

Covariance patterns between sleep health domains and distributed intrinsic functional connectivity

Received: 15 January 2023

Accepted: 25 October 2023

Published online: 06 November 2023

 Check for updates

Yulin Wang^{1,2}, Sarah Genon^{3,4}, Debo Dong^{2,3}, Feng Zhou², Chenyu Li⁵, Dahua Yu⁶, Kai Yuan⁷, Qinghua He², Jiang Qiu², Tingyong Feng², Hong Chen² & Xu Lei^{1,2}✉

Sleep health is both conceptually and operationally a composite concept containing multiple domains of sleep. In line with this, high dependence and interaction across different domains of sleep health encourage a transition in sleep health research from categorical to dimensional approaches that integrate neuroscience and sleep health. Here, we seek to identify the covariance patterns between multiple sleep health domains and distributed intrinsic functional connectivity by applying a multivariate approach (partial least squares). This multivariate analysis reveals a composite sleep health dimension co-varying with connectivity patterns involving the attentional and thalamic networks and which appear relevant at the neuromolecular level. These findings are further replicated and generalized to several unseen independent datasets. Critically, the identified sleep-health related connectome shows diagnostic potential for insomnia disorder. These results together delineate a potential brain connectome biomarker for sleep health with high potential for clinical translation.

It is increasingly recognized that sleep health (SH) is a multi-dimensional construct^{1,2}. Sleep health has been defined as “a multi-dimensional pattern of sleep-wakefulness, adapted to individual, social, and environmental demands, that promotes physical and mental well-being”³. Resting on this theoretical consideration and previous researches examining associations between different sleep measures and health outcomes^{3–5}, sleep health mainly consists of six domains of sleep and circadian functioning: Regularity in sleep, Satisfaction with sleep/sleep quality, Alertness during waking hours, Timing of sleep, Sleep Efficiency/Continuity, and Sleep Duration (Ru-SATED). Emerging studies also pointed out that there exist other domains that can be considered as relevant to the construction of

sleep health, such as sleep quality⁶, insomnia symptoms⁷, and sleep medication use⁷.

Recent studies have begun to construct a sleep health composite towards studying multiple domains of sleep^{1,8–10}. Indeed, previous research has mostly focused on single indicators of sleep health, making it difficult to provide a consistent guideline for research and practical settings⁶. A composite measure not only would provide a more comprehensive indicator of sleep health than any single standard measure^{11–13} but also would be less prone to noise and hence could present a greater neurobiological validity¹⁴. For instance, Dalmasas and colleagues reported that the Ru-SATED score was more strongly linked with self-rated health status than sleep duration

¹Sleep and Neuroimaging Center, Faculty of Psychology, Southwest University, Chongqing, China. ²Key Laboratory of Cognition and Personality, Ministry of Education, Faculty of Psychology, Southwest University, Chongqing, China. ³Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany. ⁴Institute for Systems Neuroscience, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. ⁵Sleep Center, Department of Brain Disease, Chongqing Traditional Chinese Medicine Hospital, Chongqing, China. ⁶Information Processing Laboratory, School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia, China. ⁷School of Life Science and Technology, Xidian University, Xi'an, Shanxi, China. ✉ e-mail: xlei@swu.edu.cn

alone¹⁵. Nonetheless, considerations of constructing a sleep health composite with multi-domains of sleep health rather than individual sleep metrics is still an emerging research area with many remaining gaps in the literature¹⁴. In particular, the neurobiological validity of a composite score could be importantly increased by capitalizing on its association with intrinsic functional connectivity in the population.

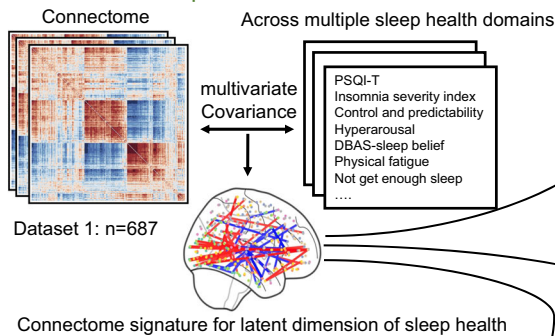
Resting-state networks are typically derived from the connectivity profile of spontaneous fluctuations in functional MRI (fMRI) signals and are thought to reflect the intrinsic brain functional connectivity^{16,17}. Converging research evidence across datasets, methods, and laboratories has agreed upon the principle resting-state networks such as the somatomotor, visual, default mode, and fronto-parietal control networks^{18,19}. In that context, resting-state functional connectivity (RSFC), such as the anti-correlations between the default mode network (DMN) and the task-positive network (TPN) was found to be a reliable biomarker to classify different sleep stages [for a review, see²⁰]. Also, RSFC was found to be influenced by sleep deprivation evidenced by the reduced connectivity within the DMN²¹, the dorsal attention network (DAN), and the auditory, visual, and motor networks [for a tabular overview, see ref. 16]. Moreover, abnormalities in functional network modules subserving hyperarousal, salience, sensory-motor, cognitive, and self-referential processes, including the limbic, thalamus, sensory-motor, fronto-parietal control, and default mode networks have been shown in insomnia disorder [for a review, see refs. 22,23]. These lines of research together suggest RSFC contains crucial information underlying the neurobiological mechanisms of different aspects of sleep health.

Actually, RSFC has been shown to be associated with some domains of sleep health such as sleep duration²⁴, sleep quality²⁵ and timing of sleep (or chronotype)²⁶. Despite this increasing interest in revealing the interplay between intrinsic brain functional connectivity and domains of sleep health, existing studies have been limited in

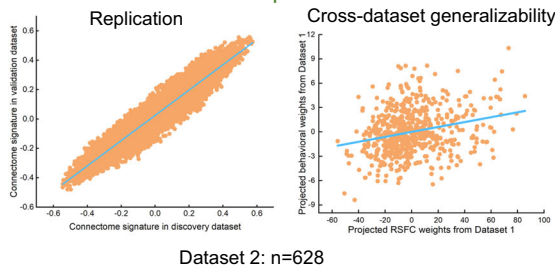
several respects. First, most have adopted a categorical approach, or only examined a single domain of sleep health, and are therefore unable to capture the heterogeneity across different sleep health domains. Second, constructions of sleep health composite were mainly driven by the conceptual Ru-SATED model, rather than being guided by the intrinsic structure of the brain and behavior features. Third, existing work in sleep health composite has often used relatively small samples (e.g., dozens of participants)²⁷. While multivariate techniques allow the examination of both multiple brain systems and sleep health domains simultaneously, such techniques usually require large samples^{28,29}. Also in line with these considerations, the high dependence and interaction across different domains of sleep health (e.g., short sleep duration is usually accompanied by lower efficiency and regularity, lower sleep satisfaction/quality) encourages a transition in sleep health research¹⁴ from categorical to dimensional approaches that integrate intrinsic brain functional connectivity and sleep health.

In this study, we addressed this research gap by relating RSFC to a large set of sleep health measures in a single integrated analysis to identify a brain connectome biomarker for sleep health. More specifically, we addressed the following four research questions (Fig. 1): (1) Can we identify a composite sleep health dimension that relates to RSFC patterns? (2) Can the obtained dimension be replicated and generalized to unseen independent datasets? (3) A growing number of studies highlights the relationships between alterations of some neurotransmitter systems including the serotonin receptors, glutamate, and γ -aminobutyric acid (GABA) and sleep disturbances or “unhealthy” sleep^{30–32}. Accordingly, we here examined whether the brain connectome underlying interindividual differences in sleep phenotype may be related to specific neurotransmitter systems (based on neurotransmitter receptor maps³³). (4) Does the sleep-health-related brain connectome have predictive utility for both the health population and insomnia disorder?

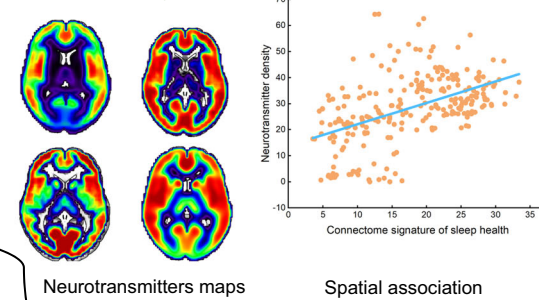
Q1: Can we identify a composite sleep health dimension that relates to RSFC patterns?



Q2: Can the obtained dimension be replicated and generalized to unseen independent datasets?



Q3: What is the neuromolecular relevance of the brain connectome underlying interindividual differences in sleep phenotype?



Q4: Does the sleep-health related brain connectome have predictive utility in health population and insomnia?

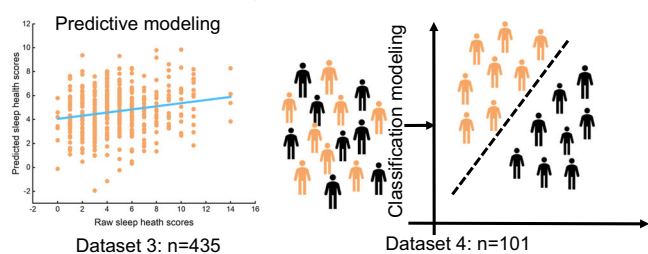


Fig. 1 | Overview of research questions and main analyses. This study aims to answer four research questions (Q1–4) to identify a brain connectome biomarker for sleep health. To do so, this study used the multivariate and machine-learning

approach (i.e., partial least square, predictive modeling, classification) with four independent datasets.

To answer these questions, we first have capitalized on a large sample of young adults from the Behavioral Brain Research Project to Chinese Personality (BBP) by applying a machine-learning technique called partial least squares (PLS)^{34–36}. As a multivariate data-driven statistical technique, PLS is capable of discovering the latent dimensions that link interindividual variability in intrinsic functional organization to interindividual variability in behavioral measures spanning multiple domains of sleep health. To comprehensively assess this latter, we included measures suggested by the Ru-SATED model, but we also incorporated several additional metrics associated with sleep health defined broadly⁶, that are: (1) beliefs, attitudes, and habits about sleep, considering that individuals with more-positive beliefs and attitudes about sleep or who adopt sleep-protective behaviors are more likely to experience better sleep health³⁷; (2) sleep deficiency which also incorporates insomnia symptoms, sleep disorders and sleep medication use, considering both sleep health and sleep deficiency can be regarded as the anchors at either end of a continuum³; (3) have one slept enough, whether one felt refreshed upon waking³⁸, the necessity of nap and the needed nap time were also incorporated to evaluate the satisfaction with sleep/sleep quality⁶; (4) tendency to engage in spontaneous waking thought³⁹, attention-related deficiency⁴⁰ and fatigue⁴¹ as supplements to assess the alertness during waking hours. To ensure that the latent dimensions linking sleep phenotype to RSFC uncovered by PLS were robust, multiple control analyses were performed. Furthermore, as multivariate approaches in high-dimensional data such as PLS are prone to overfitting, a 10-fold cross-validation was performed here to assess the generalization performance of the latent dimensions to unseen test data.

As described below, we uncovered one latent dimension of the RSFC that was highly correlated with one latent dimension of sleep health in a discovery dataset ($n = 687$). This dimension was characterized by a specific spectrum of sleep health-related measures and by specific RSFC features. This latter appeared as a neurobiologically relevant connectome since the RSFC loadings were spatially correlated with the distribution of several neurotransmitter systems relevant to sleep health, including serotonin receptors, as well as metabotropic glutamate receptor 5 (mGluR5) and the γ -Aminobutyric acid type A (GABA_A) receptor^{30,31}. Furthermore, the associative pattern was successfully replicated, and the model showed generalizability in an internal validation dataset ($n = 628$). Importantly, the identified connectome pattern shows predictive utility for sleep quality in an independent sample consisting of unrelated individuals from the Human Connectome Project (HCP) dataset ($n = 435$). Finally, this connectome pattern critically shows diagnostic potential to distinguish insomnia patients ($n = 52$) from sleep-healthy subjects ($n = 49$) with an accuracy of 79.12%. These results hence delineated a sleep health dimension whose associated functional connectome could serve as an objective neuroimaging biomarker in clinical translations, such as to assess sleep-based interventions for improving brain health.

Results

PLS model reveals one robust dimension linking sleep health and resting-state function connectivity

We sought to delineate multivariate relationships between resting-state functional connectivity and sleep health in a large discovery dataset ($n = 687$). To this end, we applied PLS, an unsupervised machine-learning technique that seeks to find covariance between two high-dimensional matrices, namely whole-brain RSFC and 36 behavioral measures spanning multiple domains of sleep health. Following preprocessing using a validated pipeline that minimizes the impact of in-scanner motion (see “Methods” section), we constructed subject-level RSFC using a 246-node parcellation system⁴² that includes 210 cortical regions and 36 subcortical regions. Prior to analysis with PLS, we regressed age, sex, handedness, and head motion out of both the RSFC and behavior data to ensure that these potential confounders did

not drive results. The input data thus consisted of 30,135 unique functional connections and 36 sleep health-related variables. The selection of the sleep health measures was based on the consideration of selecting available variables in the BBP that (1) represent central domains of SH described in the Ru SATED³ as well as in the National Sleep Foundation (NSF)’s Sleep Health Index⁶; (2) provide high consistency with previous SH studies^{5,7,11}; (3) draw a multifaceted picture of SH to a bigger extend based on its broader definition. On that basis, the chosen 36 variables were grouped into seven domains including (1) Satisfaction with sleep/ Sleep quality; (2) Alertness during waking hours; (3) Timing of sleep; (4) Sleep efficiency/ continuity; (5) Sleep duration; (6) Sleep deficiency; (7) Sleep beliefs, attitudes, and habits (see “Methods” and Table S1 for the details).

