



# The 103,200-arm acceleration dataset in the UK Biobank revealed a landscape of human sleep phenotypes

Machiko Katori<sup>a,1</sup>, Shoi Shi<sup>b,c,1</sup>, Koji L. Ode<sup>b,c</sup>, Yasuhiro Tomita<sup>b,d</sup>, and Hiroki R. Ueda<sup>a,b,c,2</sup>

Edited by Joseph Takahashi, The University of Texas Southwestern Medical Center, Dallas, TX; received September 10, 2021; accepted January 18, 2022

Human sleep phenotypes can be defined and diversified by both genetic and environmental factors. However, some sleep phenotypes are difficult to evaluate without long-term, precise sleep monitoring, for which simple yet accurate sleep measurement is required. To solve this problem, we recently developed a state-of-the-art sleep/wake classification algorithm based on wristband-type accelerometers, termed ACCEL (acceleration-based classification and estimation of long-term sleep-wake cycles). In this study, we optimized and applied ACCEL to large-scale analysis of human sleep phenotypes. The clustering of an about 100,000-arm acceleration dataset in the UK Biobank using uniform manifold approximation and projection (UMAP) dimension reduction and density-based spatial clustering of applications with noise (DBSCAN) clustering methods identified 16 sleep phenotypes, including those related to social jet lag, chronotypes (“morning/night person”), and seven different insomnia-like phenotypes. Considering the complex relationship between sleep disorders and other psychiatric disorders, these unbiased and comprehensive analyses of sleep phenotypes in humans will not only contribute to the advancement of biomedical research on genetic and environmental factors underlying human sleep patterns but also, allow for the development of better digital biomarkers for psychiatric disorders.

sleep | sleep landscape | clustering | UMAP | insomnia

Scientific and technological advancements often provide a landscape of our world and shape new ways to understand it. In the life sciences over the past 20 y, from the analysis of the whole human genome to the development of next generation sequencing, the landscape of genome science has changed dramatically (1–4). In fact, there have been many sustained efforts to understand the genetic landscape of diseases such as cancer, neurodegenerative diseases, and psychiatric disorders by conducting large-scale genome-wide association studies and comprehensive analyses of genetic variants (5, 6). In recent years, the trend of big data analysis has been moving toward a future where in-depth genetic analysis is combined with rich phenotype analysis (7). In this context, large-scale analysis of human phenotypes that considers both genetic as well as environmental factors will be important for gaining a deeper understanding of the topics of investigation.

Sleep is a physiological phenomenon that is widely conserved throughout the animal kingdom. Its basic structure is genetically conserved within species—humans are no exception to this. However, this structure can change either transiently or chronically depending on environmental factors (8). This effect of environmental factors on sleep can be observed in terms of the diversification of modern lifestyles. Humans are basically diurnal animals that are active during the day and sleep at night, but the widespread use of electricity has freed us from this constraint. It is also possible to temporarily affect wakefulness or sleep, as a way to resist genetically predetermined sleep times, through the intake of substances such as caffeine or alcohol. The way we work has also diversified as shift work and other types of work schedules have become more popular (8, 9). However, there are health trade-offs to this diversification of sleep schedules. For example, a night person tends to have less sleep on weekdays (10, 11), probably due to the social obligation of attending school or work in the morning. This might be associated with poorer concentration and academic/professional performance in the morning in such individuals (12). This leads to the concept of “social jet lag,” representing the difference in sleep duration between weekdays and weekends or holidays, which is a concern due to the potential for adverse health effects (13, 14). The diverse sleep phenotypes are not independent of mortality risks from chronic diseases such as cardiovascular disease, metabolic syndrome, and diabetes, although the complex relationships between them remain poorly understood (15).

The major effects of sleep shortage include physical effects (drowsiness, fatigue, and high blood pressure), cognitive impairment (decreased performance, attention, motivation, mental focus, and intellectual capacity), immune system dysfunction, and possible

## Significance

Human sleep phenotypes are diversified by genetic and environmental factors, and a quantitative classification of sleep phenotypes would lead to the advancement of biomedical mechanisms underlying human sleep diversity. To achieve that, a pipeline of data analysis, including a state-of-the-art sleep/wake classification algorithm, the uniform manifold approximation and projection (UMAP) dimension reduction method, and the density-based spatial clustering of applications with noise (DBSCAN) clustering method, was applied to the 100,000-arm acceleration dataset. This revealed 16 clusters, including seven different insomnia-like phenotypes. This kind of quantitative pipeline of sleep analysis is expected to promote data-based diagnosis of sleep disorders and psychiatric disorders that tend to be complicated by sleep disorders.

Author contributions: M.K., S.S., and H.R.U. designed research; M.K. and S.S. performed research; K.L.O. contributed new reagents/analytic tools; M.K. and S.S. analyzed data; and M.K., S.S., Y.T., and H.R.U. wrote the paper.

Competing interest statement: M.K., S.S., K.L.O., and H.R.U. have filed a patent application regarding the sleep/wake classification algorithm. H.R.U. is the founder and Chief Technology Officer of ACCELStars Inc.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

<sup>1</sup>M.K. and S.S. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: uedah-ky@umin.ac.jp.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2116729119/-DCSupplemental>.

Published March 18, 2022.

complications related to mental illnesses (16); 60 to 70% of adults are known to have sleep-related anxiety/problems, some of which are defined as sleep disorders (17). For example, hypersomnia is characterized by significant daytime sleepiness despite sufficient sleep at night (18). Circadian rhythm sleep disorder is another well-known sleep disorder and is characterized by a discrepancy between the actual sleep schedule and the required sleep schedule. One of the major sleep disorders is insomnia, which presents as a combination of the symptoms (difficulty in initiating sleep, difficulty in maintaining sleep, and waking up earlier than necessary) and associated daytime consequences. About 20% of the general population has insomnia symptoms three or more nights a week, of which about half have daytime consequences and meet the criteria of an insomnia diagnosis (19).

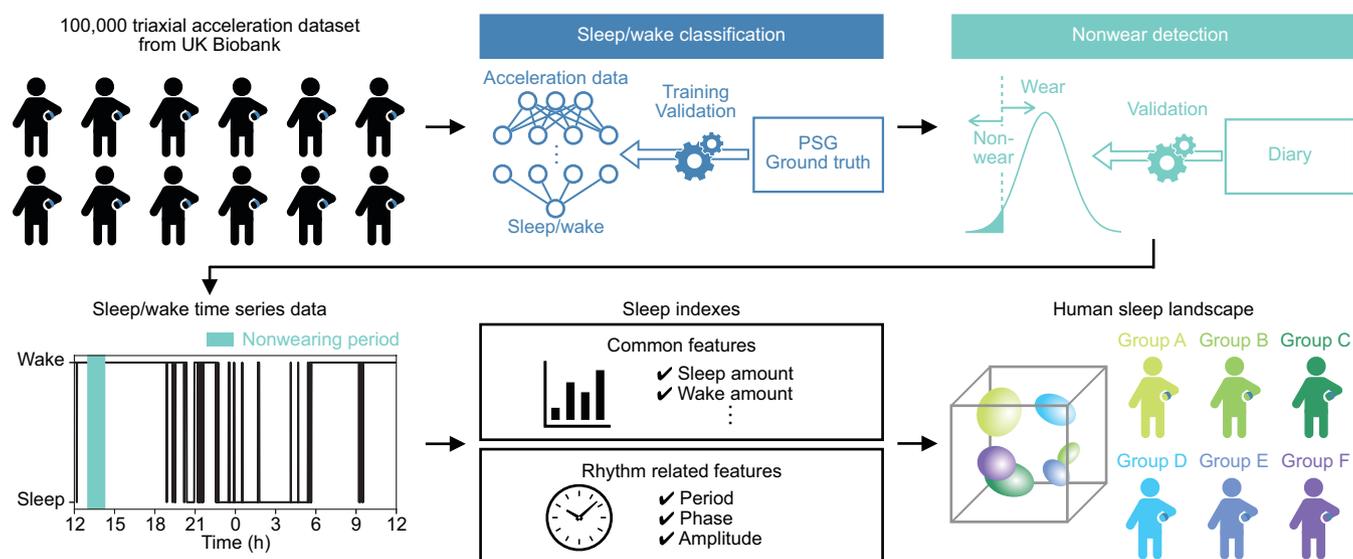
Several instruments are available to assist health care providers in evaluating complaints of insomnia. Polysomnography (PSG), which monitors such as brain activity and muscle movements, is the gold standard for evaluating sleep disorders such as insomnia, hypersomnia, and sleep apnea, and it is essential for the definitive diagnosis of these disorders. Since long-term measurements using PSG are not feasible, sleep questionnaires are also used with PSG measurement for diagnosis today (20, 21). However, these questionnaire-based estimates do not reflect the absolute values obtained by PSG. Thus, wristband accelerometers are commonly used as an alternative to PSG and sleep questionnaires. In fact, they have been used for many large-scale analyses so far (22–24) and have led to the discovery of several genes involved in sleep time regulation (25, 26). In combination with machine learning, it is now possible to capture the detailed structure of sleep, including midawake. The large-scale dataset thus obtained enabled us to systematically classify and interpret the various sleep phenotypes in modern society at a high resolution. Ultimately, large-scale measurements and automatic and quantitative classification of sleep phenotypes could reduce the workload of clinicians and may even lead to the discovery of rare phenotypes that have not received much attention so far.

In this study, as a first effort at automatic classification of sleep phenotypes, we classified sleep phenotypes based on over 100,000 accelerometer-acquired datasets in the UK Biobank (27). The large-scale acceleration data were converted to sleep/wake time series data by combining the state-of-the-art sleep/wake classification

algorithm, termed ACCEL (28), and a nonwear detection algorithm. We calculated 21 sleep indexes from sleep/wake time series data and applied manifold-based dimension reduction and clustering methods (29, 30). The systematic and unsupervised clustering of the large-scale dataset revealed 16 clusters, representing distinctive sleep phenotypes that are consistent with medically described conditions, including social jet lag-related sleep phenotypes and several insomnia-related sleep phenotypes (Fig. 1).

## Results

**Sleep/Wake Time Series Data.** In this study, we further optimized and applied the algorithm of ACCEL to extract sleep/wake time series data from acceleration data obtained using Axivity, the activity-tracking wristband with a triaxial accelerometer used in the UK Biobank project (27). The original sleep/wake classification algorithm is a machine learning-based algorithm that uses XGBoost and the power spectrum of jerk (a derivative of acceleration) as its features (28). In this study, we simultaneously acquired 27 PSG data and Axivity acceleration data and optimized the sleep/wake classification algorithm for Axivity (*SI Appendix, Table S1*). The Axivity signal had lower amplitudes in terms of jerk and the power spectrum during sleep epochs than wake epochs, as was shown in a previous study (*SI Appendix, Fig. S1 A–C*) (28). The XGBoost hyperparameters were optimized to maximize the summation of accuracy and F measure the sleep/wake classification algorithm using Bayesian optimization (*SI Appendix, Fig. S1D*). We also adopted the nonwear detection algorithm from a previous study (31); this algorithm uses the thresholds of two features, SD and range of acceleration, to predict nonwearing periods. In this study, we analyzed the distributions of these features between wearing and nonwearing periods and demonstrated that the thresholds proposed in the previous study are adaptable to Axivity (*SI Appendix, Fig. S1 E and F and Table S2*). As a result of combining the sleep/wake classification and nonwear detection algorithms, the algorithm used in this study achieved high sensitivity ( $97.20 \pm 2.38\%$ ) and specificity ( $82.19 \pm 12.03\%$ ); these are the percentages that the algorithm correctly classified sleep epochs as sleep and wake epochs as wake, respectively



**Fig. 1.** Overview. About 100,000 triaxial acceleration datasets stored in the UK Biobank were converted to the sleep/wake time series data through the sleep/wake classification and the nonwear detection algorithms. The sleep/wake time series data were then converted to 21 sleep indexes. Lastly, the landscape of human sleep phenotypes was classified by clustering methods based on the sleep indexes.

(SI Appendix, Fig. S1G). We also confirmed that our algorithm shows high performance for acceleration data with a different sampling frequency (SI Appendix, Fig. S1H and I and Table S3). Moreover, the high specificity of the algorithm allowed us to accurately detect short-term awake episodes during sleep, which had been difficult in previous studies (22–24). We also calculated two standard sleep indexes, total sleep time (TST), and wake after sleep onset (WASO). These are used to characterize sleep structures in PSG-based studies and are also often used to evaluate the performance of sleep/wake classification algorithms (SI Appendix, Fig. S2A–D) (22–24). In this study, sleep onset and offset were defined from sleep/wake time series data, during which summation of sleep time and awaking time were measured as TST and WASO. Bland–Altman plots show that our algorithm overestimated TST and WASO by only 5.89 and 1.43 min, respectively (SI Appendix, Fig. S2E and F), which are almost comparable with other previous studies (22–24).

**Sleep Indexes Extraction.** Generally, people take the longest sleep at night, but due to the increased diversity of social life, some people sleep longest during the day (32). In addition, variation in total sleep amount per day, high WASO, and low TST are sometimes considered as hallmarks of sleep disorders (33). To capture the diverse structures of sleep, we converted sleep/wake time series data to a total of 21 sleep indexes, including 17 common sleep indexes representing quantity-related features and four rhythm-related sleep indexes representing circadian rhythm-related features (Fig. 2A and SI Appendix, Table S4).

The 17 common sleep indexes include the number and length of sleep windows, which represent the time zone of dense sleep (34), and sleep time (ST) and wake time (WT), which represent sleep duration and midawake duration during sleep windows, respectively. In this study, two different sleep windows were considered to capture both long and short sleep windows according to the following procedure (Fig. 2B). Convert wake periods of less than 10 min to sleep periods and vice versa (Fig. 2B, *a* and *b*). The threshold of 10 min was determined by verifying the sensitivity and specificity of the sleep/wake classification algorithm using the pseudo-sleep/wake time series data (SI Appendix, Fig. S3A–C). Evaluate the length of time gaps between sleep periods. If the duration is less than 60 min, connect them as a sleep window (35) (Fig. 2B, *c*). According to the length of each sleep window, name it a short sleep window if shorter than a threshold or a long sleep window if longer than a threshold (Fig. 2B, *d* and SI Appendix). Fig. 2C shows an example of sleep data of 1 day from noon to the next noon (noon-to-noon data). There is one long sleep window and one short sleep window, where the threshold was set as 3.75 h. The blue area in Fig. 2C shows the long sleep window from after 9:00 PM until about 7:00 AM, including a total of 0.89 h of wake episodes (WT long) and 8.30 h of sleep episodes (ST long). The green area in Fig. 2C shows that there is a short sleep window from before 9:00 AM until about 11:00 AM, during which there is 0.97 h of wake episodes (WT short) and 1.42 h of sleep episodes (ST short). The percentage of sleep time in 24 h (sleep percentage) is 40.49%. For subjects with multiple days of measurement, mean (MN) and SD were calculated for each sleep index.

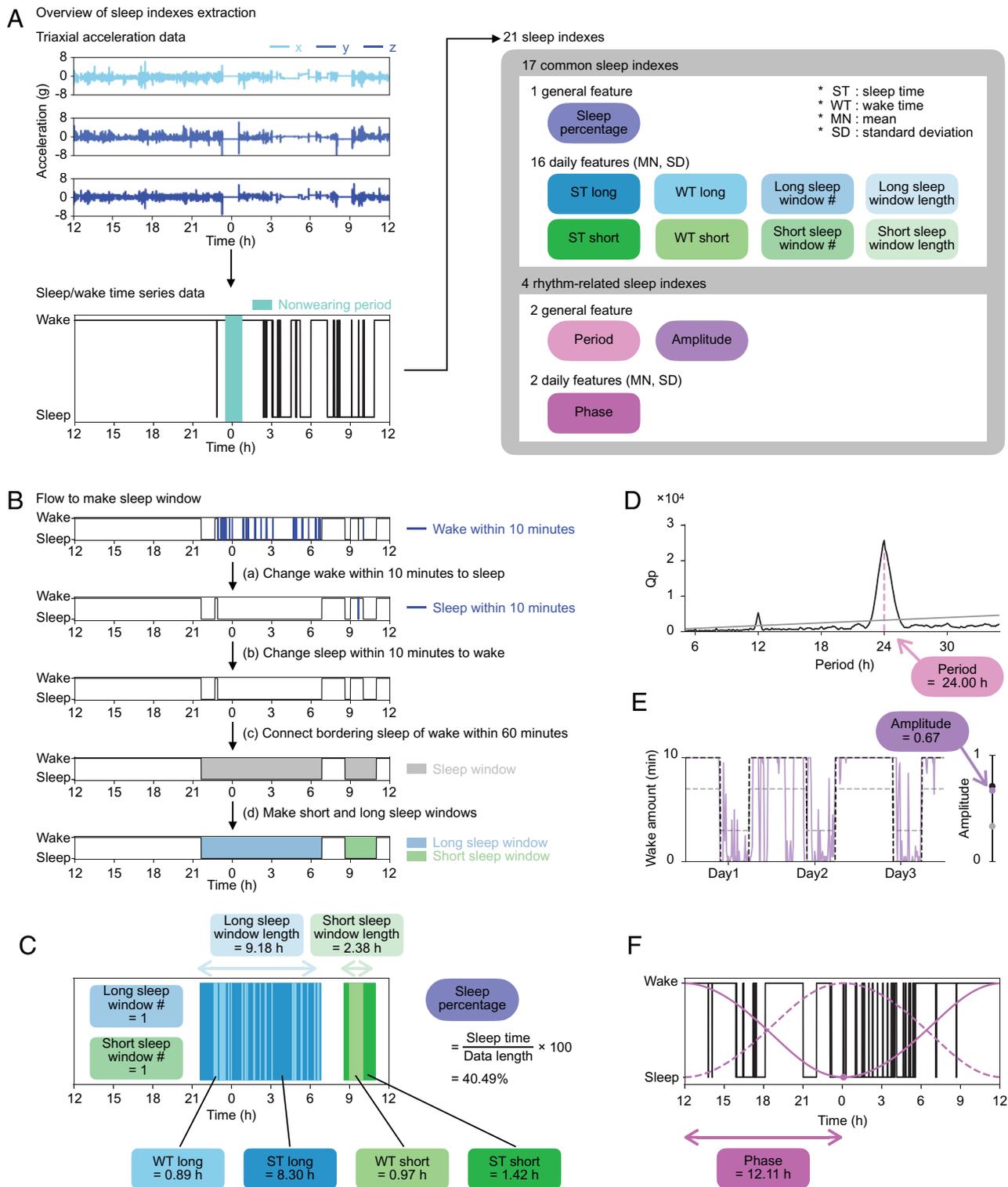
The rhythm-related sleep indexes adopted in this study (period, amplitude, and phase) are commonly used features in the field of circadian rhythm research (Fig. 2D–F) (36–39). Period was calculated as the maximum peak of chi-square periodogram (SI Appendix, Fig. S3D–F) (36). It is usually difficult to obtain the true amplitude of circadian rhythms without various perturbations (such as light illumination). Therefore, amplitude in this study was defined as a coefficient of variation (SD/mean) of wake

amount per 10 min, as in a previous study (Fig. 2E) (37), to represent the amplitude of a circadian output to sleep. In the case of data in Fig. 2D and E, period is 24.00 h, and amplitude is 0.67. To calculate phase, we used the van der Pol limit cycle, which is a classical model for calculating the circadian phase in human studies (38, 39). After fitting data, we defined phase as the duration between the minimum point on the fitted curve (the magenta point in Fig. 2F) and the last noon. In the case of data in Fig. 2F, phase is 12.11 h.

