenotypes and SDB

(DSC = 0.31), and with sleep timing domain (DSC = 0.31), respectively. The overlap between sleep duration and sleep timing domain were lower (DSC = 0.26). The lowest levels of overlap were observed for sleep duration traits and SDB domain (DSC = 0.07).

### Bipartite network analysis based on statistically significant associations between metabolites and sleep phenotypes

Fig. 6 shows a network of nodes and edges in which each node represents either a sleep phenotype domain or a metabolomic subpathway, while each edge represents one or more statistically significant associations between two nodes based on the single metabolite association analysis results (FDR-corrected  $p < 0.05$ ). The width of the edge indicates the number of statistically significant associations between two nodes. Some metabolomic subpathways have exclusive connections with one sleep domain, such as fatty acid and dicarboxylate subpathways and sleep timing domain. Others tend to have many connections with multiple sleep domains, such as sphingomyelins with sleep timing, heart rate and SDB phenotype domains. The placement of the nodes reflects their connectivity similarities—the closer two nodes are, the more similar their overall connectivity patterns are. SDB and heart rate phenotypes, for instance, are closely placed together which also corresponds to the high DSC values between the two groups (Fig. 5). Metabolomic subpathways located close to the centre of the network indicates their connectivity with multiple sleep domains, while metabolomic subpathways located at the outskirts of the network indicates their connections with one or few closely located sleep phenotype domains. For instance, metabolites from nicotine and nicotinamide metabolism and leucine, isoleucine and valine metabolism were associated with all sleep phenotype domains. Lipids (clustered on the right side of the network diagram), were more connected to SDB, HR, sleep timing and EDS than to sleep duration and self-reported insomnia symptoms (Fig. 6), which share similar connectivity patterns with glutamate metabolism, glycolysis, gluconeogenesis, and pyruvate metabolism fructose, mannose and galactose metabolism from amino acid and carbohydrate superpathways.

Treated as a bipartite network model, we further characterised the connectivity between the metabolites, grouped at the superpathway level, and sleep domains quantified by network properties. The number of connected metabolites is the lowest for energy metabolites and the highest for lipids, which corresponds to the number of connected sleep phenotypes for the two superpathways and the average connection per node (i.e., links per node) (Supplementary Figure S9). Similar trend can be observed for the mean number of shared metabolites among sleep phenotypes—the number of overlapping metabolites for sleep phenotypes is the

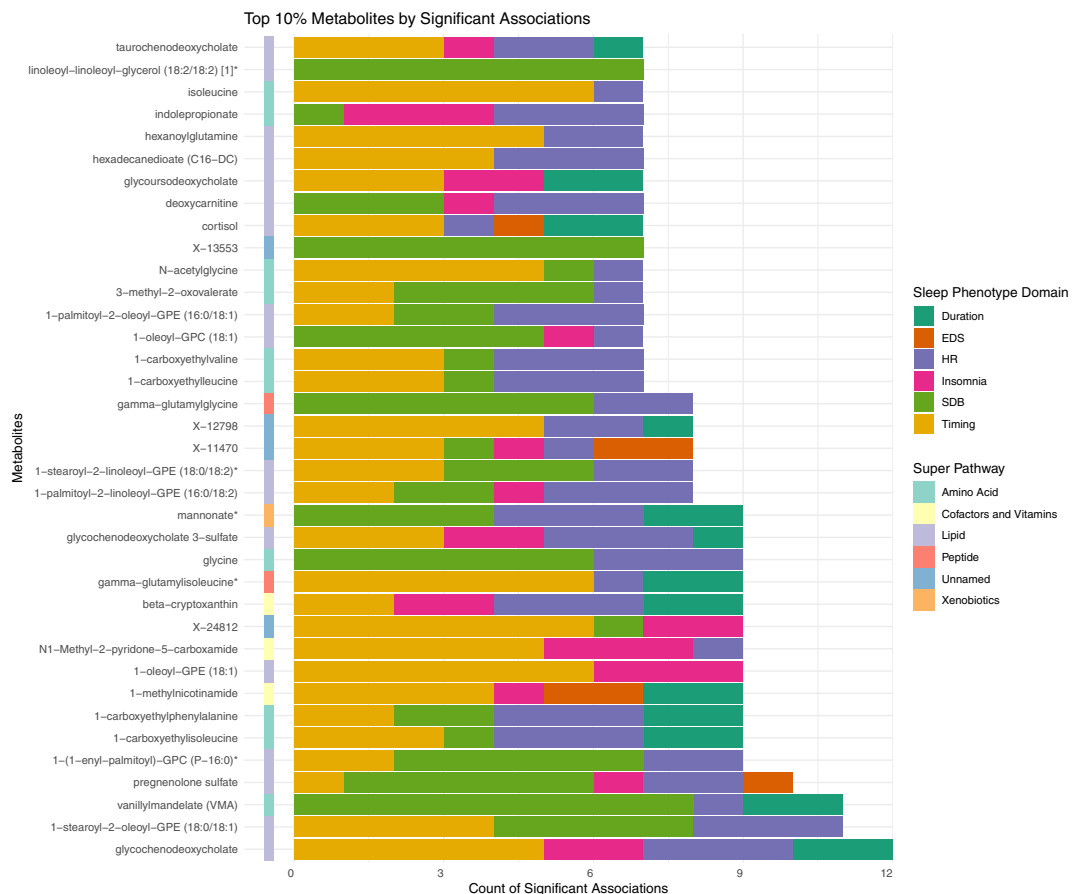


**Fig. 3: Percentage of statistically significant associations between metabolites and sleep phenotypes aggregated by superpathway and sleep phenotype domain.** Statistically significant associations are defined as FDR-corrected  $p$ -value, derived using the Benjamini-Hochberg method to control false discovery rate (FDR) for multiple testing across all metabolites within each model and sex stratum for each sleep phenotype is below 0.05; while the unadjusted two-sided  $p$  values were derived by accounting for the complex sampling design-based degrees of freedom, using adjusted standard errors to compute the  $t$ -statistic in single metabolite association analysis with each sleep phenotype as dependent variables. The sample sizes for the association analysis of each pair of metabolite and sleep phenotype are available in figshare (doi.org/10.6084/m9.figshare.27212511).

highest among lipids and lowest among energy metabolites, which suggests that more lipids are connected to the same sleep phenotype compared to metabolites from other superpathways. The mean number of shared sleep phenotypes among metabolites, however, peaked for cofactors and vitamins and carbohydrates, followed by lipids, suggesting metabolites from the former two superpathways are more likely to be connected to similar

sleep phenotypes than other superpathways, despite their low number of total metabolites.

Cluster coefficient, the number of realised links divided by the number of possible links, is the highest among carbohydrates and energy. This pattern indicates a higher degree of overlap in the neighbouring nodes among carbohydrates and energy, compared to unnamed metabolites and when combining all metabolites



**Fig. 4: Number of significant associations aggregated by sleep phenotype domain among the top 10% connected metabolites with the most statistically significant associations.** Statistically significant associations are defined as FDR-corrected  $p$ -value, derived using the Benjamini-Hochberg method to control false discovery rate (FDR) for multiple testing across all metabolites within each model and sex stratum for each sleep phenotype is below 0.05; while the unadjusted two-sided  $p$  values were derived by accounting for the complex sampling design-based degrees of freedom, using adjusted standard errors to compute the  $t$ -statistic in single metabolite association analysis with each sleep phenotype as dependent variables. The sample sizes for the association analysis of each pair of metabolite and sleep phenotype are available in figshare ([doi.org/10.6084/m9.figshare.27212511](https://doi.org/10.6084/m9.figshare.27212511)).

together regardless of superpathways, which may suggest a lower degree of diversity in terms of their connectivity patterns.