Figure 2a shows the amount of covariance explained by each latent variable (LV). Notably, only one LV (LV1) survived after permutation testing with FDR correction ($q < 0.05$) (Fig. 2b). Importantly, several control analyses were performed to ensure the robustness of the obtained LV1. See Table S2 for details. First, 10-fold cross-validation was successful; PLS components estimated from 90% of the participants successfully generalized to the remaining 10% of participants, as indicated by the significant correlation between RSFC and behavioral composite scores in the test folds (LV1, mean $r = 0.17$, permuted $p < 3.0 \times 10^{-3}$). It should be noted that to avoid data leakage issues, the adjustment for confounds and data standardization were performed within the cross-validation loop (i.e., at first, we estimated parameters of data standardization and confounds regression in the nine training folds and then applied the obtained parameters to the test fold). Second, PLS components were robust to global signal regression, total intracranial volume regression, time (hour) of acquisition regression, the pre-scanning positive and negative affect regression, body mass index (BMI) regression, and family income regression, as indicated by the high correlation ($r > 0.93$, Table S2) between saliences of original PLS and PLS with corresponding variable regression. Third, instead of regressing age, sex, handedness, and head motion from the data, these variables were added to the phenotypic data for the PLS analysis. The results were largely unchanged as indicated by the high correlation ($r > 0.92$, Table S2) between the saliences of original PLS and PLS with adding age, sex, handedness, and motion into the phenotypic data. Fourth, PLS components were not driven by the non-Gaussian distributions of the behavioral data and skewed behavioral distributions, as indicated by the high correlation ($r > 0.98$, Table S2) between saliences of original PLS and PLS with quantile normalization to improve the Gaussian distributions of the behavioral data. Fifth, the results remained largely unchanged when using a different Seitzman et al.’ Atlas⁴³ containing 300 regions for the RSFC construction, as indicated by the high correlation ($r = 0.99$ for behavior data; $r = 0.78$ for RSFC data) between loading scores of Brainconnectome Atlas and Seitzman et al.’ Atlas, see details in Supplementary results, Figs. S1 and S2. Sixth, to demonstrate the sleep health component was independent of both the circadian timing of acquisitions and the state of alertness of individuals during the resting-state fMRI acquisition, we extracted the dimension score of “Sleepiness” from the Amsterdam Resting-State questionnaire (ARSQ)2.0⁴⁴ and time (hour) of acquisition, then correlated them with both the RSFC and behavior composite scores for the discovery dataset. The results revealed no significant correlation between participants’ sleepiness and either their RSFC composite score ($r = -0.0349$, $p = 0.3605$) or their behavior composite score ($r = -0.0223$, $p = 0.5603$). Additionally, no significant relationship between the acquisition time and either the RSFC composite score ($r = 0.0395$, $p = 0.3013$) or the behavior composite score ($r = -0.0393$, $p = 0.3941$) was observed. Seventh, the first latent component was robust across analytical approaches, which was evidenced by the high correlation between the first principal component of the sleep health behavioral measures (obtained by principal component analysis) and the behavioral saliences of LV1 ($r = 0.91$, $p = 1.22 \times 10^{-14}$).

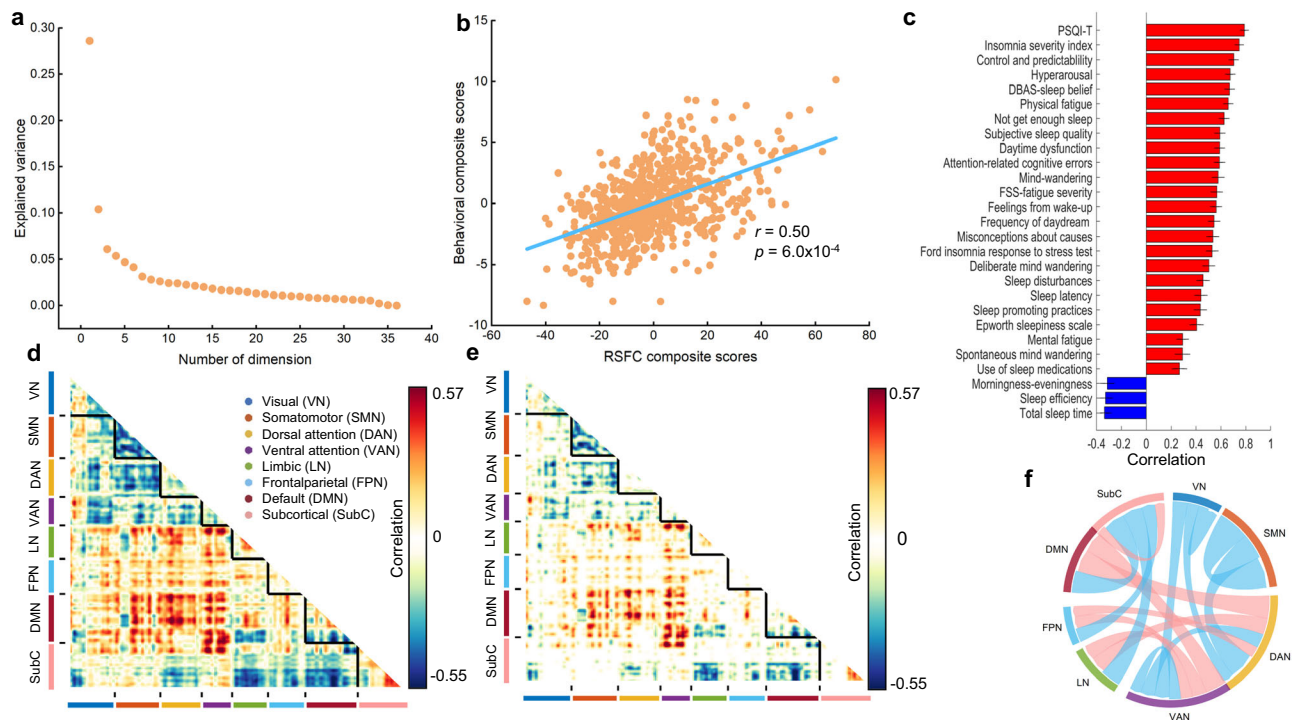


Fig. 2 | PLS reveals one robust dimension linking sleep health and resting-state function connectivity in the discovery dataset. **a** The amount of covariance explained by each latent variable (LV). Each orange dot represents a LV, only the first LV (LVI) survived after permutation testing with FDR correction ($q < 0.05$). This survived dimension (LVI) accounted for 28.6% of resting-state functional connectivity (RSFC)-behavior covariance. **b** scatter plots to illustrate the significant association between individual-specific RSFC and behavioral composite scores of participants in LVI ($r = 0.50$, two-sided, permuted $p = 6.0 \times 10^{-4}$) using Pearson correlational analysis. **c** significant 27 strongest correlations between participants' behavioral measures and their behavioral composite scores on the group level. Greater loading on LVI was associated with poorer sleep health. Error bars indicate bootstrapped standard deviation with 1000 bootstrap estimations ($n = 1000$). Behavioral measures for which higher values indicate better sleep health are colored blue. For example, sleep efficiency is colored blue because higher values indicate better sleep health. **d** unthresholded correlations between participants' RSFC data and their RSFC composite scores. Red (or blue) color indicates that

greater RSFC is positively (or negatively) associated with LVI. **e** thresholded correlations between participants' RSFC data and their RSFC composite scores (false discovery rate $q < 0.05$). The significant edges were widely distributed throughout the brain, contained a small portion of the total edges in the connectome (5956 edges total out of 30135 or 19.76%). A total of 2666 common edges (8.85% of the 30135 total edges) positively correlated with RSFC composite score and a total of 3290 common edges (10.92% of the 30135 total edges) negatively correlated with RSFC composite score. **f** correlations between participants' RSFC data and their RSFC composite scores, averaged within and between networks defined by Yeo et al's seven networks with significant bootstrapped Z-scores. The pink line represents a positive correlation while the blue line represents a negative correlation. DMN, Default mode network; PSQI-T, total score of Pittsburgh Sleep Quality Index. DBAS, Dysfunctional beliefs, and attitudes about sleep scale. FPN Fronto-parietal network, VN Visual network, SMN Somatomotor network, DAN dorsal attention network, VAN ventral attention network, LN limbic network. Source data are provided as a Source Data file.

This survived dimension (LVI) accounted for 28.6% of RSFC-behavior covariance (Fig. 2a), with significant association ($r = 0.50$, permuted $p = 6.0 \times 10^{-4}$) between RSFC and behavioral composite scores (Fig. 2b). Figure 2c shows the top correlations between LVI's behavioral composite/loading score and the 27 significant sleep health measures on the group level. A greater behavioral composite score was associated with poorer sleep health (e.g., poor sleep quality, insomnia, hyperarousal, dysfunctional beliefs and attitudes about sleep, physical fatigue, not getting enough sleep). Correlations between LVI's RSFC composite/loading scores and the RSFC data are shown in Fig. 2d (unthresholded correlations) and Fig. 2e (significant correlations). The significant edges were widely distributed throughout the brain, contained a small portion of the total edges in the connectome (5956 edges total out of 30135 or 19.76%). A total of 2666 common edges (8.85% of the 30135 total edges) positively correlated with RSFC composite score and a total of 3290 common edges (10.92% of the 30135 total edges) negatively correlated with RSFC composite score.

Figure 2f shows the significant RSFC correlations averaged within and between networks defined by Yeo et al's seven network¹⁹. Greater RSFC composite score was associated with increased RSFC within the subcortical network (SubC) mainly the thalamus (Fig. 3a and Fig. S3h), and increased RSFC between DMN and the dorsal attention network

(DAN), between DMN and the ventral attention network (VAN), between the fronto-parietal network (FPN) and DAN, between FPN and VAN (Fig. 2d, e, f and Fig. S3c, S3d, S3f, S3g). Greater RSFC composite score was associated with decreased RSFC within the Somatomotor network (SMN) and the VAN (Figs. 2d, e, f and 3a; Fig. S3b, S3d), and between the DAN and VAN, between the DAN and SMN, between the SubC and DMN, between the SubC and FPN (Figs. 2d, 2e, 2f, and 3a; Fig. S3c, S3c, S3b, S3h).

Figure 3b further demonstrates the relative importance of the regions in the obtained significant RSFC pattern by showing the top five nodes with highest weighted degree in both the positive and negative networks (Table S3). Specifically, for the positive network, the top five nodes with the greatest number of edges were primarily located in the VAN, i.e., the dorsal insular cortex and precentral gyrus (Fig. 3b). For the negative network, the top five nodes with the greatest number of edges were primarily located in the Subcortical regions, i.e., the thalamus, and somatomotor cortex (e.g., postcentral gyrus and paracentral lobule) (Fig. 3b).

Cross-dataset replicability and generalizability

The robustness of the obtained LVI was further ensured with a large replication sample ($N = 628$) from the BBP (see "Methods" for the

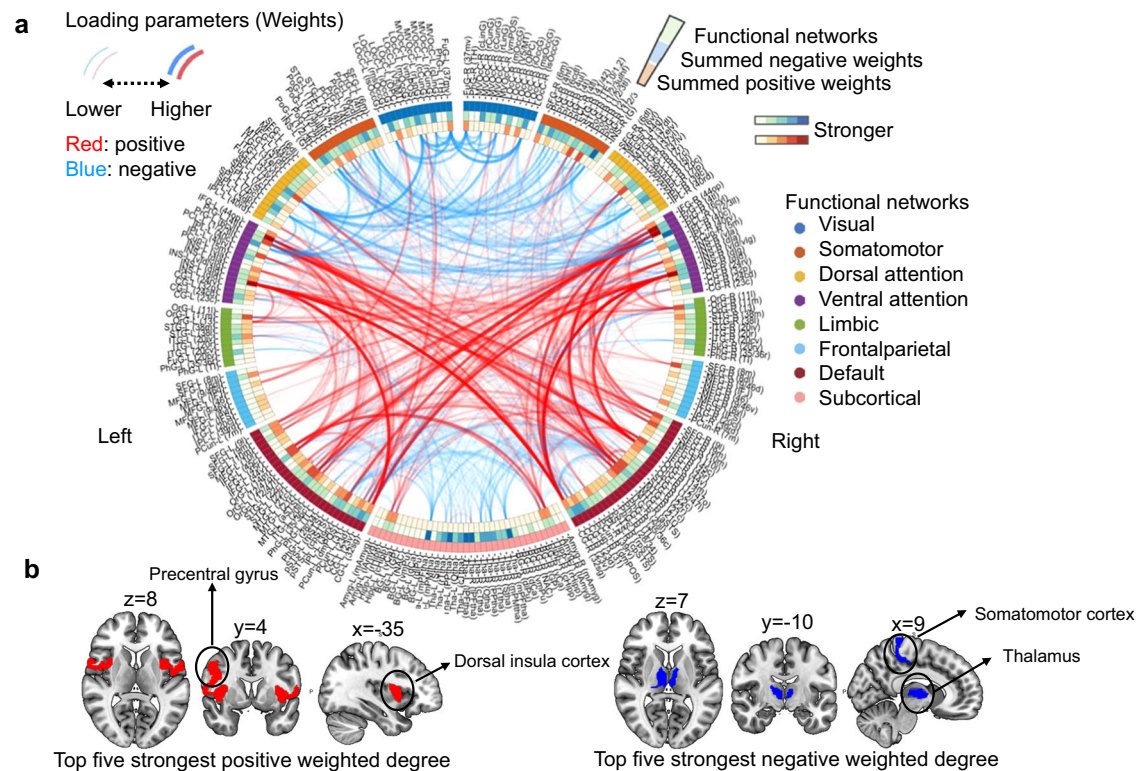


Fig. 3 | Correlations between LVI's RSFC composite/loading scores and the RSFC data. **a** circular plot representation of the correlations between LVI's RSFC composite/loading scores and the RSFC data. From outermost to innermost, the first layer of the circle represents different resting-state networks, and the second and third layers each represent the sum of positive and negative predictive weights

coming from each brain region. **b** The top five nodes with the highest degree in both the positive (left side) and negative networks (right side). The abbreviations of the brain regions within each resting-state network can be found on the website: <https://atlas.brainnetome.org/download.html>. LV, latent variable; RSFC, resting-state functional connectivity. Source data are provided as a Source Data file.

