Received: 29 March 2024 | Revised: 4 June 2024 | Accepted: 1 July 2024

DOI: 10.1111/ejn.16465

SPECIAL ISSUE ARTICLE

EJN European Journal of Neuroscience FENS WILEY

Steven A. Brown Special Issue: Dynamic Interactions of Biological Clocks, Sleep and Metabolism

An automated sleep staging tool based on simple statistical features of mice electroencephalography (EEG) and electromyography (EMG) data

Rikuhiro G. Yamada[1,2] | **Kyoko Matsuzawa**[1] | **Koji L. Ode**[3] | **Hiroki R. Ueda**[1,2,3]

[1]Laboratory for Synthetic Biology, RIKEN Center for Biosystems Dynamics Research, Osaka, Japan

[2]Department of Systems Biology, Institute of Life Science, Kurume University, Fukuoka, Japan

[3]Department of Systems Pharmacology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Correspondence
Rikuhiro G. Yamada and Hiroki R. Ueda, Laboratory for Synthetic Biology, RIKEN Center for Biosystems Dynamics Research, 565-0871 Osaka, Japan.
Email: rikuhiro@gmail.com; uedah-tky@umin.ac.jp

Funding information
Exploratory Research for Advanced Technology, Grant/Award Number: JPMJER2001; Japan Science and Technology Agency, Grant/Award Number: JPMJMS2023; Japan Society for the Promotion of Science, Grant/Award Number: JP21K06385; Human Frontier Science Program, Grant/Award Number: RGP0019/2018; MEXT Quantum Leap Flagship Program, Grant/Award Number: JPMXS0120330644; AMED, Grant/Award Number: JP20am0401011; RIKEN Center for Biosystems Dynamics Research

Edited by: Konstantinos Kompotis

Abstract

Electroencephalogram (EEG) and electromyogram (EMG) are fundamental tools in sleep research. However, investigations into the statistical properties of rodent EEG/EMG signals in the sleep–wake cycle have been limited. The lack of standard criteria in defining sleep stages forces researchers to rely on human expertise to inspect EEG/EMG. The recent increasing demand for analysing large-scale and long-term data has been overwhelming the capabilities of human experts. In this study, we explored the statistical features of EEG signals in the sleep–wake cycle. We found that the normalized EEG power density profile changes its lower and higher frequency powers to a comparable degree in the opposite direction, pivoting around 20–30 Hz between the NREM sleep and the active brain state. We also found that REM sleep has a normalized EEG power density profile that overlaps with wakefulness and a characteristic reduction in the EMG signal. Based on these observations, we proposed three simple statistical features that could span a 3D space. Each sleep–wake stage formed a separate cluster close to a normal distribution in the 3D space. Notably, the suggested features are a natural extension of the conventional definition, making it useful for experts to intuitively interpret the EEG/EMG signal alterations caused by genetic mutations or experimental treatments. In addition, we developed an unsupervised automatic staging algorithm based on these features. The developed algorithm is a valuable tool for expediting the quantitative evaluation of EEG/EMG signals so that researchers

Abbreviations: EEG, electroencephalogram; EMG, electromyogram; FFT, fast-Fourier transform; GHMM, Gaussian hidden Markov model; GMM, Gaussian mixture model; NREM, nonrapid eye movement; PSD, power spectrum density; REM, rapid eye movement.

Classification: Sleep, Data science, Electrophysiology, Neuroscience

can utilize the recent high-throughput genetic or pharmacological methods for sleep research.

KEYWORDS
automated sleep staging, EEG/EMG, sleep–wake cycle

1 | INTRODUCTION

Sleep is a fundamental process that maintains our physiological balances through the interaction of two regulatory components: sleep homeostasis and circadian clock. S.A. Brown and his group revealed that the circadian clock provisions synapse with transcripts that are later used in response to sleep–wake cycles (Brüning et al., 2019; Noya et al., 2019). They also showed that the mammalian central clock system, the suprachiasmatic nucleus, plays a role in coordinating the transition between sleep and wakefulness (Collins et al., 2020). These findings highlighted that the two regulatory components of sleep are more intermingled than was thought, demanding further identification of genetic elements and their interactions in the sleep regulatory systems (Franken & Dijk, 2024).

The spontaneous oscillatory activity in the brain during sleep and wakefulness generates a strong electrical field, which is readily measurable on the skull through electroencephalography (EEG) (Buzsáki, 2002). Together with electromyography (EMG), EEG/EMG has been a fundamental tool in investigating neural activities and their underlying molecular or genetic factors involved in sleep regulation. In recent years, genetic methodologies in mice have seen rapid and significant advancements, enabling more efficient perturbation of genes of interest (Tatsuki et al., 2016; Tone et al., 2022). However, the development of EEG/EMG analysis technology to assess the perturbation results has lagged behind.

In sleep studies, researchers typically record several days of EEG/EMG data per animal. Each record is divided into short epochs (segments) of several seconds in length, and human experts assign sleep stages—such as wakefulness, rapid-eye-movement sleep (REM sleep), or non-REM (NREM) sleep—to each epoch. The staging of tens of thousands of epochs, however, consumes a substantial amount of time for an expert.

Recently, there are tools that can automatically score the sleep stages from EEG/EMG records by machine learning (Ellen & Dash, 2021; Grieger et al., 2021; Yamabe et al., 2019). However, to build an effective machine learning model, humans must prepare a large amount of data, and because of the dependence on the prepared data, performance is not always guaranteed for data measured under different conditions. This might be a severe hindrance in exploring sleep phenotypes because EEG signals could be altered by multiple factors, such as the developmental stages of animals and mutations in sleep-related genes (Joho et al., 1999; Panagiotou et al., 2017). There was an effort to address these issues by incorporating training data from different laboratories, different genotypes, and even different species (Miladinović et al., 2019). However, it is generally challenging for humans to interpret the scoring criteria acquired in the model. This precludes researchers from investigating the reasons for automatically scored stages that differ from intuition or further examining the features in the EEG/EMG data.

Traditionally, the fast-Fourier transform (FFT) power spectrum is utilized by experts to analyse EEG time series data for sleep staging. In the NREM stage, there is a notable increase in power within the delta frequency band (below 4 Hz), contrasting with wakefulness, where an elevation in power is observed in the high-frequency gamma band (above 30 Hz) (Buzsáki et al., 1992; Franken et al., 1994). During the REM stage, the EEG frequency characteristics bear resemblance to those of wakefulness, with a distinct increase in theta power (4 to 10 Hz), accompanied by a marked reduction in EMG signal strength. These distinguishing features have been employed for decades as key indicators of the brain's neural activity throughout the sleep–wake cycle, despite the advent of modern computational methods (Rayan, Agarwal, et al., 2024).

In this study, we aimed to maintain simplicity and facilitate intuitive interpretation of the data by providing intuitive criteria based on the classical definition of sleep stages. To refine the established features to simple and intuitive features, we have acquired mouse EEG/EMG data and re-examined the spectrum power distribution across the 0- to 50-Hz frequency. Our analysis revealed a symmetrical shift in spectrum power distribution between wakefulness and NREM sleep, centred around the 20- to 30-Hz frequency band. Notably, during NREM sleep, there is an increase in low-frequency power (below 30 Hz), in contrast to an elevation in high-frequency power (30 to 50 Hz) during wakefulness. This pattern suggests that the high-frequency band, beyond traditional delta and theta frequencies, holds significant potential for sleep–wake classification. We demonstrated that a comparative analysis of low- and high-frequency power might contribute to the accuracy of sleep–wake determination. Additionally,

the utilization of the statistical principle that the aggregation of frequency powers converges towards a Gaussian distribution further enhanced the reliability of our staging methodology. Moreover, by integrating the notable rise in theta power in relation to delta power and the concurrent decrease in EMG signal intensity during REM sleep, we showed that a distinct metric (REM-metric) effectively differentiates REM sleep from wakefulness.

Based on these findings, we supposed that a three-dimensional (3D) space is defined by the three features: the low-frequency power, the high-frequency power, and the REM-metric. In this space, each epoch of EEG/EMG time-series data takes a position, forming distinct clusters according to their origin, that is, wakefulness, NREM sleep, or REM sleep. These clusters are expected to be close to Gaussian distributions. Also, the transition of stages is far from random, with a high probability of remaining in the same stage over time. This property of clusters underscores the utility of a probabilistic approach. To model transitions of stages, we employed the Gaussian hidden Markov model (Gaussian-HMM), an unsupervised, versatile, and lightweight probabilistic model. Building upon the 3D space of stage clusters modelled with Gaussian-HMM, we developed an automated tool for vigilant stage determination from EEG/EMG time-series data, named FASTER2, an advanced version of our previously reported EEG/EMG automated staging tool, FASTER (Sunagawa et al., 2013). FASTER2 implemented a novel algorithm that significantly enhances the process of sleep stage determination.

We validated FASTER2 by applying it to both publicly available EEG/EMG datasets and datasets collected from two different wildtype mice strains (C57BL/6N and DBA/2) and mice with sleep gene mutations measured in our laboratory. FASTER2, being unsupervised, does not require the preparation of teaching data and parameter optimization, hence is easily applicable to a range of sleep research purposes. Moreover, its computation is lightweight, with most of the operation time being spent on EEG/EMG data reading and for graph output. As an open-source resource, FASTER2 is readily accessible to the research community.

2 | RESULTS

2.1 | Powers of frequencies higher than 30 Hz have information to distinguish wakefulness and NREM sleep

The power spectrum derived from the EEG using the fast Fourier transformation (FFT) has been a standard metric for defining sleep stages. The delta power (<4 Hz) and the theta power (4–10 Hz) are especially reliable factors as they are prominent in NREM sleep and REM sleep, respectively. While numerous automated sleep staging tools focused on analysing frequencies below 30 Hz, including delta and theta powers (Grieger et al., 2021; Van Gelder et al., 1991; Veasey et al., 2000), studies showed that the gamma power (>30 Hz) elevates during wakefulness in rodents, suggesting that higher frequency powers can offer valuable insights for differentiating sleep stages (Buzsáki et al., 1992; Franken et al., 1994). To explore this potential, we analysed EEG/EMG signals from an 8-month-old wildtype mouse (C57BL/6N) over a 5-day period. We divided the recordings into 8-s segments (epochs) and manually assigned each to a sleep stage (NREM, Wake, or REM) (Figure 1a). During this process, we observed instances where EMG signals contradicted expected patterns, showing high signals during NREM sleep or low signals during wakefulness, which indicated the EEG signals as a more reliable metric for sleep stage classification (Figure 1b).

To make an objective metric, we subjected each epoch to FFT by using the Welch method to estimate the power spectrum density (PSD) (Virtanen et al., 2020). The PSD indicated the apparent increase of the power at the lower frequency band in NREM sleep over wakefulness (Figure 1c). In contrast, the difference in the powers at the higher frequency band was unclear as the mean values of powers were small both in NREM sleep and wakefulness (Figure 1c). Although this relatively weak power in the high-frequency domain made the high-frequency powers look less informative, we saw there was a linear relation between the mean values and the standard deviation of the powers across frequency domains (Figure S1). This linear relation indicated that the logarithmic transformation of powers makes the distribution closer to the normal distribution and statistically more tractable (Montgomery et al., 2021). Indeed, the difference in powers became apparent at the higher frequency domain after the logarithmic transformation (Figure 1d). Then, we normalized the distribution of the logarithmically transformed powers, making (mean, SD) = (0, 1) at all frequency bins over all epochs. This normalization allowed us to fairly compare powers over the whole range of frequencies and showed that the difference between NREM sleep and wakefulness in the higher frequency domain was as large as in the lower frequency domain (Figure 1e). The same logarithmic transformation and normalization showed that the PSD of EMG signals also exhibited a comparable amount of difference between NREM sleep and wakefulness at lower and higher frequency bands (Figure 1f).