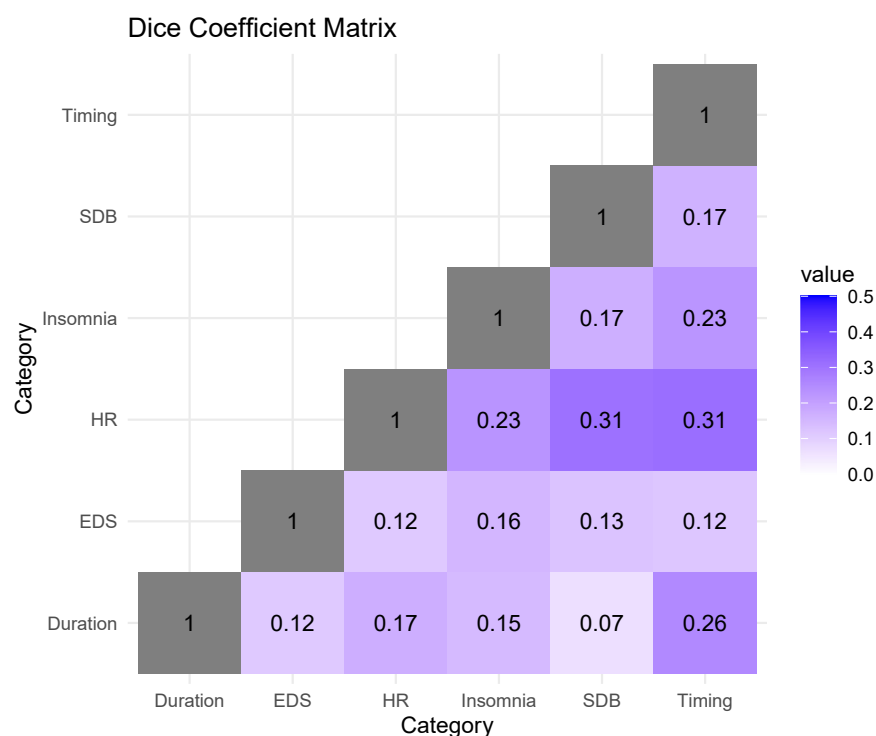
Nestedness of a graph is the property where groups of connected nodes are nested within larger groups of connected nodes. Two versions of nestedness measures were calculated. Both weighted (using the number of sleep domain-superpathway associations) and non-weighted nestedness measures generally agree, and suggest relatively high nestedness among lipids, amino acids, unnamed metabolites as well as combining all metabolites together, and relatively low nestedness among energy, carbohydrates and nucleotides (Supplementary Table S9).

Modularity  $Q$  is a measure of community structure. Modularity values are higher when there are more clusters of connected nodes in a graph. Modularity  $Q$  was the highest among xenobiotics ( $Q = 0.58$ ) and

energy ( $Q = 0.56$ ), followed by peptides ( $Q = 0.54$ ), and the lowest among carbohydrates ( $Q = 0.41$ ), suggesting that among the former three superpathways, there are groups of metabolites in the superpathway that tend to connect to the same sleep phenotype. Among carbohydrates, there are fewer such “communities” of metabolites with similar sleep domain connections.

## Discussion

In this study we performed single metabolite association analyses for a variety of sleep phenotypes including SDB, insomnia, EDS, sleep duration, sleep timing, and heart rate measures during sleep, adjusted for common demographic covariates and BMI, and identified metabolites with statistically significant associations with each sleep phenotype. Individually, none of the metabolites were strong biomarker candidates for the studied



**Fig. 5: Dice coefficient matrix of sleep phenotype domain based on associated metabolites.** Dice coefficient was calculated based on the shared associations with metabolites between any two sleep-related phenotype domain over the total significant associations for both sleep phenotype domains combined. A high value (maximum value of 1) indicates complete overlap, while a low value (minimum value of 0) indicates no overlap between two sleep phenotype domains. Statistically significant associations are defined as FDR-corrected *p*-value, derived using the Benjamini-Hochberg method to control false discovery rate (FDR) for multiple testing across all metabolites within each model and sex stratum for each sleep phenotype is below 0.05; while the unadjusted two-sided *p* values were derived by accounting for the complex sampling design-based degrees of freedom, using adjusted standard errors to compute the *t*-statistic in single metabolite association analysis with each sleep phenotype as dependent variables. The sample sizes for the association analysis of each pair of metabolite and sleep phenotype are available in figshare (doi.org/10.6084/m9.figshare.27212511).

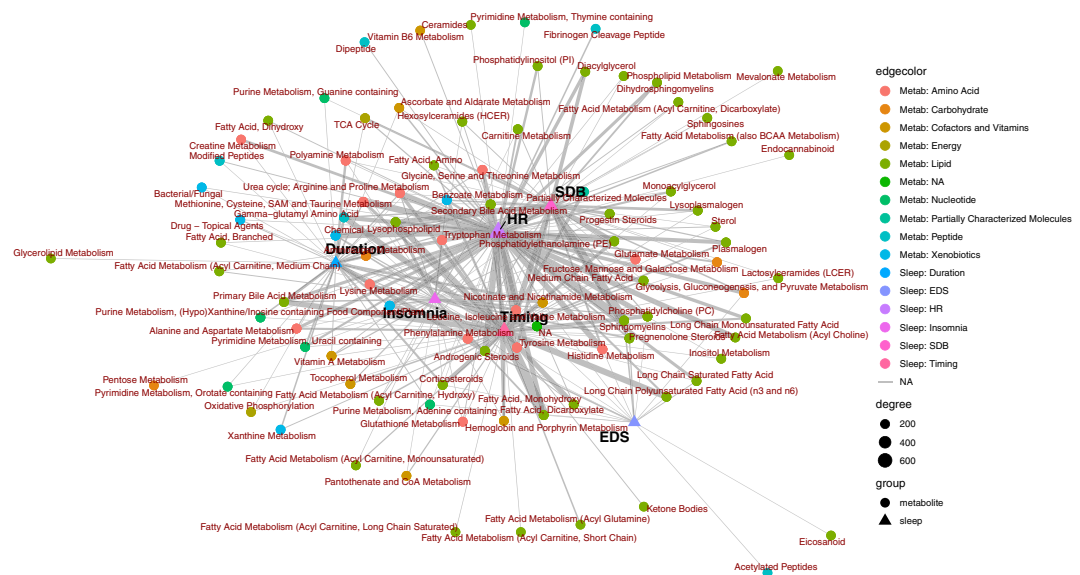
phenotypes. Taking a network analysis approach, we developed a bipartite network based on the identified associations between metabolites and sleep phenotypes. This approach characterises the connectivity patterns of metabolites from different metabolite super- and sub-pathways, and identifies “neighbouring” metabolites and sleep phenotypes that are inter-connected, as well as metabolomic subpathways that may be of interest for further study.