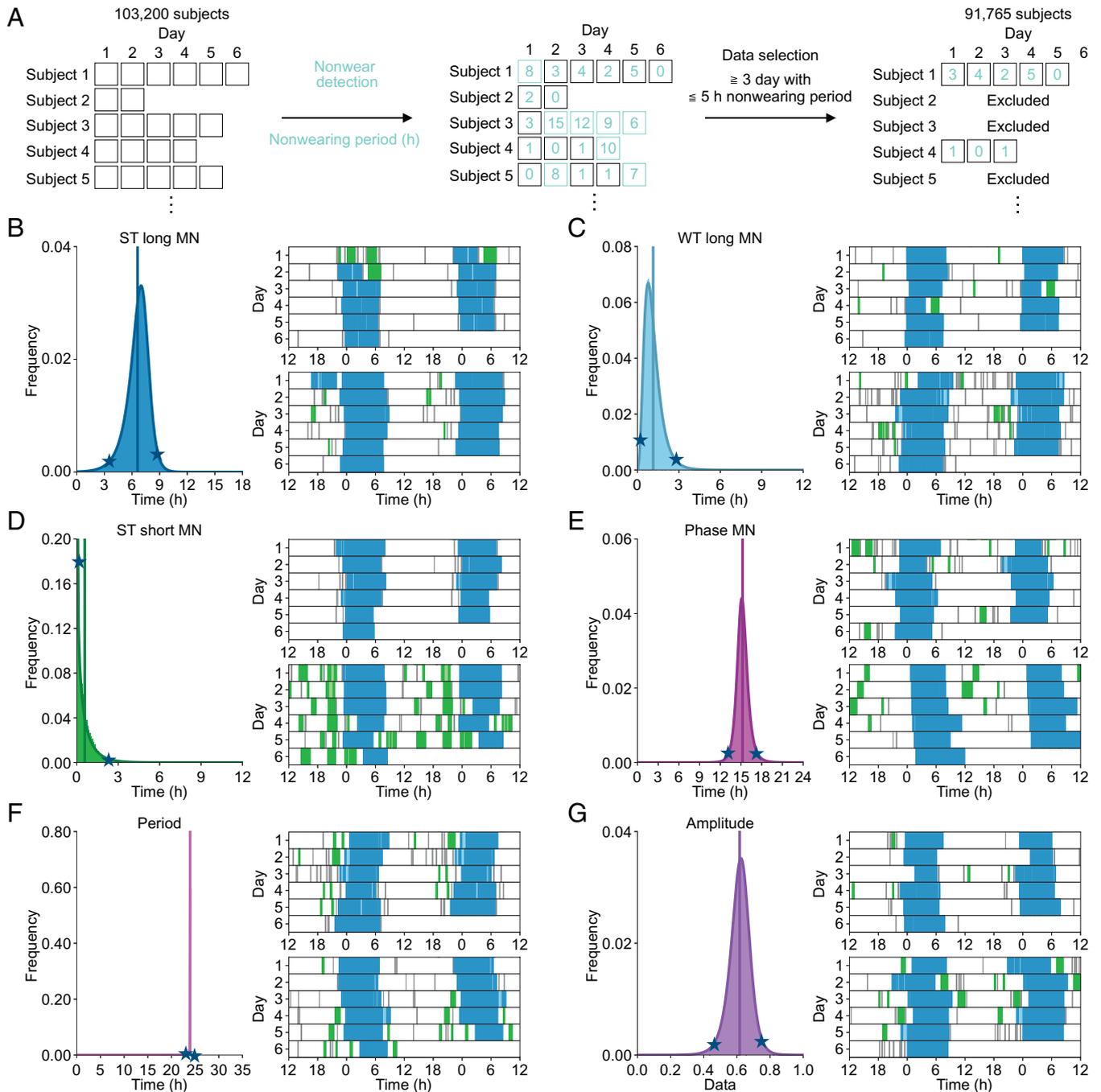
**Distribution of Sleep Indexes.** In this study, we analyzed the Axivity dataset of 103,200 subjects over 522,826 d in total, with up to 7 d of continuous measurement from the UK Biobank project (27). To calculate amplitude and period, we selected individual records from subjects with more than 3 d of continuous measurement and less than 5 h of nonwearing period (Fig. 3A and SI Appendix, Fig. S4A); 91,765 subjects met these criteria, and their acceleration data were converted into sleep/wake time series data for the calculation of 21 sleep indexes. The threshold between the long and short sleep windows was determined by setting a threshold based on the distribution of sleep window length (SI Appendix, Fig. S4B).

The correlation coefficients between the sleep indexes are shown in SI Appendix, Fig. S4C; the median of coefficients is 0.09, which indicates that there are weak correlations between the sleep indexes. We plotted distributions for each sleep index (Fig. 3B–G and SI Appendix, Fig. S4D–R), where curve fitting was conducted for all distributions, except for integral and SD features, to capture their shapes. The mean of ST long MN was 6.60 h (Fig. 3B) (age =  $62.01 \pm 7.82$ ). When we analyzed the data for subjects in their 70s, the mean of ST long MN was 6.33 h, which is close to TST in a PSG-based large-scale study ( $n = 889$ , age =  $76.28 \pm 5.47$ ) (40). The representative plots from both sides (2.28 percentile) of distribution are shown in Fig. 3B–G. We found that the representative plots where certain sleep indexes are in the upper or lower 2.28 percentiles show interesting sleep phenotypes (Fig. 3C–E), some of which are similar to sleep disorder-like phenotypes. With regard to the representative plot with high WT long MN shown in Fig. 3C, awake time during the long sleep window per day is 170.25 min, which may indicate an insomnia-like sleep phenotype with long WASO (33). Similarly, the representative plot with high ST short MN shown in Fig. 3D includes longer daytime sleep episodes (137.83 min/d), which are a feature of hypersomnia (33). The mean of phase MN was 15.22 h, where representative plots with low and high phase represent phenotypes of a morning person (subjective midnight at around 1:30 AM) and a night person (subjective midnight at around 5:30 AM), respectively (Fig. 3E). Thus, it may be possible to find a sleep phenotype that has the same characteristics as a particular sleep disorder by setting a proper threshold; however, setting a threshold can sometimes be arbitrary. In addition, the correlations among 21 sleep indexes (SI Appendix, Fig. S4C) could make classification difficult. Therefore, we decided to use all 21 sleep indexes for our classification.

**Dimension Reduction and Clustering.** For analyses of large-scale data with multiple dimensions, dimension reduction methods usually play an important role and enable us to visualize high-dimensional data in a two- or three-dimensional space. In this study, dimension reduction methods were applied to 21 sleep indexes to classify sleep phenotypes and delineate the sleep landscape (Fig. 4A). We applied four methods—principal component analysis (PCA), t-distributed stochastic



**Fig. 2.** Sleep index extraction. (A) The overview of sleep indexes extraction. Each set of axial data is shown in the three panels in *Upper Left* (row 1: x; row 2: y; row 3: z). The sleep/wake time series data are shown in the same format as in Fig. 1. Twenty-one sleep indexes converted from the sleep/wake time series data, including 17 common sleep indexes and four rhythm-related sleep indexes. The sleep indexes, calculated as a single value throughout the measurement period, were named general features (oval icons). From the daily features (rectangle icons), both MN and SD were included in the sleep indexes. (B) The procedures to make the sleep window. We changed epochs of continuous wake or sleep for less than 10 min to sleep or wake, respectively. The sleep window was created by connecting sleep epochs, ignoring waking epochs of 60 min or less. Long sleep windows (blue) and short sleep windows (green) are made based on the length of the sleep window. (C) An example of noon-to-noon data and common sleep indexes calculated for a day. (D) The result of the chi-square periodogram. The black line shows the  $Q_p$  values (a statistic of chi-square), and the gray line shows 0.01 levels of statistical significance ranging from 5.00 to 35.00 h. The pink dashed line shows the point when the difference between  $Q_p$  and the significant value is at its maximum, and its value, in this case 24.00 h, is used as the period. (E) The purple line shows wake amount per 10 min. The black and gray dashed lines represent a 24-h periodic square wave signal with 1/3 duty in the range from 0 to 10 min and from 3 to 7 min, respectively. The dots on the right bar show the amplitudes of the three lines that are calculated as the coefficient of variation  $SD/mean$ . The purple dot plotted at 0.67 is the amplitude of this example data. (F) The black line shows the sleep/wake time series data. The dashed and solid magenta lines are the van der Pol limit cycles. The dashed line is the curve with the minimum point at noon. The solid line is a fitted curve to the sleep/wake time series data, and the dot is the minimum point of this curve. The duration between the minimum point and the last noon is calculated as the phase: in this case, 12.11 h.

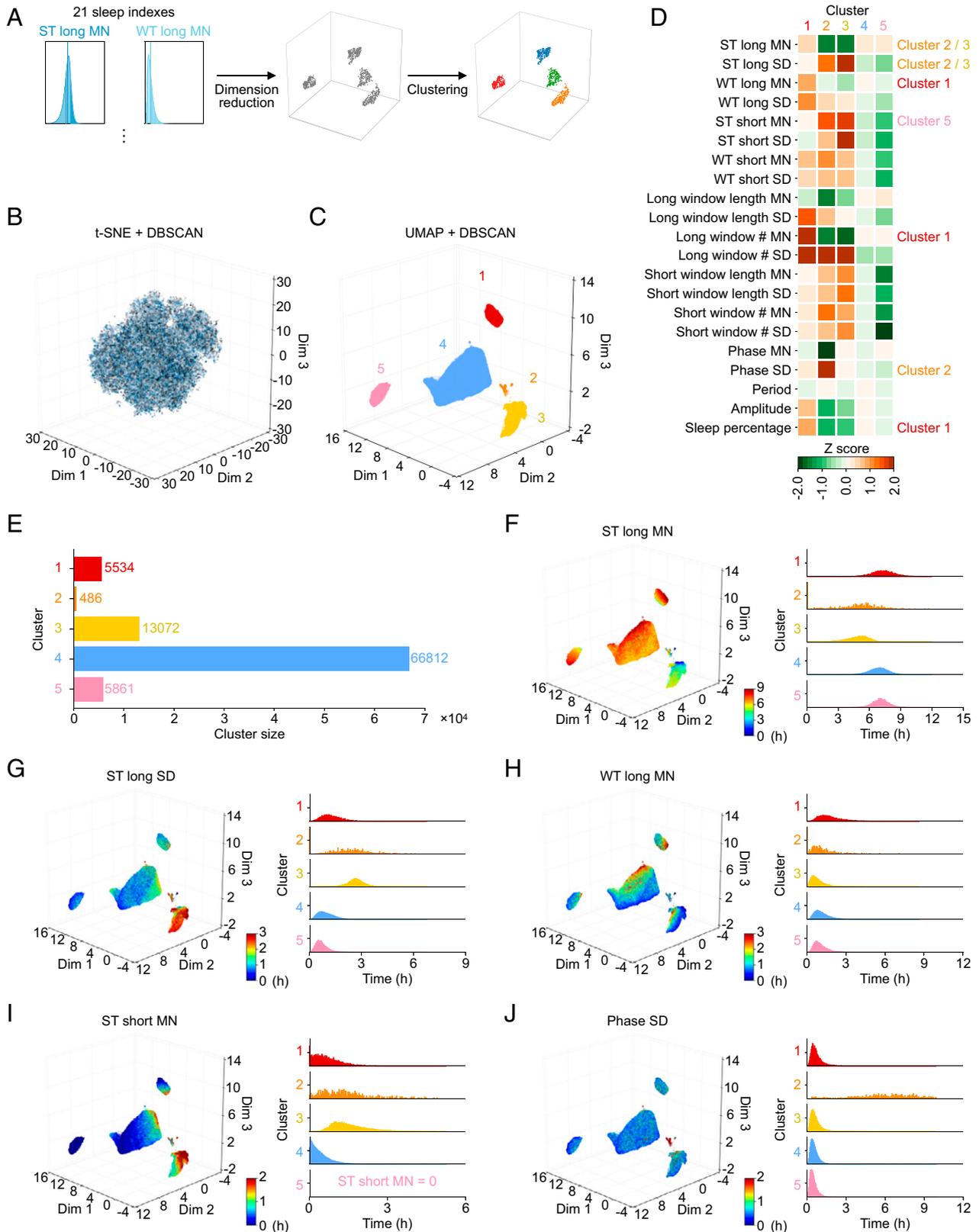


**Fig. 3.** Distribution of sleep indexes. (A) The flow of data exclusion for large-scale sleep analysis. Nonwearing periods (emerald green) were calculated for noon-to-noon data. The noon-to-noon data with less than 5 h of nonwearing period and continuing more than 3 d were used for the large-scale sleep analysis (black squares). In this schema, data 1 and 4 are included in the large-scale sleep analysis. (B–G), *Left* shows the distribution of sleep index, with the mean and the fitted curve shown as the solid lines and solid curves, respectively. The stars show the locations of representative plots (lower or upper 2.28 percentiles) shown as double plots in *Right*, where ST long, WT long, ST short, and WT short are colored the same color as the icons of sleep indexes in Fig. 2A. The sleep epochs outside long and short sleep windows are shown in gray.

neighbor embedding (t-SNE), uniform manifold approximation and projection (UMAP), and a combination of PCA and UMAP—to the 21-dimensional data and converted them to 3-dimensional data, resulting in UMAP dividing the dataset into more interpretable clusters than the other methods (Fig. 4B and C and *SI Appendix, Figs. S5 A–D and S6A*). Therefore, we selected the UMAP method in this study and further applied a clustering method known as density-based spatial clustering of applications with noise (DBSCAN).

As a result, we obtained five clusters (clusters 1 to 5) as shown in Fig. 4C. The features and population size of each cluster

shown in Fig. 4D and F–J and *SI Appendix, Fig. S5 E–T* and in Fig. 4E, respectively. Cluster 1 has higher WT long MN than other clusters (Fig. 4D and H), with more than one long sleep window per day (*SI Appendix, Fig. S5K*) and thus, high sleep percentage (*SI Appendix, Fig. S5T*), which means that the subjects in cluster 1 slept more and with long-term midawake. Since difficulty in maintaining sleep is one of major characteristics of insomnia (18, 33), we named cluster 1 “insomnia with long sleep duration and midawake.” The criteria for an insomnia diagnosis include both nocturnal sleep problems and daytime consequences, but the UK Biobank dataset used in this study does not have additional



**Fig. 4.** Clustering analysis revealed five clusters. (A) The flow of clustering. (B) The result of t-SNE and DBSCAN. Individual records are divided into many small clusters. (C) The result of UMAP and DBSCAN. Datasets are divided into five clusters named clusters 1 to 5. (D) Heat map of z score. The names of clusters are shown next to their main features. (E) The size of each cluster. The histogram is colored using the same colors as in (C). (F–J), *Left* represents the result of first clustering, where each individual record is colored corresponding to the heatmap of each sleep index. *Right* shows the histograms of the distribution of each sleep index. The scale of the y axis is the same among clusters and was set based on the range of histogram values. The histogram is colored with the same colors as in (C).

information about daytime complaints, which led to the focus being on only one aspect of insomnia, nocturnal sleep problems. Both clusters 2 and 3 are characterized by low ST long MN and

high ST long SD (Fig. 4 D, F, and G), while their phase SDs are different (Fig. 4 D and J). High phase SD in cluster 2 might capture an irregular feature of subjects' lifestyle. Another feature

of cluster 2 is that ST long MN varies between days (Fig. 4*G*), suggesting that the length and timing of sleep for subjects in cluster 2 vary from day to day. Meanwhile, cluster 3 shows normal phase SD (Fig. 4*J*) and high ST long SD (Fig. 4*G*), suggesting that the subjects in cluster 3 slept at almost the same time every day but that the sleep amount per day varied. Cluster 4 contains about 72.81% of total population (Fig. 4*E*), and its features are almost the same as those of whole datasets (Fig. 4*D*), suggesting that cluster 4 is the major cluster. Cluster 5 has zero ST short MN, while the other sleep indexes are normal, which means that the subjects in cluster 5 do not have any daytime sleep (Fig. 4*D* and *I* and *SI Appendix*, Fig. S5*O*).

We obtained five clusters related to insomnia or lifestyle. Interestingly, the sleep indexes in some clusters suggested that they could be divided even further. For example, the distribution of phase SD of cluster 2 was broad, raising a possibility that it contains more than two clusters with different phase SD features (Fig. 4*J*). Therefore, we repeated the same clustering process (i.e., dimension reduction and clustering) for the individual records contained in each cluster and obtained the next-layer clusters (Fig. 5*A*). By repeating this clustering process until a cluster is not divided into at least two clusters with their size equaling 20 or larger, we obtained 17 clusters (*SI Appendix*, Tables S5 and S6). However, there were strong correlations between the 17 clusters, suggesting that some of the clusters may be easier to understand if they were regrouped together (*SI Appendix*, Fig. S6*B*). Thus, we regrouped 17 clusters by applying Ward's method, an objective hierarchical clustering method, to their sleep indexes; regrouping across clusters 1 to 5 was not performed to maintain the first-layer relationship (Fig. 5*A* and *SI Appendix*, Fig. S6*B*). As a result, we obtained eight clusters, where clusters 1 and 5 were undivided, and clusters 2, 3, and 4 were divided into two subgroups (cluster 2a/b, cluster 3a/b, and cluster 4a/b) (Fig. 5*B* and *C* and *SI Appendix*, Figs. S7–S9 and Tables S7 and S8).

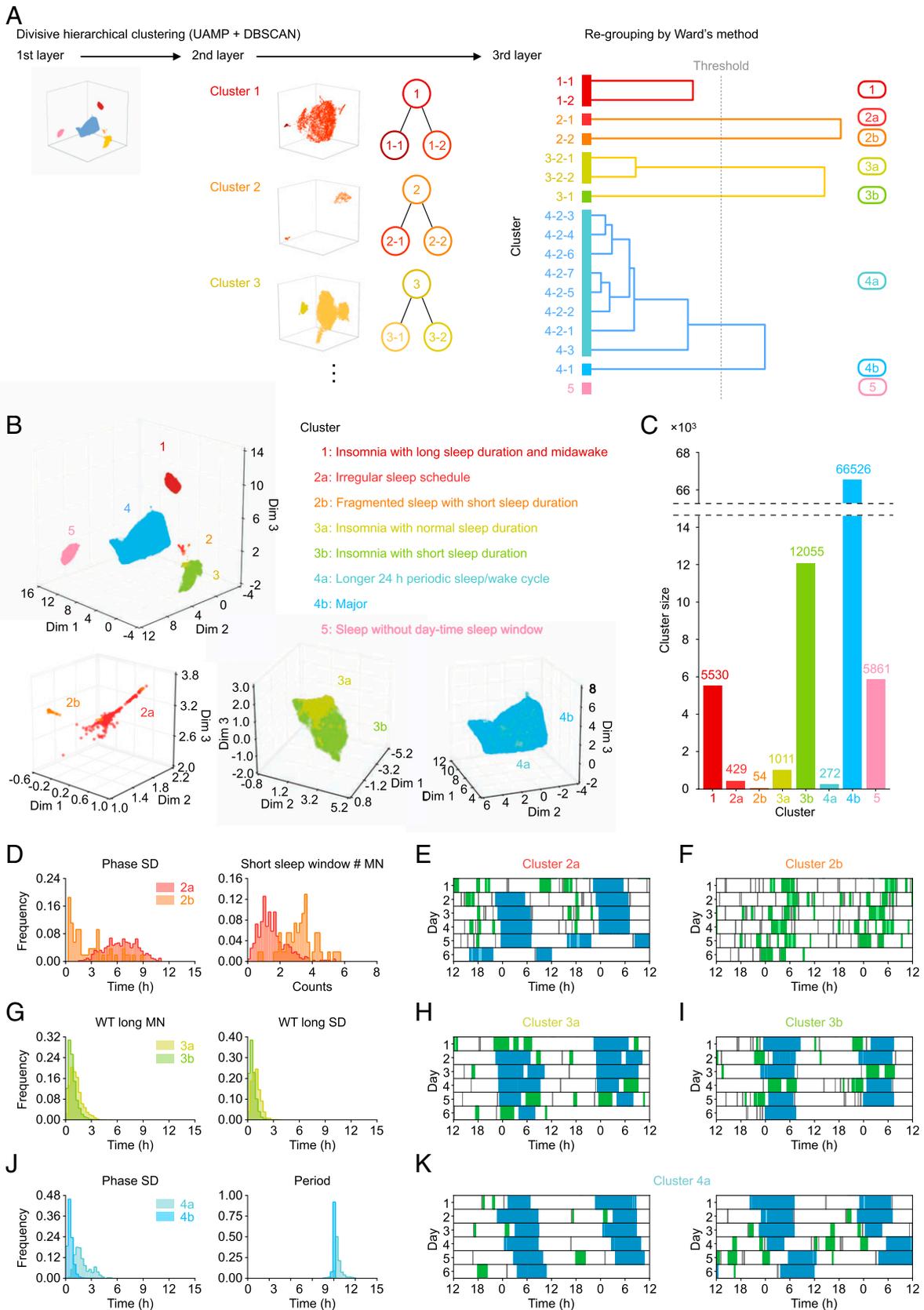
Fig. 5*B* shows the eight clusters mapping on the space of clusters 1 to 5, where clusters 2a and 2b are spatially separated but the boundaries of clusters 3a, 3b, 4a, and 4b are not obvious. Cluster 2a contains about 88.27% of individual records in cluster 2 (Figs. 4*E* and 5*C*) and shows high phase SD (Fig. 5*D*), low ST long MN (*SI Appendix*, Fig. S7*A*), and high ST long SD (*SI Appendix*, Fig. S7*B*), as does cluster 2. Therefore, we named cluster 2a “irregular sleep schedule” (Fig. 5*E* and *SI Appendix*, Fig. S8*B*). This irregular sleep phenotype may be related to the difference in sleep/wake patterns between workdays and holidays or social jet lag (13). This sleep phenotype sometimes appears in rotating shift workers who alternate daytime and night shifts (9, 41). On the other hand, cluster 2b has no long sleep window (*SI Appendix*, Fig. S7*K*) but many short sleep windows, which lead to various phases during days and low sleep percentage (Fig. 5*D* and *SI Appendix*, Fig. S7*U*). These characteristics show that the subjects in cluster 2b repeated short-term sleep (Fig. 5*F* and *SI Appendix*, Fig. S8*C*), and thus, this cluster was named “fragmented sleep with short sleep duration.” Cluster 3 was divided into clusters 3a and 3b, where almost all features of cluster 3 are preserved (e.g., low ST long MN, high ST long SD, and constant phase SD). However, cluster 3a has higher WT long MN, WT long SD, and sleep percentage than cluster 3b (Fig. 5*G* and *SI Appendix*, Fig. S7*U*), which means that there are differences in the length of midawake and sleep duration between clusters 3a and 3b (Fig. 5*H* and *I* and *SI Appendix*, Fig. S8*D* and *E*). Thus, we named clusters 3a and 3b “insomnia with normal sleep duration” and “insomnia with short sleep duration,” respectively. Cluster 4 was divided into clusters 4a and 4b, with the distribution of period varying between them (Fig. 5*J*). Cluster 4a has a

sleep/wake cycle with period apparently longer than 24 h (the mean of the period is 25.19 h), which leads to higher phase SD (Fig. 5*J*). Thus, we characterized cluster 4a as “longer 24-h periodic sleep/wake cycle,” where subjects were also identified by setting a threshold on the distribution of period (Fig. 3*F*). Notably, this apparent long period may represent a mixture of two different sleep phenotypes (Fig. 5*K*). Longer-term measurements (e.g., more than 2 wk) and more information about the subjects would allow for deeper insights. Cluster 4b is the largest group (Fig. 5*C*), and its distribution is close to that of the whole dataset, indicating that cluster 4b is the major group.