We then calculated the summations of PSD values for the lower frequency band (<20 Hz) and the higher

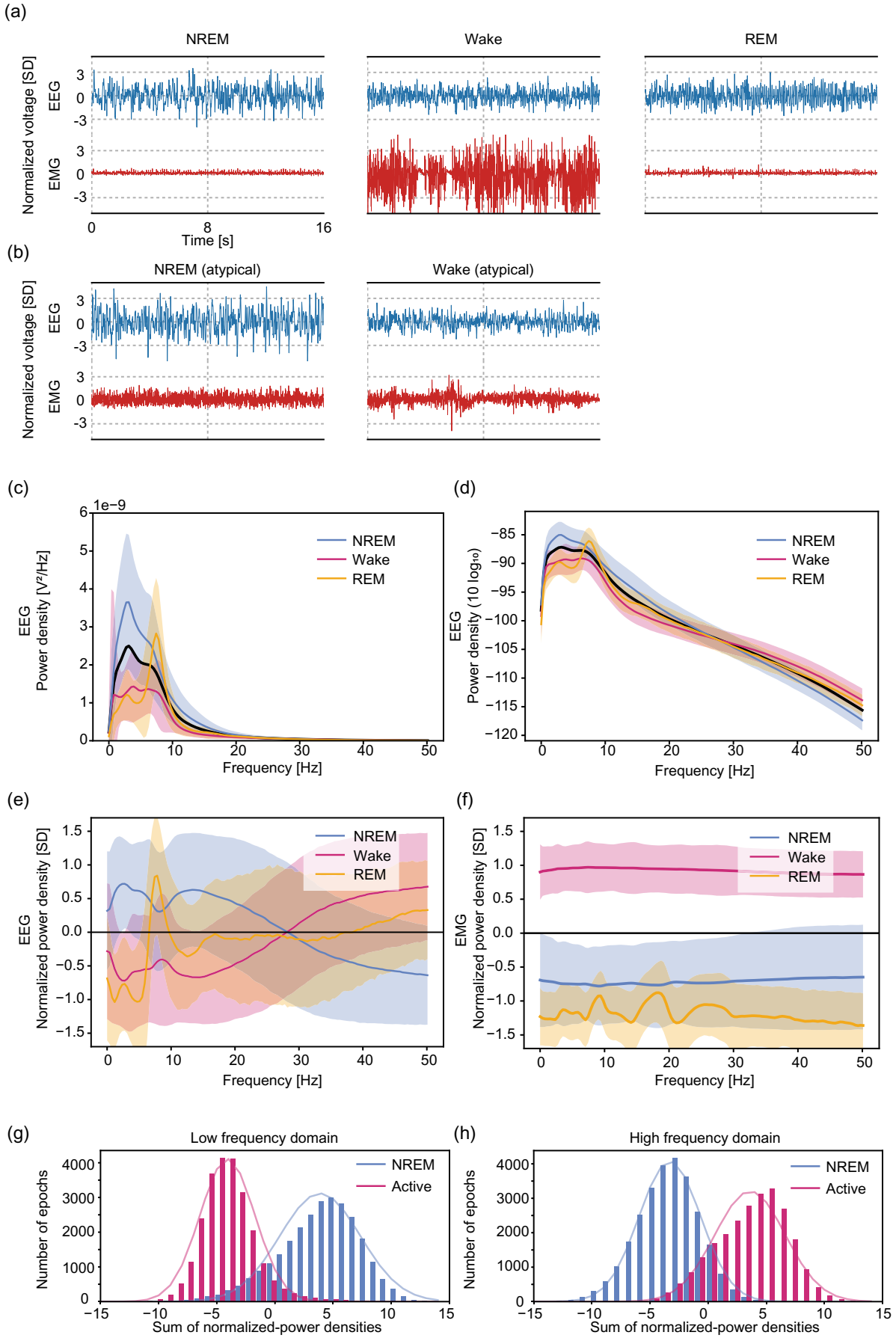


FIGURE 1 The normalized power spectrum density (PSD) of EEG shows a contrast pattern between NREM sleep and wakefulness. (a) The typical EEG and EMG signals of each sleep stage; NREM sleep, wakefulness, and REM sleep over 2 epochs (16 s). (b) The atypical EEG and EMG signals of NREM sleep and wakefulness. The whole 5 days recorded voltages were normalized to have the mean and standard deviation of 0 and 1, respectively in (a) and (b). (c) The PSD of each sleep stage. The black line is the mean of power densities across sleep stages in the 5 days recording. (d) The log-transformed PSD of each sleep stage. The black line is the mean of power densities across sleep stages. (e) The normalized PSD of each sleep stage. The mean and standard deviation of power densities across sleep stages are on the *x*-axis and ±1, respectively. (f) The normalized PSD of each sleep stage derived from EMG. The mean and standard deviation of power densities across sleep stages are on the *x*-axis and ±1, respectively. The coloured lines are means of sleep stages. The shaded area represents ±standard deviation in (c–f). (g) The epoch distribution of the sum of low-frequency (<20 Hz) power densities of NREM sleep or active brain states. (h) The epoch distribution of the sum of high-frequency (>30 Hz) power densities of NREM sleep or active brain states. The normal distribution curves corresponding to each distribution's mean and standard deviation were plotted in (g and h).

frequency band (>30 Hz). We divided each summation by the square root of the number of PSD components within the band, ensuring comparability across different frequency bands. Each frequency band contains 52 frequency components, so we expected that the distribution of summed values of epochs was close to the normal distribution given the central limit theorem. Indeed, the histograms of the summed power values of the lower and higher frequency domains seemed to obey normal distributions with different parameters associated with sleep stages (Figure 1g,h).

2.2 | The EEG signal suffices to distinguish NREM sleep from active brain state, while the EMG signal differentiates REM from wakefulness in the active states

The brain states can be roughly divided into two states: One is the NREM sleep, and the other is the active state. The active state is composed of wakefulness and REM sleep. Because both wakefulness and REM sleep generate similar EEG signals with high frequency and low amplitude, and the amount of REM sleep is relatively small (about 5%) in a day, the summed PSD values of the active brain state form a distribution that is close to normal distributions either with the lower and higher frequency domain (Figure 1g,h).

To highlight the difference in distributions of NREM sleep and the active state, we drew a scatter plot on the two axes of low-frequency and high-frequency normalized powers. Points on the scatter plot were epochs of the EEG signal, and the summed PSD values of each epoch were calculated for low-frequency and high-frequency domains. The scatter plot clearly showed that the epochs of active and NREM sleep formed distinctive clusters on both sides of the diagonal line (Figure 2a). Notably, the separation was along the line perpendicular to the diagonal line, implying that the low- and high-frequency powers are equally informative in classifying NREM sleep and the active brain state (Figure S2). Indeed, the diagonal line had a good separation of NREM sleep and the active state. The 96.8% of NREM sleep epochs and 96.0% of active epochs were under and above the diagonal line, respectively. This observation indicated that the analysis of EEG signals could achieve a practically sufficient performance without relying on EMG signals for separating NREM sleep and the active state.

When we labelled REM epochs in the two-dimensional (2D) plane, we found REM epochs were mostly in the active cluster and overlapped with the cluster of wakefulness (Figure 2a). To separate REM sleep from wakefulness, we introduced an additional metric. During REM sleep, the theta power in the EEG is higher compared with delta power (Figure 1e), and the EMG signals are lower than during NREM sleep (Figure 1f). In addition, the metric defined as *theta-frequency power/(delta-frequency power * muscle power)* was proven useful (Soltani et al., 2019). Based on these observations, we defined the REM-metric as follows:

$$REM\_metric = \frac{1}{\sqrt{n_{theta\_freq}}} \sum_{theta_frequencies} EEG_{normalized\_power} - \frac{1}{\sqrt{n_{delta\_freq}}} \sum_{delta_frequencies} EEG_{normalized\_power} - \frac{1}{\sqrt{n_{muscle\_freq}}} \sum_{muscle_frequencies} EMG_{normalized\_power}$$

where *theta_frequencies*, *delta_frequencies*, and *muscle_frequencies* are the frequency bins in the ranges from 4 to 10 Hz, from 0 to 4 Hz, and >30 Hz, respectively. n_{theta_freq}, n_{delta_freq}, and n_{muscle_freq} are the numbers of frequency bins in each range. $EEG_{normalized\_power}$ and $EMG_{normalized\_power}$ are the normalized log-transformed powers of EEG and EMG, respectively. We focused on the high-frequency domain of EMG normalized powers to define the muscle power because the mean of low-frequency powers was variable depending on frequencies (Figure 1f).

By introducing this REM-metric as the *z*-axis to the two-dimensional plane, we plotted a 3D scatter plot of epochs. This scatter plot showed a clearly separated