details). To do this, we replicated the PLS procedure conducted in the discovery dataset with the replication dataset (see “Methods” for the details). The LVI accounted for 26.3% variance and still survived after permutation testing with FDR correction ($q < 0.05$) in the replication dataset with significant association ($r = 0.44$, permuted $p = 0.006$) between RSFC and behavioral composite scores (Fig. S4). Importantly, the obtained LVI was largely replicated, evidenced by the high correlation between the behavioral salience scores in the discovery and replication dataset ($r = 0.95$, $p = 0$, Fig. 4a), between the behavioral loading scores in the discovery and replication dataset ($r = 0.99$, $p = 0$, Fig. 4c), between the RSFC salience scores in the discovery and replication dataset ($r = 0.71$, $p = 0$, Fig. 4b), between the RSFC loading scores in the discovery and replication dataset ($r = 0.95$, $p = 0$, Fig. 4d).

We further tested the cross-dataset generalizability by projecting dataset 2 onto the salience parameters learned by PLS in dataset 1. Then, we examined the correlation between the behavioral and RSFC composite scores in dataset 2. We found the obtained LVI from dataset 1 successfully generalized to dataset 2, as evidenced by a significant correlation between the obtained behavioral and RSFC composite score ($r = 0.25$, permuted $p = 2.66 \times 10^{-6}$, Fig. 4e).

Spatial correlation with neurotransmitter densities

Furthermore, we answered question3 (Fig. 1) by testing whether the sleep-health-related connectome (Fig. 2e) was spatially correlated with the distribution of several neurotransmitter systems involved in the domains of sleep health, including serotonin receptors (5-HT1a, 5-HT1b, and 5-HT2a) and transporters (5-HTT), together with metabolic glutamate receptor 5 (mGluR5) and the γ -Aminobutyric acid type A (GABA) receptor. First, density values were derived from average group maps of healthy volunteers obtained in prior

multitracer molecular imaging studies. These maps were resampled to an isotropic 2-mm spatial resolution as we did in the fMRI data. Then, we obtained the average value for each region of the Brainconnectome atlas for each Neurotransmitter density map. Next, we summed the positive and negative FC loadings (weighted) separately for each region of the Brainconnectome atlas to represent the region importance scores in positive (Fig. 5a) and negative network (Fig. 5b). Finally, spearman correlation analysis between the region importance score and receptor/transporter densities were conducted⁴⁵. Please refer to the “Methods” and Supplementary Materials for the details. After permutation testing with FDR correction ($q < 0.05$), significant associations were found between on the one hand, the summed positive network and the 5HT1a map (Fig. 6a), the 5HT2a map (Fig. 6b), the GABA map (Fig. 6c) and the mGluR5 (Fig. 6d) and on the other hand, the summed negative network and the 5HT1a (Fig. 6e) map, and the mGluR5 (Fig. 6f) map. These results suggested the PLS pattern estimated from RSFC may be relevant at the neuromolecular level in line with sleep health.

Predictive utility of the sleep-health-related connectome: for sleep quality in a healthy population

Next, considering that the most contributing measure in the LV/dimension found in the discovery ($r = 0.79$, Fig. 2c) and replication ($r = 0.78$, Fig. S4c) BBP samples were the PSQI total score, we further examine the predictive utility of the sleep-health-related connectome features for individual PSQI scores in a completely independent dataset. We here used support vector regression (SVR) to predict the sleep quality measured by the PSQI total score (see Fig. 7a) in the unrelated individuals of the HCP dataset ($n = 435$), see “Methods” for the details. It turned out that the significant edges of the sleep-health-related connectome found in the discovery sample can predict the

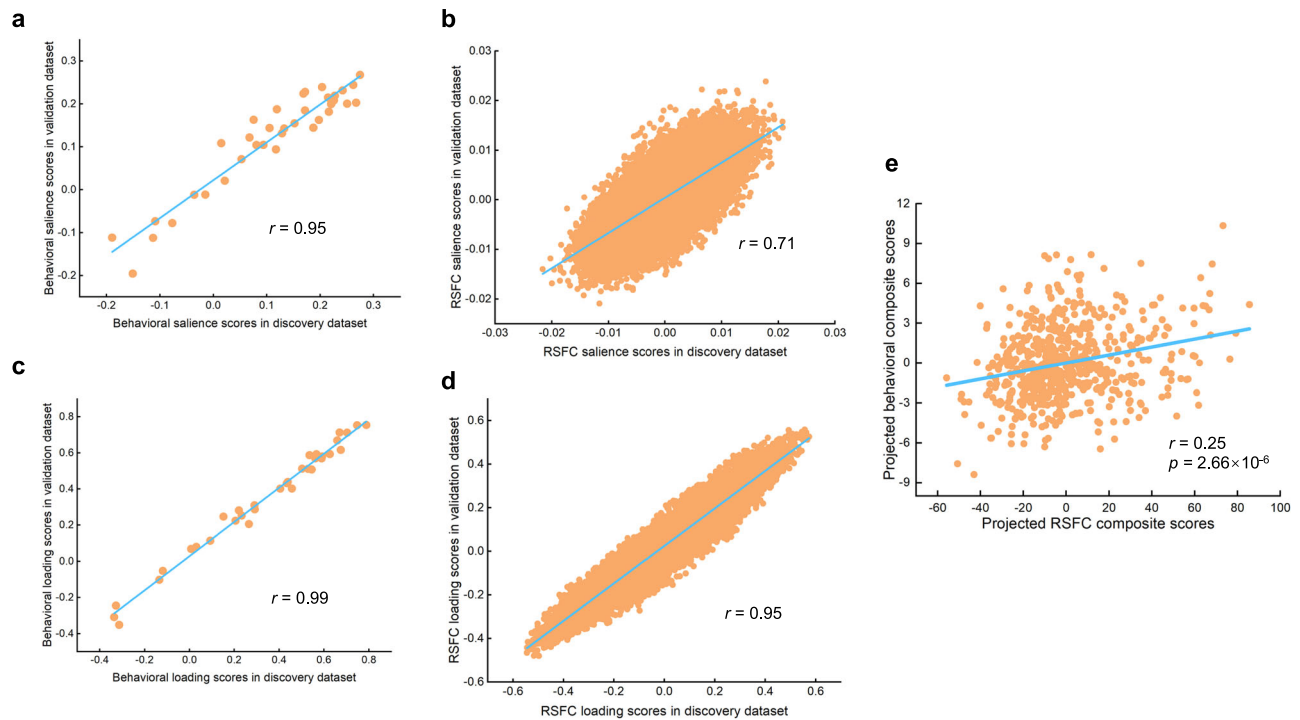


Fig. 4 | The robustness of the obtained LV1 was further ensured with a large replication dataset. **a** High correlation between the behavioral salience scores in the discovery and replication dataset ($r = 0.95$, $p = 0$). **b** High correlation between the RSFC salience scores in the discovery and replication dataset ($r = 0.71$, $p = 0$). **c** High correlation between the behavioral loading scores in the discovery and replication dataset ($r = 0.99$, $p = 0$). **d** High correlation between the RSFC loading scores in the discovery and replication dataset ($r = 0.95$, $p = 0$). **e** The Pearson correlation between the obtained behavioral and RSFC composite score ($r = 0.25$, two-sided, permuted $p = 2.66 \times 10^{-6}$) by projecting dataset 2 onto the salience parameters learned by PLS in dataset 1. LV latent variable, RSFC resting-state functional connectivity, PLS partial least squares. Source data are provided as a Source Data file.

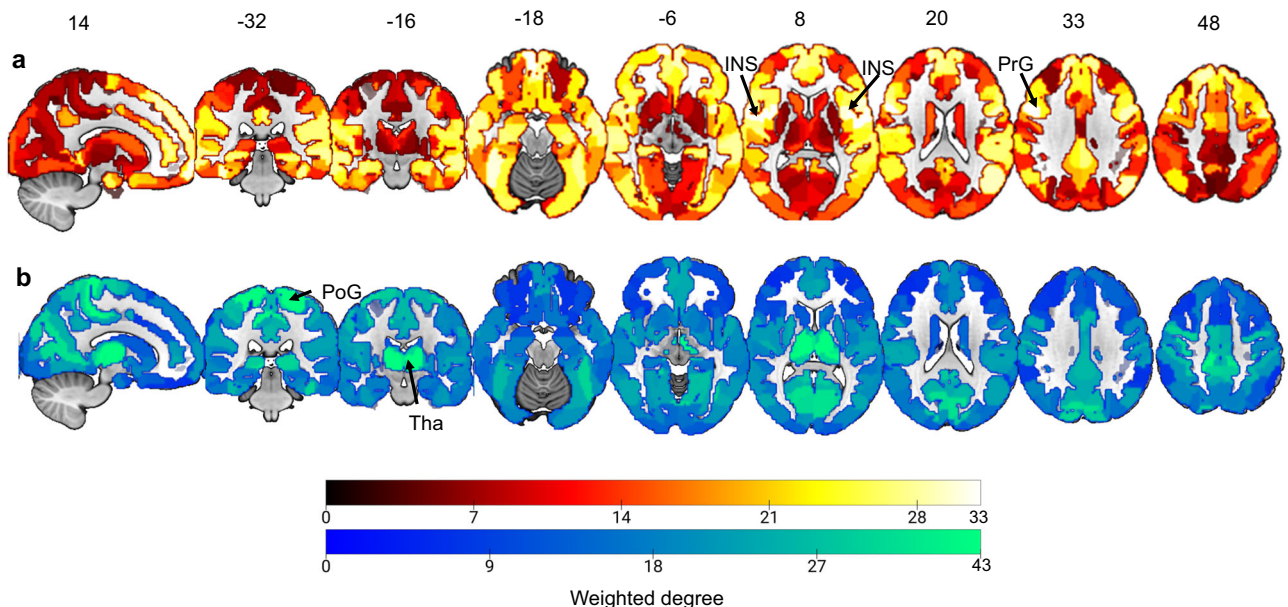


Fig. 5 | Maps of region importance scores that are covaried with sleep health dimension. The positive and negative RSFC loadings were summed separately for each region of the Brainconnectome atlas to represent the region importance scores in positive (a) and negative networks (b). RSFC, resting-state functional connectivity. INS, Insular Gyrus; PrG, Precentral Gyrus; PoG, Postcentral Gyrus; Tha, Thalamus. Source data are provided as a Source Data file.

total score of PSQI in the HCP dataset as evidenced by a significant Pearson correlation between the predicted and original PSQI (r average = 0.18, r range = 0.13–0.23 permuted $p < 0.0006$) after a 100 repeated 10-fold cross-validation (Fig. 7b, c). The Mean Absolute Error

(MAE) of the PSQI prediction using the connections identified in the PLS is 2.39. Notably, the predicted PSQI score did not show significant associations with demographic variables and head motion (Table S4), suggesting that the demographic variables and head motion did not