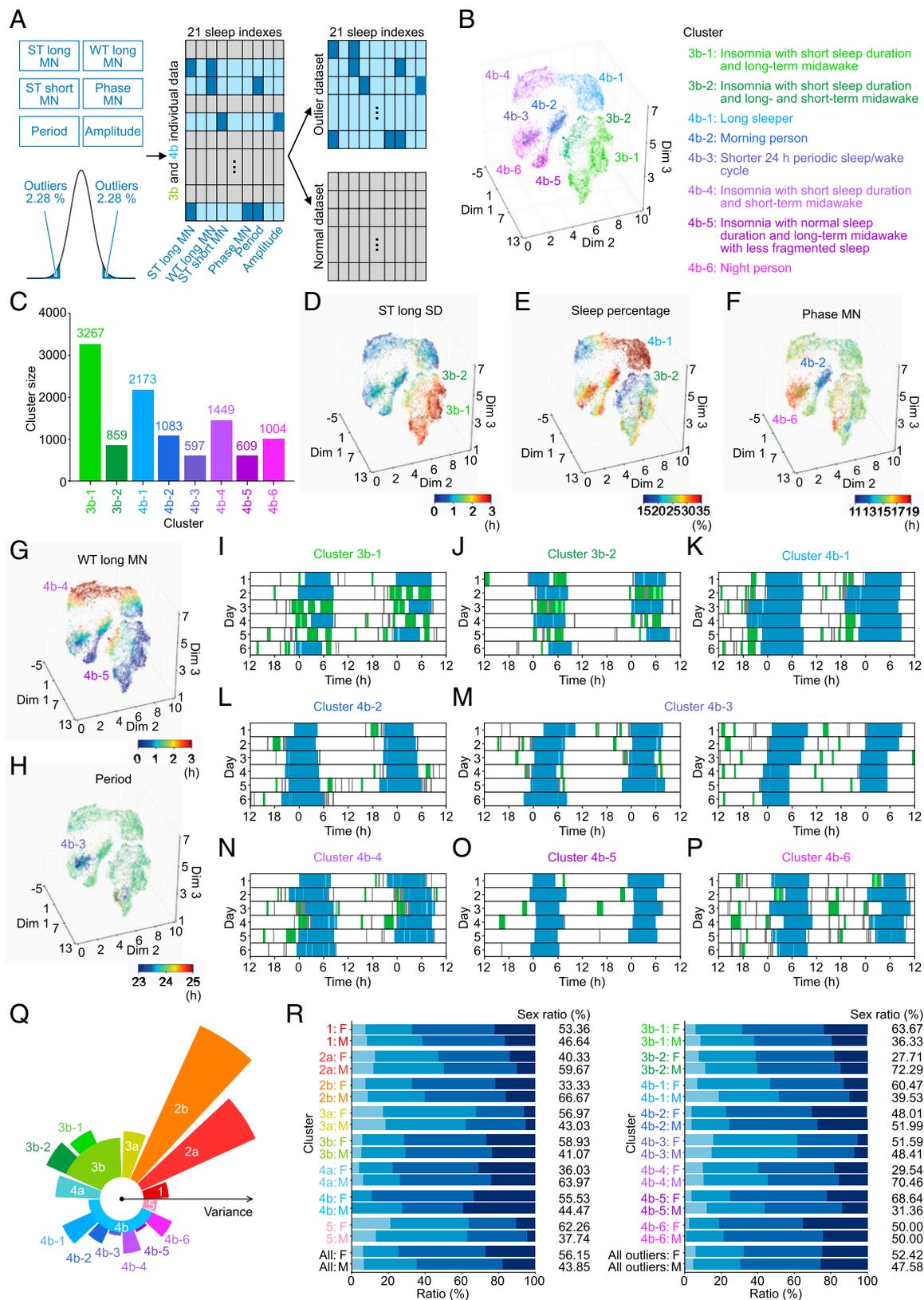
In summary, we identified eight clusters whose sleep indexes distribute smoothly (*SI Appendix*, Fig. S7), indicating that these clusters could not have been obtained by using only a threshold-based method but were revealed by using an unsupervised and unbiased clustering method like UMAP. On the other hand, groups like cluster 4a could be obtained by setting a proper threshold for the distribution of one sleep index (Fig. 3*F*), which suggests that some classifications could be revealed by using a threshold. In other words, it may be possible to understand human sleep phenotypes at a higher resolution by combining threshold-based and UMAP-based classification, especially in the understanding of sleep disorders. Although we have obtained a couple of insomnia-like clusters (Fig. 5*B*), this may be only a piece of sleep disorders, and larger clusters, like clusters 3b and 4b, may contain other small groups representing sleep disorders. Thus, if we can classify normal sleep and abnormal sleep based on some thresholds and then apply UMAP to abnormal sleep, we would obtain classifications of abnormal sleep at a higher resolution, which are expected to be related to sleep disorders.

**Outliers.** We chose six sleep indexes and set their lower or upper 2.28 percentiles as an outlier dataset in clusters 3b and 4b (Fig. 6*A*). The proportions of the outlier data in clusters 3b and 4b were 35.63 and 10.83%, respectively. These discretized datasets were further classified by using UMAP and DBSCAN, which revealed eight clusters; of these, two (clusters 3b-1 and 3b-2) were derived from cluster 3b, and the others (clusters 4b-1 to 4b-6) were derived from cluster 4b (Fig. 6*B–P* and *SI Appendix*, Figs. S10–S12 and Tables S9 and S10). We confirmed that similar clusters, clusters b and c and clusters e to j, can be obtained when the analysis is applied to the entire outlier dataset, corresponding to clusters 3b-1 and 3b-2 (clusters b and c) and clusters 4b-1 to 4b-6 (clusters e to j), respectively (*SI Appendix*, Fig. S12*E* and *F*). As shown in Fig. 6*B*, there are four more insomnia-like clusters (*SI Appendix*, Tables S11 and S12) and four more lifestyle-related clusters in these eight clusters.

Both clusters 3b-1 and 3b-2 have the insomnia-like phenotype characterized by a long sleep window divided into multiple short sleep windows just like cluster 3b (Fig. 6*I* and *J* and *SI Appendix*, Fig. S11*A* and *B*). Fig. 6*D* shows one feature of this type of insomnia-like phenotype, where clusters 3b-1 and 3b-2 have higher ST long SD (*SI Appendix*, Fig. S10*B*). Thus, clusters 3b-1 and 3b-2 can be considered as subtypes of the insomnia-like phenotype in terms of different distributions of WT long MN from cluster 3b (Fig. 6*G* and *SI Appendix*, Fig. S10*C*). The subjects in cluster 3b-1 have long-term midawake at night while having low WT long MN. These features demonstrate that the subjects in cluster 3b-1 could not fall asleep smoothly after midawake, although they sleep with normal numbers of midawake (Fig. 6*I* and *SI Appendix*, Fig. S11*A*). On the other hand, cluster 3b-2 is characterized by higher WT long MN than cluster 3b and by both long-term and short-term midawake at



**Fig. 5.** Hierarchical clustering analysis revealed eight clusters. (A) The flow of divisive hierarchical clustering. The same clustering process was repeated three times (*SI Appendix*). The 17 clusters obtained by divisive hierarchical clustering were regrouped using Ward's method and named as clusters 1, 2a, 2b, 3a, 3b, 4a, 4b, and 5. (B) *Upper* shows the result of first-layer clustering, where each individual record was colored by clusters' colors. The caption summarizes sleep phenotypes of each cluster. *Lower Left*, *Lower Center*, and *Lower Right* are the enlargement figures of clusters 2, 3, and 4, respectively. (C) The size of each cluster. Twenty-seven individual records were detected as noise by DBSCAN. (D-K) The distribution of sleep indexes of (D) clusters 2a and 2b, (G) clusters 3a and 3b, and (J) clusters 4a and 4b and representative plots of (E) cluster 2a, (F) cluster 2b, (H) cluster 3a, (I) cluster 3b, and (K) cluster 4a shown as double plots.



**Fig. 6.** Clustering analysis of the outlier dataset revealed eight clusters. (A) The flow of data selection for the outlier clustering. Blue marks the lower and upper 2.28 percentiles in six sleep indexes in *Center*. The individual records with such values colored sky blue are divided as the outlier dataset, while the remaining individual records colored gray are divided as the normal dataset. (B) The result of clustering. The outlier dataset is divided into eight clusters. (C) The size of each cluster. Four hundred fifty-eight individual records were detected as noise by DBSCAN. (D–H) The results of outlier clustering, where each individual record is colored corresponding to the heatmap of each sleep index. (I–P) Representative plots of clusters in the outlier clustering shown as double plots. (Q) The summary of whole clustering and outlier clustering. The radius of each cluster shows the L2 norm between the mean of each cluster and that of whole dataset (the black center point). (R) Sex and age proportions of whole clustering and outlier clustering.

night, which leads to insufficient sleep amount (Fig. 6J and *SI Appendix*, Figs. S10U and S11B). The phenotypes of clusters 4b-4 and 4b-5 are also subtypes of the insomnia-like phenotype, but sleep in both clusters is not fragmented as severely as in cluster 3-related clusters. The subjects in cluster 4b-4 had repeated short-term midawake at night (Fig. 6N and *SI Appendix*, Fig. S11F), which is reflected as high WT long MN (Fig. 6G) and leads to the insomnia-like phenotype with unfragmented and short sleep duration. In contrast, cluster 4b-5 is characterized by low WT long MN (Fig. 6G) and long-term midawake (Fig. 6O and *SI Appendix*, Fig. S11G), which are the same as cluster 3b-1 (Fig. 6I and *SI Appendix*, Fig. S11A). The difference between them is how fragmented their sleep is, in that sleep in cluster 4b-5 is fragmented less often.

Cluster 4b-1 is characterized by long sleep duration at night without long-term midawake, as indicated by high long sleep window length MN (*SI Appendix*, Fig. S10C and I). Sleep percentage (Fig. 6E and *SI Appendix*, Fig. S10U) is high not because of multiple sleep windows but due to one long sleep window of about 9 h (Fig. 6K and *SI Appendix*, Fig. S11C), suggesting that the phenotype of cluster 4b-1 is of a long sleeper. Hypersomnia is characterized by long sleep and daytime naps. The subjects in cluster 4b-1 only have the former characteristic, suggesting that they have no complaints about daytime naps and do not fall in the standard categorization of hypersomnia (42). Clusters 4b-2 and 4b-6 stand out in terms of phase MN, where cluster 4b-2 has lower phase MN, while cluster 4b-6 has higher phase MN than the whole dataset (Fig. 6F and *SI Appendix*, Fig. S10Q), indicating that the subjects in clusters 4b-2 and 4b-6 are morning persons and night persons, respectively (Fig. 6L and P and *SI Appendix*, Fig. S11D and H). The last cluster, cluster 4b-3, shows apparent shorter 24-h periods (Fig. 6H and *SI Appendix*, Fig. S10S) in contrast to cluster 4a. However, there are two types of sleep phenotypes (Fig. 6M) like the two types included in cluster 4a.

In summary, the large-scale sleep phenotype analysis on acceleration data from over 100,000 subjects identified 16 clusters (Fig. 6Q). We also investigated if these 16 clusters are related to the season in which they were measured or the age of the subjects. The proportions of sex, age (40s, 50s, 60s, and 70s), and measurement month in each cluster are shown in Fig. 6R and *SI Appendix*, Fig. S12G. We found that there is no obvious correlation between the cluster and the measurement season or month. Notably, the proportion of subjects in their 70s in cluster 2b is higher than that in other clusters (Fig. 6R). As age increases, WASO has been reported to gradually increase, and TST has been reported to decrease, suggesting that some elderly people might have such a phenotype (43). Therefore, the 16 clusters demonstrate various sleep phenotypes and reveal a first draft landscape of human sleep phenotypes.

## Discussion

In this study, we conducted a large-scale systematic analysis of the diversity of sleep phenotypes using about 100,000 long-term acceleration data collected by the UK Biobank (Fig. 1). The acceleration data were converted to 21 sleep indexes using a state-of-the-art sleep/wake classification algorithm and a nonwear detection algorithm (Fig. 2A). By applying UMAP and DBSCAN, we obtained 16 clusters (clusters 1, 2a, 2b, 3a, 3b, 4a, 4b, and 5 and clusters 3b-1, 3b-2, and 4b-1 to 4b-6), including insomnia-like and lifestyle-related clusters (Figs. 5B and 6B). Parts of these insomnia-like clusters were identified by focusing on the outliers

that have at least one sleep index with a value in the upper or lower 2.28 percentiles of distribution (Fig. 6B)—cluster 4b-4 represents an insomnia-like phenotype with unfragmented and short sleep duration, whereas cluster 4b-5 represents an insomnia-like phenotype with long-term midawake and less fragmented sleep. In this way, we identified several diverse sleep phenotypes and delineated a real-world sleep landscape (Fig. 6Q).

The irregular sleep schedule in cluster 2a involves a few days with delayed sleep onset compared with onset on other days, which might reflect different lifestyles on workdays and holidays. These delayed sleep schedules are also observed for rotating or fixed night shift workers, such as a sleep onset delay of about 7 h on night shift days compared with holidays and daytime workdays (9, 41), suggesting that cluster 2a could reflect the lifestyle of shift workers who work at night 1 or 2 d/wk. Most shift workers suffer from misalignments between the circadian rhythm and the required sleep schedule, which cause reduced sleep duration and low sleep quality after night work shifts (19, 44), often leading to depression and anxiety (16). Thus, quantitative detection of midawake and daytime sleep may be important in supporting the health of shift workers. Notably, shift workers represent about 16% of the working population (45), but cluster 2a accounts for only about 0.47% of the whole dataset (Fig. 5C), which may be due to the dataset used in this study being obtained over too short a period to identify shift work when the shifts are not very frequent. Measurements over a longer period (e.g., 2 wk) would be useful for more than just the identification of shift workers. Two clusters were identified based on the constant phase of sleep/wake cycle and named as “morning person” and “night person” (clusters 4b-2 and 4b-6, respectively). The question of whether the circadian clocks of these people or the phases shifted due to social constraints would be resolved by long-term measurements. If the shift is of the circadian clock, the subjects would have advanced sleep phase syndrome or delayed sleep phase syndrome, which are known to be linked to genetic factors (46, 47). Clusters 4a and 4b-3 are characterized by an apparent period that is shorter or longer than 24 h, while each of them includes two types of sleep phenotypes. In cluster 4a, for example, the sleep phenotype shown in Fig. 5K, *Left* appears to be free running, while that shown in Fig. 5K, *Right* can be considered as a difference between workdays (days 1 to 4) and holidays (days 5 and 6). In cluster 4b-3, Fig. 6M, *Left* seems to have a short period, while the apparent short period of Fig. 6M, *Right* is caused by the effects of working days; for example, days 1 to 3 were holidays followed by three continuous workdays. Long-term measurements may reveal the underlying environmental factors affecting these two phenotypes. Long circadian rhythms are observed in subjects with visual handicaps (48) or having lifestyles without sunlight exposure (49). Furthermore, using pseudo-sleep/wake time series data that included different sleep schedules per day, we verified that a 6-d measurement is not sufficient to completely distinguish between such sleep and short/long circadian rhythms (*SI Appendix*, Fig. S3F). Particularly, the period of such pseudo data tends to be underestimated, which is why such a short period cluster that is rarely observed in the light–dark cycle condition (48) appeared in this study. Of note, the pioneering large-scale study of human sleep succeeded in an extraction of several features corresponding to the diversity of human sleep phenotypes from long-term measurements (34). Winnebeck et al. (34) utilized acceleration data with more than weeks to extract time series features, which could be applied to our sleep landscape pipeline in the future. Advances in the machine learning field could drive and color the sleep landscape if we can use the sleep/wake time series data as the input instead of the sleep index. Thus, longer-term measurements and time series analyses

can better identify the sleep phenotypes of subjects, which may further improve our understanding of genetic and environmental regulation of sleep.

The sleep/wake classification algorithm used in this study reached high specificity and sensitivity, demonstrating the ability to capture short-term midawake. This advantage led to identifying insomnia-like phenotypes characterized by midawake, known as difficulty in maintaining sleep. Seven clusters related to insomnia were classified, with a wide variety of sleep duration, midawake counts, and midawake duration between them (*SI Appendix, Tables S11 and S12*). By focusing on sleep duration, these insomnia clusters could be classified into three groups—the long sleep duration (cluster 1), normal sleep duration (clusters 3a and 4b-5), and short sleep duration (other clusters) groups. Several differences between insomnia with normal sleep duration and with short sleep duration have been reported (50, 51). Interestingly, insomnia with short sleep duration has been associated with impaired neurocognitive functioning (50), while insomnia with normal sleep duration has been associated with an anxious–ruminative profile (51). Those differences demonstrate that it is important to describe subjects with insomnia in more detail based on how much they sleep. Moreover, insomnia clusters with normal or short sleep duration can be subdivided based on midawake counts. Clusters 3a and 3b and clusters 3b-1 and 3b-2, derived from cluster 3b, are characterized by fragmentation of sleep caused by more frequent midawake. Cluster 3a shows the most fragmented sleep among these clusters, but sleep duration was normal. Clusters 3b, 3b-1, and 3b-2 also show fragmented sleep, with clusters 3b-1 and 3b-2 being more fragmented. The number of awakenings is usually measured by PSG but is rarely used in studies on insomnia. Our results suggest that the number of awakenings might distinguish heterogeneous sleep phenotypes that were otherwise considered as the same phenotype (52). The comparison between clusters 3b-1 and 3b-2 demonstrates the variation of duration of each midawake. Cluster 3b-1 has 3.09 times more long-term midawake (longer than or equal to 60 min) but 1.39 times less short-term midawake (less than 60 min) than the average of all data, indicating that subjects in cluster 3b-1 could sleep deeply but could not fall asleep smoothly again once they woke up. On the other hand, cluster 3b-2 has both long-term and short-term midawake, which is why this cluster shows the shortest sleep duration among insomnia-like clusters. The phenotypes of subjects with long-term midawake (e.g., clusters 3b-1 and 3b-2) could be considered as insomnia with difficulty in both maintaining and initiating sleep (53). Both clusters 4b-4 and 4b-5 have normal midawake counts, but what this characteristic means varies between them. Cluster 4b-4 has the most amount of short-term midawake among the insomnia-like clusters. The combination of normal midawake counts and short-term midawake suggests that subjects in cluster 4b-4 wake and fall asleep repeatedly and frequently, indicating that they might not have trouble falling asleep again after midawake but have a problem in maintaining sleep. Cluster 4b-5 has normal counts of midawake and little short-term midawake, indicating that subjects in this cluster suffer from insomnia-like symptoms less often than those in other insomnia clusters. The *International Classification of Sleep Disorders*, Third Edition (18) and the *Diagnostic and Statistical Manual of Mental Disorders*, Fifth Edition (33) define insomnia as a sleep phenotype with sleep difficulty on more than three nights per week, suggesting that the subjects in cluster 4b-5 do not fully meet this criterion and could be classified as preinsomnia. In the context of insomnia, individual records for a longer period would also be useful. Taken together, seven clusters show various insomnia phenotypes, ranging from

well-known to undefined types, suggesting that quantitative, detailed, and accurate sleep analysis using accelerometers will also contribute to the classification as well as diagnosis of diseases.