Many of the metabolites identified in our study as associated with sleep phenotypes have been previously reported to be linked to various sleep phenotypes, usually concurring with our findings. Among the top 10% of connected metabolites with the highest number of associations, regardless of sleep phenotype domain, almost all metabolites were reported in prior publications (Table 3), of which 18 were related to sleep. These include some published results from our prior work in HCHS/SOL (i.e., non-independent findings). Glycine, for instance, has shown associations with multiple SDB phenotypes, and was reported to be associated with

sleep deprivation.<sup>44,75</sup> In a recent study, glycine was found to be depleted in blood among Hispanic/Latino individuals with severe SDB, likely due to microbiome changes in an oxygen-poor environment.<sup>45</sup> Four other metabolites – 1-oleoyl-GPE (18:1), 1-(1-enyl-palmitoyl)-GPC (P-16:0)\*, linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]\* and pregnenolone sulfate, previously reported to be associated with novel SDB phenotype metrics after dimension reduction in our prior work in HCHS/SOL,<sup>63</sup> were here also associated with several SDB phenotypes as well as with phenotypes from domains including insomnia, sleep timing, and heart rate during sleep.

Many of the metabolites identified in this study were also reported to be associated with comorbidities such as cardiovascular, metabolic, and neurodegenerative diseases, potentially connecting sleep to a wide range of chronic adverse health outcomes. Four metabolites listed in Table 3 belong to the group of N-lactoyl amino acids—specifically, N-lactoyl isoleucine, N-lactoyl phenylalanine, N-lactoyl leucine, and N-lactoyl valine—which were previously misidentified as 1-carboxyethyl





**Fig. 6: Network based on associations between metabolites and sleep phenotypes aggregated by subpathway and sleep phenotype domain using the Fruchterman-Reingold force-directed algorithm.** Here we present a network of nodes and edges in which each node represents either a sleep phenotype domain or a metabolomic subpathway, while each edge represents one or more statistically significant associations between two nodes based on the single metabolite association analysis results. The width of the edge (i.e., degree) indicates the number of statistically significant associations between two nodes. The placement of nodes is based on connectivity similarity between nodes, where similar connectivity is reflected as proximity of nodes. Statistically significant associations are defined as FDR-corrected *p*-value, derived using the Benjamini-Hochberg method to control false discovery rate (FDR) for multiple testing across all metabolites within each model and sex stratum for each sleep phenotype is below 0.05; while the unadjusted two-sided *p* values were derived by accounting for the complex sampling design-based degrees of freedom, using adjusted standard errors to compute the *t*-statistic in single metabolite association analysis with each sleep phenotype as dependent variables. The sample sizes for the association analysis of each pair of metabolite and sleep phenotype are available in figshare (doi.org/10.6084/m9.figshare.27212511).

amino acids by Metabolon.<sup>37</sup> N-lactoyl amino acids (lac-AAs) are an emerging class of bioactive metabolites that link amino acid and energy metabolism. They are formed through cytosolic nonspecific dipeptidase 2 (CNDP2)-mediated reverse proteolysis of lactate with amino acids.<sup>76</sup> Recent studies have associated lac-AAs with various physiological and pathological conditions, including exercise,<sup>77</sup> prediabetes and diabetes,<sup>47</sup> appetite suppression,<sup>37</sup> and poor outcomes in septic shock.<sup>78</sup> Lac-AAs may link sleep-related phenotypes and metabolic disorders such as obesity and diabetes. Cortisol, as the end product of the hypothalamic-pituitary-adrenal (HPA) axis, has shown to be associated with sleep disturbance in a bidirectional manner,<sup>79</sup> and linked to increased adiposity, insulin resistance, and dyslipidaemia.<sup>80</sup> Hexadecanedioate (C16-DC), a dicarboxylate fatty acid, was reported to be associated with blood pressure regulation<sup>65</sup> and neurodegenerative diseases.<sup>66</sup> Beta-cryptoxanthin, an antioxidant and pre-vitamin A carotenoid found in fruits and vegetables, was found positively associated with cognition (analysis was not adjusted for sleep traits) in individuals of diverse race/ethnic backgrounds.<sup>59</sup> These metabolites suggest some shared biochemical mechanisms between sleep and other chronic adverse health outcomes.

When working with complex data such as untargeted metabolomic profiling and multiple phenotypes, effectively summarise data and develop useful insights is challenging.<sup>81</sup> Different visualisation and analytical approaches have been developed to facilitate this process.<sup>82</sup> Here, we applied a systems biological approach – network analysis which has been widely used in gene expression, gene regulation, gene-disease network, and drug–drug interaction studies, on statistical relations among multiple phenotypes and metabolites. This data-driven network approach is different from knowledge-based network construction approaches<sup>83</sup> built on biochemical relations such as KEGG networks.<sup>84</sup> Here, the network was built on inferred statistical relations among metabolites and phenotypes, leveraging, for interpretation purposes, well-studied network properties metrics from other fields such as ecology, socioeconomics, neuroscience, drug-disease networks, among others.<sup>85–88</sup> One interesting observation is the relatively high nestedness among lipids and amino acids. High nestedness indicates that metabolites from these two superpathways form such a structure that metabolites with fewer connections (referred to as “degree”) are more likely to “connect”, via their mutual connections with sleep domains, with metabolites with higher