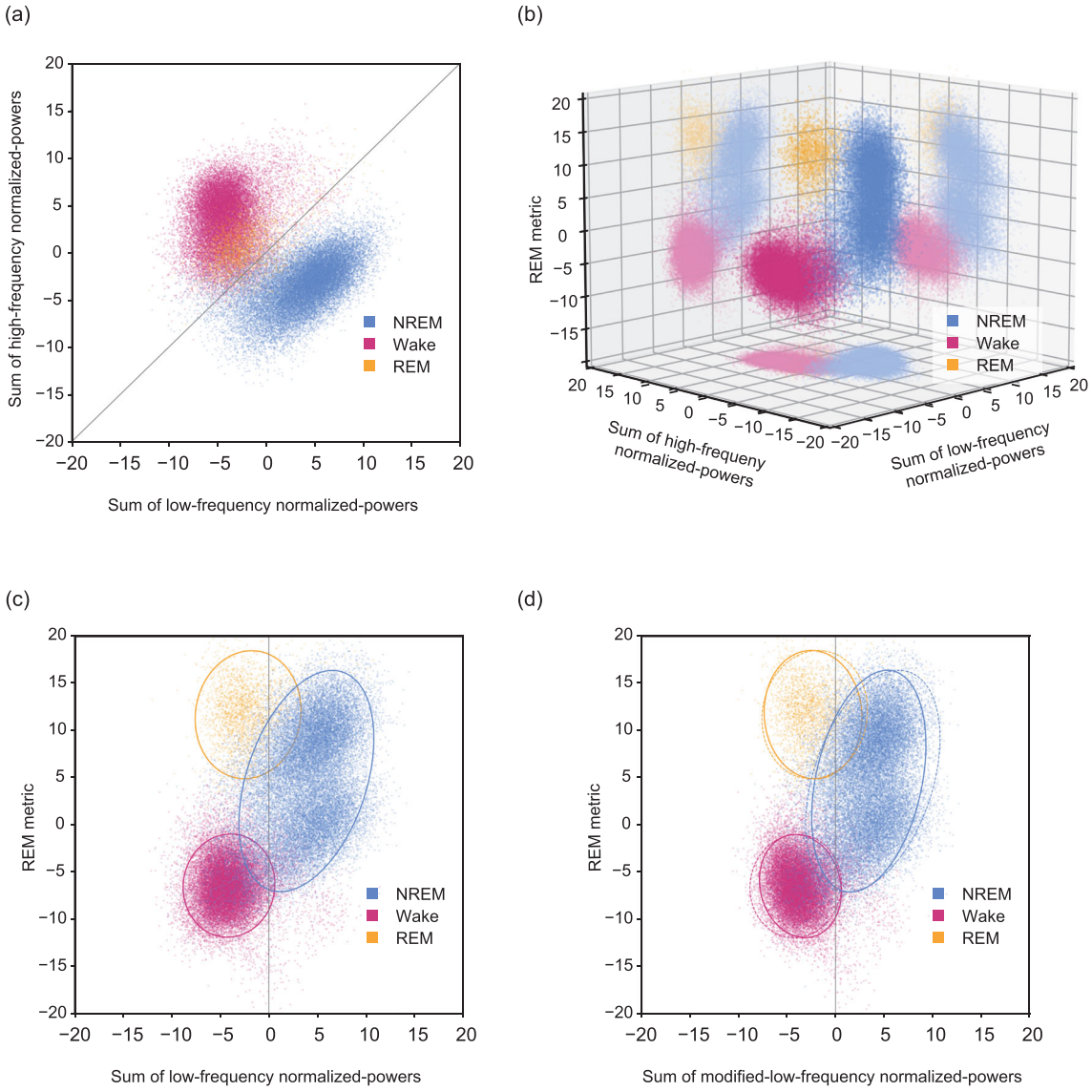
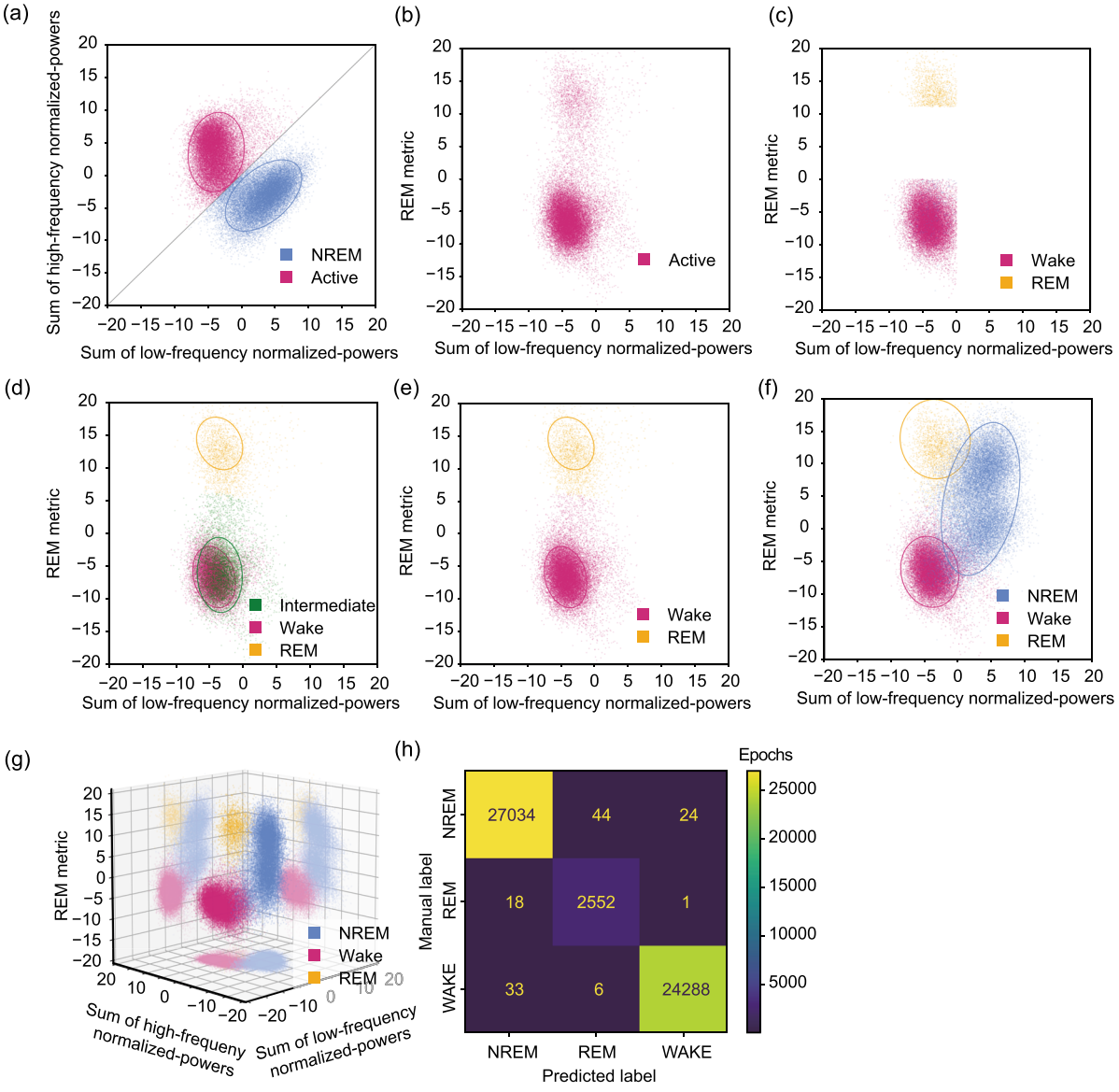
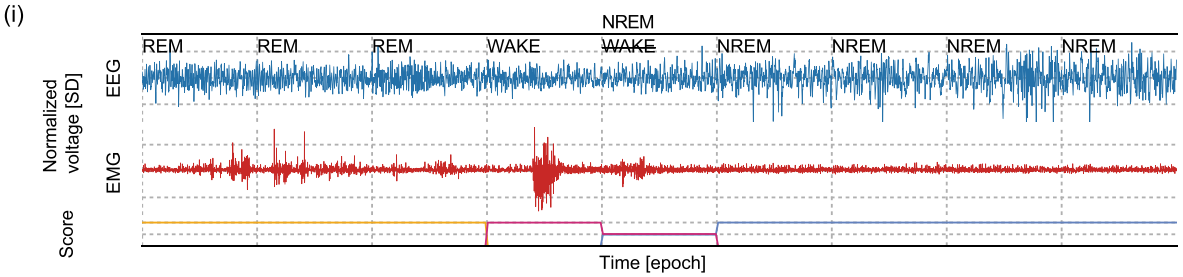


FIGURE 2 Manually annotated epochs form separated sleep-stage clusters that are close to normal distributions. (a) Three sleep-stage clusters plotted on the 2D space spanned by the two metrics; low-frequency and high-frequency normalized powers. The NREM sleep and wakefulness clusters distribute above and below the diagonal line. The REM sleep cluster largely overlaps with the wakefulness cluster and is hardly visible. (b) Three sleep-stage clusters plotted on the 3D space expanded from the 2D space of (a) by adding the REM-metric as *z*-axis. (c) Three sleep-stage clusters plotted on the 2D space spanned by the two metrics: low-frequency powers and REM-metric. The low-frequency powers include all frequency bins in the range from 0 to 20 Hz. Ellipsoids represent the contour of the 95% confidence area of normal distributions of each cluster. (d) Three sleep-stage clusters plotted on the 2D space spanned by the two metrics: modified-low-frequency powers and REM-metric. The modified-low-frequency powers are the frequency powers given by subtracting the theta frequencies (4–10 Hz) band from the low-frequency powers. Ellipsoids with solid lines represent the contour of the 95% confidence area of normal distributions of each cluster in the space. Ellipsoids with the dashed lines are the same contours plotted in (c).

REM cluster above the wakefulness cluster (Figure 2b). It is noteworthy that because the REM-metric is defined only by linear operations (subtractions) on variables from a normal distribution, it is expected that each cluster—whether REM sleep, NREM sleep, or wakefulness—would follow a normal distribution along the *z*-axis.

In the subsequent analysis, we used a modified version of the low-frequency metric. We modified the metric of *low-frequency normalized power* by excluding the theta band (from 4 to 10 Hz) to reduce the confounding effect of theta-power that rises in various brain activities (Buzsáki, 2002). The exclusion of the theta band made the variance of each stage smaller on the *x*-axis and was





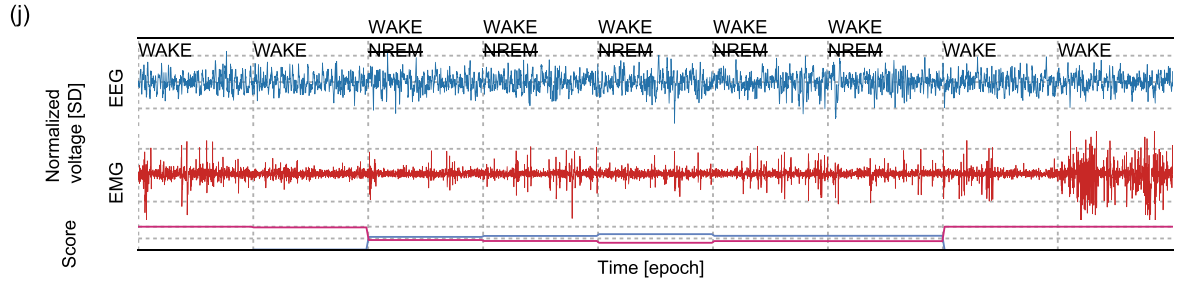


FIGURE 3 Clusters can be automatically annotated in the space of the proposed 3D metrics. (a) The diagonal line separates epochs into two clusters of NREM sleep and the active brain state. (b) The active cluster expanded into the third dimension by adding the REM-metric as the *z*-axis to the 2D space of (a). (c) The clusters we focused based on their coordinates by excluding the intermediate epochs from (b). The bottom cluster was regarded as the core of the wakefulness cluster (red), and the top cluster was regarded as the core of the REM sleep cluster (yellow). (d) The GMM estimated three clusters on the epochs in (c), that is, the REM sleep cluster (yellow), the wakefulness cluster (red), and the intermediate cluster (green). (e) Two clusters resulted from merging the intermediate cluster into the wakefulness cluster in (d), that is, the REM sleep cluster (yellow) and the wakefulness cluster (red). (f) Starting from the three clusters in of (a) and (e), that is, the NREM sleep cluster of (a), REM sleep and wakefulness clusters of (e), the Gaussian hidden Markov model estimated the REM sleep, wakefulness, and NREM sleep clusters in the 3D space by taking the transition probability into account. (g) The 3D view of the automatically staged clusters by the GHMM. The 2D plot was drawn by projecting the 3D space onto the low-frequency and REM-metric axes in (f). Ellipsoids represent the contour of the 95% confidence area of normal distributions in (a and d–f). (h) The confusion matrix derived from the comparison between manual stating and automatic staging with number of matched epochs shown at each intersection. The data consists of 5 days recording of one mouse (54,000 epochs in total). (i and j) Example epochs where human scorer corrected the automatic staging. The automatic stage label is on the left corner of each epoch, and the corrected manual stage is above the strikeout automatic stage if it was manually corrected. The EEG and EMG signals are plotted with blue and red lines, respectively. The whole five-day recorded voltages were normalized to have the mean and standard deviation of 0 and 1. Each dashed vertical bar represents epoch division with a length of 8 (s). The probabilities of wakefulness, REM sleep, and NREM sleep are plotted at the bottom with red, yellow, and blue lines, respectively. An example of correction from wakefulness to NREM sleep (i). An example of correction from NREM sleep to wakefulness (j).

expected to stabilize the performance of classifying the sleep stages (Figure 2c,d).