We propose a pipeline for drawing a landscape of sleep phenotypes using a systematic and unbiased clustering method. By linking it with information of different modalities, such as present illness, past medical history, medications, education, occupations, lifestyle habits (e.g., alcohol intakes, smoking, diet), blood biochemistry, and genomics, the human sleep landscape could become more comprehensive and accurate. Some of this information can be obtained from the UK Biobank. The UK Biobank contains a series of diagnostic data classified following the International Classification of Diseases, 10th revision categories. Thus, it would be a promising next step of this study to analyze the relationship between the subjects in each cluster and their medical history, although the dataset in the UK Biobank is not inclusive for all subjects. For example, for subjects with a medical history of sleep apnea syndrome or other sleep-related disorders, it would be interesting to use this information for an annotation of clusters, which will expand the current human sleep landscape.

We also note that accurate and long-term sleep measurement could help us to diagnose psychiatric disorders complicated by sleep disorders because detailed sleep phenotypes of such disorders vary according to the underlying psychiatric condition (54) and may work as digital biomarkers. For example, 15% of depressed patients complain of hypersomnia, and about 70% complain of insomnia with difficulty in initiating or maintaining sleep (55). A study reported that 11 first episode and neuroleptic-naïve patients showed prolonged sleep latency but normal sleep duration, indicating that these patients had insomnia with difficulty in initiating sleep (56). Interestingly, some studies reported that patients with schizophrenia show decreased TST and significant disruption in maintaining sleep (57). As psychiatric disorders, including schizophrenia, are understood as a spectrum, it may be possible to segregate psychiatric disorders at a higher resolution by classifying them based on the symptoms of complicating sleep disorders. Moreover, medication strategies depending on complicating sleep disorders may lead to better pharmacologic treatments of psychiatric disorders. In this context, utilizing the rich dataset stored in the UK Biobank including the diagnostic results of mental disease, will enable a pilot study for investigating the relationship between the human sleep landscape and psychiatric disorders.

## Materials and Methods

**Data Acquisition.** Thirty-six simultaneously measured PSG and Activity data were obtained from healthy human volunteers at the monitoring facility in the University of Tokyo and used to train and validate the sleep/wake classification algorithm. To evaluate the accuracy of the nonwear detection algorithm, we collected Activity data of 20 subjects (267 d of data in total). Wearing and nonwearing periods were judged using the time stamp records when the subjects were not wearing the device.

**The UK Biobank Dataset.** Activity data (index [ID]: 90001) recording along with information regarding sex (index [ID]: 31), year of birth (ID: 34), and month of birth (ID: 52) were downloaded from the UK Biobank. Age was calculated by taking the difference between the first day of the subject's birth month and the first day of acceleration data.

**Data Availability.** Acceleration data used for the large-scale analysis in this study were previously deposited in the UK Biobank (project ID: 48357).

**ACKNOWLEDGMENTS.** We thank K. Mitsui, T. Mano, T. Yamada, and S. Honda for discussions regarding data analysis; F. Tatsuki and T. Katsumata for discussions on sleep-related disorders; and K. Kuwajima and T. Asanuma for helping with the PSG measurements. This work was supported by Young Scientist Grants 19K16487 (to S.S.) and 21K15136 (to S.S.), Transformative Research Areas A Grants 20H05894 (to S.S.) and 20H05903 (to S.S.), Japan Science and Technology Agency (JST) Moonshot R and D–Multifaceted investigation challenge for new normal initiatives (MILLENNIA) Program Grant JPMJMS2023-25 (to S.S.), Japan Agency for Medical Research and Development–Core Research for Evolutional Science and Technology Grant JP17gm0610006 (to H.R.U.), Brain Mapping by Integrated Neurotechnologies for Disease Studies Grant JP17DM0207049 (to H.R.U.), Basic Science and Platform Technology Program for Innovative Biological Medicine Grant JP17AM0301025 (to H.R.U.), KAKENHI Grant-in-Aid from Japan Society for the Promotion of Science (JSPS) Scientific

Research S Grant 18H05270 (to H.R.U.), Human Frontier Science Program Research Grant Program Grant RGP0019/2018 (to H.R.U.), JST Exploratory Research for Advanced Technology Grant JPMJER1904 (to H.R.U.), and an intramural grant-in-aid from The Institute of Physical and Chemical Research (Japan) Center for Biosystems Dynamics Research (to H.R.U.). This work was also supported by a collaborative research project between Fukuda Lifetech Inc. and the University of Tokyo (H.R.U.).

Author affiliations: <sup>a</sup>Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-0033, Japan; <sup>b</sup>Department of Systems Pharmacology, Graduate School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan; <sup>c</sup>Laboratory for Synthetic Biology, RIKEN Center for Biosystems Dynamics Research, Osaka 565-5241, Japan; and <sup>d</sup>Sleep Center, Toranomon Hospital, Tokyo 105-8470, Japan

1. E. S. Lander *et al.*, Correction: Initial sequencing and analysis of the human genome. *Nature* **412**, 565 (2001).
2. J. C. Venter *et al.*, The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. M. Margulies *et al.*, Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
4. J. Shendure *et al.*, Accurate multiplex polony sequencing of an evolved bacterial genome. *Science* **309**, 1728–1732 (2005).
5. B. Vogelstein *et al.*, Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
6. C. Van Cauwenbergh, C. Van Broeckhoven, K. Sleegers, The genetic landscape of Alzheimer disease: Clinical implications and perspectives. *Genet. Med.* **18**, 421–430 (2016).
7. C. Bycroft *et al.*, The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
8. T. M. McMenamin, A time to work: Recent trends in shift work and flexible schedules. *Mon. Labor Rev.* **130**, 3–15 (2007).
9. D. Fischer, C. Vetter, C. Oberlinner, S. Wegener, T. Roenneberg, A unique, fast-forwards rotating schedule with 12-h long shifts prevents chronic sleep debt. *Chronobiol. Int.* **33**, 98–107 (2016).
10. J. Geiger-Brown, A. Trinkoff, V. E. Rogers, The impact of work schedules, home, and work demands on self-reported sleep in registered nurses. *J. Occup. Environ. Med.* **53**, 303–307 (2011).
11. J. Geiger-Brown *et al.*, Sleep, sleepiness, fatigue, and performance of 12-hour-shift nurses. *Chronobiol. Int.* **29**, 211–219 (2012).
12. I. W. Savvig, S. Pallesen, A. Wilhelmsen-Langeland, H. Molde, B. Bjorvatn, Prevalence and correlates of delayed sleep phase in high school students. *Sleep Med.* **13**, 193–199 (2012).
13. M. Wittmann, J. Dinich, M. Meroow, T. Roenneberg, Social jetlag: Misalignment of biological and social time. *Chronobiol. Int.* **23**, 497–509 (2006).
14. T. Roenneberg, K. V. Allebrandt, M. Meroow, C. Vetter, Social jetlag and obesity. *Curr. Biol.* **22**, 939–943 (2012).
15. B. M. Hittler, G. L. Gillespie, Identifying shift worker chronotype: Implications for health. *Ind. Health* **56**, 512–523 (2018).
16. C. L. Drake, T. Roehrs, G. Richardson, J. K. Walsh, T. Roth, Shift work sleep disorder: Prevalence and consequences beyond that of symptomatic day workers. *Sleep* **27**, 1453–1462 (2004).
17. M. M. Ohayon, Prevalence of DSM-IV diagnostic criteria of insomnia: Distinguishing insomnia related to mental disorders from sleep disorders. *J. Psychiatr. Res.* **31**, 333–346 (1997).
18. AAOs Medicine, *International Classification of Sleep Disorders* (American Academy of Sleep Medicine, ed. 3, 2014).
19. M. M. Ohayon, Epidemiology of insomnia: What we know and what we still need to learn. *Sleep Med. Rev.* **6**, 97–111 (2002).
20. D. J. Buysse, C. F. Reynolds III, T. H. Monk, S. R. Berman, D. J. Kupfer, The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Res.* **28**, 193–213 (1989).
21. C. H. Bastien, A. Vallières, C. M. Morin, Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med.* **2**, 297–307 (2001).
22. A. Kosmadopoulos, C. Sargent, D. Darwent, X. Zhou, G. D. Roach, Alternatives to polysomnography (PSG): A validation of wrist actigraphy and a partial-PSG system. *Behav. Res. Methods* **46**, 1032–1041 (2014).
23. M. de Zambotti *et al.*, Measures of sleep and cardiac functioning during sleep using a multi-sensory commercially-available wristband in adolescents. *Physiol. Behav.* **158**, 143–149 (2016).
24. O. Walch, Y. Huang, D. Forger, C. Goldstein, Sleep stage prediction with raw acceleration and photoplethysmography heart rate data derived from a consumer wearable device. *Sleep (Basel)* **42**, 1–19 (2019).
25. H. S. Dashti *et al.*, Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported by accelerometer-derived estimates. *Nat. Commun.* **10**, 1100 (2019).
26. S. E. Jones *et al.*, Genetic studies of accelerometer-based sleep measures yield new insights into human sleep behaviour. *Nat. Commun.* **10**, 1585 (2019).
27. C. Sudlow *et al.*, UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
28. K. L. Ode *et al.*, A jerk-based algorithm ACCEL for the accurate classification of sleep-wake states from arm acceleration. *iScience* **25**, 103727 (2022).
29. E. Martin, K. Hans-Peter, S. Jörg, X. Xiaowei, A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD* **96**, 226–231 (1996).
30. L. McInnes, J. Healy, J. Melville, UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv [Preprint] (2018). <https://arxiv.org/abs/1802.03426v2> (Accessed 14 October 2019).
31. V. T. van Hees *et al.*, Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One* **8**, e61691 (2013).
32. J. Van den Bulck, Television viewing, computer game playing, and Internet use and self-reported time to bed and time out of bed in secondary-school children. *Sleep* **27**, 101–104 (2004).
33. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, Washington, DC, ed. 5, 2013).
34. E. C. Winnebeck, D. Fischer, T. Leise, T. Roenneberg, Dynamics and ultradian structure of human sleep in real life. *Curr. Biol.* **28**, 49–59.e5 (2018).
35. V. T. van Hees *et al.*, Estimating sleep parameters using an accelerometer without sleep diary. *Sci. Rep.* **8**, 12975 (2018).
36. P. G. Sokolove, W. N. Bushell, The chi square periodogram: Its utility for analysis of circadian rhythms. *J. Theor. Biol.* **72**, 131–160 (1978).
37. G. A. Sunagawa *et al.*, Mammalian reverse genetics without crossing reveals Nr3a as a short-sleeper gene. *Cell Rep.* **14**, 662–677 (2016).
38. D. B. Forger, M. E. Jewett, R. E. Kronauer, A simpler model of the human circadian pacemaker. *J. Biol. Rhythms* **14**, 532–537 (1999).
39. R. E. Kronauer, D. B. Forger, M. E. Jewett, Quantifying human circadian pacemaker response to brief, extended, and repeated light stimuli over the photopic range. *J. Biol. Rhythms* **14**, 500–515 (1999).
40. T. Blackwell, S. Ancoli-Israel, S. Redline, K. L. Stone, Osteoporotic Fractures in Men (MrOS) Study Group, Factors that may influence the classification of sleep-wake by wrist actigraphy: The MrOS Sleep Study. *J. Clin. Sleep Med.* **7**, 357–367 (2011).
41. M. Juda, C. Vetter, T. Roenneberg, Chronotype modulates sleep duration, sleep quality, and social jet lag in shift-workers. *J. Biol. Rhythms* **28**, 141–151 (2013).
42. C. Vermet, I. Arnulf, Idiopathic hypersomnia with and without long sleep time: A controlled series of 75 patients. *Sleep* **32**, 753–759 (2009).
43. W. Moraes *et al.*, Effects of aging on sleep structure throughout adulthood: A population-based study. *Sleep Med.* **15**, 401–409 (2014).
44. C. Vetter, D. Fischer, J. L. Matera, T. Roenneberg, Aligning work and circadian time in shift workers improves sleep and reduces circadian disruption. *Curr. Biol.* **25**, 907–911 (2015).
45. S. M. W. Rajaratnam, M. E. Howard, R. R. Grunstein, Sleep loss and circadian disruption in shift work: Health burden and management. *Med. J. Aust.* **199**, S11–S15 (2013).
46. Y. Xu *et al.*, Functional consequences of a CK1delta mutation causing familial advanced sleep phase syndrome. *Nature* **434**, 640–644 (2005).
47. T. Ebisawa *et al.*, Association of structural polymorphisms in the human period3 gene with delayed sleep phase syndrome. *EMBO Rep.* **2**, 342–346 (2001).
48. J. S. Emens, A. J. Lewy, B. J. Lefler, R. L. Sack, Relative coordination to unknown “weak zeitgebers” in free-running blind individuals. *J. Biol. Rhythms* **20**, 159–167 (2005).
49. Anonymous, London air-raid shelters revisited. *BMJ* **2**, 414–415 (1941).
50. J. Fernandez-Mendoza *et al.*, Insomnia with objective short sleep duration is associated with deficits in neuropsychological performance: A general population study. *Sleep* **33**, 459–465 (2010).
51. J. Fernandez-Mendoza *et al.*, Sleep misperception and chronic insomnia in the general population: Role of objective sleep duration and psychological profiles. *Psychosom. Med.* **73**, 88–97 (2011).
52. J. D. Edinger *et al.*, American Academy of Sleep Medicine Work Group, Derivation of research diagnostic criteria for insomnia: Report of an American Academy of Sleep Medicine Work Group. *Sleep* **27**, 1567–1596 (2004).
53. V. Pillai, T. Roth, C. L. Drake, The nature of stable insomnia phenotypes. *Sleep (Basel)* **38**, 127–138 (2015).
54. V. C. Abad, C. Guilleminault, Sleep and psychiatry. *Dialogues Clin. Neurosci.* **7**, 291–303 (2005).
55. J. Rodin, G. McAvay, C. Timko, A longitudinal study of depressed mood and sleep disturbances in elderly adults. *J. Gerontol.* **43**, 45–53 (1988).
56. J. M. Monti, D. Monti, Sleep in schizophrenia patients and the effects of antipsychotic drugs. *Sleep Med. Rev.* **8**, 133–148 (2004).
57. D. J. Kupfer, T. P. Detre, G. Foster, G. J. Tucker, J. Delgado, The application of Delgado’s telemetric mobility recorder for human studies. *Behav. Biol.* **7**, 585–590 (1972).

1

## 2 **Supplementary Information for**

3 **The 103,200-arm acceleration dataset in the UK Biobank revealed a landscape of human**  
4 **sleep phenotypes**

5 **Machiko Katori, Shoi Shi, Koji L. Ode, Yasuhiro Tomita and Hiroki R. Ueda**

6 **Hiroki R. Ueda**

7 **E-mail: uedah-ky@umin.ac.jp**

### 8 **This PDF file includes:**

9     Supplementary text

10    Figs. S1 to S12 (not allowed for Brief Reports)

11    Tables S1 to S12 (not allowed for Brief Reports)

12    SI References

## 13 Supporting Information Text

### 14 Materials and Methods

15 **Data acquisition.** In this study, PSG data and Axivity (AX3, a wristband-type triaxial accelerometer, Axivity Ltd., Hoults  
16 Yard, UK) triaxial acceleration data were obtained in the University of Tokyo. Thirty-six PSG data and triaxial accelerometer  
17 data of one-night measurement were used to train and validate the sleep/wake classification algorithm. To evaluate the accuracy  
18 of nonwear detection algorithm, we collected Axivity data of 20 subjects (267 days data in total). Wearing and nonwearing  
19 periods were judged using the timestamp records when the subjects were not wearing the device.

20 Participants (cumulative counts) who reported no present and past diagnosis of any sleep disorders, mental illnesses, and  
21 neurodegenerative diseases were recruited. No participants declared taking medicines that may affect sleep, such as cold  
22 medicine. More than two measurements per participant were allowed. The study protocols were approved by the Research  
23 Ethics Committee, Graduate School of Medicine, the University of Tokyo (No. 2018174NI and No. 11281). Informed consent  
24 was obtained from every subject for each measurement. More information about the subjects has been provided in Table S1,  
25 Table S2, and Table S3.

26 PSG data were recorded using a wireless portable PSG system (SOMNOscreen plus, SOMNOmedics GmbH, Germany)  
27 at the monitoring facility in the University of Tokyo. Sleep staging for every 30 seconds with PSG records was conducted  
28 manually by trained experts (Fukuda Denshi Co. Ltd, Japan). The sampling rate and dynamic range of Axivity were set to 50  
29 or 100 Hz and  $\pm 8$  g at 10-bit resolution, respectively.

30 **Sleep/wake classification algorithm.** The sleep/wake classification algorithm used in this study is based on a machine learning-  
31 based algorithm created for a triaxial accelerometer (ACCEL) (1). To validate and optimize the parameters of original  
32 algorithm to the Axivity data, 27 PSG and Axivity data were used. The original algorithm uses the power spectrum of  
33 jerk as input data (Fig. S1A) and XGBoost with six hyperparameters: *learning\_rate*, *gamma*, *colsample\_bytree*, *subsample*,  
34 *max\_depth*, and *min\_child\_weight*. These parameters were optimized by using Bayesian optimization (Fig. S1D). Each  
35 parameter was allowed to have values within the range, [0, 1], [0, 5], [0.01, 1], [0.01, 1], [1, 30], and [1, 30], respectively.  
36 The parameter set of six hyperparameters was evaluated by leave-one-out cross-validation, where 26 data were used for  
37 training XGBoost with a parameter set, and 1 data was used for validation. The performance of algorithm was evaluated  
38 based on the summation of accuracy and F measure. We used Bayesian optimization and iterated 2,000 times to tune the  
39 algorithm. F measure was calculated as  $2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$ , where wake was calculated as true and  
40 sleep as false. The optimal parameters were as follows: *learning\_rate* = 0.07, *gamma* = 4.88, *colsample\_bytree* = 0.86,  
41 *subsample* = 0.93, *max\_depth* = 23, and *min\_child\_weight* = 5. The program for Bayesian optimization was adopted from  
42 <https://github.com/fmfn/BayesianOptimization>.

43 The performance of sleep/wake classification algorithm was calculated by leave-one-out cross-validation. The predicted  
44 sleep/wake time series data were compared to PSG-based sleep/wake time series data (ground truth) epoch to epoch, and the  
45 average scores are 93.24% (accuracy), 85.69% (F measure), 97.20% (sensitivity), 82.19% (specificity), and 81.03% (Cohen's  
46 kappa). Sensitivity and specificity represent the performance in terms of sleep and wake detection, respectively. The performance  
47 of sleep/wake classification algorithm was also evaluated for WASO and TST. We calculated TST in predicted sleep/wake time  
48 series data by calculating the total sleep time between sleep onset and sleep offset, where the first sleep epoch that sustained  
49 more than 15 minutes was defined as the sleep onset and the last sleep epoch as sleep offset. The performance of TST/WASO  
50 calculation under varying thresholds of sleep onset from 0 to 30 minutes with 2.5-minute increments is shown in Fig. S2A-D.  
51 WASO was calculated as the length of wake duration between sleep onset and sleep offset.

52 **Nonwear detection algorithm.** Nonwearing period was predicted based on the standard deviation and the value range of  
53 each accelerometer axis calculated for every 60-minute block shifted by 15 minutes (2). This algorithm detects a block as a  
54 nonwearing period if the standard deviation is less than the threshold in at least two axes or if the range is less than the  
55 threshold in at least two axes. Twenty triaxial accelerometer data (267 days in total) were acquired for validation of nonwear  
56 detection (Table S2). For these measurements, the participants were asked to keep the records of the time when they were not  
57 wearing the device, which was used as the ground truth. The standard deviation and range of each axis were calculated for  
58 every 60-minute period. The process to make blocks for determining thresholds was referred to that in the previous study (2).  
59 The threshold values proposed in the previous study (2) were adapted in this study (Fig. S1E, F).

60 **Verification of algorithms and thresholds for UK Biobank analysis.** In order to evaluate the effect of difference in sampling  
61 frequency of Axivity on the performance of sleep/wake classification, triaxial acceleration data sampled at 100 Hz and PSG  
62 data were obtained (n = 12, Table S3). To compare the performances, the 100 Hz triaxial acceleration data were converted to  
63 50 Hz before being input into the sleep/wake classification algorithm trained by 50 Hz data (Fig. S1H). The method converting  
64 100 Hz data to 50 Hz data was used in the analyses of the UK Biobank dataset (Fig. S1H).