Metabolite	Super pathway	Sub pathway	Prior publications
Glycine	Amino Acid	Glycine, Serine and Threonine Metabolism	Sleep deprivation, <sup>44</sup> OSA <sup>45</sup>
N-acetyl glycine	Amino Acid	Glycine, Serine and Threonine Metabolism	Sleep restriction <sup>46</sup>
N-lactoyl isoleucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism	Early onset of T2D, <sup>47</sup> mitochondrial disease <sup>48</sup>
N-lactoyl leucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism	Diabetic retinopathy, <sup>49</sup> mitochondrial disease <sup>48</sup>
N-lactoyl valine	Amino Acid	Leucine, Isoleucine and Valine Metabolism	Early onset of T2D, <sup>47</sup> mitochondrial disease <sup>48</sup>
3-methyl-2-oxovalerate	Amino Acid	Leucine, Isoleucine and Valine Metabolism	Impaired fasting glucose <sup>50</sup>
Isoleucine	Amino Acid	Leucine, Isoleucine and Valine Metabolism	Sleep disturbances, circadian rhythms, DM, CVD, obesity <sup>51</sup>
N-lactoyl phenylalanine	Amino Acid	Phenylalanine Metabolism	Metformin, body weight and food intake, <sup>37,52</sup> mitochondrial disease <sup>48</sup>
Indole propionate	Amino Acid	Tryptophan Metabolism	T2D <sup>53,54</sup>
Vanillyl mandelate (VMA)	Amino Acid	Tyrosine Metabolism	Sleep deprivation <sup>55</sup>
1-methylnicotinamide	Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism	Sleep restriction <sup>56</sup>
N1-Methyl-2-pyridone-5-carboxamide	Cofactors and Vitamins	Nicotinate and Nicotinamide Metabolism	Sleep restriction <sup>56</sup>
Beta-cryptoxanthin	Cofactors and Vitamins	Vitamin A Metabolism	Sleep duration, <sup>57</sup> sleepiness and sleep disturbance <sup>58</sup> cognitive function <sup>59</sup>
Deoxycarnitine	Lipid	Carnitine Metabolism	Cognitive impairment and dementia <sup>60</sup>
Cortisol	Lipid	Corticosteroids	Night time awakenings, <sup>61</sup> CPAP treatment <sup>62</sup>
Linoleoyl-linoleoyl-glycerol (18:2/18:2) [1]*	Lipid	Diacylglycerol	SDB <sup>63</sup>
Hexanoyl glutamine	Lipid	Fatty Acid Metabolism (Acyl Glutamine)	chronic fatigue syndrome <sup>64</sup>
Hexadecanedioate (C16-DC)	Lipid	Fatty Acid, Dicarboxylate	Blood pressure regulation, <sup>65</sup> neurodegenerative diseases <sup>66</sup>
1-oleoyl-GPE (18:1)	Lipid	Lysophospholipid	SDB <sup>63</sup>
1-(1-enyl-palmitoyl)-GPC (P-16:0)*	Lipid	Lysoplasmalogen	SDB <sup>63</sup>
1-palmitoyl-2-linoleoyl-GPE (16:0/18:2)	Lipid	Phosphatidylethanolamine (PE)	Infant sepsis <sup>67</sup>
1-palmitoyl-2-oleoyl-GPE (16:0/18:1)	Lipid	Phosphatidylethanolamine (PE)	Blood pressure <sup>68</sup>
1-stearoyl-2-linoleoyl-GPE (18:0/18:2)*	Lipid	Phosphatidylethanolamine (PE)	
1-stearoyl-2-oleoyl-GPE (18:0/18:1)	Lipid	Phosphatidylethanolamine (PE)	OSA <sup>24</sup>
Pregnenolone sulfate	Lipid	Pregnenolone Steroids	SDB, <sup>63</sup> REM sleep <sup>69</sup>
Glycochenodeoxycholate	Lipid	Primary Bile Acid Metabolism	Poor sleep quality <sup>70</sup>
Glycochenodeoxycholate 3-sulfate	Lipid	Primary Bile Acid Metabolism	OSA <sup>29</sup>
Taurochenodeoxycholate	Lipid	Primary Bile Acid Metabolism	OSA, <sup>26</sup> sleep midpoint <sup>27</sup>
Glycoursodeoxycholate	Lipid	Secondary Bile Acid Metabolism	Bed time <sup>71</sup>
Gamma-glutamyl isoleucine*	Peptide	Gamma-glutamyl Amino Acid	Sleep midpoint, <sup>27</sup> wake time and sleep midpoint <sup>72</sup>
Gamma-glutamyl glycine	Peptide	Gamma-glutamyl Amino Acid	Myocardial infarction <sup>73</sup>
Mannonate*	Xenobiotics	Food Component/Plant	Undernutrition <sup>74</sup>

N-lactoyl isoleucine, N-lactoyl leucine, N-lactoyl valine, N-lactoyl phenylalanine were relabelled from the original annotation (previously 1-carboxyethylisoleucine, 1-carboxyethylleucine, 1-carboxyethylvaline, 1-carboxyethylphenylalanine) based on.<sup>37</sup> Abbreviations. CPAP: continuous positive airway pressure; OSA: obstructive sleep apnoea; REM: repetitive eye movement; SDB: sleep disordered breathing; T2D: type 2 diabetes.

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

degree of connections, rather than with other metabolites with a similar low degree of connectiveness. An intuitive depiction of such network is a “core-periphery” structure, in which a “core” of nodes is connected with other nodes, where a “periphery” of nodes tends to only

connect with the nodes in the core. Contextually, among lipids and amino acids associated with sleep phenotypes, higher nestedness, compared to other superpathways, implies that these superpathways are more likely to have a subset of metabolites that play a “key”

role forming connections with many, and the same, sleep phenotypes, rather than an evenly distributed network where metabolites form connections with various sleep phenotypes from different domains in a random manner. Metabolites with high connectiveness from these two super-pathways likely have roles in shared biological processes across multiple sleep phenotypes, especially considering 26 out of 32 top 10% connected metabolites are either amino acids or lipids. Modularity metrics, on the other hand, measure how well a network can be partitioned into clusters or compartments in which there are dense connections internally and sparser connections with other clusters.<sup>89</sup> Among xenobiotics, energy and peptide metabolites associated with sleep phenotypes, relatively higher modularity compared to other superpathways suggests the existence of such subgroups with distinct relationships with sleep phenotypes.

HR domain had the highest percentage of metabolite associations of these assessed. HR during sleep reflects activity of the autonomic nervous system, and is influenced by cardiac function, sleep stage (i.e., lowest in deep restorative sleep and highest and most variable in REM sleep and wakefulness during the sleep period), and sleep apnea-related heart rate response.<sup>90,91</sup> Notably, both low and elevated heart rate response to SDB events are associated with biomarkers of cardiovascular diseases, while elevated heart rate response to SDB events is predictive of incident fatal and non-fatal CVD.<sup>92</sup> The finding that HR during sleep associated with hundreds of metabolites, from all superpathways, may reflect the multiple biological mechanisms that underlie this phenotype, as well as support the sleep-related HR as a marker of multiple biologically processes that may be targeted for interventions. Sleep timing also was associated with a relatively high number of metabolite associations, supportive of growing data implicating timing-related behaviours on metabolic outcomes,<sup>93</sup> as well as the correlation of sleep timing with other sleep domains.<sup>94</sup> In contrast, fewer associations were observed between SDB phenotypes and metabolites. This may be because that SDB measured by traditional metrics such as the Apnea-Hypopnea Index (AHI) may poorly characterise disease that is influenced by multiple mechanistic pathways.<sup>95</sup>

An analytic choice that we made that is worth discussing is missing metabolite data imputation. Here, we imputed missing values of metabolites with no more than 25% missing values. For non-xenobiotic metabolites, we selected the imputation method based on an empirical investigation of the proportion of replicated associations between batches, and the selected method was multiple imputation that included all metabolite and other measures (covariates and lab values) that have strong associations with metabolite levels. It is natural to question whether this may somehow bias results: for example, is it possible that, for instance, using diabetes

and BMI (among the rest) to impute metabolite values will lead to metabolite values that are “too reflective” of BMI and diabetes and therefore will somehow generate spurious metabolite associations with sleep measures that are associated with BMI and diabetes? The answer is that this is very unlikely. Generation of metabolite values that are overly similar to BMI or diabetes (in this example) by the predictive mean matching function suggests overfitting to the values of these covariates. Such overfitting will reduce the likelihood of replication of associations in the second batch, rather than increase it. Further, the predictive mean matching has some randomness due to sampling, further limiting overfitting to the variables used in imputation. Finally, it is important to note that the correlation between metabolites and covariates, as with sleep, is an inherent characteristic of this biological signal (i.e. it is a “feature, not a bug”), as are the metabolite associations with sleep measures. Therefore, leveraging this characteristic is useful.