2.3 | The Gaussian mixture model estimates NREM, REM, and wakefulness clusters

In the previous section, we saw that each cluster of REM, NREM, and wakefulness forms a distribution close to the normal distribution. Based on the observations, we intended to estimate the parameters of each normal distribution by using the Gaussian mixture model. To standardize these summed PSD values, such as the *low-frequency-*, *high-frequency-*, and *muscle normalized-power*, we divided each by the square root of the number of PSD components within the band, ensuring comparability across different frequency bands. By using these three metrics, every epoch of EEG/EMG signals was plotted in the 3D space, and epochs formed a mixture of clusters of REM sleep, NREM sleep, and wakefulness that is each close to a normal distribution.

However, the sizes of each normal distribution are not balanced; the REM cluster constitutes only small number of the total epochs, and there is significant overlap between the distributions. These properties of the unbalanced mixture distribution of REM sleep, NREM sleep, and wakefulness posed a difficulty in estimating the parameters of each component distribution. To robustly estimate each distribution, we introduce a priori knowledge. Namely, REM sleep and wakefulness are both derived from the active state, and their epochs exhibit higher *high-frequency power* than *low-frequency power*. Conversely, NREM sleep epochs show opposite pattern. Therefore, clusters of active state epochs and NREM sleep epochs should be roughly separated by a diagonal line passing through the origin on a two-dimensional plane, as we saw in the previous section (Figures 2a and 3a). All epochs were then projected onto the axis perpendicular to the diagonal line, and the Gaussian mixture model (GMM) classified the NREM sleep and active state's epochs on the axis (Figure S3a).

Next, to separate the REM sleep cluster within the active cluster from wakefulness, we introduced the REM-metric as defined in the previous section to expand the active cluster into the third-dimension space. Given the EMG component of the REM-metric and its susceptibility to contamination (e.g., cardiac signals), the REM sleep cluster sometimes extends into an intermediate zone between REM sleep and wakefulness (Figure 3b). Also, the distribution of the wakefulness cluster varies depending on the animal's behaviour (e.g., moving around, staying still, and eating). Furthermore, the number of epochs in the REM cluster is extremely small compared with the number of epochs in the wakefulness cluster (usually about 1/10). These characteristics of REM sleep and wakefulness clusters make the automatic separation a challenging task. In this study, we decided to start with roughly estimating the "cores" of the wakefulness and REM sleep clusters. First, we excluded epochs within the intermediate zone, only focusing on reliable wakefulness and REM sleep epochs to estimate clusters based on their coordinates (Figure 3c). The core parts of each cluster appear in approximately the same positions in the 3D space regardless of the experimental

conditions because the frequency power is normalized. The centre of the wakefulness cluster can be roughly identified around *(x, y, z) = (−5, −5, −5 to −10)*, and the centre of the REM sleep cluster around *(0, 0, 15 to 20)*. Based on these empirical values, we focused on epochs whose coordinates were *REM-metric < 0 and low-frequency power < 0* as the wakefulness and also focused on epochs whose coordinates were *REM-metric > rem-floor* as REM sleep. The constant *rem-floor* was defined as $\sqrt{n_{theta\_freqs}} + \sqrt{n_{n\_muscle\_freqs}}$ based on the definition of the REM-metric, expecting the mean of data-frequency power to be zero, means of theta-frequency power and muscle-power are higher and lower than at least 1 SD, respectively, in REM sleep (Figures 3c and S3c). We then estimated a GMM with three clusters, namely, the wakefulness cluster, the REM sleep cluster, and an intermediate cluster, using the core coordinates *(x, y, z) = (−5, −5, −10)*, *(0, 0, 20)*, and *(0, 0, 0)*, respectively, as initial values (Garreta & Moncecchi, 2013). We estimated three clusters because, as mentioned, the wakefulness cluster might consist of multiple overlapping clusters reflecting its different behavioural states, such as moving around and staying still (Figures 3d and S3d). Finally, we merged the wakefulness and intermediate clusters and designated them as a wakefulness cluster. With this procedure, we were able to consistently estimate the wakefulness and REM clusters (Figures 3e and S3e).

Having identified distinct clusters for wakefulness, REM sleep, and NREM sleep, we faced the challenge of defining precise cluster borders. The GMM delineates these borders based on the probability distance from each distribution's centre, leading to artificial and sometimes unnatural transitions between sleep stages (Figure 3a). To incorporate the generally low transition probability between stages, we used the Gaussian hidden Markov model (GHMM) (Lebedev, 2016). GHMM is a versatile model that outputs random variables according to normal distributions for each state, that is, REM sleep, NREM sleep, or wakefulness, while transitioning stages according to the Hidden Markov model. We started from the three clusters estimated by the GMM and optimized the GHMM. During this process, we applied constraints to ensure that each cluster does not significantly exceed its boundaries inherent from its definition; that is, the REM sleep cluster should reside in the area where the low-frequency power is less than zero, and the REM-metric is more than zero. Also, we assumed that the principal axes of the 95% confidence area of NREM and wakefulness clusters should not cross the diagonal line (Figure 3a). Through GHMM, we calculated the likelihood of each epoch belonging to a specific sleep stage, assigning stages based on the highest likelihood. This process resulted in labelled epochs that formed three clusters corresponding to REM sleep, NREM sleep, and wakefulness, with nuanced intermingling at cluster boundaries (Figure 3g,h and Table 1). We developed this algorithm in Python, naming the developed tool as FASTER2.

2.4 | FASTER2 is robust and accurate with a practical sufficiency

To evaluate the practical applicability of FASTER2, we labelled each epoch using FASTER2 and evaluated how well these stage labels matched with labelling conducted by a human expert. First, we reanalysed EEG/EMG signals from eight mice measured in our laboratory (Tone et al., 2022). The eight mice were composed of four normal sleep phenotype mice and four long sleep phenotype mice. All mice (C57BL/6N) were injected with adeno-associated virus (AAVs) carrying wildtype or mutant *CamkIIβ* gene. The signals were divided into epochs of 8 s in length, and each epoch was labelled with FASTER2 then corrected with visual inspection for the segment of 24 h. The manual labels were compared with the automatic stages of FASTER2. We found that FASTER2 achieved sufficient accuracy (Table 2).

Next, to evaluate the robustness of our method, we focused on a different inbred strain mouse and used the EEG/EMG data of three wildtype DBA/2Cr mice measured in our laboratory. The signals were divided into epochs of 8 s in length, and each epoch was labelled with FASTER2 then corrected with visual inspection for the segment of 48 h. Interestingly, we found occasional occurrence of brief spindle episodes in EEG, even during wakefulness (Ryan, 1984). The presence of spindles suggested that they should be classified as NREM sleep (Fuentealba & Steriade, 2005). FASTER2 aligned with the expectation and staged those epochs as NREM sleep when the epoch was dominated by the spindle episode, even if it was an isolated epoch surrounded by wakefulness epochs (Table 3).

Furthermore, to confirm that the developed algorithm in FASTER2 was not overly optimized for our laboratory environment (amplifiers or animal lineage), we applied FASTER2 to publicly available rat EEG/EMG data (Ellen & Dash, 2021). Because the dataset was composed of EEG 2ch + EMG 1ch, we simply extracted one channel of EEG signal, converting the EEG/EMG signals to EEG 1ch + EMG 1ch data format, which conforms to the FASTER2. The dataset also contained manual staging. Therefore we could evaluate the FASTER2 without introducing our unconscious bias in the manual staging. This evaluation confirmed that FASTER2 was highly accurate

TABLE 1 Summary of evaluation results of FASTER2's each analysis step.

Mouse ID	NREM-rec	NREM-prec	REM-rec	REM-prec	Active or wake-rec	Active or wake-prec	Accuracy	F1 score	Kappa
Active and NREM separation (Figure 3a)	0.974	0.961	-	-	0.960	0.973	0.967	0.967	0.934
REM and wake separation (coordinates, Figure 3c)	-	-	0.602	0.983	0.984	0.983	0.926	0.938	0.638
REM and wake separation (GMM, Figure 3e)	-	-	0.943	0.958	1.000	0.969	0.968	0.955	0.823
Three stage separation (Figure 3d + NREM)	0.974	0.961	0.827	0.958	0.969	0.969	0.965	0.964	0.935
Three stage separation (GHMM, Figure 3f,g)	0.997	0.998	0.993	0.981	0.998	0.999	0.998	0.998	0.996

Note: The recall scores and the precision scores of each sleep stage (NREM sleep, REM sleep, and wakefulness) were based on the prediction by the FASTER2's each analysis step, given the manual staging as the ground truth. Accu and kappa are the accuracy and Cohen's kappa, respectively. F1 score calculated metrics for each label and find their average weighted by support (the number of true instances for each label). The data recorded for 5 days (54,000 epochs) from one C57BL/6N mice was used for this evaluation.

TABLE 2 Summary of evaluation results of FASTER2 with manually annotated stages of genetically modified C57BL/6N mice.

Mouse ID	NREM-rec	NREM-prec	REM-rec	REM-prec	Wake-rec	Wake-prec	Accuracy	F1 score	Kappa
Normal_01	0.980	0.980	0.938	0.852	0.985	0.997	0.980	0.980	0.964
Normal_02	0.967	0.990	0.983	0.853	0.993	0.991	0.981	0.981	0.966
Normal_03	0.978	0.990	0.990	0.848	0.991	0.996	0.985	0.985	0.973
Normal_04	0.975	0.981	0.938	0.817	0.985	0.996	0.978	0.978	0.959
Long_01	0.978	0.993	0.993	0.878	0.986	0.984	0.982	0.982	0.964
Long_02	0.948	0.991	0.988	0.857	0.982	0.933	0.961	0.962	0.926
Long_03	0.974	0.993	0.998	0.912	0.982	0.957	0.977	0.978	0.953
Long_04	0.970	0.990	0.997	0.843	0.982	0.989	0.977	0.977	0.958

Note: The recall scores and the precision scores of each stage (NREM sleep, REM sleep, and wakefulness) were based on the prediction by FASTER2, given the manual staging as the ground truth. Kappa is Cohen's kappa. F1 score calculated metrics for each label and find their average weighted by support (the number of true instances for each label). The data recorded for 3 days and 24 h of it (10,800 epochs for each mouse) were used for this evaluation.

TABLE 3 Summary of evaluation results of FASTER2 with manually annotated stages of a different inbred strain (DBA/2Cr).

Mouse ID	NREM-rec	NREM-prec	REM-rec	REM-prec	Wake-rec	Wake-prec	Accuracy	F1 score	Kappa
DBA2_1	0.999	0.988	0.992	0.994	0.984	1.000	0.993	0.993	0.987
DBA2_2	1.000	0.967	0.990	0.996	0.969	1.000	0.983	0.982	0.968
DBA2_3	1.000	0.981	0.999	1.000	0.975	1.000	0.990	0.990	0.981

Note: The recall scores and the precision scores of each stage (NREM sleep, REM sleep, and wakefulness) were based on the prediction by FASTER2, given the manual staging as the ground truth. Kappa is Cohen's kappa. F1 score calculated metrics for each label and find their average weighted by support (the number of true instances for each label). The data recorded for 4 days and 48 h of it (21,600 epochs [8 s/epoch] for each mouse) were used for this evaluation.

regardless of laboratory environment and animal lineage, demonstrating that FASTER2 is a robust and versatile tool (Table 4).