65 To evaluate the performance of sleep episode detection, epochs were randomly sampled from the triaxial acceleration data  
66 and connected to create a pseudo sleep/wake time series data (Fig. S3A). To eliminate bias due to epochal selection, 1,000  
67 validations were performed on each of the two data groups. Pseudo sleep/wake time series data containing 10 hours of sleep  
68 or wake were created, and the pseudo wake or sleep data were inserted in between, respectively. The length of the inserted  
69 data was varied from 0 to 100 minutes at a 1-minute step. Sensitivity when sleep pulse pseudo data were classified by the  
70 sleep/wake classification algorithm is shown in Fig. S3B. Specificity when wake pulse pseudo data were classified is shown in

71 Fig. S3C. Both scores reach 90% the first time when the threshold is 10 minutes when the threshold is changed from 0 minute  
72 to 100 minutes by a 5-minute step. These results demonstrate that epochs over 10 minutes, which are continuously classified as  
73 sleep or wake, are sufficiently reliable as characteristics of sleep patterns.

74 **UK Biobank dataset.** Axivity triaxial acceleration data (ID: 90001) recording as up to seven-day continuous measurements,  
75 along with information regarding sex (ID: 31), year of birth (ID: 34), and month of birth (ID: 52), were downloaded from the  
76 UK Biobank. Age was calculated by taking the difference between the first day of the subject’s birth month and the first day  
77 of acceleration data. In this analysis, one day was defined as the sleep/wake time series data from a noon to the next noon and  
78 named noon-to-noon data. Individual records from 103,200 subjects were analyzed in this study.

79 The triaxial acceleration data were converted into sleep/wake time series data as follows. The nonwear detection algorithm  
80 was applied to the all dataset. The noon-to-noon data containing less than five hours of nonwearing period were considered  
81 as qualified noon-to-noon data. This criterion was decided by referencing previous studies (3–5). The maximum counts of  
82 continuous qualified noon-to-noon data ranged from zero to six as shown in Fig. S4A, and individual records with two continuous  
83 noon-to-noon data showed the lowest ratio. We selected subjects with more than three continuous qualified noon-to-noon data  
84 and used their records (91,765 individual records) for the analysis of sleep index (Fig. 3A).

85 **Extraction of common sleep indexes.** Due to the low sensitivity and specificity of determining wake episodes and sleep episodes  
86 shorter than 10 minutes (see also **Verification of algorithms and thresholds for UK Biobank analysis**), sleep episodes shorter  
87 than 10 minutes were converted to wake episodes and vice versa. If the length of the time gaps between sleep periods was less  
88 than 60 minutes, we connected them as a sleep window (4). According to the length of each sleep window, we named them as  
89 a short sleep window or a long sleep window. The threshold between the short and long sleep windows was determined by  
90 plotting the distribution of length of sleep window and fitting the distribution with the exponential curve and the Gaussian  
91 curve (Fig. S4B). The intersection point of two curves, 3.75 hours, was determined as the threshold. Then, non-qualified  
92 noon-to-noon data were removed and 21 sleep indexes were calculated for qualified noon-to-noon data (522,826 days). ST and  
93 WT within each long or short sleep window were measured as ST long mean, WT long mean, ST short mean, and WT short  
94 mean in each qualified noon-to-noon data. Each type of value between multiple noon-to-noon data was summarized as the  
95 mean (MN) and the standard deviation (SD) within each individual record. Each window was considered to belong to the day  
96 with the middle of the window to avoid double counts. Finally, sleep percentage was determined as the percentage of sleep  
97 amount during each individual record.

98 **Extraction of rhythm-related sleep indexes.** **Period:** The chi-square periodogram (6) was applied to the sleep/wake time series  
99 data of qualified noon-to-noon data in the range 5 to 35 hours with a 0.1-hour step. The significance level of chi-square  
100 periodogram was determined as 0.01, as on the reference studies (6, 7). The maximum significant point was determined as  
101 period of sleep/wake oscillation (Fig. 2D). If there was no significant peak, period was set to zero. **Amplitude:** Sleep/wake time  
102 series data were divided into 10-minute blocks. The wake duration was calculated in every block without nonwearing periods.  
103 The standard deviation and mean of the wake duration between blocks were calculated, and a coefficient of variation (standard  
104 deviation/mean) was defined as amplitude as previously described (8) (Fig. 2E). **Phase:** Kronauer’s model based on the van  
105 der Pol limit cycle was used to extract sleep-wake cycles phases. Sleep/wake time series data were converted to binary data  
106 (sleep as 0 and wake as 1), where the nonwearing period was considered as missing value. In order to fit the limit cycle curve to  
107 the binary data, the limit cycle oscillating at 0~1 was shifted by 30 seconds, and the mean square error from the binary data  
108 was calculated to determine the best-fitted curve. The duration between the minimum point of the best-fitted curve and the  
109 last noon was determined as phase of each day. The fitting of phase is calculated for a 24-hour period, regardless of the value  
110 of period estimated in sleep-index analysis. Phase MN and SD were calculated to summarize phase features (Fig. 2F).

111 **Finding the function form to fit each sleep index.** To find the function form that fits the distribution of each sleep index,  
112 we tested all functions implemented in `scipy.stats` to the distribution of sleep indexes except for the number-related sleep  
113 indexes (long sleep window # and short sleep window #) and SD features (Fig. 3B-G, Fig. S4G, I, M, R). The ten curves,  
114 “rv\_continuous”, “rv\_histogram”, “levy\_stable”, “ncx2”, “crystalball”, “geninvgauss”, “vonmises”, “gausshyper”, “chi”, and  
115 “powerlaw”, were excluded due to errors during calculation. Each curve was fitted using maximum likelihood estimation for  
116 each sleep index, and the curve with the lowest Akaike information criterion score was selected as the best-fitted curve for each  
117 sleep index. Details about the fitting curves are available at <https://docs.scipy.org/doc/scipy/reference/stats.html>.

118 **Clustering process for whole dataset and outlier dataset.** The clustering processes consisted of three steps: z score normaliza-  
119 tion, dimension reduction, and clustering (Fig. 4A). Z score normalization was performed on each sleep index to normalize the  
120 mean and standard deviation of each sleep index to zero and one, respectively. We tested four dimension reduction methods:  
121 PCA, t-SNE, UMAP (9), and a combination of PCA and UMAP (Fig. 4B, C, Fig. S5A-D). In the first three approaches, the  
122 21-dimensional z score data were compressed to three-dimensional data. In the last approach, the 21-dimensional z score data  
123 were converted to 21 principal components (PCs) by PCA, and then  $n$  PCs were further compressed to three-dimensional data  
124 by UMAP (Fig. S5D). DBSCAN was used as the clustering method.

125 To perform hierarchical clustering, the individual records within a cluster were applied to the clustering process. Z score  
126 normalization was conducted in every clustering process to normalize the target individual records. The repetition of this  
127 clustering process continued as long as at least two of the newly generated clusters had a size equaling 20 or larger.

128 The mean of each sleep index was calculated in each identified cluster (Fig. S6B). The mean was regarded as the characteristic  
129 of each cluster. We divided clusters obtained in hierarchical clustering into five groups depending on Clusters 1 to 5 generated  
130 by the first clustering process. Ward's method was used to measure the distances among clusters within their mother cluster  
131 (i.e. Clusters 1 to 5), where the weight was determined by the each cluster size (Fig. 5A). The distribution of distances  
132 among clusters was represented as a histogram ranging from the minimum distance to maximum distance. The bin size was  
133 determined by increasing the number of bins until the gap point, a bin with zero value, occurred. The clusters were merged if  
134 their distances were lower than the median value of gap (Fig. 5A).

135 An outlier dataset was selected by considering six sleep indexes: ST long MN, WT long MN, ST short MN, phase MN,  
136 period, and amplitude (Fig. 6A). Individual records with abnormal values (upper or lower 2.28 percentiles) in at least one of  
137 six sleep indexes were selected as an outlier dataset from Clusters 3b and 4b (Fig. 6B) or all clusters (Fig. S12E, F).

138 **Hyperparameter setting in clustering.** Default parameters in the UMAP package were used for all analyses. This package  
139 is available at <https://umap-learn.readthedocs.io/en/latest/index.html>. Among the hyperparameters,  $n\_neighbors$ , (default: 15)  
140 and  $min\_dist$  (default: 0.1) manage structures in a high- or low-dimensional field. We tested different values of those two  
141 parameters with the whole data (Fig. S6A).

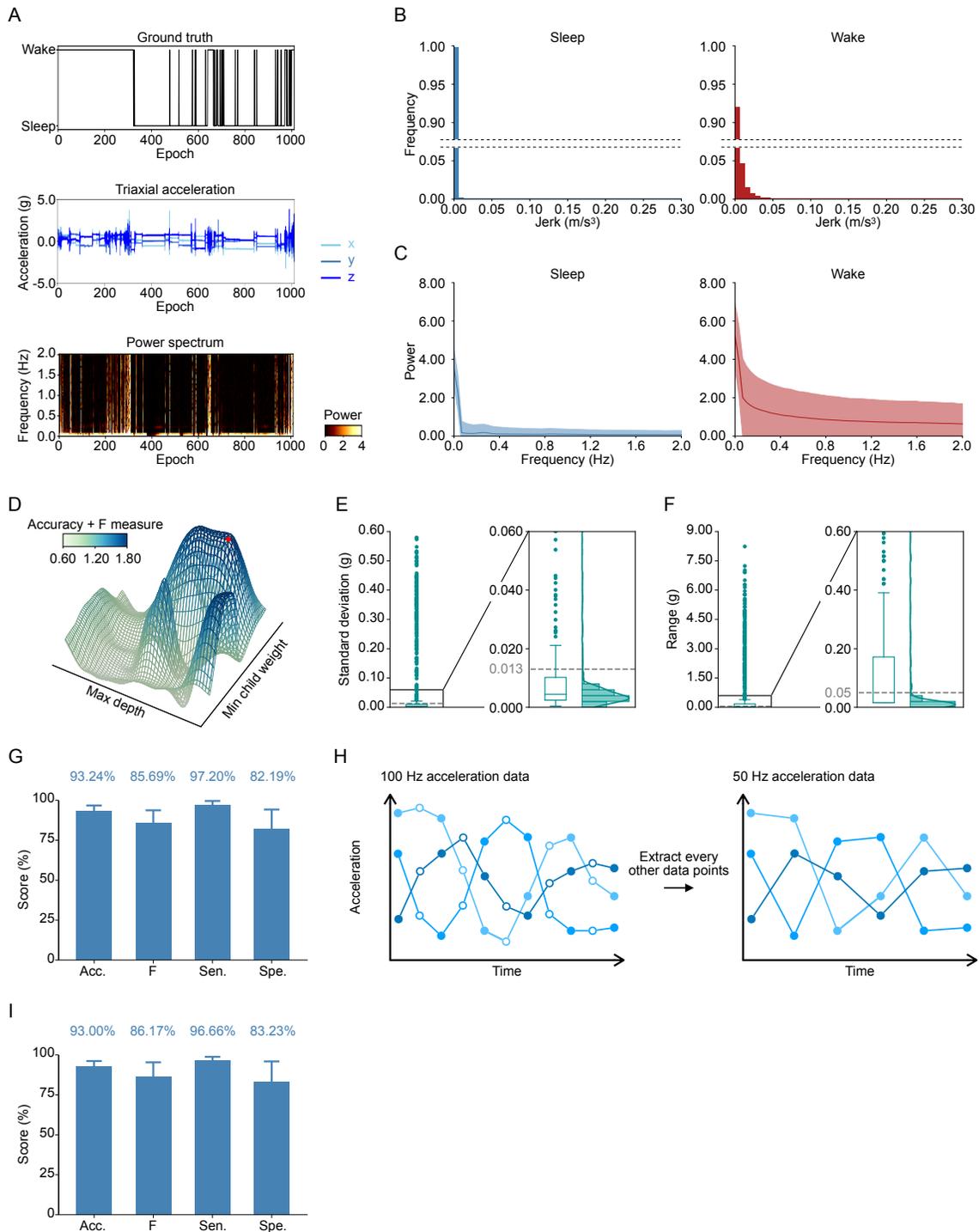
142 The DBSCAN package implemented in sklearn was used with the default values in all parameters except for  $min\_samples$ ,  
143 which is the parameter to form a dense region (10, 11). The default value, 5, was used in the whole clustering, and 99 or 115  
144 were used in the outlier clustering of Clusters 3b and 4b or of all clusters.

145 **Representative plot.** Two types of representative plot were shown in this study. The first type focused on one sleep index.  
146 Individual record with abnormal values (upper or lower 2.28 percentiles) in the focused sleep index were selected as candidates.  
147 Euclidean distances from the mean of the whole dataset were calculated using 21 sleep indexes, and the individual record that  
148 was closest to the mean was shown as the representative plot (Fig. 3B-G). The other type was the representative plot in each  
149 cluster. Distances from the mean of cluster data were calculated, and representative plots were displayed in ascending order of  
150 distance (Fig. 5E, F, H, I, K, Fig. 6I-P, Fig. S8, Fig. S11).

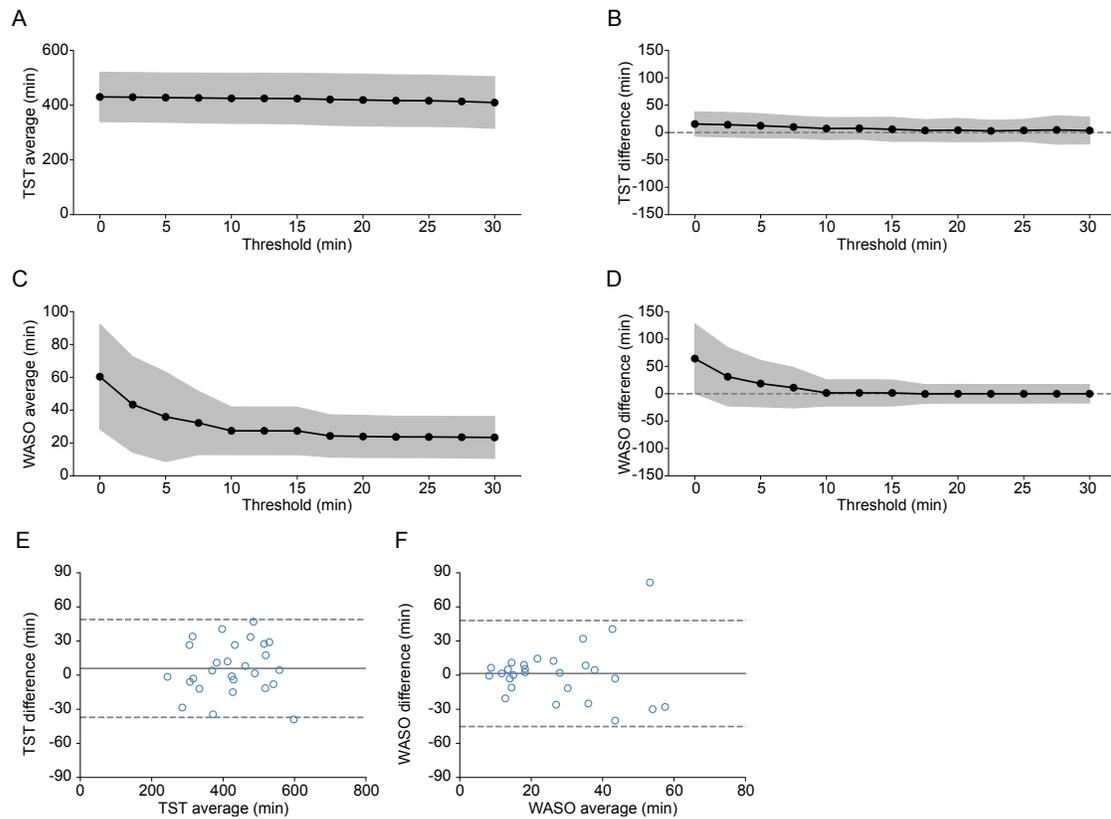
151 **Analysis of age and sex information.** The sex percentage was calculated for each cluster, the whole dataset, and the outlier  
152 dataset from Clusters 3b and 4b. Age was categorized into four groups, the 40s, 50s, 60s, and 70s.

## 153 References

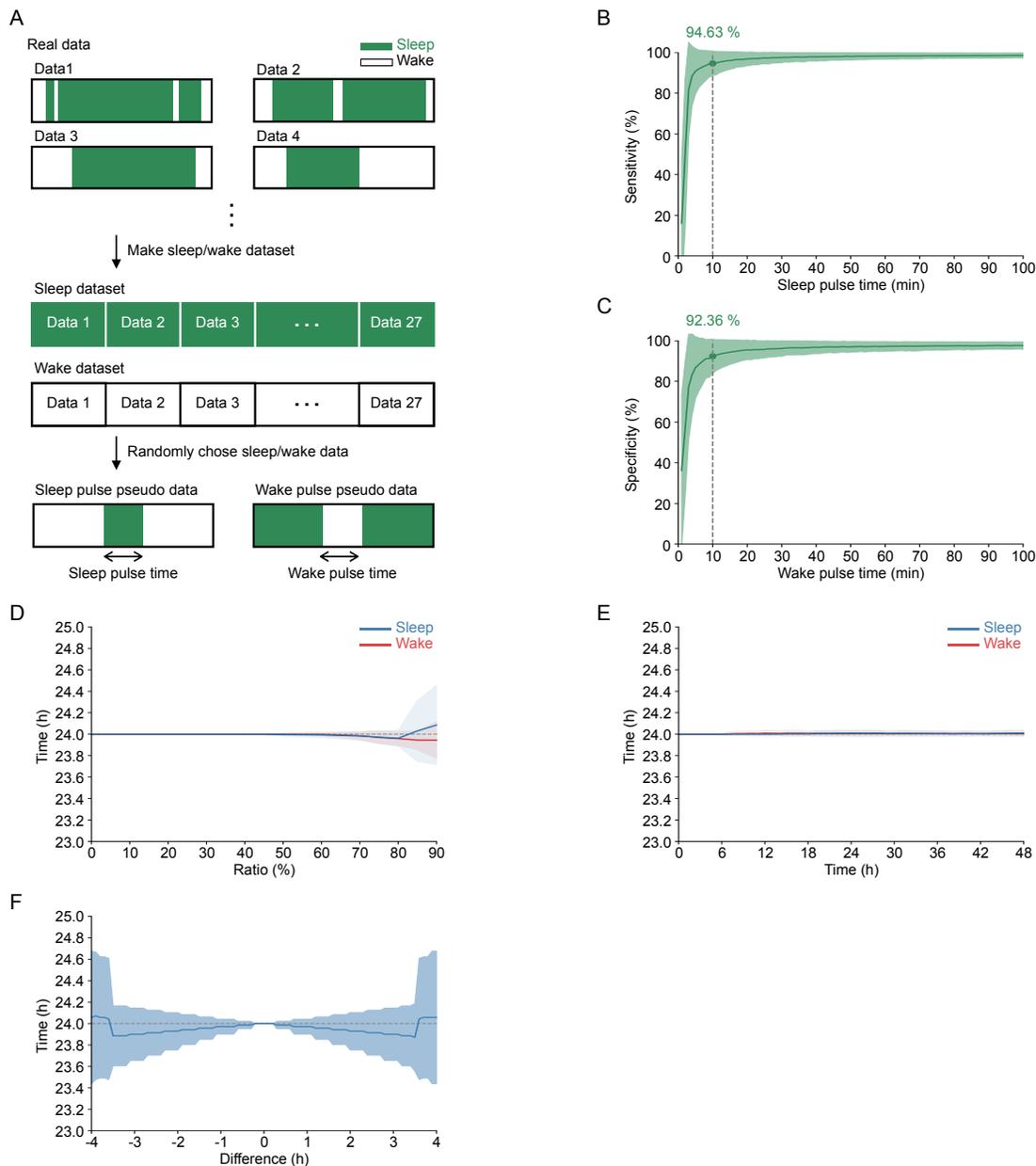
- 154 1. KL Ode, et al., A jerk-based algorithm ACCEL for the accurate classification of sleep-wake states from arm acceleration.  
155 *iScience* **25**, 103727 (2022).
- 156 2. VT van Hees, et al., Separating Movement and Gravity Components in an Acceleration Signal and Implications for the  
157 Assessment of Human Daily Physical Activity. *PLoS One* **8**, e61691 (2013).
- 158 3. HS Dashti, et al., Genome-wide association study identifies genetic loci for self-reported habitual sleep duration supported  
159 by accelerometer-derived estimates. *Nat. Commun.* **10**, 1100 (2019).
- 160 4. VT van Hees, et al., Estimating sleep parameters using an accelerometer without sleep diary. *Sci. Rep.* **8**, 12975 (2018).
- 161 5. A Doherty, et al., Large Scale Population Assessment of Physical Activity Using Wrist Worn Accelerometers: The UK  
162 Biobank Study. *PLoS One* **12**, e0169649 (2017).
- 163 6. PG Sokolove, WN Bushell, The chi square periodogram: Its utility for analysis of circadian rhythms. *J. Theor. Biol.* **72**,  
164 131-160 (1978).
- 165 7. YL Huang, et al., Age-associated difference in circadian sleep-wake and rest-activity rhythms. *Physiol. Behav.* **76**, 597-603  
166 (2002).
- 167 8. GA Sunagawa, et al., Mammalian Reverse Genetics without Crossing Reveals Nr3a as a Short-Sleeper Gene. *Cell Rep.* **14**,  
168 662-677 (2016).
- 169 9. L McInnes, J Healy, J Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv*  
170 (2018).
- 171 10. E Martin, K Hans-Peter, S Jörg, X Xiaowei, A Density-Based Algorithm for Discovering Clusters in Large Spatial  
172 Databases with Noise. *Kdd* **96**, 226-231 (1996).
- 173 11. E Schubert, J Sander, M Ester, HP Kriegel, X Xu, DBSCAN Revisited, Revisited: Why and How You Should (Still) Use  
174 DBSCAN. *ACM Trans. Database Syst.* **42**, 1-21 (2017).