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

correlation patterns between them. Finally, the recorded sex data was not verified, and it is possible that the sex of a few individuals was mislabelled.

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

#### Contributors

Ying Z performed association analyses, network analyses, and data visualisation. BWS developed the R/shiny app for visualisation of results from single metabolite association analysis. Yu Z performed data imputation and sensitivity analyses. DAW provided code for modelling circular sleep phenotypes. BY, QQ, and EB designed and established the metabolomics ancillary studies. MA, LA-S, MD, RK, and JC participated in HCHS/SOL study design, recruitment, and operations. Ying Z and TS drafted the manuscript. BWS, Yu Z, DAW, BY, QQ, MA, LA-S, EB, MD, RK, JC, and SR critically reviewed and approved the manuscript. TS supervised the work. Data was accessed and verified by Ying Z, Yu Z, and TS. All authors read and approved the final version of the manuscript.

#### Data sharing statement

HCHS/SOL data are available through application to the data base of genotypes and phenotypes (dbGaP) accession phs000810. HCHS/SOL metabolomics data are available via data use agreement with the HCHS/SOL Data Coordinating Center (DCC) at the University of North Carolina at Chapel Hill, see collaborators website: <https://sites.csc.unc.edu/hchs/>. The metabolite association data generated in this study are provided in figshare (doi.org/10.6084/m9.figshare.27212511) and available for interactive data visualisation via a shiny app [https://bidmc-cardiology-2024.shinyapps.io/sleep\\_metabolite\\_viewer\\_v1/](https://bidmc-cardiology-2024.shinyapps.io/sleep_metabolite_viewer_v1/). The code used in this work has been deposited in the public repository [https://github.com/yzhang104/HCHS\\_SOL\\_Sleep\\_Metabolomics\\_Atlas.git](https://github.com/yzhang104/HCHS_SOL_Sleep_Metabolomics_Atlas.git).

#### Declaration of interests

Dr. Argos reports receiving grants R01ES033883, U01HG013275, U2RTW010122, and R01CA255082, with payments made to the institution. Dr. Redline reports receiving consultant fees from Eli Lilly, with payments made to self, unpaid participation in the Data Safety Monitoring Board or Advisory Board of Apnimed, and unpaid leadership or fiduciary roles in the Alliance for Sleep Apnea Partners and in the National Sleep Foundation. Dr. Sofer reports receiving a grant from the National Human Genome Research Institute, with payments made to the institution.

#### Acknowledgements

The authors thank the staff and participants of HCHS/SOL for their important contributions. Investigators website: <http://www.csc.unc.edu/hchs/>. This work was supported by National Heart, Lung, and Blood Institute (NHLBI) grants R01HL161012 to TS, R35HL135818 to SR, National Institute on Aging grant R01AG80598 to TS. DS was supported by NHLBI grant K99HL166700. Support for metabolomics

data was graciously provided by the JLH Foundation (Houston, Texas) and by NHLBI grant R01HL141824. The Hispanic Community Health Study/Study of Latinos is a collaborative study supported by contracts from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN2682013000011/N01-HC-65233), University of Miami (HHSN2682013000041/N01-HC-65234), Albert Einstein College of Medicine (HHSN2682013000021/N01-HC-65235), University of Illinois at Chicago (HHSN2682013000031/N01-HC-65236 Northwestern Univ), and San Diego State University (HHSN2682013000051/N01-HC-65237). The following Institutes/Centres/Offices have contributed to the HCHS/SOL through a transfer of funds to the NHLBI: National Institute on Minority Health and Health Disparities, National Institute on Deafness and Other Communication Disorders, National Institute of Dental and Craniofacial Research, National Institute of Diabetes and Digestive and Kidney Diseases, National Institute of Neurological Disorders and Stroke, NIH Institution-Office of Dietary Supplements.

#### Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2024.105507>.

#### References

- 1 Chattu VK, Manzar MD, Kumary S, Burman D, Spence DW, Pandi-Perumal SR. The global problem of insufficient sleep and its serious public health implications. *Healthcare (Basel)*. 2018;7(1). <https://doi.org/10.3390/healthcare7010001>.
- 2 Worley SL. The extraordinary importance of sleep: the detrimental effects of inadequate sleep on health and public safety drive an explosion of sleep research. *P T*. 2018;43(12):758–763.
- 3 Ramar K, Malhotra RK, Carden KA, et al. Sleep is essential to health: an American Academy of Sleep Medicine position statement. *J Clin Sleep Med*. 2021;17(10):2115–2119. <https://doi.org/10.5664/jcsm.9476>.
- 4 Ioachimescu OC. From seven sweethearts to life begins at eight thirty: a journey from life's simple 7 to life's essential 8 and beyond. *J Am Heart Assoc*. 2022;11(21):e027658. <https://doi.org/10.1161/JAHA.122.027658>.
- 5 Buysse DJ. Sleep health: can we define it? Does it matter? *Sleep*. 2014;37(1):9–17. <https://doi.org/10.5665/sleep.3298>.
- 6 Chung J, Goodman M, Huang T, et al. Multi-dimensional sleep and mortality: the multi-ethnic study of atherosclerosis. *Sleep*. 2023;46(9). <https://doi.org/10.1093/sleep/zsaa048>.
- 7 Lee S, Mu CX, Wallace ML, et al. Multidimensional sleep health problems across middle and older adulthood predict early mortality. *J Gerontol A Biol Sci Med Sci*. 2024;79(3). <https://doi.org/10.1093/geron/glad258>.
- 8 Lee S, Mu CX, Wallace ML, et al. Sleep health composites are associated with the risk of heart disease across sex and race. *Sci Rep*. 2022;12(1):2023. <https://doi.org/10.1038/s41598-022-05203-0>.
- 9 Sofer T, Kurniansyah N, Murray M, et al. Genome-wide association study of obstructive sleep apnoea in the Million Veteran Program uncovers genetic heterogeneity by sex. *eBioMedicine*. 2023;90:104536. <https://doi.org/10.1016/j.ebiom.2023.104536>.
- 10 Madrid-Valero JJ, Sánchez-Romera JF, Martínez-Selva JM, Ordoñana JR. Phenotypic, genetic and environmental architecture of the components of sleep quality. *Behav Genet*. 2022;52(4-5):236–245. <https://doi.org/10.1007/s10519-022-10111-0>.
- 11 Wang H, Lane JM, Jones SE, et al. Genome-wide association analysis of self-reported daytime sleepiness identifies 42 loci that suggest biological subtypes. *Nat Commun*. 2019;10(1):3503. <https://doi.org/10.1038/s41467-019-11456-7>.
- 12 Sweetman A, Lack L, McEvoy RD, et al. Bi-directional relationships between co-morbid insomnia and sleep apnea (COMISA). *Sleep Med Rev*. 2021;60:101519. <https://doi.org/10.1016/j.smrv.2021.101519>.
- 13 Amatoury J, Azarbarzin A, Younes M, Jordan AS, Wellman A, Eckert DJ. Arousal intensity is a distinct pathophysiological trait in obstructive sleep apnea. *Sleep*. 2016;39(12):2091–2100. <https://doi.org/10.5665/sleep.6304>.
- 14 Keenan BT, Kim J, Singh B, et al. Recognizable clinical subtypes of obstructive sleep apnea across international sleep centers: a cluster analysis. *Sleep*. 2018;41(3). <https://doi.org/10.1093/sleep/zsx214>.