FASTER2 assumes that epochs originating from NREM sleep and the active state are distributed on either side of the diagonal line on the two-dimensional plane, which is spanned by the low- and high-frequency normalized power (Figure 3a). Therefore, it was anticipated that the estimation of the Gaussian mixture model might fail when epochs originating from one brain state predominate. That is, if the time series of the EEG signal is biased towards either NREM sleep or the active state, the overall distribution also shifts accordingly, which may cause the centres of NREM sleep and the active state clusters to deviate from the diagonal of the two-dimensional plane (Figure S4a). To correct this, we

TABLE 4 Summary of evaluation results of FASTER2 with manually annotated stages in the publicly available dataset.

Rat ID	NREM-rec	NREM-prec	REM-rec	REM-prec	Wake-rec	Wake-prec	Accuracy	F1 score	Kappa
rat_01	0.976	0.873	0.949	0.762	0.864	0.996	0.919	0.919	0.855
rat_02	0.971	0.868	0.939	0.700	0.860	0.996	0.913	0.914	0.843
rat_03	0.972	0.891	0.927	0.728	0.857	0.996	0.918	0.919	0.853
rat_04	0.980	0.916	0.960	0.765	0.876	0.993	0.933	0.934	0.881
rat_05	0.965	0.921	0.935	0.688	0.889	0.976	0.930	0.931	0.871

Note: The recall scores and the precision scores of each stage (NREM sleep, REM sleep, and wakefulness) were based on the prediction by FASTER2, given the manual staging as the ground truth. Kappa is Cohen's kappa. F1 score calculated metrics for each label and find their average weighted by support (the number of true instances for each label). The data recorded for 1 days and 24 h of it (21,600 epochs [4 s/epoch] for each mouse) were used for this evaluation.

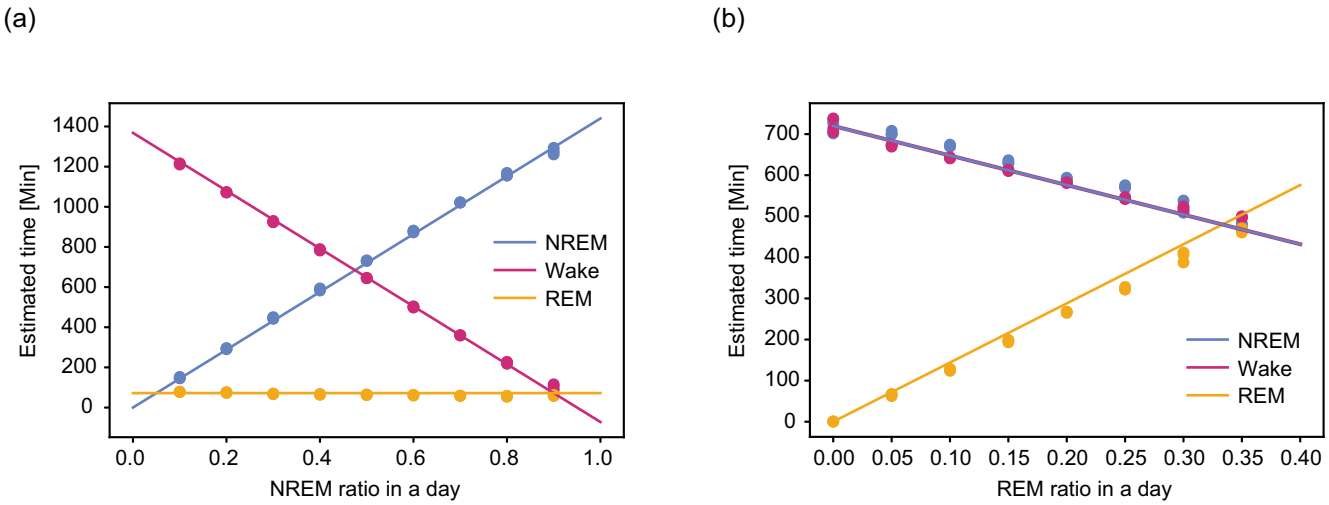


FIGURE 4 FASTER2 can correctly estimate the abnormal amount of sleep-stages. (a) The estimated time of sleep-stages (NREM sleep, wakefulness, or REM sleep) in a day by varying the NREM sleep time. The solid line is the expected time according to the various NREM sleep ratios in a day. The ratio of REM sleep in a day was fixed to 0.05. The dot is the time of sleep-stages estimated by FASTER2. (b) The estimated time of sleep-stages in a day by varying the REM sleep time from 0 to 0.35. The ratio of NREM sleep to wakefulness was fixed to 1. The solid line is the expected time according to the various REM sleep ratios in a day. The dot is the time of sleep-stages estimated by FASTER2.

projected all epochs onto the axis perpendicular to the diagonal line and estimated two normal distributions for NREM sleep and the active state clusters, ensuring that the centre of the means of these normal distributions lies on the diagonal line (Figures S4b and S4c). To ascertain the level of bias our implementation could correct, we generated synthetic EEG/EMG data by sampling NREM sleep and the active state signals at various ratios from actual data and applied FASTER2. We found that up to about 80% bias towards either state did not affect the estimation (Figure 4a), indicating that FASTER2 is applicable even when NREM sleep exceeds 1000 min per day. Moreover, we assessed FASTER2's performance in estimating REM sleep stages. Considering reports of mutants with minimal REM sleep detection (Niwa et al., 2018), we explored scenarios with REM constituting between 0% and 30% of a day. By incorporating an exception-handling mechanism to operate GHMM solely with NREM and wakefulness in the absence of REM, we ensured accurate stage determination across the entire range (Figure 4b). These findings demonstrated FASTER2's practicality for studying not only wildtype but also mutant mice with diverse sleep phenotypes.

3 | DISCUSSION

3.1 | The spectrum power of NREM sleep and the active state crosses around 20–30 Hz

It is well-established that low-frequency power increases during NREM sleep. Similarly, the elevation of high-frequency power, such as gamma waves, during wakefulness is recognized. Our present analysis revealed a distinct shift in the spectrum power on the high-

frequency side, akin to that observed on the low-frequency end, by normalizing the log-transformed spectrum power across all frequencies to a mean and variance of 0 and 1, respectively, based on sleep stages. Consequently, high-frequency normalized power is prevalent in the active state, including wakefulness and REM sleep, while low-frequency normalized power is dominant in brains during NREM sleep. The intersection of these high and low-frequency power predominancies occurs around 20–30 Hz. Intriguingly, this pattern was also observed in rat data acquired externally, indicating that the utilization of both low and high-frequency powers from the power spectrum effectively distinguishes between NREM sleep and the active state across different species and measurement systems. The observation of the enhanced high-frequency power (30–50 Hz) during wakefulness compared with sleep might suggest that functional brainwaves contribute to the high-frequency power component. For example, Gamma waves, typically in the range of 30 to 100 Hz, are known to be involved in higher cognitive functions such as attention and perception and in controlling the timing of sensory input in the cortex (Cardin et al., 2009). The increase in high-frequency power during wakefulness, as identified by our method, aligns with these findings. However, it should be noted that cortical oscillations in this band also occur during sleep (Csicsvari et al., 2003; Steriade et al., 1996), indicating that high-frequency functional brainwaves like Gamma waves are not definitive feature of wakefulness.

These observations suggest that not only the delta power but the entire frequency range of the low-frequency normalized power contained useful information, and the information given by the low-frequency power can be corroborated with the high-frequency power. While many existing automatic sleep staging tools focus on low-frequency powers (Grieger et al., 2021; Miladinović et al., 2019; Van Gelder et al., 1991; Veasey et al., 2000), studies use high-frequency EEG powers such as Gamma, in addition to the Delta, Theta, Alpha, Beta bands along with EMG signals. Interestingly, a study indicated the possibility of identifying sleep substages by training a machine learning model (Boltzmann machine), in addition to traditional sleep stages (Katsageorgiou et al., 2018). Notably, they mentioned in another work that by using a log transformation, the distribution of EEG power becomes closer to a normal distribution, making various analytical methods easier to use (Katsageorgiou et al., 2015). However, because machine learning models and principal component analysis were employed in their works, the combinations of frequency bands to focus on might change depending on dataset. This dependency on the dataset makes it difficult for humans to interpret the obtained sleep stage classification results. In FASTER2, we endeavoured to define features as simply as possible and in line with classical definitions to make the classification results easily interpretable by humans. To achieve this, instead of relying on a single classification method or machine learning model, we designed and implemented a stepwise classification process involving the separation of the active state and NREM sleep using a diagonal line in the feature space, the separation of wakefulness and REM sleep from the active state using GMM, and the separation of NREM sleep, wakefulness, and REM sleep using GHMM.

In FASTER2, a similar bandwidth to the low-frequency power was used for the high-frequency power, which was in the range of approximately 30–50 Hz. Therefore, if EEG/EMG data was recorded with a sampling frequency of 100 Hz or higher, which gives a Nyquist frequency of higher than 50 Hz, FASTER2 can be applied. The sampling frequency of 100 Hz or higher is a relatively standard frequency for EEG/EMG. Hence, we suppose most of the data recorded so far conform to this requirement. However, even if the sampling frequency is over 100 Hz, FASTER2 cannot perform to its full potential if the 30- to 50-Hz bandwidth is filtered out, except with filters strictly limited to removing power-line noise at 50 or 60 Hz.

3.2 | The sampling frequency and epoch length for analysis can be flexibly changed

FASTER2 can be applied to EEG/EMG data obtained in different laboratories with parameters used so far in the laboratory. We developed FASTER2 primarily using data obtained at sampling frequencies of 100 or 128 Hz from C57BL/6N mice in 8-s epochs. However, these parameters, such as sampling frequency and epoch length, often vary by laboratory. For example, the public data that we used for Table 4 was recorded at a sampling frequency of 250 Hz from SD rats with an epoch length of 4 s (Ellen & Dash, 2021). To accommodate data with those different parameters, FASTER2 uses the Welch method for PSD calculation, dividing one epoch into multiple overlapping windows, performing FFT on each window, and averaging the FFTs from each window. In addition, we set the FFT window width to a constant (2.56 s) regardless of the sampling frequency or epoch length to ensure that the frequency bins of the resulting PSD are constant. Thus, changing the sampling frequency and epoch length does not require any adjustments to the following FASTER2 processes, such as scattering each epoch into the 3D space and estimating parameters of normal distributions of stage clusters in the 3D space. A similar process might also be beneficial for analysing various types of

datasets in addition to EEG/EMG, such as acceleration data recorded on the wrist for human activity (Maczák et al., 2023). Additionally, because FASTER2 normalizes power distribution at each frequency bin and uses the summed values of the normalized powers for determining sleep stages, it is robust to differences in the sensitivity at any frequency, which can vary depending on the design and implementation of recording equipment. These properties of FASTER2 allow it to be applicable to data obtained under various conditions in different laboratories without intricate tunings.