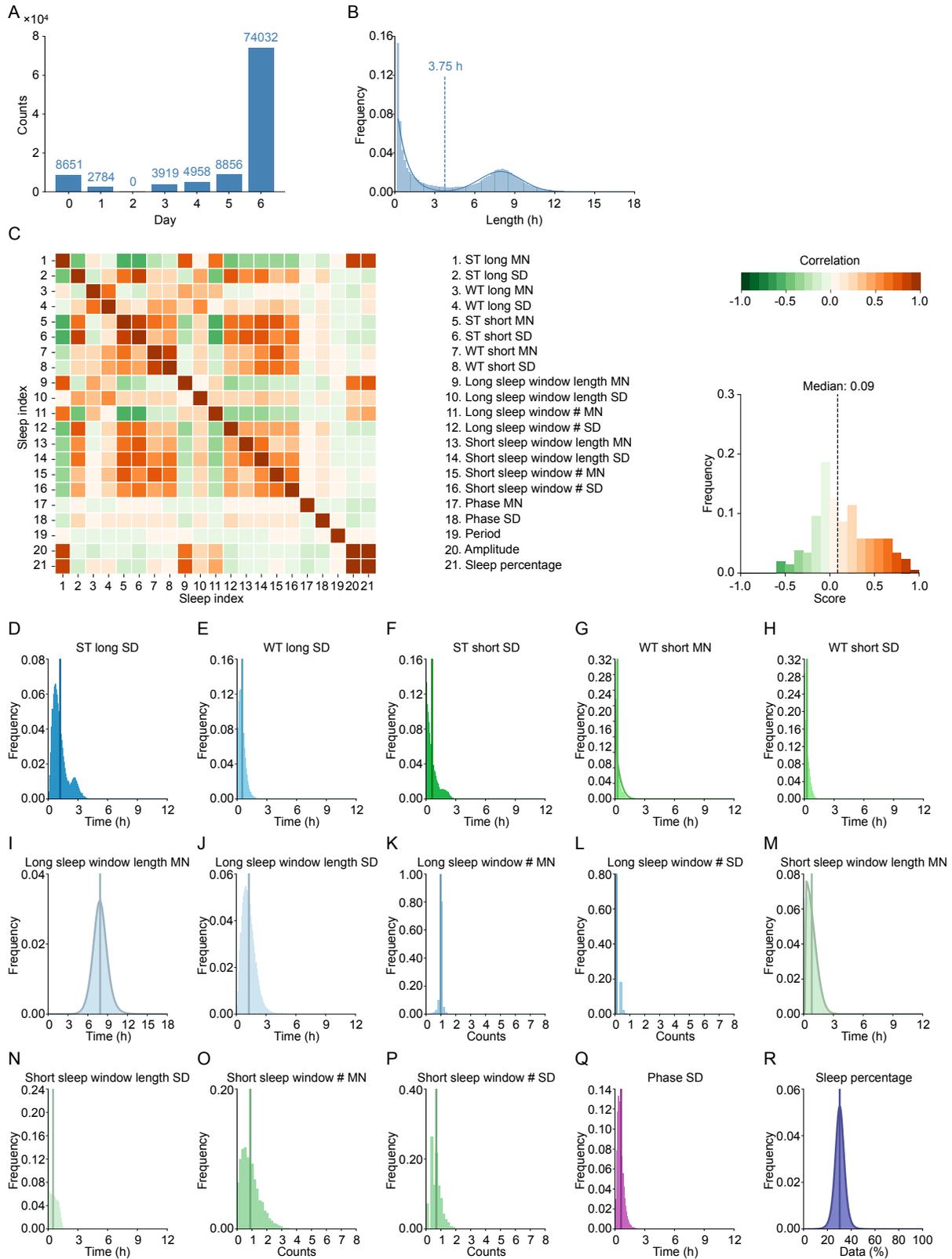
**Fig. S1.** Validation of sleep/wake classification algorithm. (A) An example of ground truth, triaxial acceleration, and power spectrum. In the middle panel, three axes of raw acceleration data are plotted in different colors. The power spectrum in the bottom panel is shown by heatmap, where the lighter color shows a higher value. (B) The histogram of jerk among sleep and wake epochs. (C) The power spectrum among sleep and wake epochs. The solid line and the shadowed area shows the mean and the standard deviation. (D) Results of hyperparameter tuning of XGBoost used in the sleep/wake classification algorithm as 3D plot of summation of accuracy and F measure with two hyperparameters as x and y axes. The max depth and min child weight represent the maximum tree depth for base learners and the minimum summation of instance weight (hessian) needed in a child, respectively. The red circle shows the point of maximum score, whose parameter values were used for the sleep/wake classification algorithm. (E, F) The histogram and distribution of standard deviation and range in nonwearing periods. The gray dashed lines show the threshold used in this study. (G) The performance of sleep/wake classification algorithm. (H) Flow of thinning process for converting 100 Hz to 50 Hz acceleration data. The left panel shows acceleration data sampled by 100 Hz, where each dot shows data point. Data points plotted as white dots are excluded, then the 100 Hz acceleration data is converted to 50 Hz acceleration data shown in the right panel. (I) The result when the sleep/wake classification algorithm trained by 50 Hz acceleration data were applied to reprocessed 100 Hz acceleration data. The 100 Hz acceleration data were converted to 50 Hz acceleration data by the thinning process shown in (H) as the preprocessing, and then the algorithm was applied to the converted acceleration data. Acc: Accuracy. F: F measure. Sen: Sensitivity. Spe: Specificity.



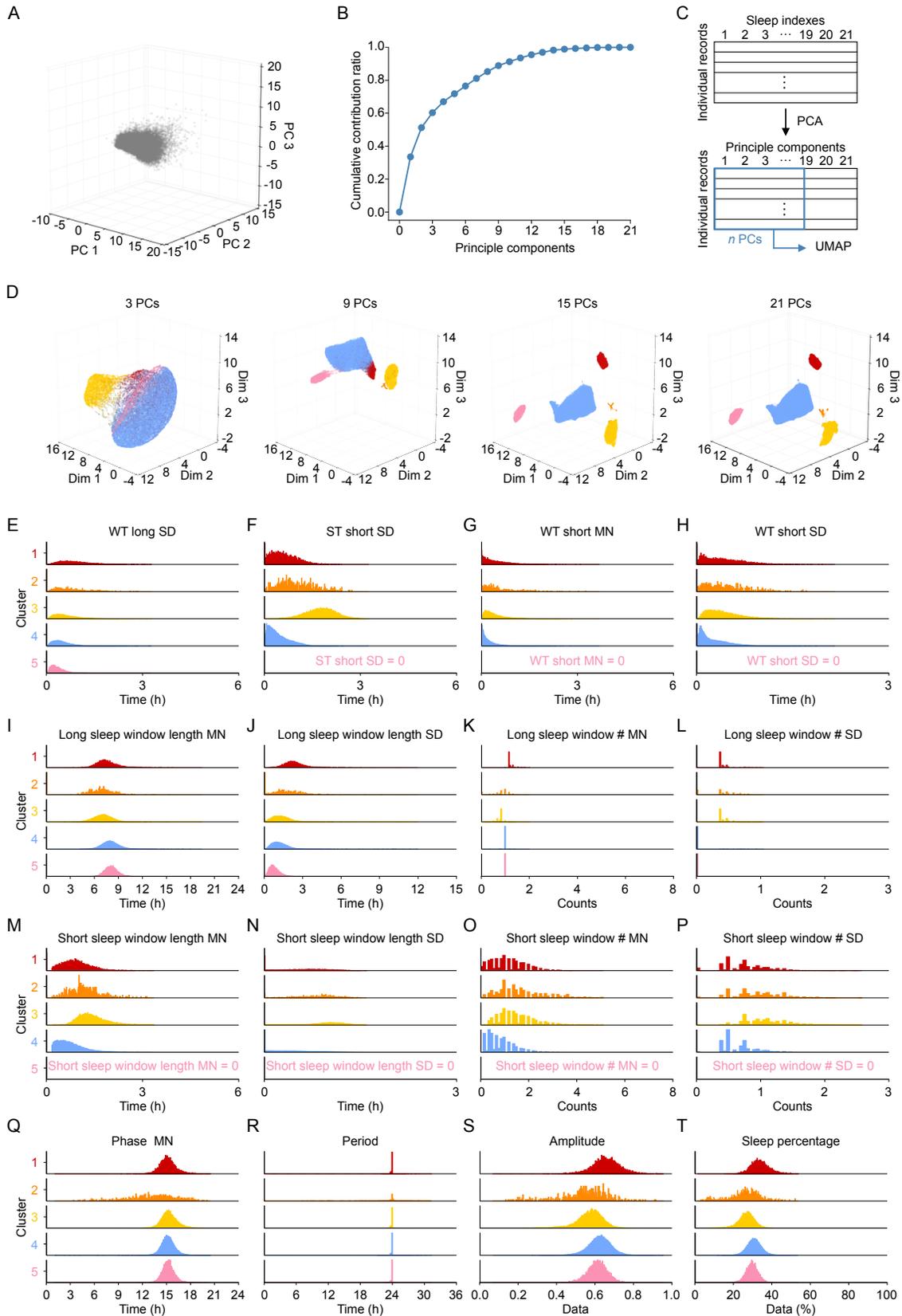
**Fig. S2.** Evaluation of performance of sleep/wake classification algorithm by TST and WASO. (A-D) The results of TST and WASO under varying threshold used to detect sleep onset. The threshold was changed from 0 to 30 minutes by a 2.5-minute step. The black dot and solid line shows the mean, and the shaded area shows the standard deviation among subjects. (E, F) Bland-Altman plots of TST and WASO. The solid gray lines show the mean, and the dashed lines represent the mean  $\pm$  1.95 standard deviation.



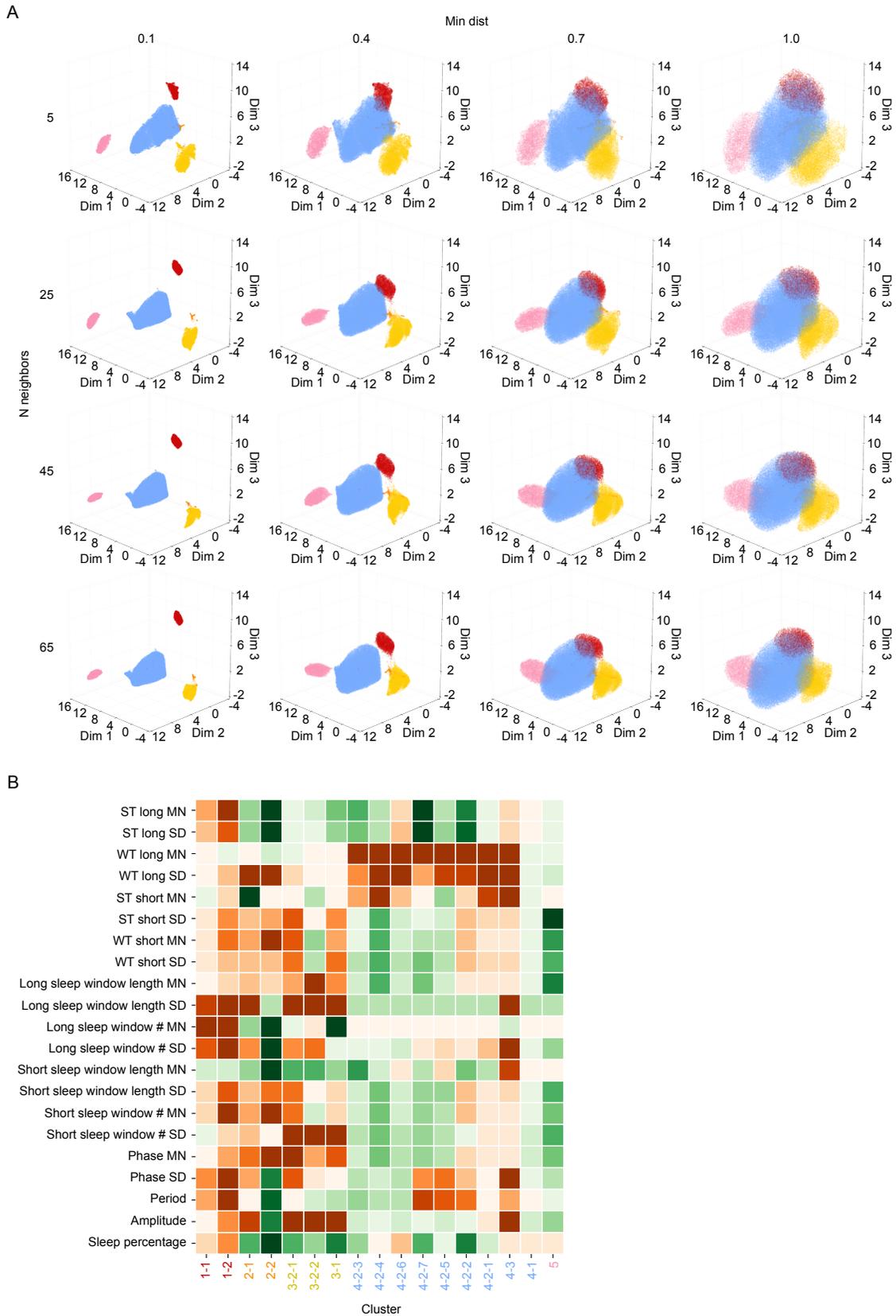
**Fig. S3.** Validation of sleep indexes extraction. (A) Flow diagram to make two types of pseudo sleep/wake time series data, namely, sleep pulse pseudo data and wake pulse pseudo data. The sleep data colored by green and wake data colored by white were grouped as a sleep dataset and wake dataset. Pseudo sleep/wake time series data were made by randomly choosing sleep and wake epochs from the sleep and wake dataset and connecting them. The sleep pulse pseudo data was 10 hours data and consisted of wake epochs except for sleep epochs lasting some time (defined as sleep pulse time) at the middle of data as shown in the left bottom panel. On the opposite, the wake pulse pseudo data consisted of sleep epochs except for wake epochs lasting some time (defined as wake pulse time) as shown in the right bottom panel. (B, C) The sensitivity and specificity of sleep/wake classification algorithm with sleep pulse pseudo data and wake pulse pseudo data. The sleep pulse time and wake pulse time were changed from 0 to 100 minutes at a 1-minute step. All conditions were conducted 1,000 times to eliminate bias due to epoch selection. The solid line and the shaded area shows the mean and the standard deviation. (D, E) The validation of interpolating nonwearing periods as wake or sleep. The sleep/wake time series data was generated by connected six noon-to-noon data consisted of 11 hours wake followed by eight hours sleep and five hours wake as the standard data. Nonwearing periods were inserted into the standard data in two ways. The first way is randomly choosing epochs and converting them to nonwearing periods. A percentage of nonwearing periods was changed from 0 to 90% by a 5% step. The sleep/wake time series data with random nonwearing periods was generated 100 times in each condition and interpolated to wake or sleep periods, then the chi-square periodogram was applied to these data. Period was about 24 hour with the nonwearing percent lower than 50% in both interpolations to sleep and wake (D). The second way is choosing continuous epochs and converting them to nonwearing periods. The duration was changed from 0 to 48 hours by a 1-hour step. The sleep/wake time series data was generated 100 times by changing the start point of nonwearing periods in each condition and interpolated to wake or sleep periods, then the chi-square periodogram was applied to these data. Period was also about 24 hour in both interpolations (E). (F) The validation of effect of different sleep schedules between workdays and holidays on the chi-square periodogram. To analyze how the different sleep schedules between workdays and holidays affected a period, we set some days as holidays and shifted the 8-hour sleep window. The shift of sleep window was changed from -4 to 4 hours by a 0.1-hour step. In each condition, seven types of holidays were simulated: only the first day, the continuous two days (the first and second, second and third, third and fourth, fourth and fifth, and fifth and last days), and the last day. The simulation sleep/wake time series data with holidays tended to have less than 24-hour period, which was regarded as the effect of difference between workdays and holidays.



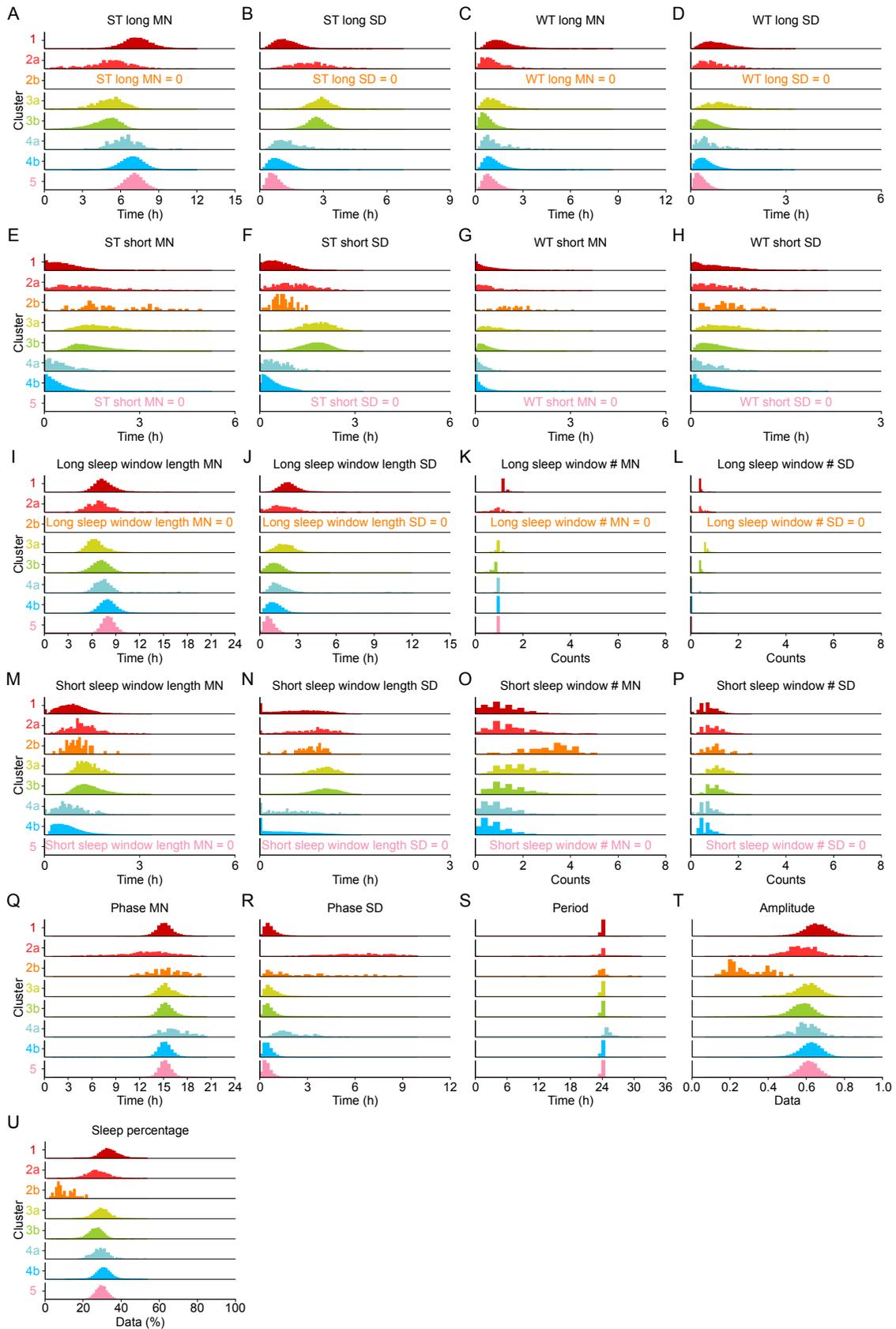
**Fig. S4.** Results of large-scale sleep analysis. (A) Distribution of maximum continuous days. Individual records with more than three continuous days were used for the large-scale sleep analysis. (B) Distribution of length of sleep windows. The solid curves shows the exponential curve fitted to the distribution and the Gaussian curve fitted to the distribution with a restriction that the peak of curve and the right ridge were the same. The intersection points of two curves, 3.75 hours, was determined to divide sleep windows into long sleep windows and short sleep windows. (C) The left heatmap shows the correlation coefficient between the sleep indexes. The right histogram shows the distribution of correlation coefficient. (D-R) Distribution of sleep indexes. The solid line shows the mean. The solid curves shown in (G), (I), (M), and (R) are the curves fitted to each distribution.