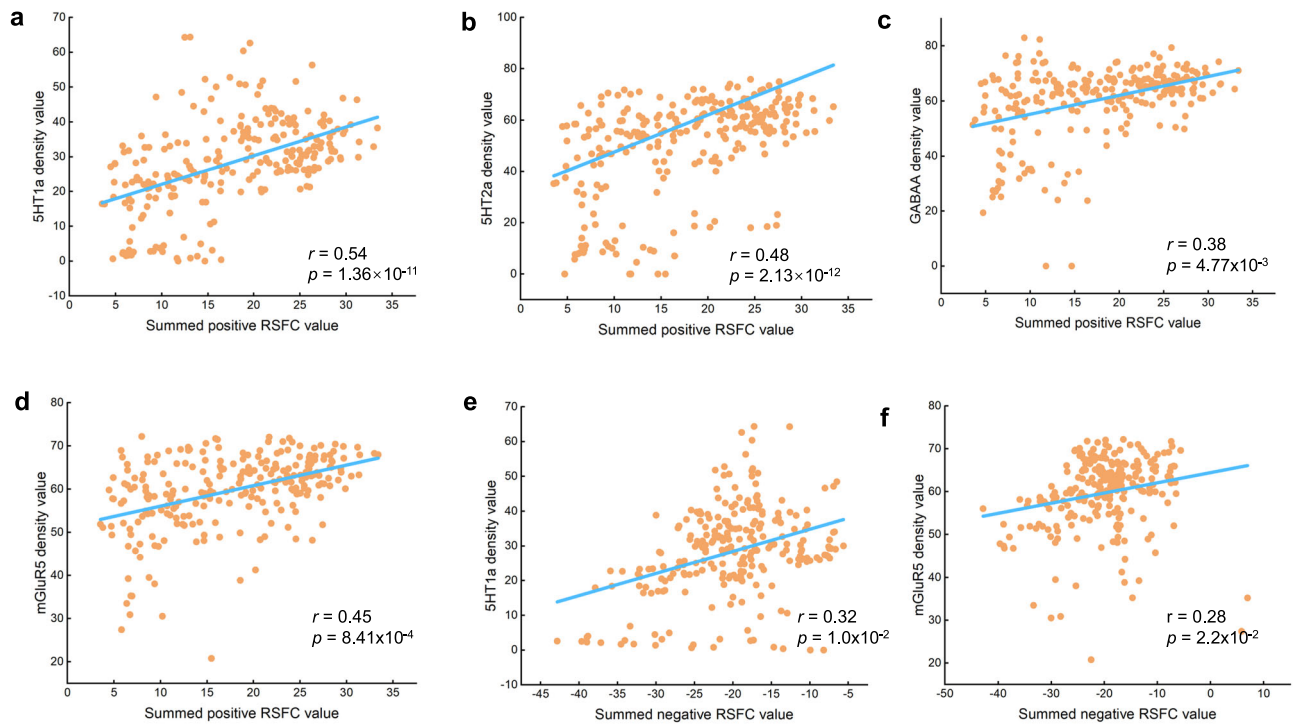


Fig. 6 | Spatial correlation with neurotransmitter densities. The RSFC loadings (weighted) were spatially correlated with the distribution of several neurotransmitter systems potentially involved in domains of sleep health with Pearson correlational analysis, specifically, the density value of 5HT1a (**a**, $r = 0.54$, two-sided, $p = 1.36 \times 10^{-11}$), 5HT2a (**b**, $r = 0.48$, two-sided, $p = 2.13 \times 10^{-12}$), GABA (**c**, $r = 0.38$, two-sided, $p = 4.77 \times 10^{-3}$), mGluR5 (**d**, $r = 0.45$, two-sided, $p = 8.41 \times 10^{-4}$) were found to be significantly correlated with the summed positive network; the density value of 5HT1a (**e**, $r = 0.32$, two-sided, $p = 1.0 \times 10^{-2}$), mGluR5 (**f**, $r = 0.28$, two-sided, $p = 2.2 \times 10^{-2}$) were found to be also significantly correlated with the summed negative network. RSFC, resting-state functional connectivity. Source data are provided as a Source Data file.

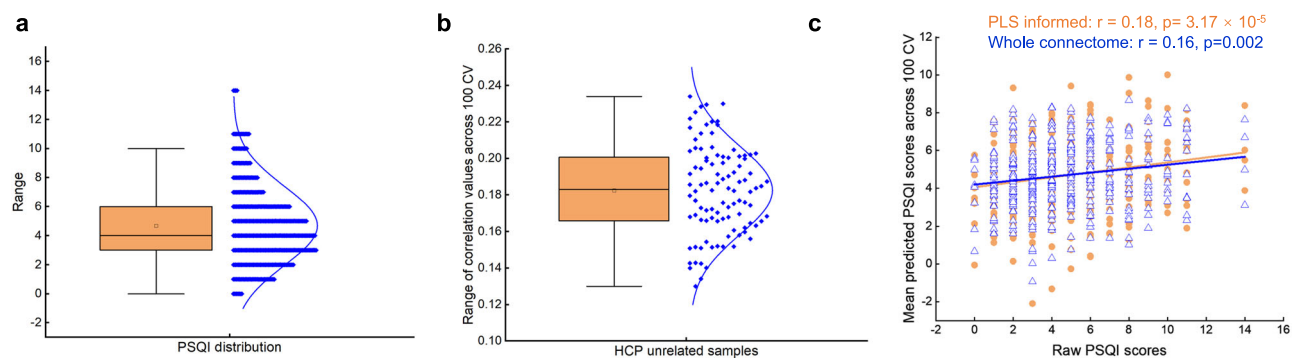


Fig. 7 | Predictive utility of the sleep-health connectome. **a** The distribution of the total score of the Pittsburgh Sleep Quality Index (PSQI) in the Human Connectome Project (HCP) dataset ($n = 435$). For the box plot, minima = 0, maxima = 14, center = 4.68, median = 4, 25th percentile = 3, 75th percentile = 6. **b** The distribution of the correlation value between the predicted (by the RSFC spatial pattern of LV1) and original PSQI after a 100 repeated 10-fold cross-validation ($n = 100$). For the box plot, minima = 0.13, maxima = 0.23, center = 0.18, median = 0.18, 25th percentile = 0.17, 75th percentile = 0.20. **c** The prediction of PSQI total score in the HCP dataset using the significant edges of the RSFC spatial pattern obtained by PLS analysis in dataset 1 as shown in the yellow correlational graph with a Pearson correlation between the original PSQI and the averaged predicted PSQI after a 100 repeated 10-fold cross-validation ($r = 0.18$, two-sided, permuted $p = 3.17 \times 10^{-5}$); the prediction of PSQI total score in the HCP dataset using the whole-brain connectome as depicted in the blue correlational graph with a Pearson correlation between the original PSQI and the averaged predicted PSQI after a 100 repeated 10-fold cross-validation ($r = 0.16$, two-sided, permuted $p = 0.002$). PLS partial least squares, RSFC resting-state functional connectivity, LV latent variable. Source data are provided as a Source Data file.

significantly affect the prediction. Importantly, we also examined the prediction performance of models using the whole-brain connectome. Our results here showed that models using the specific sleep-health-related connectome yield similar prediction performance with whole-brain connectome models (after a 100 repeated 10-fold cross-validation, r average = 0.16, r range = 0.10–0.22, permuted $p = 0.002$, MAE = 2.33 (the blue correlational graph in Fig. 7c)) hence demonstrating that the identified sleep-health-related connectome captures

sufficient and relevant information for individual prediction of sleep phenotype in the adult population.

Predictive utility of the sleep-health-related connectome: for sleep disorders

Furthermore, we investigated the possibility of distinguishing insomnia patients from sleep-healthy controls based on the sleep-health-related connectome. To accomplish this, we utilized a Gaussian radial

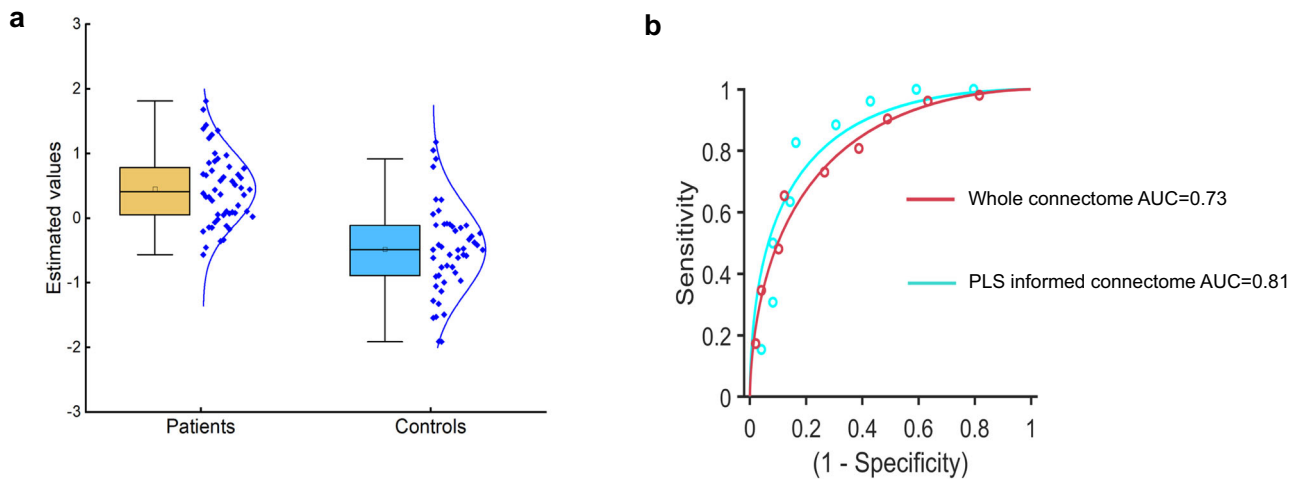


Fig. 8 | the sleep-health-related connectome also has diagnostic potential for insomnia disorder. **a** The mean estimated label values to index the classification performance for insomnia patients ($n = 52$) and health controls ($n = 49$) group across a 100 repeated 10-fold cross-validation strategy. For the box plot of the patients group: minima = -0.56 , maxima = 1.82 , center = 0.46 , median = 0.41 , 25th percentile = 0.06 ; for the box plot of the health controls: minima = -1.91 , maxima = 1.18 , center = -0.48 , median = -0.49 , 25th percentile = -0.89 , 75th percentile = -0.11 . **b** the green receiver operating characteristic (ROC) curve

depicts the classification performance using the significant edges of the RSFC spatial pattern obtained by PLS analysis in dataset4, suggesting the sleep-health-related connectome also has diagnostic potential to distinguish insomnia patients from sleep-healthy subjects with an average accuracy across 100 CV of 79.73%, permuted $p = 2.39 \times 10^{-3}$; the red ROC curve depicts the classification performance using whole-brain connectome with an average accuracy of 72.62%, permuted $p = 7.53 \times 10^{-3}$. LV latent variable, RSFC resting-state functional connectivity. Source data are provided as a Source Data file.

basis function (RBF) kernel support vector machine (SVM) classifier which was implemented using the LIBSVM toolbox⁴⁶ on our classification dataset (dataset4) containing 52 insomnia patients and 49 sleep-healthy subjects (see “Methods” and Table S6 for the sample details). The classifier performance was validated using a 100-repeated 10-fold cross-validation strategy (see “Methods” for details). Classification performance that had an accuracy higher than 70% was considered to be meaningful. It turned out that the sleep-health-related connectome also has diagnostic potential to distinguish insomnia patients from sleep-healthy subjects in dataset4 with an average accuracy across 100 CV of 79.73% (permuted $p = 2.39 \times 10^{-3}$) with 78.85% sensitivity, 83.67% specificity and area under the curve (AUC) = 0.81 (Fig. 8a, green receiver operating characteristic (ROC) curve in Fig. 8b). Importantly, we tested to which extent the sleep-health-related connectome does as well or better than using the whole-brain connectome for the classification performance. It turned out the whole-brain connectome-based classification performance was lower as indicated by an average accuracy across 100 CV of 72.62% (permuted $p = 7.53 \times 10^{-3}$) with 78.85% sensitivity, 67.35% specificity and AUC = 0.73 (red ROC curve in Fig. 8b).

Moreover, the classification performance based on the sleep-health-related connectome can also be generalized to an external classification dataset containing 35 insomnia patients and sleep-healthy subjects with an average accuracy across 100 CV of 78.29% (permuted $p = 3.89 \times 10^{-3}$) with 80.00% sensitivity, 77.14% specificity and area under the curve (AUC) = 0.79 (Supplementary Fig. S5a, green receiver operating characteristic (ROC) curve in Fig. S5b). Sample details about this external classification dataset can be found in Supplementary Methods. We also tested to which extent the sleep-health-related connectome does as well or better than using the whole-brain connectome for the classification performance in the external classification dataset. It turned out the whole-brain connectome-based classification performance was lower as indicated by an average accuracy across 100 CV of 74.27% (permuted $p = 6.27 \times 10^{-3}$) with 74.29% sensitivity, 77.14% specificity and AUC = 0.75 (red ROC curve in Fig. S5b). These results together suggest that the sleep-health-related connectome feature sets have specific classification power for sleep

health phenotype at the individual level and could therefore be seen as useful biomarkers.

Discussion

There is growing interest toward using a conceptual framework that articulates sleep health as multi-facet^{21,47}. Although the importance of RSFC in multi-facets of sleep health is increasingly recognized, the degree to which RSFC is associated with sleep health as a multivariate dimension remains largely unknown. Leveraging a unique dataset including resting-state fMRI and behavioral assessments spanning multiple domains of sleep health, we found robust correlated patterns of RSFC and sleep health measures that could be represented in one dimension. The derived pattern estimated from RSFC shows neurobiological relevance and predictive utility for sleep phenotype in healthy populations. Critically, the RSFC spatial pattern of this dimension additionally shows diagnostic potential to distinguish insomnia patients from sleep-healthy subjects. Together these results delineate the RSFC-guided dimension of sleep health, which could serve as a foundation to develop a reliable brain connectome biomarker for sleep health with high potential for clinical translation and ultimately for the diagnosis, prognosis, treatment, and prevention of sleep health-related problems.