3.3 | The FASTER2 is based on the classical definition of sleep stages

This study captured simple statistical properties of EEG/EMG signals in the sleep–wake cycle. We normalized the log-scaled PSDs derived from EEG and EMG signals to make the distribution at each frequency conform to a mean of 0 and a standard deviation of 1 over all the epochs of the recorded data. Then, we spread each epoch into 3D space spanned by the three axes of the low-frequency normalized power, high-frequency normalized power, and the REM-metric. The low-frequency normalized power, high-frequency normalized power, and REM-metrics consist of 37, 52, and 78 normalized PSD values. Because only the linear operations are used in calculating these three metrics, the resulting metric values are expected to approach a normal distribution according to the central limit theorem. This property of the metrics made the subsequent analysis simple. We could expect that clusters of epochs of each sleep stage distribute like a Gaussian distribution in the 3D space. This statistical property of clusters allowed us to apply the useful, popular, and established methods in the subsequent analysis that assume the input dataset is subject to the Gaussian distribution. In addition to the preferable statistical property, the three metrics well separated the clusters of each sleep stage.

Importantly, we could tell where epoch clusters originating from each stage should be located in the 3D space, based on the classical definition of NREM sleep, wakefulness, and REM sleep. NREM sleep is characterized by a higher power in the low frequency compared with the high frequency. The centre of the epoch cluster originating from NREM sleep should be located in the region where $y < x$ on the xy plane within the 3D space where the x-axis and y-axis are for the low frequency and high frequency, respectively. In contrast, because a higher power in the high frequency characterizes wakefulness compared with the low-frequency, the centre of the cluster originating from wakefulness should be located in the region where $y > x$ on the xy plane within the 3D space. During wakefulness, strong EMG signals are measured due to muscle activity for body movement. This activity causes the centre of the wakefulness cluster to be located below the xy plane ($z < 0$); that is, the centre is below the average of all epochs in the z direction. Because REM sleep is defined by EEG waveforms similar to wakefulness, the centre of the cluster originating from REM sleep should be located in the region where $y > x$, which is similar to the wakefulness cluster or near $y = x$. However, during REM sleep, a significant decrease in EMG signals is observed due to muscle atonia, positioning the centre of the REM sleep cluster above the xy plane ($z > 0$) in the 3D space. Furthermore, because the definition of the z-axis includes the relative values of theta power and delta power, epochs corresponding to REM sleep are clearly separated above the wakefulness cluster along the z-axis.

We can approximately predict where each cluster forms a Gaussian distribution in the 3D space according to the classical definition of sleep stages. However, this approximation can be somewhat coarse for actual data. The fitting of two Gaussian distributions to the wakefulness cluster suggests that the wakefulness state can be divided into multiple states. However, the wakefulness and intermediate clusters are close to each other and largely overlap (Figures 3d and S3d), and their separation are not consistent across different measurements. Particularly for the NREM sleep cluster, it occasionally appears divided into two or more distributions along the z-axis. This is because the NREM sleep cluster often has a relatively large spread in the z-direction, corresponding to the occurrence of NREM sleep with stable yet strong EMG signals. The EMG signals during NREM sleep, unlike during wakefulness, have a constant amplitude, suggesting that these EMG signals during NREM sleep reflect muscle tension for maintaining posture rather than body movement. Although this deviation of the NREM sleep cluster from the predicted distribution might seem to confound FASTER2, the NREM sleep cluster still resides in the expected region in the 3D space; $y < x$. In the region, we can safely regard the distorted NREM sleep cluster as a single normal distribution.

On the other hand, if the distribution of EEG/EMG data deviates from the expectation in regard to the position of centres and the shape of spreads, it may indicate that the EEG/EMG signals do not conform to the classical definitions of sleep stages or simply that the signal quality is unacceptably low due to measurement noise or disconnection of signal cables. In any case, the deviation of clusters indicates that FASTER2 cannot provide a reliable judgment of stages, and even experienced human experts would struggle with making judgments on such data. Thus, by simply looking at the scatter plot output

by FASTER2, it is possible to overview the overall quality of the EEG/EMG data.

FASTER2 makes determinations in accordance with classical definitions, making it easy for human researchers to understand the rationale behind the decision results. A possible source of difference in the determination is the transitioning epoch. For example, animals briefly move to change their posture during sleep. This kind of brief movement during sleep might be differently judged by the human scorer and FASTER2. In most cases, our method aligned with humans even for these transition epochs; therefore, we may feel uncertainty about amending the automatic stage result because these epochs looked ambiguous even to experts (Figure 3i). The other potential difference came from epochs where mice were performing repetitive movements, such as grooming over a long period. Such epochs occasionally exhibit an EEG pattern intermediate between NREM sleep and wakefulness, but human experts may be able to judge the epoch as wakefulness by watching the simultaneously recorded video or relatively large variation in the EMG signal (Figure 3j). In either case, the rationale behind FASTER2's judgment is easily interpretable. In the rat data, the recall for wakefulness and the precision for REM sleep were relatively low (Table 4). Upon specific examination of epochs where manual judgment and FASTER2's judgment diverged, many were cases where EMG signals were partially detected within the epoch, confounding FASTER2's judgment of REM sleep from wakefulness. These observations might suggest that incorporating EMG signals into judgment more aggressively can improve the performance of FASTER2. However, it is challenging to introduce an objective standard for judging stages solely based on the EMG signals. Furthermore, these wrong judgments occur relatively rarely because FASTER2 uses HMM, which takes the stage transition probability into account, to suppress the sporadic transition of stages when the EEG/EMG signals are ambiguous. Indeed, when compared with expert judgments, FASTER2's scores are sufficiently high. Hence, FASTER2 can be regarded as consistently providing objective judgment based on probability distributions according to the classical definition, accepting some errors on epochs that exhibit as ambiguously intermediate signals as human experts may also wonder without any additional support, such as video.

In our previous report, we used the probability-density-function-clustering method after principal component analysis (PCA) (Sunagawa et al., 2013). We input concatenated power spectrums of EEG and EMG to the PCA to find a combination of informative frequency powers both from EEG and EMG. This process was based on an intuition that the EMG is a reliable signal in distinguishing NREM sleep from wakefulness because the body is active during wakefulness while resting during sleep. However, it is not rare to see high EMG signals during NREM sleep (Figure 1b). The high EMG activity observed during NREM sleep is often stable in amplitude and continues for tens of seconds to minutes. This relatively long duration of high EMG signals indicates that mice may need to keep muscle tension in order to preserve a comfortable sleeping posture in the environment, including their head attached to the EEG/EMG cable. It is not a trivial task to computationally distinguish this stable muscle tension in sleep from the normal muscle activity in wakefulness. Therefore, we decided not to use the EMG signal for distinguishing a NREM sleep brain from the active brain in the present study. Additionally, we used the classical definition as a priori knowledge to form the feature space instead of using PCA, and we summed powers to make a cluster close to a normal distribution in the feature space. These points successfully contributed to make the EEG/EMG dataset more tractable with simple but functional methods such as GHMM and significantly improved performance.

We emphasized providing judgment criteria based on intuitive and classical indicators. The delta power during NREM sleep, the dominance of high-frequency components above 30 Hz during wakefulness, the relative increase in theta power during REM sleep, and the significant decrease in EMG signals during REM sleep are all metrics long recognized in the field of sleep research as characteristics of each sleep stage. In FASTER2, we redefined these metrics and achieved fully automatic sleep stage classification based on such intuitive indicators. This contrasts with metrics such as the skewness and kurtosis of EEG signal distribution, which are somewhat distant from the researchers' intuition (Karasinski et al., 1994). Unlike many existing automatic sleep stage classification tools, which require training even if they use familiar metrics such as delta power (Gross et al., 2009), FASTER2 does not require the preparation of training data. The metrics adopted in FASTER2 are not novel per se. However, they are based on familiar metrics like delta power and theta power, with slight statistical refinements in preprocessing to accommodate differences in experimental conditions such as sampling frequency. This approach eliminates the need for pretraining or threshold adjustments, significantly improving user convenience.

Various algorithms for automatic sleep staging were developed. Our method has the following four features: (1) Addressing the Need for Standardization: Our method proposes a standardized way to analyse rodent EEG/EMG data, addressing the weakness of criteria in defining sleep stages. This is a significant step forward,

given the current reliance on human expertise for visual inspection, which is time-consuming and prone to variability. (2) Statistical Features: By identifying three simple statistical features that can represent sleep–wake stages in a 3D space, our method demonstrated that the classification process could achieve high accuracy depending on the simple features. This approach not only aligns with traditional definitions but also allows for intuitive interpretation by experts. (3) Unsupervised Automatic Staging: The development of an unsupervised automatic staging algorithm based on these features is a major advancement. Unlike supervised machine learning models that require extensive labelled data for training, our unsupervised approach eliminates the need to prepare large amounts of annotated data. This property of our method makes it more adaptable to various experimental conditions and genetic backgrounds. (4) Broad Validation: The validation of our method on both publicly available datasets and those collected from our laboratory demonstrates its applicability and robustness across different experimental setups. This broad validation supports the generalizability of our method. Of course, these points have been at least partly addressed by existing methods. For example, tools have been proposed that quantify the extent of interexpert variation and can be used consistently across various laboratories regardless of the expertise or animal species (Miladinović et al., 2019). Additionally, automated stage detection tools using widely accepted indicators such as Delta, Theta, Alpha, Beta, and Gamma have also been proposed (Katsageorgiou et al., 2018). Furthermore, tools that provide universal classification results independent of training data by adopting an unsupervised learning approach have been proposed (Sunagawa et al., 2013). More details on various methods are summarized in comprehensive reviews (Rayan, Agarwal, et al., 2024; Rayan, Szabo, & Genzel, 2024). We believe that a recent trend in the development of automated sleep staging tools is the use of machine learning models. This direction might be promising for subdividing sleep and wake stages to perform more detailed phenotypic analyses, and it is also a very interesting study from a data science perspective. However, automatically constructed features by machine learning models do not necessarily align with familiar indicators in the field of sleep research. Additionally, because the models vary depending on the training dataset, it is anticipated that the desired performance may not be achieved across different researchers. We believe that in the field of sleep research, it is first important to classify the three most basic sleep stages (wakefulness, REM sleep, and NREM sleep) independently of experimental conditions. Tools like FASTER2, built from features designed in line with familiar indicators, are expected to be useful in many instances. Once basic sleep stage classification is achieved, it is expected to facilitate the development of more detailed stage classification methods using machine learning models and other approaches.

3.4 | The FASTER2 can be a basis for further analyses

In practical situations, researchers perform subsequent analyses based on the annotated sleep stages. FASTER2 comprises several utility tools that provide basic follow-up analyses, including summaries of each stage times in a day, time series of the hourly stage time over the recording period, and transition probabilities among sleep stages. If the measurement has control and test mice for comparison, FASTER2 performs the basic statistical test, such as a t-test between them.

FASTER2 is a suitable tool to analyse the delta-power dynamics because it can stably annotate stages for a lengthy EEG/EMG data, which is usually 5 days long or longer to include a couple of days of the basal recording followed by sleep deprivation and the recovery period. The delta-power represents sleepiness, and its dynamics is an important measure of sleep phenotype (Borbély, 1982). FASTER2 has a utility tool to estimate parameters of delta power dynamics, such as the time constants of increase (τ_i) and decrease (τ_d) of delta power (Franken et al., 2001), providing the research community with potential avenues for further research such as analysing the delta power dynamics as a sleep phenotype.

There is also unexplored potential in the presented methods. Our method consists of three simple statistical features in line with familiar classical indicators. In the 3D space spanned by these three features, distributions of sleep stage clusters are located in consistent positions regardless of the experimental conditions. Studying how distributions change in this space due to genetic mutations or drug treatments would be interesting. For example, epochs of deep sleep after sleep deprivation are expected to appear far from the REM sleep cluster in the NREM sleep cluster. On the other hand, epochs of anaesthesia-induced sleep would occupy a unique position that physiological perturbations such as sleep deprivation cannot reach. If this is the case, researchers can test the gene function to respond to anaesthesia by seeing the scatter plots of mice having a mutation in the gene.