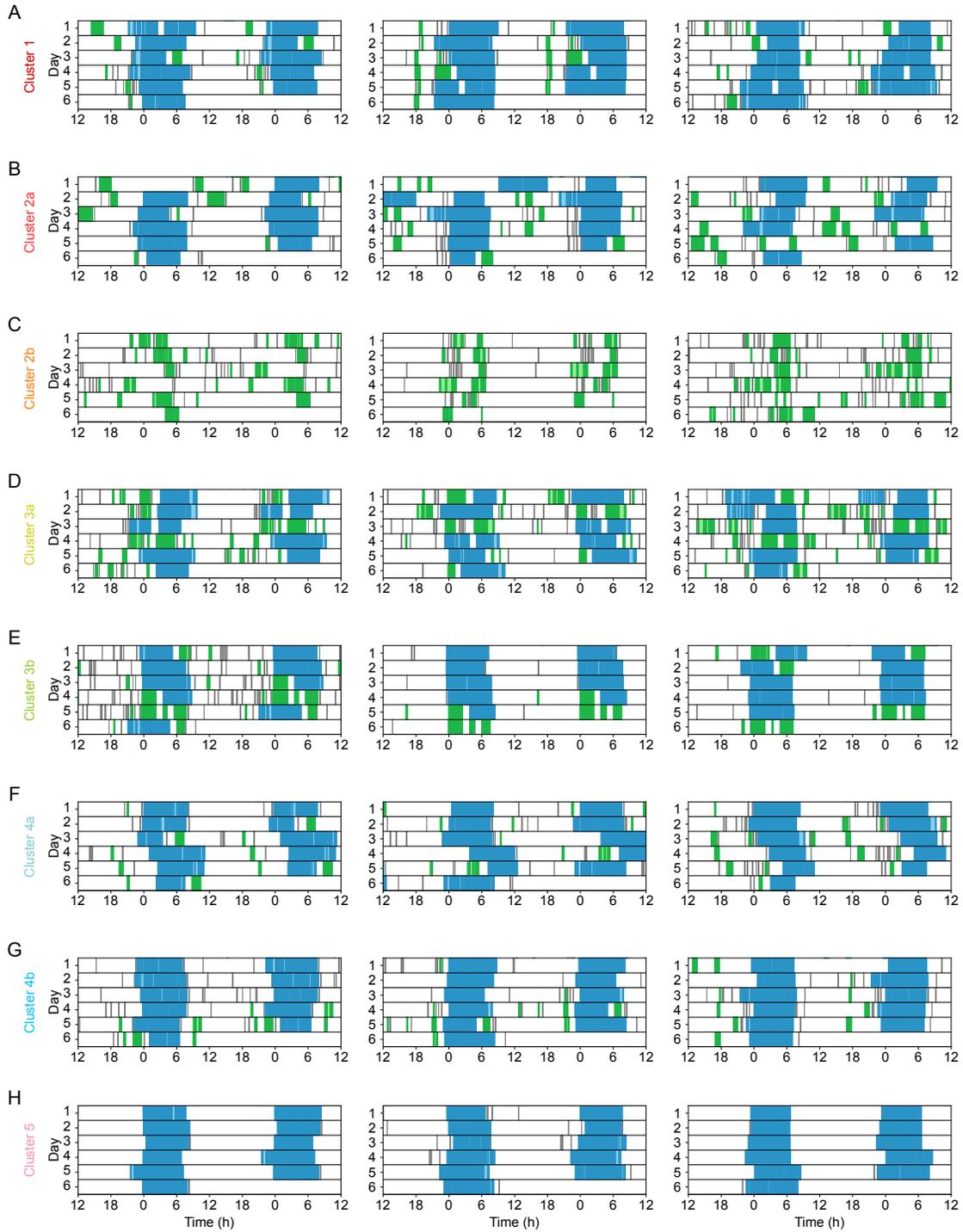
**Fig. S5.** Results of first clustering. (A) The results of PCA of the whole dataset. (B) Cumulative contribution ratio of principle components. (C) The flow of clustering method combining PCA and UMAP. 21 sleep indexes are converted to 21 principle components (PCs) by PCA.  $n$  PCs are then used to UMAP to express  $n$ -dimensional data to three-dimensional data. (D) The results of clustering, where 3, 9, 15, or 21 PCs were used for UMAP. (E-T) Distribution of sleep indexes divided into five clusters. The label of y axis is the cluster name.



**Fig. S6.** Effect of hyperparameters and hierarchical clustering result. (A) Effect of hyperparameters on the first clustering. Two hyperparameters of UMAP,  $n\_neighbors$  and  $min\_dist$ , were changed. Individual records are plotted as the clusters' colors. (B) Heatmap of z score of sleep indexes.



**Fig. S7.** Distribution of sleep indexes within clusters obtained in the whole clustering. (A-U) Distribution of sleep indexes divided into eight. The label of y axis is the cluster name.



**Fig. S8.** Representative plots within clusters of the whole clustering. (A-H) Representative plots shown as double plot. The label of x axis is time.

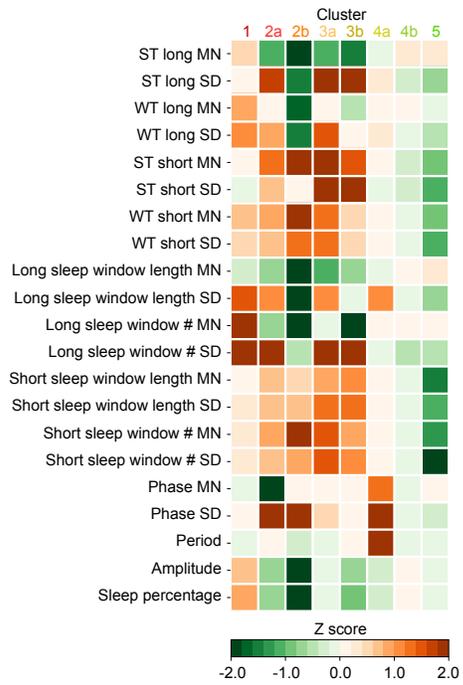
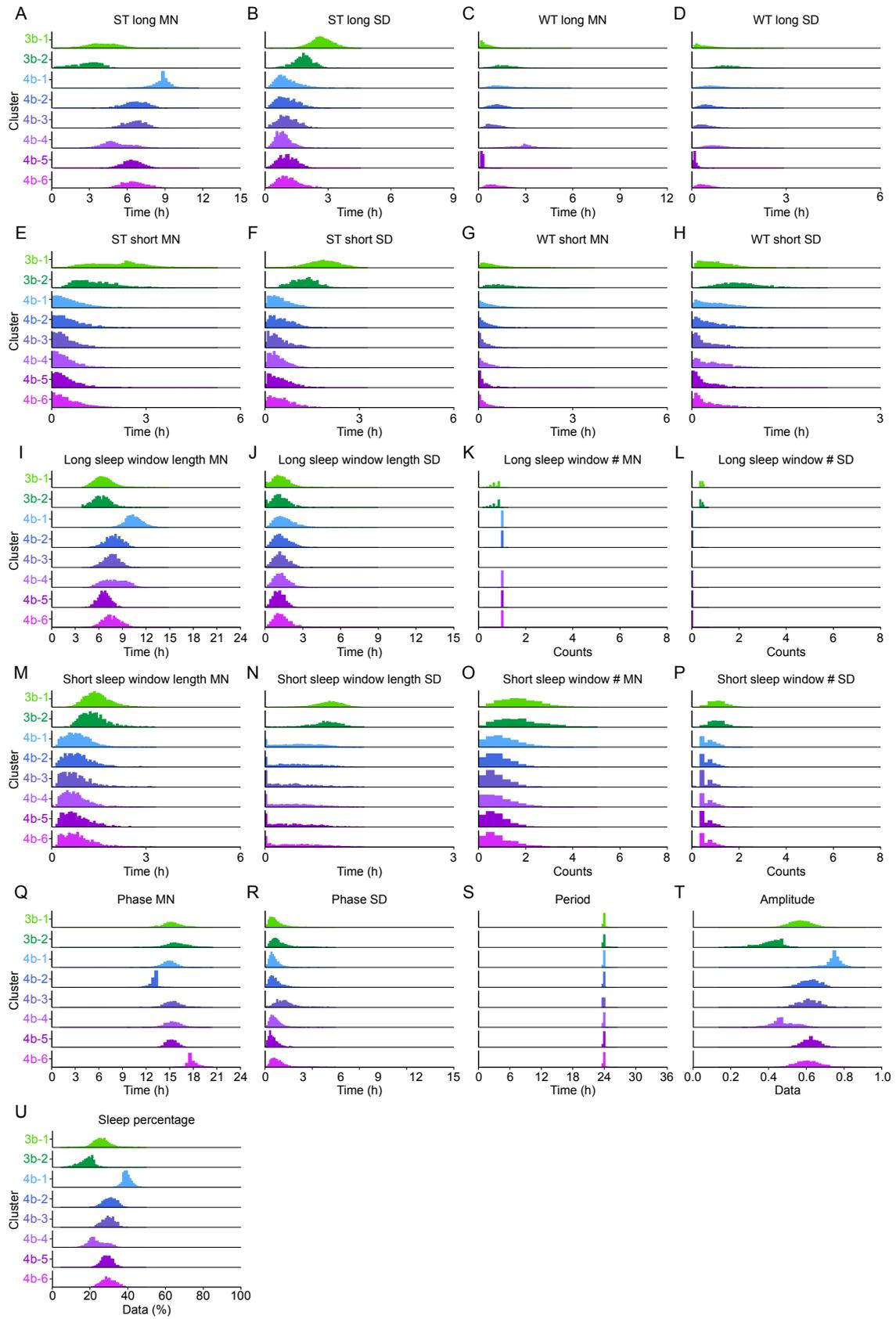
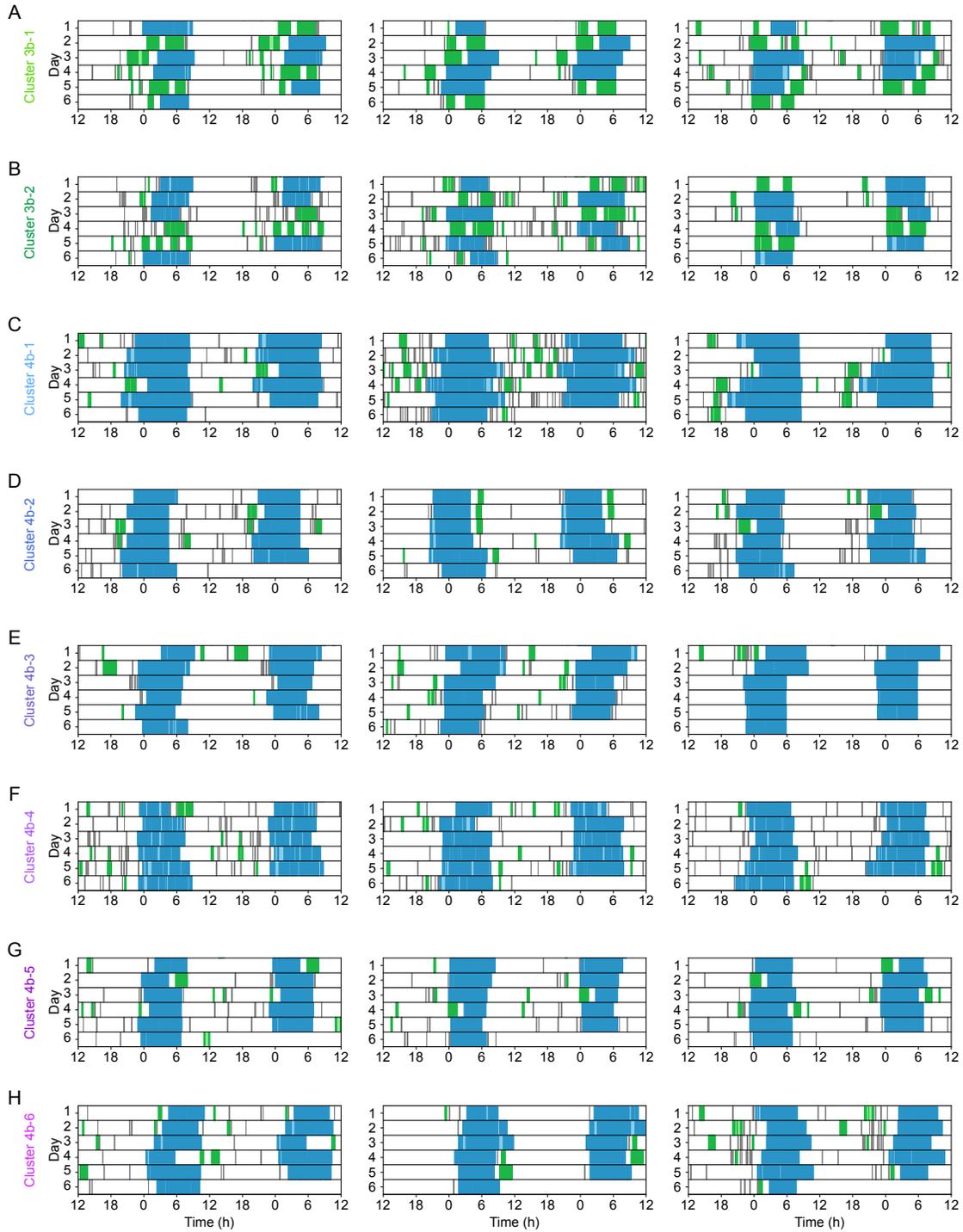


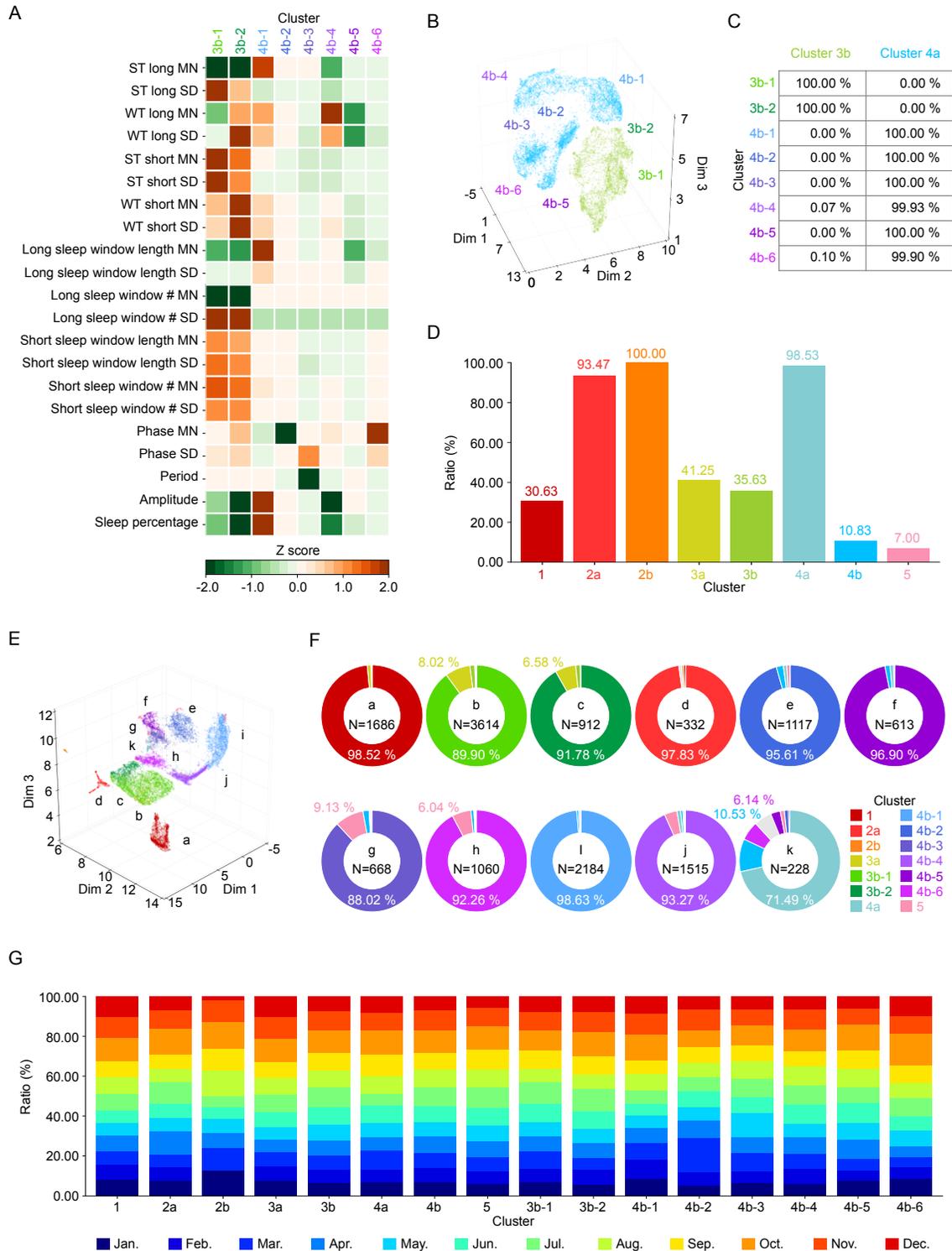
Fig. S9. Heatmap of z score of sleep indexes of each cluster in the whole clustering.



**Fig. S10.** Distribution of sleep indexes within clusters of the outlier clustering. (A-U) Distribution of sleep indexes divided into eight. The label of y axis is the cluster name.



**Fig. S11.** Representative plots within clusters of the outlier clustering. (A-H) Representative plots shown as double plot. The label of x axis is time.



**Fig. S12.** Supportive data of clusters of outlier clustering. (A) Heatmap of z score of sleep indexes. (B) The result of outlier clustering, where individual records are colored by the color of whole clustering. (C) Ratio of Clusters 3b and 4b contained in each cluster defined in the outlier clustering. (D) Ratio of the outlier dataset contained in each cluster defined in the whole clustering. (E) The result of clustering using the outlier dataset from all clusters. Individual records plotted by the clusters' colors were divided into 11 clusters named from a to k. (F) Ratio of individual records contained in clusters a to k. (G) Distribution of month in which the Activity recording was started.

**Table S1. PSG and Axivity recordings (50Hz)**

Subject	Sex	Age
1	F	22
2	M	22
3	F	22
4	M	19
5	F	24
6	M	25
7	M	24
8	F	22
9	M	23
10	M	20
11	F	21
12	M	23
13	F	25
14	M	26
15	F	21
16	M	23
17	F	24
18	M	20
19	F	23
20	F	24
21	F	24
22 *	F	23
23 †	F	22
24 ‡	M	21
25	M	23
26	F	24
27	M	22

List of subjects recruited for one-night PSG measurements. All participants wore the Axivity device (50 Hz sampling frequency). The subjects with any symbol (\*, †, ‡) also wore Axivity recorded by 100 Hz sampling frequency and the same subject shown in Table S3. F indicates female and M indicates male in all tables.

**Table S2. Long-term Axivity recordings for nonwear detection**

Subject	Sex	Age	Recording Days
1	M	25	14
2	M	29	14
3	F	23	14
4	F	54	14
5	F	33	14
6	M	30	14
7	M	23	14
8	M	22	14
9	M	21	14
10	F	23	9
11	M	27	14
12	F	23	14
13	F	23	14
14	M	21	14
15	M	38	14
16	M	28	14
17	F	22	14
18	F	24	14
19	F	31	7
20	M	27	14

List of subjects recruited for long-term Axivity recording to validate the nonwear detection algorithm. All participants were asked to record timestamps when they wore or took off Axivity.

**Table S3. PSG and Axivity recordings (100Hz)**

Subject	Sex	Age
1 *	F	23
2 †	F	22
3 ‡	M	21
4	M	27
5	F	23
6	F	21
7	M	20
8	M	24
9	M	21
10	F	21
11	F	21
12	M	21

List of subjects recruited for one-night PSG measurements. All participants wore the Axivity device (100 Hz sampling frequency). The subjects with any symbol (\*, †, ‡) also wore Axivity recorded by 50 Hz sampling frequency and the same subject shown in Table S1.