In the present study, the measurement of sleep health composite was guided by both behavior and brain functional connectivity. The obtained dimension of sleep health was associated with a unique cluster of 27 out of the 36 selected SH measures mainly including sleep quality, the severity of insomnia, degree of hyperarousal, degree of dysfunctional beliefs and attitudes about sleep, degree of physical fatigue, and degree of subjective sleep lack/deprivation, etc. (Fig. 2c). This significant cluster of SH variables were still fallen within the proposed seven domains and largely overlapped with the domains of sleep health examined in previous studies^{17,36} but draw a multifaceted picture of SH to a bigger extent based on its broader definition². The present study thus can be regarded as a step forward in measuring and quantifying sleep health composite by incorporating whole-brain resting-state functional connectivity². To this end, the present study validated the theory that sleep health can be operated and constructed as a

composite dimension and that sleep health domains do not exist in isolation^{9,12}.

Notably, four out of six measures in the domain of beliefs, attitudes, and habits about sleep⁴⁸ were found to be important measurements in the significant latent dimension, with the measure of diminished control and predictability contributing the most. Existing evidence suggests that individuals with more positive beliefs and attitudes about sleep are more likely to experience better sleep health^{37,49}. On the other hand, previous studies have proposed that the dysfunctional beliefs about sleep (e.g., diminished control and predictability of sleep) can result in the misperception of sleep duration (e.g., underestimate of sleep duration)⁵⁰, especially in people who are suffering from insomnia^{23,51}. One possible explanation is that dysfunctional beliefs may alter the constructive process involved in remembering the amount of sleep obtained, thereby contributing to misperception⁵¹. Also recent theories claimed that people tend to gather evidence that confirms our beliefs rather than evidence that challenges them⁵², so it is crucial to disprove the dysfunctional beliefs to improve sleep health. In correspondence, Cognitive Behavioral Therapy for Insomnia (CBT-I) emphasized the importance of cognitive restructuring and sleep hygiene to alter the dysfunctional beliefs, attitudes, and habits about sleep as its two main components⁵³. Taken together, future studies investigating sleep health should pay more attention to the domain of beliefs, attitudes, and habits about sleep.

The latent dimension of sleep health was associated with a unique spatial pattern of resting-state functional connectivity, namely a positive and negative network pattern. For the positive network pattern, poorer sleep health was associated with hyper-connectivity between the network involved in internally oriented attention (DMN) and the network responsible for modulation of the goal-directed attention (DAN), between DMN and the network involved in processing of salience (VAN), between the network involved in external goal-directed regulation (FPN) and DAN, between FPN and VAN as well as hyper-connectivity within the subcortical network (SubC) mainly the thalamus. The heightened coupling between the higher-order and attentional networks may suggest a decreasing dedifferentiation of the high-order system to the middle attentional system^{19,54} while the increased FC mainly within the thalamus (a brain system responsible for gating information) further indicates sustained ascending arousal input from the thalamus¹⁶ in poor sleep health individual phenotype. Notably, the positive network consisted of the highest-degree nodes in the left dorsal insular cortex and precentral gyrus. As key hubs of the salience network, these brain regions play a crucial role in cognitive and emotional processes^{23,55}. Impaired connectivity patterns between these areas and other regions could underlie cognitive, vigilance, and perception dysfunctions, as well as subjective distress and sleep complaints⁵⁶.

For the negative network pattern, poorer sleep health was also associated with hypo-connectivity between the two attentional networks, between DMN and SubC mainly the thalamus, between FPN and SubC mainly the thalamus, between DAN and the network involved in sensory and motor perception (SMN), as well as hypo-connectivity within the VAN and SMN. The reduced coupling between and within the attentional networks and the SMN may mainly suggest the collapse of top-down orienting of attention, bottom-up salience processing^{77,58} and sensory and motor perception²² while the reduced connectivity between the high-order networks and thalamus further reflects the reduction in the upward transmission of information to the higher-order networks by the thalamus as a relay station⁵⁹ in the condition of poor sleep health. Moreover, the negative network consisted of the highest-degree nodes in the thalamus and somatomotor cortex. Both the thalamus⁵⁹ and the somatomotor cortex²² play a crucial role in sleep physiology. Impaired connectivity patterns between these areas and other regions could suggest the dysfunction of relaying sensory and motor signals to the cerebral cortex and regulating sleep, alertness, and consciousness.

Both the positive and negative network patterns were spatially correlated with the distribution of several neurotransmitter systems involved in sleep health, including the serotonin receptors, the glutamate receptor (mGluR5), and the GABAA receptor. Previous studies have proven that: (1) serotonin functions predominantly to promote wakefulness and to inhibit REM (rapid eye movement) sleep (REMS)³¹; (2) γ -aminobutyric acid (GABA) is increasingly recognized as an important inhibitory neurotransmitter for the initiation and maintenance of sleep^{30,60}; (3) glutamate is the primary excitatory neurotransmitter in the central nervous system, its altered levels were found in people who suffered from insomnia disorders^{23,32,61}. Taken together, the linked distribution of these neurotransmitter systems and the RSFC spatial pattern found in the present study further suggest that the derived pattern estimated from RSFC is neurobiologically relevant with regard to neuromolecular systems engaged in sleep health.

Critically, the RSFC spatial pattern of this dimension also has diagnostic potential to distinguish insomnia patients from sleep-healthy subjects with an accuracy of 79.73% and 78.29% in dataset4 and the external classification dataset respectively. While a previous study utilized a “resting-state” fMRI support vector machine (SVM) classifier that achieved higher classification accuracy between IDs and HCs, it included 5 min of wakefulness, as well as data from sleep stages 1 (S1), 2 (S2), and 3 (S3) for each subject⁶². However, this approach could not exclude the possibility of shared individualized characteristics across multiple sessions from the same subject, which could inflate the classification performance. Importantly, our classification results demonstrated that the obtained brain connectome pattern also contained pathophysiological relevance for insomnia disorder. Insomnia is a common, distressing, and clinically complex symptom, that causes difficulty initiating sleep; frequent awakening; or early-morning awakening with daytime dysfunction^{23,32}. Despite its high prevalence, insomnia often goes under-diagnosed and untreated, resulting in general fatigue and decreased productivity⁶³. Based on the RSFC spatial pattern-based SVM classifier derived from the obtained sleep health dimension, we will be able to aid in the diagnosis of insomnia disorder. To this end, these findings may suggest that insomnia is a condition with a multifaceted pathophysiology spanning over deficits in multiple domains of sleep health. Ultimately, the present study can serve as a foundation to identify targets that may enable sleep-based interventions for improving both brain health and clinical outcomes.

The strength of the present study includes the utilization of a large sample, advanced multivariate methods, and replication of results in an independent sample. The component we identified with RSFC not only showed macromolecular relevance but also had diagnostic potential, generalized well to unrelated individuals in the HCP dataset, and was robust across alternative methodological strategies. Nonetheless, our work has several limitations. First, here we sought to delineate linked dimensions of sleep health and intrinsic brain functional connectivity, uncovering dimensions of sleep health that are guided by and linked to underlying brain connectome properties. Although we included the largest sleep health domains in the neuroimaging studies of sleep health, to the best of our knowledge, this approach necessarily is limited by the available variables measuring sleep health in the BBP, which did not represent the full domains of sleep health. Future studies can incorporate more sleep-related variables (i.e., sleep regularity) to depict the sleep health dimension. Second, our current analysis only considered functional connectivity and behavioral measures of sleep health. Future research could incorporate rich multi-modal imaging data, objective sleep measurements with actigraphy and polysomnography, biochemics, and genomics. Third, the present study only used subjective measures to control for the state of alertness of individuals during the resting-state fMRI acquisition. Future research could use more objective measurements such as EEG, camera, or eye-tracking to monitor alertness/drowsiness states during resting-state fMRI scanning. Fourth, the sample mainly

consisted of young adults in pattern recognition and middle-aged adults in the prediction and classification test, leaving the RSFC-guided sleep health dimension in other age groups such as children, adolescents, and the elderly unknown. Recently, the definition of what constitutes good sleep health was adapted to pediatrics⁶⁴ as well as the elderly⁶⁵. Hence, it is of great importance for future studies to also examine the RSFC-informed sleep health dimension in both younger and older populations. Fifth, it should be noted that causality between sleep health and its associated RSFC component cannot be established in this cross-sectional study. Future studies can use a longitudinal design to test whether worse sleep health leads to alterations of this sleep-health-related connectome.

In summary, in this study, we discovered and replicated multivariate patterns of intrinsic functional connectivity that are highly correlated with a dimension of sleep health in a large sample of young adults. This dimension was composed of a unique cluster of sleep health and unique features of resting-state functional connectivity, namely a positive and negative network pattern. Both the positive and negative network patterns were spatially correlated with the distribution of several neurotransmitter systems involved in sleep health. Moreover, the obtained positive and negative network patterns also contained pathophysiological relevance for insomnia disorder. These results thus delineate the connectivity-guided dimension of sleep health, which could serve as a foundation for developing brain connectome-based biomarkers in sleep symptomatology.

Methods

Participants

Discovery and replication datasets (dataset 1 and 2). The main dataset comprises individuals from the Behavioral Brain Research Project of Chinese Personality (BBP)⁶⁶, which was launched in September 2019 (still in progress) to recruit participants from that year's freshmen at Southwest University, Chongqing, China. In total, 1369 participants completed the cross-sectional neuroimaging protocol as well as the behavioral measures of sleep health. To create two independent samples for discovery and replication analyses, we split the participants according to their data collection periods (September–December 2019 [discovery dataset], namely freshman enrolled in the year 2019; September–December 2020 [replication dataset] and freshman enrolled in the year 2020). Specifically, a discovery dataset ($n = 712$) and a replication dataset ($n = 657$) were created. Of the discovery dataset, 25 participants were excluded due to excessive head motion during scanning (e.g., with a mean framewise displacement [FD] larger than 0.3 mm), resulting in a final sample of 687 participants (mean age 18.96, $SD = 0.95$; 233 males and 454 females). Applying the same exclusion criteria to the replication dataset produced 628 participants (mean age 19.16, $SD = 1.03$; 208 males and 420 females). See Table 1 for detailed demographics of each dataset. The two datasets were confirmed to also have similar basic demographic variables, i.e., age, sex, and race (Table 1), as well as head motion (Table 1). Participants were compensated for 50 Chinese Yuan for their participation. Full informed consent from each participant was obtained by BBP Consortium, and research procedures and ethical guidelines were followed in compliance with Southwest University (SWU) institutional review board approval.

HCP external validation dataset (dataset 3). The dataset 3 was selected from the 1200 Subjects Release of the Human Connectome Project (HCP). This dataset provides high-quality behavioral/demographic and imaging data from healthy young adults (<https://www.humanconnectome.org/>)⁶⁷. From the original HCP dataset, we first included a group of participants ($n = 1084$) for which L-R resting-state fMRI data as well as PSQI scores were available. From this group, we further excluded data from 46 participants based on the exclusion criteria indicated as follows: (1) participants with missing values on

Table 1 | Demographic characteristics of participants in the discovery and replication dataset

	Discovery dataset (N = 687)	Replication dataset (N = 628)
Age, mean (SD), years	18.96(0.95)	19.16(1.03)
Female	454(66.08)	420(66.88)
Handedness-Right	606(88.21)	575(91.56)
BMI, mean (SD)	21.23(2.87)	21.85(3.19)
Race		
Han	569(82.82)	527(83.92)
Other	118(17.18)	101(16.08)
Framewise displacement (FD), mean (SD)	0.10(0.05)	0.10(0.05)
Total brain volume, mean (SD), mm ³	1541331.47(138185.45)	1539323.87(158514.00)
Family income per year (yuan, RMB)		
<5000	125(18.20%)	93(14.81%)
5001 – 25000	277(40.32%)	262(41.72%)
25001 – 45000	108(15.72%)	96(15.29%)
45001 – 65000	55(8.01%)	66(10.51%)
65001 – 85000	39(5.68%)	36(5.73%)
85001 – 105000	45(6.55%)	31(4.94%)
>105,000	38(5.53%)	44(7.01%)
PANAS-P, mean (SD)	28.96(6.12)	29.45(6.20)
PANAS-N, mean (SD)	17.66(5.97)	17.39(6.00)

PANAS Positive and negative affect schedule, SD standard deviation.

demographic variables such as age, sex, education, BMI, and race or family information; (2) participants with a history of hyper/hypothyroidism or history of other endocrine problems; (3) women who had recently given birth; and (4) participants having a mean FD more than 0.3 mm. Importantly, to exclude the influence of shared genetic and environmental factors, we randomly kept one subject from each family⁶⁸, resulting in 435 final unrelated subjects. See Table S5 for detailed demographics of each sample. Full informed consent from each participant was obtained by the Washington University–University of Minnesota (WU–Minn) HCP Consortium, and research procedures and ethical guidelines were followed in compliance with WU institutional review board approval.