- 15 Zinchuk AV, Jeon S, Koo BB, et al. Polysomnographic phenotypes and their cardiovascular implications in obstructive sleep apnoea. *Thorax*. 2018;73(5):472–480. <https://doi.org/10.1136/thoraxjnl-2017-210431>.
- 16 Zhang L, Wei T-T, Li Y, et al. Functional metabolomics characterizes a key role for N-acetylneuraminic acid in coronary artery diseases. *Circulation*. 2018;137(13):1374–1390. <https://doi.org/10.1161/CIRCULATIONAHA.117.031139>.
- 17 Buerger T, Steinfeldt J, Ruyoga G, et al. Metabolomic profiles predict individual multidisease outcomes. *Nat Med*. 2022;28(11):2309–2320. <https://doi.org/10.1038/s41591-022-01980-3>.
- 18 Nikolic SB, Sharman JE, Adams MJ, Edwards LM. Metabolomics in hypertension. *J Hypertens*. 2014;32(6):1159–1169. <https://doi.org/10.1097/HJH.0000000000000168>.
- 19 Griffin JL, Atherton H, Shockcor J, Atzori L. Metabolomics as a tool for cardiac research. *Nat Rev Cardiol*. 2011;8(11):630–643. <https://doi.org/10.1038/nrcardio.2011.138>.
- 20 Yu B, Flexeder C, McGarrah RW, et al. Metabolomics identifies novel blood biomarkers of pulmonary function and COPD in the general population. *Metabolites*. 2019;9(4). <https://doi.org/10.3390/metabo9040061>.
- 21 Depner CM, Cogswell DT, Bisesi PJ, et al. Developing preliminary blood metabolomics-based biomarkers of insufficient sleep in humans. *Sleep*. 2020;43(7). <https://doi.org/10.1093/sleep/zsz321>.
- 22 He S, Granot-Hershkovitz E, Zhang Y, et al. Blood metabolites predicting mild cognitive impairment in the study of Latinos: investigation of neurocognitive aging (HCHS/SOL). *Alzheimer's Dement (Amst)*. 2022;14(1):e12259. <https://doi.org/10.1002/dad2.12259>.
- 23 Fritz J, Huang T, Depner CM, et al. Sleep duration, plasma metabolites, and obesity and diabetes: a metabolome-wide association study in US women. *Sleep*. 2023;46(1). <https://doi.org/10.1093/sleep/zsac226>.
- 24 Zhang Y, Ngo D, Yu B, et al. Development and validation of a metabolite index for obstructive sleep apnea across race/ethnicities. *Sci Rep*. 2022;12(1):21805. <https://doi.org/10.1038/s41598-022-26321-9>.
- 25 Lin BM, Zhang Y, Yu B, et al. Metabolome-wide association study of estimated glomerular filtration rates in Hispanics. *Kidney Int*. 2022;101(1):144–151. <https://doi.org/10.1016/j.kint.2021.09.032>.
- 26 Pinilla L, Benítez ID, Santamaria-Martos F, et al. Plasma profiling reveals a blood-based metabolic fingerprint of obstructive sleep apnea. *Biomed Pharmacother*. 2022;145:112425. <https://doi.org/10.1016/j.biopha.2021.112425>.
- 27 Xiao Q, Derkach A, Moore SC, et al. Habitual Sleep and human plasma metabolomics. *Metabolomics*. 2017;13(5). <https://doi.org/10.1007/s11306-017-1205-z>.
- 28 Humer E, Pieh C, Brandmayr G. Metabolomics in sleep, insomnia and sleep apnea. *Int J Mol Sci*. 2020;21(19). <https://doi.org/10.3390/ijms21197244>.
- 29 Xu H, Zheng X, Qian Y, et al. Metabolomics profiling for obstructive sleep apnea and simple snorers. *Sci Rep*. 2016;6:30958. <https://doi.org/10.1038/srep30958>.
- 30 Lebkuhen A, Carvalho VM, Venturini G, et al. Metabolomic and lipidomic profile in men with obstructive sleep apnoea: implications for diagnosis and biomarkers of cardiovascular risk. *Sci Rep*. 2018;8(1):11270. <https://doi.org/10.1038/s41598-018-29727-6>.
- 31 Mohit, Tomar MS, Sharma D, et al. Emerging role of metabolomics for biomarker discovery in obstructive sleep apnea. *Sleep Breath*. 2023;27(4):1247–1254. <https://doi.org/10.1007/s11325-022-02730-y>.
- 32 Lavigne LM, Kalsbeek WD, Sorlie PD, et al. Sample design and cohort selection in the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20(8):642–649. <https://doi.org/10.1016/j.annepidem.2010.05.006>.
- 33 Sorlie PD, Avilés-Santa LM, Wassertheil-Smolter S, et al. Design and implementation of the Hispanic Community Health Study/Study of Latinos. *Ann Epidemiol*. 2010;20(8):629–641. <https://doi.org/10.1016/j.annepidem.2010.03.015>.
- 34 Pirzada A, Cai J, Heiss G, et al. Evolving science on cardiovascular disease among Hispanic/Latino adults: JACC international. *J Am Coll Cardiol*. 2023;81(15):1505–1520. <https://doi.org/10.1016/j.jacc.2023.02.023>.
- 35 Redline S, Sotres-Alvarez D, Loredó J, et al. Sleep-disordered breathing in hispanic/Latino individuals of diverse backgrounds. The Hispanic Community Health Study/Study of Latinos. *Am J Respir Crit Care Med*. 2014;189(3):335–344. <https://doi.org/10.1164/rccm.201309-1735OC>.
- 36 Gonzalez S, Strizich G, Isasi CR, et al. Consent for use of genetic data among US hispanics/Latinos: results from the Hispanic Community Health Study/Study of Latinos. *Ethn Dis*. 2021;31(4):547–558. <https://doi.org/10.18865/ed.31.4.547>.
- 37 Scott B, Day EA, O'Brien KL, et al. Metformin and feeding increase levels of the appetite-suppressing metabolite Lac-Phe in humans. *Nat Metab*. 2024;6(4):651–658. <https://doi.org/10.1038/s42255-024-01018-7>.
- 38 Kim S, SenGupta A. Regressions involving circular variables: an overview. In: Chattopadhyay AK, Chattopadhyay G, eds. *Statistics and its Applications: Platinum Jubilee Conference, Kolkata, India, December 2016*. 244.
- 39 Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91(434):473. <https://doi.org/10.2307/2291635>.
- 40 Liu Y, Chen S, Li Z, Morrison AC, Boerwinkle E, Lin X. ACAT: a fast and powerful p value combination method for rare-variant analysis in sequencing studies. *Am J Hum Genet*. 2019;104(3):410–421. <https://doi.org/10.1016/j.ajhg.2019.01.002>.
- 41 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B*. 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- 42 Sorensen JT. A method of establishing group of equal amplitude in plant sociobiology based on similarity of species content and its application to analyses of the vegetation on Danish commons. 1948.
- 43 Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Software Pract Exp*. 1991;21(11):1129–1164. <https://doi.org/10.1002/spe.4380211102>.
- 44 Gou X-J, Cen F, Fan Z-Q, Xu Y, Shen H-Y, Zhou M-M. Serum and brain metabolomic variations reveal perturbation of sleep deprivation on rats and ameliorate effect of total ginsenoside treatment. *Int J Genomics*. 2017;2017:5179271. <https://doi.org/10.1155/2017/5179271>.
- 45 Elgart M, Zhang Y, Zhang Y, et al. Anaerobic pathogens associated with OSA may contribute to pathophysiology via amino-acid depletion. *eBioMedicine*. 2023;98:104891. <https://doi.org/10.1016/j.ebiom.2023.104891>.
- 46 Bell LN, Kilkus JM, Booth JN, Bromley LE, Imperial JG, Penev PD. Effects of sleep restriction on the human plasma metabolome. *Physiol Behav*. 2013;122:25–31. <https://doi.org/10.1016/j.physbeh.2013.08.007>.
- 47 Sharma S, Dong Q, Haid M, et al. Role of human plasma metabolites in prediabetes and type 2 diabetes from the IMI-DIRECT study. *Diabetologia*. 2024;30. <https://doi.org/10.1007/s00125-024-06282-6>.
- 48 Gucek M, Sack MN. Proteomic and metabolomic advances uncover biomarkers of mitochondrial disease pathophysiology and severity. *J Clin Invest*. 2021;131(2). <https://doi.org/10.1172/JCI145158>.
- 49 Fernandes Silva L, Hokkanen J, Vangipurapu J, Oravilahti A, Laakso M. Metabolites as risk factors for diabetic retinopathy in patients with type 2 diabetes: a 12-year follow-up study. *J Clin Endocrinol Metab*. 2023;109(1):100–106. <https://doi.org/10.1210/clinem/dgad452>.
- 50 Menni C, Fauman E, Erte I, et al. Biomarkers for type 2 diabetes and impaired fasting glucose using a nontargeted metabolomics approach. *Diabetes*. 2013;62(12):4270–4276. <https://doi.org/10.2337/db13-0570>.
- 51 Li H, Seugnet L. Decoding the nexus: branched-chain amino acids and their connection with sleep, circadian rhythms, and cardiometabolic health. *Neural Regen Res*. 2024;20(5):1350–1363. <https://doi.org/10.4103/NRR.NRR-D-23-02020>.
- 52 Xiao S, Li VL, Lyu X, et al. Lac-Phe mediates the effects of metformin on food intake and body weight. *Nat Metab*. 2024;6(4):659–669. <https://doi.org/10.1038/s42255-024-00999-9>.
- 53 Wang Z, Peters BA, Yu B, et al. Gut microbiota and blood metabolites related to fiber intake and type 2 diabetes. *Circ Res*. 2024;134(7):842–854. <https://doi.org/10.1161/CIRCRESAHA.123.323634>.
- 54 Hu Y, Li J, Wang B, et al. Interplay between diet, circulating indolepropionate concentrations and cardiometabolic health in US populations. *Gut*. 2023;72(12):2260–2271. <https://doi.org/10.1136/gutjnl-2023-330410>.
- 55 Müller HU, Riemann D, Berger M, Müller WE. The influence of total sleep deprivation on urinary excretion of catecholamine metabolites in major depression. *Acta Psychiatr Scand*. 1993;88(1):16–20. <https://doi.org/10.1111/j.1600-0447.1993.tb03407.x>.
- 56 Sengupta A, Rhoades SD, Kim EJ, et al. Sleep restriction induced energy, methylation and lipogenesis metabolic switches in rat liver.