**Table S4. Sleep indexes used in this paper**

Sleep index	Name	Feature type
ST long MN	Sleep time long mean	Daily, Common Index
ST long SD	Sleep time long SD	Daily, Common Index
WT long MN	Wake time long mean	Daily, Common Index
WT long SD	Wake time long SD	Daily, Common Index
ST short MN	Sleep time short mean	Daily, Common Index
ST short SD	Sleep time short SD	Daily, Common Index
WT short MN	Wake time short mean	Daily, Common Index
WT short SD	Wake time short SD	Daily, Common Index
Long sleep window length MN	Long sleep window length mean	Daily, Common Index
Long sleep window length SD	Long sleep window length SD	Daily, Common Index
Long sleep window # MN	Long sleep window number mean	Daily, Common Index
Long sleep window # SD	Long sleep window number SD	Daily, Common Index
Short sleep window length MN	Short sleep window length mean	Daily, Common Index
Short sleep window length SD	Short sleep window length SD	Daily, Common Index
Short sleep window # MN	Short sleep window number mean	Daily, Common Index
Short sleep window # SD	Short sleep window number SD	Daily, Common Index
Sleep percentage	Sleep percentage	General, Common Index
Phase MN	Phase mean	Daily, Rhythm-related index
Phase SD	Phase SD	Daily, Rhythm-related index
Period	Period	General, Rhythm-related index
Amplitude	Amplitude	General, Rhythm-related index

**Table S5. Means of sleep indexes among the 17 clusters generated by divisive hierarchical clustering**

Sleep index / Cluster	1-1	1-2	2-1	2-2	3-1	3-2-1	3-2-2	4-1
ST long MN (h)	7.32	7.91	5.16	0.00	4.73	5.15	5.76	6.88
ST long SD (h)	1.28	2.06	2.50	0.00	2.65	2.92	2.98	0.95
WT long MN (h)	1.77	2.76	1.20	0.00	0.79	1.26	0.90	1.15
ST long SD (h)	0.87	1.51	0.79	0.00	0.58	1.00	0.63	0.48
ST short MN (h)	0.66	1.16	1.33	2.23	1.53	1.78	1.11	0.42
ST short SD (h)	0.60	0.95	1.09	0.73	1.75	1.83	1.96	0.44
WT short MN (h)	0.42	0.83	0.49	1.18	0.41	0.64	0.15	0.22
ST short SD (h)	0.40	0.67	0.44	0.58	0.42	0.58	0.28	0.25
Long sleep window length MN (h)	7.53	7.42	7.00	0.00	7.13	6.51	6.54	8.03
Long sleep window length SD (h)	2.32	2.99	1.94	0.00	1.26	1.94	2.13	1.20
Long sleep window # MN	1.21	1.45	0.91	0.00	0.77	0.98	1.02	1.00
Long sleep window # SD	0.39	0.77	0.46	0.00	0.40	0.63	0.61	0.00
Short sleep window length MN (h)	0.91	1.15	1.21	1.05	1.45	1.38	2.90	0.76
Short sleep window length SD (h)	0.60	0.83	0.82	0.80	1.01	1.02	0.31	0.43
Short sleep window # MN	1.14	1.70	1.50	3.29	1.45	1.84	0.44	0.82
Short sleep window # SD	0.77	1.06	0.95	1.03	1.07	1.23	0.76	0.66
Phase MN (h)	15.20	15.61	12.92	15.34	15.37	15.38	14.80	15.20
Phase SD (h)	0.72	1.03	6.46	2.63	0.67	0.86	0.63	0.56
Period (h)	23.94	23.87	23.99	23.79	23.94	23.92	23.94	23.93
Amplitude	0.66	0.71	0.57	0.28	0.57	0.61	0.61	0.62
Sleep percentage (%)	33.82	38.61	27.57	10.20	26.52	29.50	28.84	30.84

Sleep index / Cluster	4-2-1	4-2-2	4-2-3	4-2-4	4-2-5	4-2-6	4-2-7	4-3	5
ST long MN (h)	6.24	4.85	5.68	6.61	6.45	7.46	5.09	6.97	7.08
ST long SD (h)	1.46	1.07	1.00	1.09	1.00	0.91	1.08	2.75	0.70
WT long MN (h)	1.14	1.95	0.70	0.87	2.04	0.80	2.28	1.71	1.02
ST long SD (h)	0.53	0.75	0.32	0.44	0.92	0.38	0.90	1.20	0.34
ST short MN (h)	0.75	0.82	0.34	0.02	0.26	0.27	0.16	0.79	0.00
ST short SD (h)	0.76	0.60	0.33	0.05	0.32	0.30	0.19	0.76	0.00
WT short MN (h)	0.32	0.45	0.14	0.01	0.09	0.14	0.06	0.34	0.00
ST short SD (h)	0.35	0.42	0.17	0.02	0.11	0.18	0.08	0.37	0.00
Long sleep window length MN (h)	7.38	6.80	6.38	7.48	8.49	8.26	7.36	9.71	8.09
Long sleep window length SD (h)	1.75	1.50	1.13	1.19	1.60	1.10	1.51	5.26	0.80
Long sleep window # MN	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.95	1.00
Long sleep window # SD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.47	0.00
Short sleep window length MN (h)	1.00	0.94	0.63	0.16	0.65	0.67	0.34	1.00	0.00
Short sleep window length SD (h)	0.68	0.77	0.29	0.00	0.30	0.27	0.12	0.68	0.00
Short sleep window # MN	1.09	1.40	0.84	0.13	0.56	0.66	0.56	1.11	0.00
Short sleep window # SD	0.86	0.91	0.67	0.25	0.58	0.60	0.64	0.83	0.00
Phase MN (h)	16.79	15.62	16.08	17.11	14.48	15.98	15.39	17.07	15.22
Phase SD (h)	2.04	1.60	1.24	2.05	1.55	1.69	1.16	2.89	0.48
Period (h)	25.42	25.03	24.77	24.96	25.26	25.17	24.72	25.03	23.93
Amplitude	0.60	0.50	0.56	0.59	0.57	0.66	0.47	0.64	0.61
Sleep percentage (%)	29.62	24.09	25.53	27.89	28.26	32.59	22.30	32.85	29.68

**Table S6. Standard deviations of sleep indexes among the 17 clusters generated by divisive hierarchical clustering**

Sleep index / Cluster	1-1	1-2	2-1	2-2	3-1	3-2-1	3-2-2	4-1	
ST long MN (h)	1.15	1.40	1.84	0.00	1.23	1.16	0.73	0.93	
ST long SD (h)	0.57	1.06	1.01	0.00	0.50	0.55	0.35	0.45	
WT long MN (h)	0.92	1.18	0.90	0.00	0.51	0.68	0.39	0.60	
ST long SD (h)	0.47	0.59	0.52	0.00	0.36	0.48	0.29	0.29	
ST short MN (h)	0.52	0.68	0.92	1.18	0.69	0.70	0.35	0.37	
ST short SD (h)	0.40	0.54	0.57	0.31	0.47	0.47	0.49	0.33	
WT short MN (h)	0.40	0.50	0.47	0.56	0.35	0.46	0.09	0.25	
ST short SD (h)	0.31	0.35	0.32	0.32	0.27	0.33	0.16	0.23	
Long sleep window length MN (h)	1.22	1.35	1.53	0.00	1.15	1.12	0.61	1.07	
Long sleep window length SD (h)	0.74	1.10	1.28	0.00	0.63	0.76	0.59	0.58	
Long sleep window # MN	0.09	0.17	0.27	0.00	0.12	0.12	0.09	0.00	
Long sleep window # SD	0.04	0.07	0.22	0.00	0.05	0.07	0.06	0.00	
Short sleep window length MN (h)	0.48	0.40	0.49	0.35	0.48	0.35	0.42	0.46	
Short sleep window length SD (h)	0.38	0.32	0.34	0.20	0.26	0.20	0.22	0.36	
Short sleep window # MN	0.70	0.75	0.90	1.02	0.74	0.80	0.13	0.55	
Short sleep window # SD	0.35	0.38	0.37	0.44	0.34	0.40	0.18	0.27	
Phase MN (h)	1.14	1.69	3.19	2.11	1.06	1.28	0.87	0.89	
Phase SD (h)	0.47	0.69	1.90	2.26	0.48	0.61	0.29	0.32	
Period (h)	0.22	0.35	4.71	3.71	0.26	0.25	0.10	0.14	
Amplitude	0.07	0.10	0.10	0.10	0.07	0.07	0.05	0.06	
Sleep percentage (%)	4.92	5.88	6.59	4.81	4.27	4.51	3.01	3.83	
Sleep index / Cluster	4-2-1	4-2-2	4-2-3	4-2-4	4-2-5	4-2-6	4-2-7	4-3	5
ST long MN (h)	0.73	0.81	0.50	0.26	0.37	0.40	0.49	2.27	0.75
ST long SD (h)	0.56	0.33	0.25	0.53	0.48	0.35	0.56	0.92	0.33
WT long MN (h)	0.65	0.95	0.32	0.25	0.46	0.32	0.43	1.12	0.49
ST long SD (h)	0.36	0.38	0.13	0.17	0.26	0.22	0.42	0.85	0.20
ST short MN (h)	0.45	0.46	0.29	0.03	0.10	0.20	0.15	0.56	0.00
ST short SD (h)	0.38	0.23	0.13	0.06	0.09	0.16	0.15	0.48	0.00
WT short MN (h)	0.30	0.22	0.09	0.01	0.06	0.14	0.04	0.27	0.00
ST short SD (h)	0.25	0.17	0.12	0.03	0.06	0.16	0.05	0.28	0.00
Long sleep window length MN (h)	1.04	0.77	0.46	0.33	0.67	0.48	0.62	4.44	0.81
Long sleep window length SD (h)	0.73	0.59	0.28	0.62	0.50	0.39	0.71	4.00	0.40
Long sleep window # MN	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.00
Long sleep window # SD	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.00
Short sleep window length MN (h)	0.40	0.32	0.17	0.17	0.21	0.39	0.14	0.55	0.00
Short sleep window length SD (h)	0.33	0.21	0.16	0.01	0.14	0.20	0.17	0.43	0.00
Short sleep window # MN	0.55	0.49	0.57	0.12	0.23	0.38	0.25	0.67	0.00
Short sleep window # SD	0.30	0.30	0.28	0.24	0.14	0.21	0.17	0.41	0.00
Phase MN (h)	1.88	0.90	1.24	1.62	0.85	1.29	0.80	2.24	0.78
Phase SD (h)	0.90	0.34	0.30	0.79	0.34	0.52	0.28	1.18	0.27
Period (h)	0.91	0.43	0.35	0.53	0.74	0.96	0.22	2.00	0.13
Amplitude	0.05	0.05	0.03	0.02	0.03	0.03	0.03	0.12	0.05
Sleep percentage (%)	3.22	2.41	1.63	1.10	1.51	1.99	1.63	8.35	3.13

**Table S7. Means of sleep indexes**

Sleep index / Cluster	1	2a	2b	3a	3b	4a	4b	5
ST long MN (h)	7.33	5.16	0.00	5.15	4.73	6.37	6.88	7.08
ST long SD (h)	1.29	2.50	0.00	2.92	2.65	1.46	0.95	0.70
WT long MN (h)	1.78	1.20	0.00	1.26	0.79	1.26	1.15	1.02
ST long SD (h)	0.88	0.79	0.00	1.00	0.58	0.63	0.48	0.34
ST short MN (h)	0.67	1.33	2.23	1.78	1.53	0.58	0.42	0.00
ST short SD (h)	0.60	1.09	0.73	1.83	1.75	0.57	0.44	0.00
WT short MN (h)	0.43	0.49	1.18	0.64	0.41	0.26	0.22	0.00
ST short SD (h)	0.40	0.44	0.58	0.58	0.41	0.28	0.25	0.00
Long sleep window length MN (h)	7.53	7.00	0.00	6.51	7.13	7.78	8.03	8.09
Long sleep window length SD (h)	2.33	1.94	0.00	1.94	1.26	2.06	1.20	0.80
Long sleep window # MN	1.21	0.91	0.00	0.98	0.77	0.99	1.00	1.00
Long sleep window # SD	0.40	0.46	0.00	0.63	0.40	0.07	0.00	0.00
Short sleep window length MN (h)	0.91	1.21	1.05	1.38	1.46	0.84	0.76	0.00
Short sleep window length SD (h)	0.61	0.82	0.80	1.02	1.01	0.52	0.43	0.00
Short sleep window # MN	1.14	1.50	3.29	1.84	1.45	0.94	0.82	0.00
Short sleep window # SD	0.78	0.95	1.03	1.23	1.07	0.76	0.66	0.00
Phase MN (h)	15.20	12.92	15.34	15.38	15.37	16.46	15.20	15.22
Phase SD (h)	0.73	6.46	2.63	0.86	0.67	1.97	0.56	0.48
Period (h)	23.94	23.99	23.79	23.92	23.94	25.19	23.93	23.93
Amplitude	0.66	0.57	0.28	0.61	0.57	0.60	0.62	0.61
Sleep percentage (%)	33.90	27.57	10.20	29.50	26.52	29.41	30.84	29.68

**Table S8. Standard deviations of sleep indexes**

Sleep index / Cluster	1	2a	2b	3a	3b	4a	4b	5
ST long MN (h)	1.16	1.84	0.00	1.16	1.23	1.26	0.93	0.75
ST long SD (h)	0.59	1.01	0.00	0.55	0.50	0.82	0.45	0.33
WT long MN (h)	0.93	0.90	0.00	0.68	0.51	0.81	0.60	0.49
ST long SD (h)	0.48	0.52	0.00	0.48	0.36	0.52	0.29	0.20
ST short MN (h)	0.52	0.92	1.18	0.70	0.69	0.48	0.37	0.00
ST short SD (h)	0.40	0.57	0.31	0.47	0.47	0.41	0.33	0.00
WT short MN (h)	0.40	0.47	0.56	0.46	0.35	0.27	0.25	0.00
ST short SD (h)	0.31	0.32	0.32	0.33	0.27	0.24	0.23	0.00
Long sleep window length MN (h)	1.23	1.53	0.00	1.12	1.15	2.09	1.07	0.81
Long sleep window length SD (h)	0.75	1.28	0.00	0.76	0.63	2.13	0.58	0.40
Long sleep window # MN	0.10	0.27	0.00	0.12	0.12	0.08	0.00	0.00
Long sleep window # SD	0.06	0.22	0.00	0.07	0.05	0.17	0.00	0.00
Short sleep window length MN (h)	0.48	0.49	0.35	0.35	0.49	0.46	0.46	0.00
Short sleep window length SD (h)	0.38	0.34	0.20	0.20	0.26	0.38	0.36	0.00
Short sleep window # MN	0.70	0.90	1.02	0.80	0.74	0.59	0.55	0.00
Short sleep window # SD	0.35	0.37	0.44	0.40	0.34	0.34	0.27	0.00
Phase MN (h)	1.16	3.19	2.11	1.28	1.06	1.81	0.89	0.78
Phase SD (h)	0.47	1.90	2.26	0.61	0.48	0.93	0.32	0.27
Period (h)	0.23	4.71	3.71	0.25	0.26	1.09	0.14	0.13
Amplitude	0.07	0.10	0.10	0.07	0.07	0.08	0.06	0.05
Sleep percentage (%)	4.97	6.59	4.81	4.51	4.27	4.98	3.83	3.13

**Table S9. Means of sleep indexes (outliers)**

Sleep index / Cluster	3b-1	3b-2	4b-1	4b-2	4b-3	4b-4	4b-5	4b-6
ST long MN (h)	4.05	2.80	8.77	6.68	6.61	5.14	6.40	6.52
ST long SD (h)	2.72	1.82	1.05	1.07	1.07	0.88	1.06	1.11
WT long MN (h)	0.53	1.66	1.70	1.24	1.02	2.85	0.25	1.10
ST long SD (h)	0.45	1.21	0.76	0.54	0.44	0.79	0.11	0.44
ST short MN (h)	2.09	1.39	0.59	0.56	0.44	0.44	0.47	0.55
ST short SD (h)	1.85	1.24	0.50	0.53	0.46	0.42	0.50	0.54
WT short MN (h)	0.47	0.84	0.39	0.26	0.20	0.31	0.17	0.24
ST short SD (h)	0.40	0.77	0.38	0.28	0.22	0.33	0.20	0.26
Long sleep window length MN (h)	6.60	6.34	10.47	7.91	7.63	7.99	6.66	7.62
Long sleep window length SD (h)	1.24	1.18	1.58	1.36	1.29	1.31	1.10	1.32
Long sleep window # MN	0.69	0.69	1.00	1.00	1.00	1.00	1.00	1.00
Long sleep window # SD	0.43	0.43	0.00	0.00	0.00	0.00	0.00	0.00
Short sleep window length MN (h)	1.47	1.35	0.84	0.87	0.79	0.77	0.79	0.84
Short sleep window length SD (h)	1.00	0.97	0.54	0.52	0.40	0.46	0.44	0.50
Short sleep window # MN	1.82	1.78	1.12	0.93	0.82	0.96	0.80	0.92
Short sleep window # SD	1.11	1.09	0.75	0.69	0.65	0.71	0.65	0.70
Phase MN (h)	15.38	15.88	14.93	12.94	15.38	15.43	15.26	17.92
Phase SD (h)	0.77	0.92	0.58	0.67	1.21	0.67	0.53	0.86
Period (h)	23.95	23.96	23.94	23.93	23.18	23.90	23.95	23.92
Amplitude	0.57	0.41	0.75	0.62	0.61	0.48	0.62	0.61
Sleep percentage (%)	26.06	18.13	39.60	30.61	29.80	23.79	28.98	29.89

**Table S10. Standard deviations of sleep indexes (outliers)**

Sleep index / Cluster	3b-1	3b-2	4b-1	4b-2	4b-3	4b-4	4b-5	4b-6
ST long MN (h)	1.28	1.10	0.68	0.90	0.83	1.08	0.75	0.95
ST long SD (h)	0.53	0.47	0.55	0.51	0.46	0.37	0.42	0.51
WT long MN (h)	0.41	0.70	0.91	0.49	0.47	0.72	0.13	0.55
ST long SD (h)	0.33	0.40	0.45	0.28	0.27	0.37	0.07	0.26
ST short MN (h)	0.87	0.66	0.49	0.43	0.35	0.37	0.39	0.47
ST short SD (h)	0.51	0.39	0.35	0.35	0.33	0.28	0.34	0.38
WT short MN (h)	0.36	0.55	0.36	0.24	0.21	0.30	0.19	0.26
ST short SD (h)	0.25	0.33	0.28	0.23	0.21	0.26	0.20	0.24
Long sleep window length MN (h)	1.22	1.13	1.12	1.05	0.97	1.47	0.77	1.09
Long sleep window length SD (h)	0.69	0.67	0.85	0.64	0.57	0.59	0.44	0.62
Long sleep window # MN	0.16	0.18	0.00	0.02	0.00	0.00	0.00	0.00
Long sleep window # SD	0.05	0.06	0.01	0.03	0.00	0.01	0.00	0.00
Short sleep window length MN (h)	0.43	0.45	0.43	0.47	0.48	0.44	0.42	0.47
Short sleep window length SD (h)	0.23	0.25	0.35	0.37	0.35	0.36	0.35	0.37
Short sleep window # MN	0.82	0.91	0.66	0.55	0.47	0.64	0.48	0.60
Short sleep window # SD	0.36	0.35	0.31	0.28	0.29	0.31	0.27	0.29
Phase MN (h)	1.37	1.36	0.93	1.04	0.96	1.08	0.66	0.81
Phase SD (h)	0.61	0.60	0.32	0.42	0.57	0.37	0.32	0.57
Period (h)	0.42	0.40	0.12	0.16	0.31	0.18	0.13	0.17
Amplitude	0.06	0.07	0.05	0.05	0.05	0.06	0.04	0.06
Sleep percentage (%)	4.10	3.87	2.64	3.42	3.30	4.55	2.84	3.90

**Table S11. Summary of insomnia-like clusters**

Cluster	Sleep duration	Fragmented	Long-term awake	Short-term awake
1	↑	↑↑	↑	↑
3a	-	↑↑↑	↑↑	↑
3b	↓	↑	↑	-
3b-1	↓	↑↑	↑↑↑	↓
3b-2	↓↓↓	↑↑	↑	↑↑
4b-4	↓↓	-	↓	↑↑↑
4b-5	-	-	↑↑	↓↓

The arrows represent how far each sleep-related indicator in the upper row is from the mean of the whole dataset in terms of seven levels (from ↓↓↓ to ↑↑↑). “Fragmented” parameter represents the summation of the number of long sleep window MN and short sleep window MN. “Long-term midawake” represents how midawake duration increases if the threshold used to connect sleep episodes to make sleep windows is 120 minutes.

“Short-term midawake” represents the summation of WT long MN and WT short MN. The actual values are shown in Table S12.

**Table S12. Values of insomnia-like clusters**

Cluster	Sleep duration (%)	Fragmented	Long-term awake (min)	Short-term awake (h)
1	33.90	2.35	1.08	2.21
3a	29.50	2.82	1.91	1.90
3b	26.52	2.21	1.60	1.20
3b-1	26.06	2.51	2.44	1.00
3b-2	18.13	2.47	1.15	2.50
4b-4	23.79	1.96	0.47	3.15
4b-5	28.98	1.80	1.74	0.43
all mean	30.33	1.87	0.79	1.38
all sd	4.34	0.65	8.48	0.77