Classification dataset (dataset 4). The classification dataset involved 109 participants, including 56 patients with insomnia disorder (ID) and 53 healthy controls (HC). Part of the dataset was collected from the Sleep Center, Department of Brain Disease of Chongqing Traditional Chinese Medicine Hospital (CTCMH), which included 25 ID and 36 HC participants (referred as the CTCMH dataset), and the remaining part of the dataset is from the Sleep and Neuroimaging Center, Southwest University (referred as SNIC dataset). All the participants received the MRI scanning at the brain imaging center of Southwest University. Patients with insomnia disorder were diagnosed by experienced hospital psychiatrists (D.G. and CY.L.) according to the International Classification of Sleep Disorders: Diagnostic and Coding Manual, 3rd ed., and insomnia symptoms have lasted at least three nights a week for more than 3 months. All the health controls met the following criteria: a good sleep habit and a good sleep onset and/or maintenance; a regular dietary habit; no consumption of any stimulants, medications, alcohol, and cocaine for at least 3 months before the study; no history of neurological or psychological disorders; lower scores of Pittsburgh Sleep Quality Index (PSQI) than 7⁶⁹. Participants with any findings of pathological brain MRI as well as ineligibility for MRI scanning (any type of metal implant) were excluded from the

study. Eight participants were excluded due to excessive head motion during the scanning (e.g., with a mean FD larger than 0.3 mm), resulting in a final sample of 52 ID patients (mean age = 44.00 years old, SD = 12.39) and 49 HC (mean age = 42.10 years old, SD = 15.93). Participants were compensated for 80 Chinese Yuan for their participation. The research projects were approved by the SWU and CTCMH institutional review boards, and written informed consent was obtained from each participant in accordance with the Declaration of Helsinki. The two groups were matched for age, sex, and head motion (Table S6). The participants' demographic characteristics are summarized in Table S6.

Sleep health-related assessments in the discovery and replication dataset

A large set of sleep health measures (Table 2 and Table S1) were included in the current study based on the consideration to select

available variables in the Behavioral Brain Research Project of Chinese Personality (BBP) that 1) represent central domains of SH described in the Ru SATED³ as well as in the National Sleep Foundation (NSF)'s Sleep Health Index⁶; 2) provide high consistency with previous SH studies^{15,70}; 3) draw a multifaceted picture of SH to a bigger extend based on its broader definition by including variables that are sleep deficiency related, nighttime- and daytime-related, quantitative and qualitative informed. On that basis, the chosen 36 variables were grouped into seven domains: (1) Satisfaction with sleep/sleep quality (e.g., PSQI total score, Not get enough sleep, Feelings from wake-up, Necessity of nap, Needed nap time, Subjective sleep quality); (2) Alertness during waking hours (e.g., Epworth sleepiness scale, Mind-wandering, Spontaneous mind wandering, Deliberate mind wandering, Frequency of daydream, Fatigue severity, Physical fatigue, Mental fatigue, Attention-related cognitive errors, Valid sleep cue reaction time(RT), Invalid sleep cue RT, Valid sleep cue accuracy (ACC), Invalid

Table 2 | Mean and standard deviation (SD) of the sleep health measures in the discovery and replication dataset

Domains of sleep health	Measures	Discovery dataset (N = 687) Mean (SD)	Replication dataset (N = 628) Mean (SD)
Satisfaction with sleep/ Sleep quality	Not get enough sleep	2.11 (0.74)	2.07 (0.75)
	Feelings from wake-up	2.35 (0.82)	2.32 (0.76)
	Necessity of nap	3.09 (0.89)	2.98 (0.91)
	Needed nap time	4.22 (1.14)	4.19 (1.15)
	PSQI-T (otal score)	4.97 (2.44)	5.33 (2.58)
	Subjective sleep quality	1.16 (0.83)	1.10 (0.84)
Alertness during waking hours	Epworth sleepiness scale	8.91 (3.38)	8.92 (3.47)
	Mind-wandering	13.06 (3.67)	12.84 (3.77)
	Spontaneous mind wandering	19.11 (4.88)	18.36 (4.80)
	Deliberate mind wandering	18.11 (4.23)	17.85 (4.49)
	Attention-related cognitive errors	30.32 (6.80)	30.60 (7.87)
	Frequency of daydream	33.19 (9.07)	31.28 (8.78)
	Fatigue severity	38.96 (8.53)	38.89 (8.88)
	Physical fatigue	4.21 (2.41)	3.95 (2.53)
	Mental fatigue	2.63 (1.17)	2.66 (1.22)
	Valid sleep cue RT	515.85 (71.61)	521.43 (83.61)
	Invalid sleep cue RT	538.14 (76.47)	553.40 (90.26)
	Valid sleep cue ACC	0.95 (0.06)	0.95 (0.06)
	Invalid sleep cue ACC	0.94 (0.07)	0.94 (0.07)
Timing of sleep	Morningness-eveningness questionnaire	13.60 (2.84)	13.22 (2.84)
Sleep efficiency/ continuity	Sleep efficiency	0.94 (0.06)	0.93 (0.07)
	Sleep latency	0.68 (0.79)	0.73 (0.80)
	Wake-up times	0.30 (0.69)	0.36 (0.80)
Sleep duration	Total sleep time	6.74 (0.87)	6.86 (0.94)
Sleep deficiency	Insomnia Severity Index	7.60 (4.25)	7.47 (4.31)
	Hyperarousal	31.90 (9.50)	32.12 (9.72)
	Ford insomnia response to stress test	20.42 (5.30)	20.78 (5.26)
	Sleep disturbances	0.86 (0.50)	0.89 (0.49)
	Use of sleep medications	0.03 (0.243)	0.06 (0.35)
	Daytime dysfunction	1.38 (0.82)	1.75 (0.83)
Sleep beliefs, attitudes, and habits	DBAS-total score	128.74 (33.44)	129.75 (32.84)
	Misconceptions about causes	5.73 (1.72)	5.79 (1.66)
	Diminished control and predictability	3.42 (1.43)	3.43 (1.47)
	Unrealistic sleep expectations	4.67 (1.63)	4.68 (1.72)
	Misattribution of the consequences	3.02 (1.59)	3.13 (1.68)
	Faulty beliefs about sleep-promoting practices	4.00 (1.26)	4.03 (1.28)

RT reaction time, ACC, accuracy, DBAS Dysfunctional Beliefs and Attitudes about Sleep Scale.

sleep cue ACC); (3) Timing of sleep (e.g., Morningness-eveningness questionnaire); (4) Sleep efficiency/ continuity (e.g., Sleep efficiency, Sleep latency, Wake-up times); (5) Sleep duration (e.g., Total sleep time); (6) Sleep deficiency (e.g., Insomnia severity index, Hyperarousal, Ford insomnia response to stress test, Sleep disturbances, Use of sleep medications, Daytime dysfunction); (7) Sleep beliefs, attitudes, and habits (e.g., Dysfunctional beliefs and attitudes about sleep scale (DBAS)-total score, Misconceptions about causes, Diminished control and predictability, Unrealistic sleep expectations, Misattribution of the consequences, Faulty beliefs about sleep promoting practices). For the specific descriptions about the chosen variables, see Table 2 and Supplementary Table S1.

Data acquisition and image preprocessing

The discovery and replication sample

Data acquisition. All MR Images were acquired using a 3T Siemens Prisma-fit scanner with a standard 32-channel head coil located at Southwest University. A high-resolution T1-weighted structural image was obtained using a three-dimensional gradient sequence in order to facilitate alignment of individual subject images into a common space (repetition time (TR) = 2530 ms, time of echo (TE) = 2.98 ms, field of view (FOV) = 256 × 256 mm², thickness = 1 mm, voxel size = 0.5 × 0.5 × 1 mm³, flip angle = 7°, resolution matrix = 256 × 256, slices = 192, slice oversampling = 33.3%, phase-encoding direction = AC × PC). Approximately 8 min of rs-fMRI data containing 240 volumes were acquired for each subject using a blood oxygen level-dependent (BOLD-weighted) sequence (TR = 2000 ms; TE = 30 ms; slices = 62; slice thickness = 2 mm; FOV = 224 × 224 mm²; flip angle = 90°; resolution matrix = 112 × 112; voxel size = 2 × 2 × 2 mm³; phase-encoding direction = PC × AC). In order to minimize motion, prior to data acquisition participants' heads were stabilized in the head coil using one foam pad over each ear and a third over the top of the head. During the resting-state scanning, a fixation cross was displayed as images were acquired. Participants were instructed to stay awake, keep their eyes open, fixate on the displayed crosshair, and remain still.

Image preprocessing. The preprocessing steps on all collected neuroimaging data were performed using the publicly available CONN functional connectivity toolbox (version 20.b; <https://www.nitrc.org/projects/conn>), together with SPM12 (Wellcome Department of Cognitive Neurology, London, UK; <http://www.fil.ion.ucl.ac.uk/spm>). Functional images were (1) slice time corrected, (2) underwent motion correction and susceptibility artifact correction based on field map, (3) warped into Montreal Neurological Institute (MNI) standard space using the diffeomorphic Anatomical Registration Through exponentiated Lie Algebra (Dartel) approach to realign the 3D anatomical data into Montreal Neurological Institute space⁷¹, (4) smoothed spatially with a Gaussian kernel of 6 mm full width at half maximum (FWHM). Next, functional images further underwent denoising steps using the anatomical component-based correction (aCompCor) method⁷⁰. Specifically, noise signals, including signals from cerebrospinal fluid and white matter (WM) (five principal components), and movement parameters (six motion parameters, six temporal derivatives, and their squares), and linear trend were removed from the images as confounds⁷². Subsequently, data scrubbing was implemented to address head motion concerns. The bad time points were regarded as regressors defined as volumes with FD power > 0.5 mm as well as the two succeeding volumes and one preceding volume to reduce the spillover effect of head motion⁷². Finally, functional images were filtered using a bandpass filter (0.008–0.09 Hz) to reduce the effects of very low-frequency drifts and high-frequency noises.

HCP sample

Data acquisition. Structural (T1 and T2 images, required for pre-processing functional neuroimaging data) and functional MRI data were collected at Washington University on the Siemens 3T Connectome Skyra scanner using a multi-band sequence. The structural images were 0.7 mm isotropic. The rs-fMRI data were 2 mm isotropic with TR = 0.72 s. Full details of the acquisition parameters for the HCP data can be found elsewhere⁷³. Two sessions of rs-fMRI data were collected on consecutive days for each subject, and each session consisted of one or two runs. The length of each rs-fMRI scan was 14.4 min (1200 frames). Here, the analyses were restricted to individuals for whom the left–right phase-encoding scans for the rs-fMRI session were completed and available. For rs-fMRI data acquisition, participants were asked to lie with eyes open, with a “relaxed” fixation on a white cross (on a dark background), think of nothing in particular, and not fall asleep. Details of the data collection can be found elsewhere⁷⁴. Details about behavioral measures including PSQI can be found in HCP S1200 Data Dictionary (https://db.humanconnectome.org/data/projects/HCP_1200) and⁷⁵.

Image preprocessing

We adopted the preprocessed data provided by the Human Connectome Project (HCP S1200 release), for which also the spatial normalization to the MNI152 template had already been performed before download. Details of the structural and functional data preprocessing can be found in the HCP S1200 manual⁷⁶, and we used version 3.21 of the HCP preprocessing pipeline. Consistent with the BBP dataset, the HCP downloaded functional images were further smoothed, and underwent regression of motion and non-relevant signals, including linear trend, Friston 24 head motion parameters, white matter (CompCor, 5 principal components), and CSF signal (CompCor, 5 principal components), scrubbed and filtered.

Classification sample

Data acquisition. All MR Images in the SNIC dataset were acquired using a 3T scanner (Magnetom TIM-Trio, Siemens, Erlangen, Germany) with a standard head coil located at Southwest University. In order to minimize motion, prior to data acquisition participants' heads were stabilized in the head coil using one foam pad over each ear and a third over the top of the head. High-resolution T1-weighted anatomical images were collected using 3D spoiled gradient recalled (3D SPGR) sequence (TR/TE = 8.5/3.4 ms, flip angle = 12°, resolution matrix = 512 × 512, FOV = 240 × 240 mm², with a voxel size of 1 × 1 × 1 mm³, 176 slices, 1 mm thickness). Approximately 5 min of rs-fMRI data containing 204 volumes were acquired for each subject using an echo-planar imaging (EPI) sequence (TR/TE = 1500/29 ms, flip angle = 90°, resolution matrix = 64 × 64, voxel size = 3 × 3 × 3 mm³, FOV = 192 × 192 mm², axial slices = 25, thickness/ gap = 5/0.5 mm). The first four volumes were discarded to ensure steady-state longitudinal magnetization. During the resting-state scanning, participants were instructed to fix on a crosshair in the center of the black background screen without thinking intentionally in the mind and keep as motionless as possible⁷⁷.

All MR Images in the CTCMH dataset were acquired by using the same scanner as the discovery and replication dataset. The scanning parameters are also the same. During the resting-state scanning, all participants were instructed to fix on a crosshair in the center of black background screen without thinking intentionally in the mind and keep as motionless as possible.