- Int J Biochem Cell Biol.* 2017;93:129–135. <https://doi.org/10.1016/j.biocel.2017.08.014>.
- 57 Deng M-G, Liu F, Wang K, Liang Y, Nie J-Q, Liu J. Relationship between dietary carotenoid intake and sleep duration in American adults: a population-based study. *Nutr J.* 2023;22(1):68. <https://doi.org/10.1186/s12937-023-00898-x>.
  - 58 Beydoun MA, Gamaldo AA, Canas JA, et al. Serum nutritional biomarkers and their associations with sleep among US adults in recent national surveys. *PLoS One.* 2014;9(8):e103490. <https://doi.org/10.1371/journal.pone.0103490>.
  - 59 Granot-Herskovitz E, He S, Bressler J, et al. Plasma metabolites associated with cognitive function across race/ethnicities affirming the importance of healthy nutrition. *Alzheimers Dement.* 2023;19(4):1331–1342. <https://doi.org/10.1002/alz.12786>.
  - 60 de Oliveira Otto MC, Li XS, Wang Z, et al. Longitudinal associations of plasma TMAO and related metabolites with cognitive impairment and dementia in older adults: the cardiovascular health study. *J Alzheimers Dis.* 2022;89(4):1439–1452. <https://doi.org/10.3233/JAD-220477>.
  - 61 Rösler L, van Kesteren E-J, Leerssen J, et al. Hyperarousal dynamics reveal an overnight increase boosted by insomnia. *J Psychiatr Res.* 2024;179:279–285. <https://doi.org/10.1016/j.jpsychires.2024.09.032>.
  - 62 Ken-Dror G, Fry CH, Murray P, Fluck D, Han TS. Changes in cortisol levels by continuous positive airway pressure in patients with obstructive sleep apnoea: meta-analysis of 637 individuals. *Clin Endocrinol.* 2021;95(6):909–917. <https://doi.org/10.1111/cen.14573>.
  - 63 Zhang Y, Yu B, Qi Q, et al. Metabolomic profiles of sleep-disordered breathing are associated with hypertension and diabetes mellitus development. *Nat Commun.* 2024;15(1):1845. <https://doi.org/10.1038/s41467-024-46019-y>.
  - 64 Germain A, Barupal DK, Levine SM, Hanson MR. Comprehensive circulatory metabolomics in ME/CFS reveals disrupted metabolism of acyl lipids and steroids. *Metabolites.* 2020;10(1). <https://doi.org/10.3390/metabo10010034>.
  - 65 Menni C, Graham D, Kastenmüller G, et al. Metabolomic identification of a novel pathway of blood pressure regulation involving hexadecanedioate. *Hypertension.* 2015;66(2):422–429. <https://doi.org/10.1161/HYPERTENSIONAHA.115.05544>.
  - 66 Chen H, Qiao J, Wang T, Shao Z, Huang S, Zeng P. Assessing causal relationship between human blood metabolites and five neurodegenerative diseases with GWAS summary statistics. *Front Neurosci.* 2021;15:680104. <https://doi.org/10.3389/fnins.2021.680104>.
  - 67 Wang L, Cha X, Zhang Z, Qian J. Discrimination of serum metabolomics profiles in infants with sepsis, based on liquid chromatography-mass spectrometer. *BMC Infect Dis.* 2023;23(1):46. <https://doi.org/10.1186/s12879-023-07983-w>.
  - 68 Sun Y, Zhang R, Tian L, et al. Novel metabolites associated with blood pressure after dietary interventions. *Hypertension.* 2024;81(9):1966–1975. <https://doi.org/10.1161/HYPERTENSIONAHA.124.22999>.
  - 69 Darnaudéry M, Bouyer JJ, Pallarés M, Le Moal M, Mayo W. The promnesic neurosteroid pregnenolone sulfate increases paradoxical sleep in rats. *Brain Res.* 1999;818(2):492–498. [https://doi.org/10.1016/S0006-8993\(98\)01338-9](https://doi.org/10.1016/S0006-8993(98)01338-9).
  - 70 Huang T, Zeleznik OA, Poole EM, et al. Habitual sleep quality, plasma metabolites and risk of coronary heart disease in postmenopausal women. *Int J Epidemiol.* 2019;48(4):1262–1274. <https://doi.org/10.1093/ije/dyy234>.
  - 71 Xiang X, Chen J, Zhu M, Gao H, Liu X, Wang Q. Multiomics revealed the multi-dimensional effects of late sleep on gut microbiota and metabolites in children in northwest China. *Nutrients.* 2023;15(20). <https://doi.org/10.3390/nu15204315>.
  - 72 Gordon-Dseagu VLZ, Derkach A, Xiao Q, Williams I, Sampson J, Stolzenberg-Solomon RZ. The association of sleep with metabolic pathways and metabolites: evidence from the Dietary Approaches to Stop Hypertension (DASH)-sodium feeding study. *Metabolomics.* 2019;15(4):48. <https://doi.org/10.1007/s11306-019-1472-y>.
  - 73 Li D-H, Wu Q, Lan J-S, et al. Plasma metabolites and risk of myocardial infarction: a bidirectional Mendelian randomization study. *J Geriatr Cardiol.* 2024;21(2):219–231. <https://doi.org/10.26599/1671-5411.2024.02.002>.
  - 74 Preidis GA, Keaton MA, Campeau PM, Bessard BC, Conner ME, Hotez PJ. The undernourished neonatal mouse metabolome reveals evidence of liver and biliary dysfunction, inflammation, and oxidative stress. *J Nutr.* 2014;144(3):273–281. <https://doi.org/10.3945/jn.113.183731>.