In the current development of FASTER2, we focused on achieving stable classification of the familiar three stages. However, the possibility of further subdividing the known wakefulness and NREM sleep stages remains an intriguing subject for future investigation. The

wakefulness cluster and NREM cluster can be separated into multiple subclusters because they are often observed that these clusters consist of multiple clusters (Figure 3). More detailed analyses to determine if these distributions correspond to states such as quiet wakefulness, active wakefulness, or light and deep NREM sleep would be a future avenue of research. Additionally, by adding more indicators, the separability of phases like phasic/tonic REM sleep might also be explored (Simor et al., 2020). Furthermore, incorporating an algorithm to detect spindles would enable the detection of transitional states from NREM sleep to wakefulness or REM sleep (Uygun et al., 2019). It is also notable that these analyses will facilitate our understanding of the correspondence between rodent sleep stages and human sleep stages (Rayan, Szabo, & Genzel, 2024).

FASTER2 is publicly available and can be freely used by research communities. We developed FASTER2 to provide sleep researchers with a practical tool for EEG/EMG data analysis, hoping this tool can be a first choice for various researchers and allows comparisons of their results with this common platform and facilitating sleep research. The codes for described analyses from staging to delta power dynamics are available at https://github.com/OrganismalSystemsBiology/faster2.

4 | METHODS

4.1 | Animal preparation and data collection

For evaluation of the FASTER2 performance, we used a C57BL/6N mouse (CLEA Japan, Japan) at 34 weeks old at the recording and three DBA/2Cr mice (SLC Japan, Japan) at 12 weeks old at the recording. Mice were housed in a light–dark controlling rack (Nippon Medical & Chemical Instruments, Japan). All mice were given food and water ad libitum and maintained at ambient temperature and humidity. The light was controlled under 12-h light and 12-h dark cycle. Under anaesthesia, using the mixture of three anaesthetic agents given by intraperitoneal injection: 0.75 mg/kg medetomidine (Domitol, Nippon Zenyaku Kogyo Co., Ltd., Japan), 4 mg/kg midazolam (Dormicum, Astellas Pharma Inc., Japan), 5 mg/kg butorphanol (Vetorphale, Meiji Seika Pharma Co., Ltd., Japan), the C57BL/6N mouse was implanted with electroencephalogram (EEG) and electromyogram (EMG) electrodes for EEG/EMG recordings. To monitor EEG signals, two stainless steel EEG recording screws of 1.0 mm in diameter and 2.0 mm in length were implanted on the skull of the cortex (anterior, +1.5 mm; right, +1.5 mm from bregma or lambda). EMG activity was monitored by stainless steel, Teflon-coated wires with 0.33 mm in diameter (AS633, Cooner Wire, U.S.A) placed into the trapezius muscle. The EEG and EMG wires were soldered to a miniature connector with four pins in 2-mm pitch (Hirose Electric Co., Ltd., Japan). Finally, the electrode assembly was fixed to the skull with dental cement (Unifast III, GC Corporation, Japan). For recovery from anaesthesia, 1.5 mg/kg atipamezole (Antisedan, Nippon Zenyaku Kogyo Co., Ltd., Japan) was used as a medetomidine antagonist. After a 10-day recovery period, the mice were placed in experimental cages with a connection of recording cable. The EEG/EMG signals were amplified (Biotex, Japan) and filtered (EEG, 1–60 Hz; EMG, 5–128 Hz), then digitized at a sampling rate of 128 Hz, and recorded using VitalRecorder software (KISSEI Comtec Co., Ltd., Japan). The recorded kcd file was converted to an EDF file (European Data Format) by SLEEPSIGN software (KISSEI Comtec Co., Ltd., Japan). Both the manual and the automated sleep staging were carried out in the originally developed software.

We anaesthetized the DBA/2Cr animals with isoflurane and implanted telemetry transmitters (Model F20-EET, DSI, USA) for EEG/EMG recording. Two stainless-steel screws (1-mm diameter) were soldered to the wires of telemetry transmitters and inserted through the skull of the cortex (anteroposterior, +1.0 mm; right, +1.5 mm from bregma or lambda) and served as EEG electrodes. Two other wires from the transmitter were placed into trapezius muscles serving as EMG electrodes. Animals were allowed at least 10 days to recover from surgery. EEG/EMG data were recorded wirelessly with food and water available ad libitum. The EEG/EMG data collecting system consisted of transmitters, an analogue-digital converter and a recording computer. Sampling rate was 100 Hz for both EEG and EMG. Gold Acquisition (version 4.00, DSI, USA) was used for the recording. The EEG/EMG data were converted to ASCII format, and both the manual and the automated sleep staging were carried out in the originally developed software. All experimental procedures and housing conditions were approved by the Institutional Animal Care and Use Committee of RIKEN Kobe Branch, and all the animals were cared for and treated humanely in accordance with the Institutional Guidelines for Experiments using Animals.

We also used the EEG/EMG data that was reported from our laboratory. These data came from eight mice that were injected with adeno-associated virus carrying a mutated *CamkIIβ* gene driven by a *CamkIIα* promoter or wildtype *CamkIIβ* gene as a control (Tone et al., 2022). In addition, we downloaded publicly available EEG/EMG data of Sprague–Dawley rats together with the manual stage for 24 h (Ellen & Dash, 2021).

4.2 | Data analysis

We used Python (version 3.10.11) on a desktop PC (Windows10, Intel Core i7-7700, 64-GB memory) for all analysis. For both manual and automated staging, EEG and EMG voltage data were normalized to have mean of zero and standard deviation of 1 along the time. When an epoch has missing values of voltage, we try to recover the epoch by replacing the missing values with values in the same epoch by selecting from smaller indices. In case more than half of the values are missing in an epoch, the epoch is marked as unknown and excluded from the subsequent analysis. The recorded voltage time series of EEG and EMG data were scaled by a single scalar over the recording period to normalize the mean and variance of individual mice to (0, 1). This normalization ensured a consistent plot range on the voltage axis in the later analysis but did not redistribute the power in the power spectrum. The normalized time series is divided into epochs (usually with 8 s length). The power spectrum density (PSD) was then calculated for each epoch by using the scipy.signal.welch() function (Virtanen et al., 2020). Because this function implements the Welch method for estimating the PSD, it further divides an epoch into segments with a specified number of data points (*nperseg*). We used the default value of 256 for nperseg to make segments with an overlap of 50%. The leftovers after the last segment were not evaluated. We specified the number of data points for FFT (*nfft*) to|*256 * sampling_frequency/100*| so that the length of an FFT is around 2.56 s regardless of the sampling frequency. When the sampling frequency is higher than 100 Hz, FFT data points are zero-padded. This assures we obtain powers in the frequency range of 0–50 Hz at ∼0.39-Hz resolution, by taking the first 129 values returned from the function.

The powers were log-transformed with base of 10 and then normalized at each frequency bins so that the mean and standard deviation are respectively 0 and 1 across all epochs. When a normalized power is extremely deviated (more than 3 SD) from the mean, it is replaced with a value selected randomly from a normal distribution (mean, SD) = (0, 1). The normalized powers of EEG and EMG constitute three metrics, that is, the low-frequency power, high-frequency power, and REM-metric. The low-frequency power is composed of EEG powers in the frequency range from 0 to 20 Hz excluding the range of theta powers (from 4 to 10). The high-frequency power is composed of EEG powers in the frequency range from 30 to 50 Hz. The REM-metric is given by *theta-frequency power* − *delta-frequency power* − *muscle power*. The theta-frequency is composed of EEG powers in the frequency range from 4 to 10 Hz. The delta-power is composed of EEG powers in the frequency range from 0 to 4 Hz. The muscle power is composed of EMG powers in the frequency range from 30 to 50 Hz. Each of the high, low, theta, delta, and muscle power is a sum of powers in the corresponding frequency range divided by the squared number of the summed powers. The number of summed powers in each of the low, high, theta, delta, and muscle power are respectively 37 (*n_low_freq*), 52 (*n_high_freq*), 15 (*n_theta*), 11 (*n_delta*), and 52 (*n_muscle_freq*) in the present study.

The clusters of epochs are spread into the 3D space spanned by the three metrics; low-frequency power, high-frequency power, and REM-metric. To classify the NREM sleep and active clusters, epochs are projected on the 2D plane of low-frequency power (*x*-axis), high-frequency power (*y*-axis). For centring the two clusters on the diagonal line ($y = x$), the epochs are projected onto the separation line whose slope is (1, −1), which is perpendicular to the diagonal line, and then the parameters of mean and SD of clusters are estimated on the separation line by using the Bayesian Gaussian mixture model of sklearn (Garreta & Moncecchi, 2013). The all epoch are then shifted along the separation line so that the centre of cluster means is on the diagonal line. In case the estimation of parameters is not converged, another separation line with a different slope (0, −0.5) is tested. This exceptional process may allow for better handling of EEG data, in case it has limited information in the high-frequency power, by focusing on the low-frequency power more than the normal process. If this alternative separation line does work either, the process falls back to the normal separation axis.

The epochs left side of the diagonal line are separated into wakefulness and REM sleep in the 3D space. For the separation, we first selected epochs that are reliably either from wakefulness or REM sleep by focusing on epoch with the condition: (low-frequency power < 0) and ((*REM-metric* < *0*) or (*REM-metric* > *rem-floor*)). The condition: *low-frequency power* < *0* excludes epochs contaminating from NREM sleep, *REM-metric* < *0* selects epochs from wakefulness, and *REM-metric* > *rem-floor* selects epochs from REM-sleep. The constant *rem-floor* is defined as *sqrt(n_theta)* + *sqrt(n_muscle_freq)* based on the definition of the REM-metric, expecting the mean of *data-frequency power* is zero, means of *theta-frequency power* and *muscle-power* are higher and lower than at least 1 SD, respectively in REM sleep. We then applied the Gaussian Mixture Model of sklearn to the selected epochs with the intention to refine REM-sleep and wakefulness clusters. Here we estimated three normal distributions instead of two, with three sets of initial means for REM sleep, wakefulness, and intermediate clusters. This is because the REM-metric includes EMG signal and the

EMG signal apparently originates from various states such as NREM sleep wake and active wake (walking, running, and eating etc.) in the wakefulness, one normal distribution often cannot capture the distribution of epochs from wakefulness. We incorporated epochs of the additional intermediate cluster into wakefulness, because the intermediate cluster positions around *REM-metric* ~ 0, it is more likely from wakefulness rather than from REM-sleep.

By using the estimated parameters (means and covariances) of three clusters as initial parameters of REM sleep, NREM sleep, and wakefulness clusters, we applied GHMM of hmmlearn (Lebedev, 2016) to the whole epochs in the 3D space. In the estimation process, we fixed the means and updated covariances of each clusters along with the transition probabilities across clusters. Because we fixed the means of clusters, we could keep the coverage over the reliable epochs of each cluster and update the shape of clusters by taking the transition probabilities into account. In the updating process, we constrained principal axes of the REM-sleep cluster to stay within the borders of *REM-metric* > 0, *low-frequency power* < 0. Also, we constrained principal axes of NREM sleep and wakefulness cluster to stay below and above the diagonal line, respectively. The length of the principal axis corresponds to 95% confidence area of the estimated normal distribution. In the exceptional case where any effective REM cluster was not found, that is, means of clusters found in the process of REM and wakefulness separation were both *REM-metric* < 0, we estimated the two clusters of wakefulness and NREM sleep by applying GHMM to the whole epochs in the 3D space.