Image preprocessing

The preprocessing steps on all collected neuroimaging data in the classification sample were performed using fMRIprep 21.0.1^{78,79}

(RRID:SCR_016216), which is based on Nipype 1.6.1⁸⁰(RRID:SCR_002502). Details of the structural and functional data preprocessing can be found in the supplementary materials.

Resting-state functional connectivity construction

To construct RSFC, the preprocessed data in the discovery and replication dataset were first parcellated using the Brainnetome Atlas⁴², which includes 210 cortical regions and 36 subcortical regions. For each participant, the BOLD time course of each node was extracted by taking the mean across voxels. Pearson correlation coefficient (r) between the time courses of each pair of nodes was calculated. A Fisher's r -to- z transformation was performed to improve the normality of the correlation coefficients, which resulted in a 246×246 symmetric functional connectivity matrix with 30,135 $[(245 \times 246)/2]$ edges for each participant.

Partial least squares analysis

We used PLS analysis to examine the relationship between resting-state functional connectivity (RSFC) and sleep health (SH) measures (Fig. 1) in the discovery dataset. PLS analysis is a multivariate statistical technique that derives latent variables (LVs), by finding weighted patterns of variables from two given datasets that maximally covary with each other^{36,81}. In the present analysis, 1 variable set corresponded to RSFC and the other to behavioral measures spanning multiple domains of sleep health. The two variable sets were correlated with each other across participants, and the resulting covariance matrix was subjected to singular value decomposition to identify the latent brain-behavior dimensions. Specifically, each LV is comprised of an RSFC pattern at the node level ("RSFC saliences") and an SH profile ("SH saliences"). Individual-specific RSFC and SH composite scores for each LV were obtained by linearly projecting the RSFC and SH measures of each participant onto their respective saliences. See the Supplementary Methods section "Partial least squares analysis" for mathematical details. Before the PLS analysis, we regressed out the confounding effects from both RSFC and behavior data including mean FD, age, handedness, and sex.

Inference and validation of the statistical model were performed using nonparametric methods including: (1) statistical significance of overall patterns was assessed by permutation tests using 1000 permutations for behavioral data; (2) the importance (measured as loading scores) of feature (RSFC, sleep health measures) was evaluated by bootstrap resampling; (3) the generalizability of each LV was assessed by 10-fold cross-validation with 200 repetitions. Mathematical details of the analysis and inferential methods are described in the Supplementary Methods section "Partial least squares analysis" and results. False discovery rate (FDR) correction ($q < 0.05$) was applied to all analyses.

Control and reliability analyses

We further tested whether LVs obtained in the discovery dataset were robust to (1) global signal regression, (2) total intracranial volume (including gray matter, white matter, and cerebrospinal fluid) regression, (3) time (hour) of acquisition regression, (4) the pre-scanning positive and negative affect regression, (5) BMI regression, (6) family income regression, (7) adding confounding variables (age, sex, handedness and head motion) into the phenotypic data for the PLS analysis, (8) non-Gaussian distributions of the behavioral data with quantile normalization, as well as (9) using a different brain (Seitzman et al.'s) Atlas⁴³ containing 300 regions for the RSFC construction. To assess the robustness of each LV, we computed Pearson's correlations between RSFC (or SH) saliences obtained in each of the eight-control analysis and RSFC (or SH) saliences from the original PLS analysis. Please refer to (able S2 for the results).

Internal validation

As an internal validation of the obtained LVs or dimensions in the discovery dataset, we first replicated the PLS analysis described in the

section "Partial least squares analysis", and "Control and Reliability Analyses" with the replication dataset. Then, Pearson's correlation coefficient between the behavioral salience scores in the discovery and replication dataset, between the behavioral loading scores in the discovery and replication dataset, between the RSFC salience scores in the discovery and replication dataset, between the RSFC loading scores in the discovery and replication dataset were computed respectively.

We further tested the cross-dataset generalizability by projecting the dataset 2 onto the salience parameters learned by the PLS analysis in dataset 1. Then, we examined the correlation between the behavioral and RSFC composite scores in dataset 2. The significance of the correlation value was further assessed by a permutation test (behavioral data of dataset 2 shuffled 1000 times).

External prediction validation

Considering that the PSQI total score contributes the most in the LV/dimension found in the discovery BBP sample ($r = 0.79$), we further tested whether the RSFC spatial pattern of this dimension observed in the discovery dataset can predict the sleep quality of unrelated individuals in the HCP dataset as an external prediction validation (Fig. 1). In particular, we adopted the support vector regression (SVR), the most widely used algorithm in multivariate neuroimaging research, to examine the predictive performance of the significant RSFC pattern in the LV found in the discovery sample on sleep quality in the HCP dataset. Specifically, we used all the edges of the significant RSFC pattern as the features. The 10-fold cross-validation was used to determine the optimal parameter C for SVR. Specifically, the optimal parameter was finally selected according to the prediction performance after the inner 10-fold cross-validation was performed for each parameter ($[0.1, 1]$ with a step of 0.1) in turn and was further utilized in the outer loop to accomplish the final prediction. All subjects were divided into 10 subsets by 10-fold cross-validation (outer loop). Nine of them were used as training sets and model fitting was performed. The fitted model parameters and the remaining one subset were used as the test set to generate predicted values. This process was repeated 10 times to generate predicted values for all subjects. In order to obtain stable predictive performance, we repeated the above prediction pipeline 100 times to generate 100 predicted scores for each participant and further averaged these predicted scores to obtain robust estimates. Pearson's correlation coefficient between the predicted average and actual values was computed to provide final estimates of predictive performance. The significance was evaluated by randomly shuffling the PSQI values 1000 times and running the above prediction pipeline for each time to obtain a null distribution of correlation coefficients between the predicted and actual values.

Spatial correlation with neurotransmitter densities

We further test whether the FC loadings from the PLS analysis were spatially correlated with the distribution of several neurotransmitter systems potentially involved in domains of sleep health. Based on the literature^{31,32,60,61,82}, serotonin receptors (5-HT1A, 5-HT1B, and 5-HT2A) and transporters (serotonin reuptake transporter [5-HTT]), together with metabotropic glutamate receptor 5 (mGluR5) and the γ -Aminobutyric acid type A (GABAA) receptor were investigated. Density values were derived from average group maps of healthy volunteers (5-HT1a: $n = 35$; 5-HT1b: $n = 36$; 5-HT2a: $n = 29$; 5-HTT: $n = 100$; mGluR5: $n = 22$; GABAA: $n = 11$) obtained in prior multitracer molecular imaging studies³³ (details were provided in the Table S7). These maps were resampled to an isotropic 2-mm spatial resolution as we did in the fMRI data. Then, we obtained the average value of each region of the Brainconnectome atlas for each Neurotransmitter density map, i.e., a 246×1 matrix. Further, we summed the positive and negative FC loadings separately for each region of the Brainconnectome atlas to represent the region importance scores in positive (Fig. 6a) and

negative network (Fig. 6b). Finally, we conducted a Spearman correlation analysis between the region importance score and receptor/transporter densities calculated for these regions⁴⁵. To establish the statistical significance of a spatial correlation against chance, we conducted spatial permutation tests to obtain a null distribution of correlation coefficients for region importance scores and neurotransmitter densities extracted from 5000 permutations while accounting for the spatial autocorrelation of brain regions^{83,84}. The *p*-value was determined by empirically observed spatial similarity values compared to the null distribution. The significance level was set at FDR corrected $p < 0.05$.

Clinical diagnosis validation

We further evaluated the possibility of distinguishing the insomnia patients ($n = 52$) from sleep-healthy controls ($n = 49$) in the classification dataset based on the RSFC spatial pattern of the found LV/dimension (details about the classification sample were given in the section of participants). Specifically, we used all the edges of the significant RSFC pattern as the features. To accomplish this, we utilized a Gaussian radial basis function (RBF) kernel SVM classifier which was implemented using the LIBSVM toolbox⁴⁶. We used the recommended exponentially growing sequences of parameters of C and γ by LIBSVM group, i.e., $C (2^{-5}, 2^{-3}, \dots, 2^{15})$ and $\gamma (2^{-15}, 2^{-13}, \dots, 2^3)$ based on a practical guide to SVM⁸⁵. The classifier performance was validated using a 10-fold cross-validation strategy. Specifically, all subjects were divided into 10 subsets by 10-fold cross-validation. Nine of them were used as training sets and model fitting was performed. The fitted model parameters and the remaining one subset were used as the test set to generate classified labels. This process was repeated 10 times in sequence to generate classified labels for all subjects, and then calculate the classification accuracy. Regarding the optimal values for the RBF kernel parameters C and γ , they were tuned based on the training set using a grid search strategy based on a 10-fold cross-validation (inner loop). Then the SVM classifier was trained by applying the optimized parameters to the training set, based on which the testing set was finally classified. In order to obtain stable classification performance, we repeated the above pipeline 100 times to generate 100 classification performance values and further averaged these values to obtain robust estimates.

Statistical significance was evaluated by using 1000 permutation tests with a threshold of $p < 0.05$. In Brief, labels of participants were randomly shuffled 1000 times and split into the training and testing sets. Classification performance that had an accuracy of higher than 70% was considered to be meaningful. Moreover, we replicated the classification procedure conducted in the classification dataset with an external classification dataset to distinguish insomnia patients ($n = 35$) from sleep-healthy controls ($n = 35$), please refer to Supplementary Methods for more details.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The HCP consortium database used in this study was freely available at the following accessible link: <https://db.humanconnectome.org/>. The BBP sample data are available under restricted access for it is still an ongoing project, access can be obtained by contacting a joint team (Pls: Q.H.H., J.Q., T.Y.F., H.C., and X.L.). The significant FC weights from the BBP discovery cohort are available at GitHub: https://github.com/wangyulinatUGent/Sleep_Health_Dimension. The classification sample used in this study is available upon request from the corresponding author. The PET/SPECT data from the prior vivo molecular imaging studies³³ used in this study are available at the following accessible link:

https://github.com/netneurolab/hansen_receptors/tree/main. Source data are provided in this paper.

Code availability

We used the Matlab code from <https://github.com/danizoeller/myPLS>⁸⁶, based on Krishnan and associates' work³⁵ to implement the PLS calculation. The code for spatial permutation testing can be found at https://github.com/frantisekvasa/rotate_parcellation. The codes for SVR and SVM analysis are openly available at <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>⁴⁶. The brain maps were presented using the MRIcroGL toolbox (<https://www.nitrc.org/projects/mricrogl>). The connectome maps were presented using code from (<https://github.com/cocoanlab/cocoanCORE/tree/master/Visualization>)⁸⁷.

References

1. 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. Heal.* **7**, 699–707 (2021).
2. Grandner, M. A. & Fernandez, F. X. The translational neuroscience of sleep: a contextual framework. *Sci. (80-)*. **374**, 568–573 (2021).
3. Buysse, D. J. Sleep health: can we define it? does it matter? *Sleep* **37**, 9–17 (2014).
4. Wallace, M. L. et al. Which sleep health characteristics predict all-cause mortality in older men? An application of flexible multi-variable approaches. *Sleep* **41**, 1–12 (2018).
5. Furihata, R. et al. An aggregate measure of sleep health is associated with prevalent and incident clinically significant depression symptoms among community-dwelling older women. *Sleep* **40**, zsw075 (2017).
6. Knutson, K. L. et al. The National Sleep Foundation's Sleep Health Index. *Sleep. Heal.* **3**, 234–240 (2017).
7. Schiel, J. E. et al. Associations Between Sleep Health and Amygdala Reactivity to Negative Facial Expressions in the UK Biobank Cohort. *Biol. Psychiatry* 1–8 <https://doi.org/10.1016/j.biopsych.2022.05.023> (2022).
8. Furihata, R. et al. A composite measure of sleep health is associated with symptoms of depression among Japanese female hospital nurses. *Compr. Psychiatry* **97**, 152151 (2020).
9. Lee, S. et al. Sleep health composites are associated with the risk of heart disease across sex and race. *Sci. Rep.* **12**, 1–11 (2022).
10. Lee, S. & Lawson, K. M. Beyond single sleep measures: A composite measure of sleep health and its associations with psychological and physical well-being in adulthood. *Soc. Sci. Med.* **274**, 113800 (2021).
11. Brindle, R. C., Yu, L., Buysse, D. J. & Hall, M. H. Empirical derivation of cutoff values for the sleep health metric and its relationship to cardiometabolic morbidity: Results from the Midlife in the United States (MIDUS) study. *Sleep* **42**, 1–9 (2019).
12. Dong, L., Martinez, A. J., Buysse, D. J. & Harvey, A. G. A composite measure of sleep health predicts concurrent mental and physical health outcomes in adolescents prone to eveningness. *Sleep. Heal.* **5**, 166–174 (2019).
13. Brandolim Becker, N. et al. Sleep health assessment: a scale validation. *Psychiatry Res.* **259**, 51–55 (2018).
14. Katulka, E. K., Berube, F. R. & D'Agata, M. N. Dreaming of better health: quantifying the many dimensions of sleep. *Sleep* **43**, zsz275 (2021).
15. Dalmases, M. et al. Assessing sleep health in a European population: results of the catalan health survey 2015. *PLoS ONE* **13**, 1–13 (2018).
16. Krause, A. J. et al. The sleep-deprived human brain. *Nat. Rev. Neurosci.* **18**, 404–418 (2017).
17. Smith, S. M. et al. A positive-negative mode of population covariation links brain connectivity, demographics and behavior. *Nat. Neurosci.* **18**, 1565–1567 (2015).