  - 75 Ma W-N, Zhou M-M, Gou X-J, et al. Urinary metabolomic study of chlorogenic acid in a rat model of chronic sleep deprivation using gas chromatography-mass spectrometry. *Int J Genomics.* 2018;2018:1361402. <https://doi.org/10.1155/2018/1361402>.
  - 76 Jansen RS, Addie R, Merckx R, et al. N-lactoyl-amino acids are ubiquitous metabolites that originate from CNDP2-mediated reverse proteolysis of lactate and amino acids. *Proc Natl Acad Sci U S A.* 2015;112(21):6601–6606. <https://doi.org/10.1073/pnas.1424638112>.
  - 77 Li VL, He Y, Contrepois K, et al. An exercise-inducible metabolite that suppresses feeding and obesity. *Nature.* 2022;606(7915):785–790. <https://doi.org/10.1038/s41586-022-04828-5>.
  - 78 Rogers RS, Sharma R, Shah HB, et al. Circulating N-lactoyl-amino acids and N-formyl-methionine reflect mitochondrial dysfunction and predict mortality in septic shock. *Metabolomics.* 2024;20(2):36. <https://doi.org/10.1007/s11306-024-02089-z>.
  - 79 El Mili N, Ahabrach H, Cauli O. Hair cortisol concentration as a biomarker of sleep quality and related disorders. *Life.* 2021;11(2). <https://doi.org/10.3390/life11020081>.
  - 80 Ortiz R, Kluwe B, Lazarus S, Teruel MN, Joseph JJ. Cortisol and cardiometabolic disease: a target for advancing health equity. *Trends Endocrinol Metab.* 2022;33(11):786–797. <https://doi.org/10.1016/j.tem.2022.08.002>.
  - 81 Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted metabolomics strategies-challenges and emerging directions. *J Am Soc Mass Spectrom.* 2016;27(12):1897–1905. <https://doi.org/10.1007/s13361-016-1469-y>.
  - 82 Alonso A, Marsal S, Julià A. Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng Biotechnol.* 2015;3:23. <https://doi.org/10.3389/fbioe.2015.00023>.
  - 83 Galeano J, Pastor JM, Iriando JM. Weighted-Interaction Nestedness Estimator (WINE): a new estimator to calculate over frequency matrices. *Environ Model Software.* 2009;24(11):1342–1346. <https://doi.org/10.1016/j.envsoft.2009.05.014>.
  - 84 Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28(1):27–30. <https://doi.org/10.1093/nar/28.1.27>.
  - 85 Ulrich W, Almeida-Neto M, Gotelli NJ. A consumer's guide to nestedness analysis. *Oikos.* 2009;118(1):3–17. <https://doi.org/10.1111/j.1600-0706.2008.17053.x>.
  - 86 König MD, Tessone CJ, Zenou Y. Nestedness in networks: a theoretical model and some applications. *Theor Econ.* 2014;9(3):695–752. <https://doi.org/10.3982/TE1348>.
  - 87 Centeno EGZ, Moreni G, Vriend C, Douw L, Santos FAN. A hands-on tutorial on network and topological neuroscience. *Brain Struct Funct.* 2022;227(3):741–762. <https://doi.org/10.1007/s00429-021-02435-0>.
  - 88 Rintala TJ, Ghosh A, Fortino V. Network approaches for modeling the effect of drugs and diseases. *Brief Bioinformatics.* 2022;23(4). <https://doi.org/10.1093/bib/bbac229>.
  - 89 Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2004;69(2 Pt 2):026113. <https://doi.org/10.1103/PhysRevE.69.026113>.
  - 90 Aldredge JL, Welch AJ. Variations of heart rate during sleep as a function of the sleep cycle. *Electroencephalogr Clin Neurophysiol.* 1973;35(2):193–198. [https://doi.org/10.1016/0013-4694\(73\)90176-4](https://doi.org/10.1016/0013-4694(73)90176-4).
  - 91 Yang C-K, Jordan AS, White DP, Winkelman JW. Heart rate response to respiratory events with or without leg movements. *Sleep.* 2006;29(4):553–556. <https://doi.org/10.1093/sleep/29.4.553>.
  - 92 Azarbarzin A, Sands SA, Younes M, et al. The sleep apnea-specific pulse-rate response predicts cardiovascular morbidity and mortality. *Am J Respir Crit Care Med.* 2021;203(12):1546–1555. <https://doi.org/10.1164/rccm.202010-3900OC>.
  - 93 Duffy JF, Abbott SM, Burgess HJ, et al. Workshop report. Circadian rhythm sleep-wake disorders: gaps and opportunities. *Sleep.* 2021;44(5). <https://doi.org/10.1093/sleep/zsaa281>.
  - 94 Chung J, Goodman M, Huang T, Bertisch S, Redline S. Multidimensional sleep health in a diverse, aging adult cohort: concepts, advances, and implications for research and intervention. *Sleep Health.* 2021;7(6):699–707. <https://doi.org/10.1016/j.sleh.2021.08.005>.
  - 95 Redline S, Azarbarzin A, Peker Y. Obstructive sleep apnoea heterogeneity and cardiovascular disease. *Nat Rev Cardiol.* 2023;20(8):560–573. <https://doi.org/10.1038/s41569-023-00846-6>.