18. Power, J. D. et al. Functional network organization of the human brain. *Neuron* **72**, 665–678 (2011).
19. Yeo, B. T. T. et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).
20. Tagliazucchi, E. & van Someren, E. J. W. The large-scale functional connectivity correlates of consciousness and arousal during the healthy and pathological human sleep cycle. *Neuroimage* **160**, 55–72 (2017).
21. Wang, Y., Liu, H., Hitchman, G. & Lei, X. Module number of default mode network: Inter-subject variability and effects of sleep deprivation. *Brain Res.* **1596**, 69–78 (2015).
22. Fasiello, E. et al. Functional connectivity changes in insomnia disorder: a systematic review. *Sleep. Med. Rev.* **61**, 101569 (2022).
23. Someren, E. J. W. Van Brain mechanisms of insomnia: new perspectives on causes and consequences. *Psychol. Rev.* **101**, 995–1046 (2021).
24. Killgore, W. D. S. Self-reported sleep correlates with prefrontal-amygdala functional connectivity and emotional functioning. *Sleep* **36**, 1597–1608 (2013).
25. Cheng, W., Rolls, E. T., Ruan, H. & Feng, J. Functional connectivities in the brain that mediate the association between depressive problems and sleep quality. *JAMA Psychiatry* **75**, 1052–1061 (2018).
26. Tian, Y., Chen, X., Xu, D., Yu, J. & Lei, X. Connectivity within the default mode network mediates the association between chronotype and sleep quality. *J. Sleep Res.* **29**, e12948 (2020)..
27. Redline, S. & Purcell, S. M. Sleep and Big Data: harnessing data, technology and analytics for monitoring sleep and improving diagnostics, prediction, and interventions- an era for “Sleep-Omics?”. *Sleep* **44**, 1–5 (2021).
28. Genon, S., Eickhoff, S. B. & Kharabian, S. Linking interindividual variability in brain structure to behaviour. *Nat. Rev. Neurosci.* **23**, 307–318 (2022).
29. Bzdok, D. & Yeo, B. T. T. Inference in the age of big data: Future perspectives on neuroscience. *Neuroimage* **155**, 549–564 (2017).
30. Plante, D. T., Jensen, J. E. & Winkelman, J. W. The role of GABA in primary insomnia. *Sleep* **35**, 741–742 (2012).
31. Monti, J. M. Serotonin control of sleep-wake behavior. *Sleep. Med. Rev.* **15**, 269–281 (2011).
32. Kay, D. B. & Buysse, D. J. Hyperarousal and beyond: New insights to the pathophysiology of insomnia disorder through functional neuroimaging studies. *Brain Sci.* **7**, 23 (2017).
33. Hansen, J. Y. et al. Mapping neurotransmitter systems to the structural and functional organization of the human neocortex. *Nat. Neurosci.* <https://doi.org/10.1038/s41593-022-01186-3> (2021).
34. Mihalik, A. et al. Multiple holdouts with stability: improving the generalizability of machine learning analyses of brain–behavior relationships. *Biol. Psychiatry* **87**, 368–376 (2020).
35. Krishnan, A., Williams, L. J., McIntosh, A. R. & Abdi, H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. *Neuroimage* **56**, 455–475 (2011).
36. Dong, D. et al. Linking cerebellar functional gradients to transdiagnostic behavioral dimensions of psychopathology. *NeuroImage Clin.* **36**, 103176 (2022).
37. Irish, L. A., Kline, C. E., Gunn, H. E., Buysse, D. J. & Hall, M. H. The role of sleep hygiene in promoting public health: a review of empirical evidence. *Sleep. Med. Rev.* **22**, 23–36 (2015).
38. Ramlee, F., Sanborn, A. N. & Tang, N. K. Y. What sways people’s judgment of sleep quality? A quantitative choice- making study with good and poor sleepers. *Sleep* **40**, zsx091 (2017).
39. C, A. L. & Berntsen, D. Sleeping poorly is robustly associated with a tendency to engage in spontaneous waking thought. *Conscious. Cogn.* **105**, 103401 (2022).
40. Woods, H., Marchetti, L. M., Biello, S. M. & Espie, C. A. The clock as a focus of selective attention in those with primary insomnia: An experimental study using a modified Posner paradigm. *Behav. Res. Ther.* **47**, 231–236 (2009).
41. Dawson, D. & McCulloch, K. Managing fatigue: It’s about sleep. *Sleep. Med. Rev.* **9**, 365–380 (2005).
42. Fan, L. et al. The human brainnetome atlas: a new brain atlas based on connectonal architecture. *Cereb. Cortex* **26**, 3508–3526 (2016).
43. Seitzman, B. A. et al. A set of functionally-defined brain regions with improved representation of the subcortex and cerebellum. *Neuroimage* **206**, 116290 (2020).
44. Alexander Diaz, B. et al. The ARSQ 2.0 reveals age and personality effects on mind-wandering experiences. *Front. Psychol.* **5**, 1–8 (2014).
45. Dukart, J. et al. JuSpace: a tool for spatial correlation analyses of magnetic resonance imaging data with nuclear imaging derived neurotransmitter maps. *Hum. Brain Mapp.* **42**, 555–566 (2021).
46. Chang, C. C. & Lin, C. J. LIBSVM: a Library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27 (2011).
47. Nicolazzo, J. et al. Sleep symptomatology is associated with greater subjective cognitive concerns: Findings from the community-based Healthy Brain Project. *Sleep* **4533**, 1–27 (2020).
48. Morin, C. M., Vallières, A. & Ivers, H. Dysfunctional beliefs and attitudes about sleep (DBAS): Validation of a brief version (DBAS-16). *Sleep* **30**, 1547–1554 (2007).
49. Sella, E., Carbone, E., To, E. & Borella, E. Personality traits and sleep quality: The role of sleep-related beliefs. *Pers. Individ. Dif.* **156**, 109770 (2020).
50. Means, M. K., Edinger, J. D., Glenn, D. M. & Fins, A. I. Accuracy of sleep perceptions among insomnia sufferers and normal sleepers. *Sleep. Med.* **4**, 285–296 (2003).
51. Harvey, A. G. & Tang, N. K. Y. Misperception of sleep in insomnia: a puzzle and a resolution. *Psychol. Bull.* **138**, 77–101 (2012).
52. Bromberg-Martin, E. S. & Sharot, T. The Value of Beliefs. *Neuron* **106**, 561–565 (2020).
53. Koffel, E. A., Koffel, J. B. & Gehrman, P. R. A meta-analysis of group cognitive behavioral therapy for insomnia. *Sleep. Med. Rev.* **19**, 6–16 (2015).
54. Huntenburg, J. M., Bazin, P. & Margulies, D. S. Large-scale gradients in human cortical organization. *Trends Cogn. Sci.* **22**, 21–31 (2018).
55. Seeley, W. W. The salience network: a neural system for perceiving and responding to homeostatic demands. *J. Neurosci.* **39**, 9878–9882 (2019).
56. Uddin, L. Q. Salience processing and insular cortical function and dysfunction. *Nat. Rev. Neurosci.* **16**, 55–61 (2014).
57. Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L. & Raichle, M. E. Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proc. Natl Acad. Sci.* **103**, 10046–10051 (2006).
58. Kim, H. Involvement of the dorsal and ventral attention networks in oddball stimulus processing: a meta-analysis. *Hum. Brain Mapp.* **35**, 2265–2284 (2014).
59. Kim, N., Won, E., Cho, S. E., Kang, C. K. & Kang, S. G. Thalamocortical functional connectivity in patients with insomnia using resting-state fMRI. *J. Psychiatry Neurosci.* **46**, E639–E646 (2021).
60. Morgan, P. T. et al. Cortical GABA levels in primary insomnia. *Sleep* **35**, 807–814 (2012).
61. Watanabe, K. et al. Genome-wide meta-analysis of insomnia prioritizes genes associated with metabolic and psychiatric pathways. *Nat. Genet.* **54**, 1125–1132 (2022).
62. He, D., Ren, D., Guo, Z. & Jiang, B. Insomnia disorder diagnosed by resting-state fMRI-based SVM classifier. *Sleep. Med.* **95**, 126–129 (2022).
63. Angelova, M., Karmakar, C., Zhu, Y. E., Drummond, S. P. A. & Ellis, J. Automated method for detecting acute insomnia using multi-night actigraphy data. *IEEE Access* **8**, 74413–74422 (2020).

64. Inhulsen, M.-B. M. R. Measuring sleep health in primary school-aged children: a systematic review of instruments and their content validity. *Sleep* **45**, zsac215 (2022).
65. Tanaka, H. & Shirakawa, S. Sleep health, lifestyle and mental health in the Japanese elderly: Ensuring sleep to promote a healthy brain and mind. *J. Psychosom. Res.* **56**, 465–477 (2004).
66. Wang, H. et al. Distinct neural responses of morningness and eveningness chronotype to homeostatic sleep pressure revealed by resting-state functional magnetic resonance imaging. *CNS Neurosci. Ther.* **28**, 1439–1446 (2022).
67. Essen et al. NeuroImage The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, 62–79 (2013).
68. Sabuncu, M. R. et al. Morphometricity as a measure of the neuroanatomical signature of a trait. *Proc. Natl Acad. Sci. USA* **113**, E5749–E5756 (2016).
69. Beattie, L., Espie, C. A., Kyle, S. D. & Biello, S. M. How are normal sleeping controls selected? A systematic review of cross-sectional insomnia studies and a standardized method to select healthy controls for sleep research. *Sleep. Med.* **16**, 669–677 (2015).
70. Muschelli, J. et al. Reduction of motion-related artifacts in resting state fMRI using aCompCor. *Neuroimage* **96**, 22–35 (2014).
71. Ashburner, J. A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113 (2007).
72. Satterthwaite, T. D. et al. An improved framework for confound regression and filtering for control of motion artifact in the pre-processing of resting-state functional connectivity data. *Neuroimage* **64**, 240–256 (2013).
73. Kamil, U. et al. Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *Neuroimage* **80**, 80–104 (2013).
74. Smith, S. M. et al. Resting-state fMRI in the Human Connectome Project. *Neuroimage* **80**, 144–168 (2013).
75. Barch, D. M. et al. Function in the human connectome: Task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
76. Glasser, M. F. et al. The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
77. Dong, X. et al. Rest but busy: aberrant resting-state functional connectivity of triple network model in insomnia. *Brain Behav.* **8**, 1–9 (2018).
78. Esteban, O. et al. fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
79. Esteban, O. et al. fMRIPrep: a robust preprocessing pipeline for functional MRI (22.0.2). *Zenodo* <https://doi.org/10.5281/zenodo.7117719> (2022).
80. Gorgolewski, K. et al. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* **5**, 13 (2011).
81. Kirschner, M. et al. Latent clinical-anatomical dimensions of schizophrenia. *Schizophr. Bull.* **46**, 1426–1438 (2020).
82. Monti, J. M. & Monti, D. The involvement of dopamine in the modulation of sleep and waking. *Sleep. Med. Rev.* **11**, 113–133 (2007).
83. Burt, J. B., Helmer, M., Shinn, M., Anticevic, A. & Murray, J. D. Generative modeling of brain maps with spatial autocorrelation. *Neuroimage* **220**, 117038 (2020).
84. Markello, R. D. & Misisic, B. Comparing spatial null models for brain maps. *Neuroimage* **236**, 118052 (2021).
85. Hsu, C.-W., Chang, C.-C. & Lin, C.-J. *A Practical Guide to Support Vector Classification*. 1–16. (Department of Computer Science, National Taiwan University, 2003).
86. Zöllner, D. et al. Disentangling resting-state BOLD variability and PCC functional connectivity in 22q11.2 deletion syndrome. *Neuroimage* **149**, 85–97 (2017).
87. Lee, J. J. et al. A neuroimaging biomarker for sustained experimental and clinical pain. *Nat. Med.* **27**, 174–182 (2021).

Acknowledgements

This work has been funded by the National Key Research and Development Program of China (2021YFC2501500, X.L.), the National Natural Science Foundation of China (31971028, X.L.; 82202247, D.B.D.; 32300861, Y.L.W.), and 2021 International Exchange Program and Introduction Project Funding awarded to Y.L.W.

Author contributions

Y.L.W. conceived and implemented the study, carried out the data analysis, and wrote the manuscript. X.L. supervised the work. S.G. supported the data analysis, especially the multivariate methods, and reviewed and edited the manuscript. D.B.D. supported the calculation of the spatial correlation between the sleep-health-related connectome and the distribution of neurotransmitter systems. Both D.B.D. and F.Z. helped with the machine-learning methods and commented on the manuscript. C.Y.L. acquired and shared dataset4. K.Y. and D.H.Y. acquired and shared the external classification dataset. Q.H.H., J.Q., T.Y.F., H.C., and X.L. funded, designed, and implemented the Behavioral Brain Research Project to Chinese Personality (BBP) project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-023-42945-5>.

Correspondence and requests for materials should be addressed to Xu Lei.

Peer review information *Nature Communications* thanks Huazhan Yin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023