4.3 | Manual annotation of sleep stages

In manual annotation, the EEG signal and the EMG signal were displayed on-screen simultaneously and each 8-s epoch was staged by visual inspection as NREM sleep, REM sleep and wakefulness by the following criteria. NREM sleep was characterized by high-amplitude EEG signals with relatively wide peak-to-peak interval reflecting the enhanced power at the delta-frequency band (0.5–4 Hz) with low or stable EMG power. REM sleep was characterized by low-amplitude EEG with modest width of peak-to-peak interval reflecting the enhanced power at the theta-frequency band (4–10 Hz) with very low EMG power due to the muscle atonia. The wakefulness was characterized by low-amplitude EEG with narrow width of peak-to-peak interval reflecting the enhanced power at the high-frequency band (30 Hz) with high phasic EMG signals.

4.4 | Preparation of the test data

To test the robustness against the biases of sleep stages in the EEG/EMG signals, we synthesized EEG/EMG data of 1 day length from an actual 5 days recording of C57BL/6N mouse. We randomly selected with replacement corresponding number of epochs from the original 5 days data according to the manually annotated stages and shuffled them to compose the 1-day data.

We applied FASTER2 to mice EEG/EMG data obtained in our previous study (Tone et al., 2022). We selected a dataset consisting of eight mice expressing WT *CamkIIβ* or T287D mutant under the *CamkIIα*promoter. The four WT mice exhibited a normal sleep phenotype; NREM sleep was 649 ± 46 min (mean ± SD; $n = 4$), and REM sleep was 85 ± 4 min (mean ± SD; $n = 4$) a day. The four T287D mice exhibited a long sleep phenotype; NREM was 874 ± 54 min (mean ± SD; $n = 4$), and REM sleep was 118 ± 3 min (mean ± SD; $n = 4$) a day. The 24 h of the second day in the 5 days recording was manually staged and compared with the stages given by FASTER2.

We applied FASTER2 to rat EEG/EMG dataset that is publicly available. This dataset contains raw EEG/EMG signals of nine groups of 44 Sprague–Dawley rats, manual staging for 24 h recordings with 4-s epoch. Out of the nine groups, we picked up one group referred as bobmarley in the original paper. Because the original data contains two EEG channels and one EMG channel for a rat, we arbitrarily selected one EEG channel to make a pair of EEG/EMG signals for each rat and combined the rats in the group into one edf file by using a custom Python script with the MNE (Gramfort et al., 2013) and pyedflib (Nahrstaedt & Lee-Messer, 2019). The edf file was then analysed by FASTER2 with setting the epoch length to 4 s. The accuracy was evaluated based on the manually annotated stages in the original paper.

AUTHOR CONTRIBUTIONS

Hiroki R. Ueda, Koji L. Ode, and Rikuhiro G. Yamada designed the study. Rikuhiro G. Yamada performed the study. Kyoko Matsuzawa performed EEG/EMG surgeries and data analyses. Hiroki R. Ueda, Koji L. Ode, and Rikuhiro G. Yamada wrote the manuscript.

ACKNOWLEDGEMENTS

We thank all the lab members at RIKEN BDR and the University of Tokyo, in particular, Y. Uranyu, E. Matsushita, N. Hori, and Y. Sugihara, for their kind help in preparing the materials, breeding, and housing mice. We thank G. Sunagawa for providing EEG/EMG data of DBA/2 mice. This work was also supported by grants from Exploratory Research for Advanced

Technology (ERATO) grant (JPMJER2001, H.R.U.), Moonshot R&D (JPMJMS2023, R.G.Y.) from the Japan Science and Technology Agency (JST), KAKENHI Grant-in-Aid from Japan Society for the Promotion of Science (Scientific Research C, JP21K06385, R.G.Y.), Human Frontier Science Program (HFSP) Research Grant Program (HFSP RGP0019/2018, H.R.U.), MEXT Quantum Leap Flagship Program (MEXT QLEAP, Grant Number JPMXS0120330644, H.R.U.), AMED Innovative Drug Discovery and Development (Grant Number JP20am0401011 H.R.U.), and the intramural Grant-in-Aid from the RIKEN Center for Biosystems Dynamics Research to H.R.U.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/ejn.16465.

DATA AVAILABILITY STATEMENT

The software developed in this study is openly available at https://github.com/OrganismalSystemsBiology/faster2.

ORCID

Rikuhiro G. Yamada https://orcid.org/0009-0003-0788-3176

REFERENCES

Borbély, A. A. (1982). A two process model of sleep regulation. *Human Neurobiology*, *1*, 195–204.

Brüning, F., Noya, S. B., Bange, T., Koutsouli, S., Rudolph, J. D., Tyagarajan, S. K., Cox, J., Mann, M., Brown, S. A., & Robles, M. S. (2019). Sleep-wake cycles drive daily dynamics of synaptic phosphorylation. *Science*, *366*, eaav3617. https://doi.org/10.1126/science.aav3617

Buzsáki, G. (2002). Theta oscillations in the hippocampus. *Neuron*, *33*, 325–340. https://doi.org/10.1016/S0896-6273(02)00586-X

Buzsáki, G., Horváth, Z., Urioste, R., Hetke, J., & Wise, K. (1992). High-frequency network oscillation in the hippocampus. *Science*, *256*, 1025–1027. https://doi.org/10.1126/science.1589772

Cardin, J. A., Carlén, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.-H., & Moore, C. I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature*, *459*, 663–667. https://doi.org/10.1038/nature08002

Collins, B., Pierre-Ferrer, S., Muheim, C., Lukacsovich, D., Cai, Y., Spinnler, A., Herrera, C. G., Wen, S., Winterer, J., Belle, M. D. C., Piggins, H. D., Hastings, M., Loudon, A., Yan, J., Földy, C., Adamantidis, A., & Brown, S. A. (2020). Circadian VIPergic neurons of the suprachiasmatic nuclei sculpt the sleep-wake cycle. *Neuron*, *108*, 486–499.e5. https://doi.org/10.1016/j.neuron.2020.08.001

Csicsvari, J., Jamieson, B., Wise, K. D., & Buzsáki, G. (2003). Mechanisms of gamma oscillations in the hippocampus of the behaving rat. *Neuron*, *37*, 311–322. https://doi.org/10.1016/S0896-6273(02)01169-8

Ellen, J. G., & Dash, M. B. (2021). An artificial neural network for automated behavioral state classification in rats. *PeerJ*, *9*, e12127. https://doi.org/10.7717/peerj.12127

Franken, P., Chollet, D., & Tafti, M. (2001). The homeostatic regulation of sleep need is under genetic control. *The Journal of Neuroscience*, *21*, 2610–2621. https://doi.org/10.1523/JNEUROSCI.21-08-02610.2001

Franken, P., & Dijk, D.-J. (2024). Sleep and circadian rhythmicity as entangled processes serving homeostasis. *Nature Reviews. Neuroscience*, *25*, 43–59. https://doi.org/10.1038/s41583-023-00764-z

Franken, P., Dijk, D. J., Tobler, I., & Borbély, A. A. (1994). High-frequency components of the rat electrocorticogram are modulated by the vigilance states. *Neuroscience Letters*, *167*, 89–92. https://doi.org/10.1016/0304-3940(94)91034-0

Fuentealba, P., & Steriade, M. (2005). The reticular nucleus revisited: Intrinsic and network properties of a thalamic pacemaker. *Progress in Neurobiology*, *75*, 125–141. https://doi.org/10.1016/j.pneurobio.2005.01.002

Garreta, R., & Moncecchi, G. (2013). *Learning scikit-learn: Machine learning in Python*. Packt Publishing.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*, 267. https://doi.org/10.3389/fnins.2013.00267

Grieger, N., Schwabedal, J. T. C., Wendel, S., Ritze, Y., & Bialonski, S. (2021). Automated scoring of pre-REM sleep in mice with deep learning. *Scientific Reports*, *11*, 12245. https://doi.org/10.1038/s41598-021-91286-0

Gross, B. A., Walsh, C. M., Turakhia, A. A., Booth, V., Mashour, G. A., & Poe, G. R. (2009). Open-source logic-based automated sleep scoring software using electrophysiological recordings in rats. *Journal of Neuroscience Methods*, *184*, 10–18. https://doi.org/10.1016/j.jneumeth.2009.07.009

Joho, R. H., Ho, C. S., & Marks, G. A. (1999). Increased γ- and decreased δ-oscillations in a mouse deficient for a potassium channel expressed in fast-spiking interneurons. *Journal of Neurophysiology*, *82*, 1855–1864. https://doi.org/10.1152/jn.1999.82.4.1855

Karasinski, P., Stinus, L., Robert, C., & Limoge, A. (1994). Real-time sleep-wake scoring in the rat using a single EEG channel. *Sleep*, *17*, 113–119. https://doi.org/10.1093/sleep/17.2.113

Katsageorgiou, V.-M., Lassi, G., Tucci, V., Murino, V., & Sona, D. (2015). Sleep-stage scoring in mice: The influence of data preprocessing on a system's performance. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Vol. 2015) (pp. 598–601). IEEE.

Katsageorgiou, V.-M., Sona, D., Zanotto, M., Lassi, G., Garcia-Garcia, C., Tucci, V., & Murino, V. (2018). A novel unsupervised analysis of electrophysiological signals reveals new sleep substages in mice. *PLoS Biology*, *16*, e2003663. https://doi.org/10.1371/journal.pbio.2003663

Lebedev, S. (2016). *hmmlearn*. GitHub. (https://github.com/hmmlearn/hmmlearn)

Maczák, B., Horváth, C. G., Bódizs, R., & Vadai, G. (2023). Revealing the generality of 1/f noise based spectral characteristics of human activity across different datasets. In *2023 International Conference on Noise and Fluctuations (ICNF)* (pp. 1–4). IEEE.

Miladinović, Đ., Muheim, C., Bauer, S., Spinnler, A., Noain, D., Bandarabadi, M., Gallusser, B., Krummenacher, G., Baumann, C., Adamantidis, A., Brown, S. A., & Buhmann, J. M. (2019). SPINDLE: End-to-end learning from EEG/EMG to extrapolate animal sleep scoring across experimental settings, labs and species. *PLoS Computational Biology*, *15*, e1006968. https://doi.org/10.1371/journal.pcbi.1006968

Montgomery, D. C., Peck, E. A., & Geoffrey Vining, G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.

Nahrstaedt, H. & Lee-Messer, C. (2019) holgern/pyedflib 2017. URL: https://github.com/holgern/pyedflib (accessed 21 September 2019).

Niwa, Y., Kanda, G. N., Yamada, R. G., Shi, S., Sunagawa, G. A., Ukai-Tadenuma, M., Fujishima, H., Matsumoto, N., Masumoto, K.-H., Nagano, M., Kasukawa, T., Galloway, J., Perrin, D., Shigeyoshi, Y., Ukai, H., Kiyonari, H., Sumiyama, K., & Ueda, H. R. (2018). Muscarinic acetylcholine receptors Chrm1 and Chrm3 are essential for REM sleep. *Cell Reports*, *24*, 2231–2247.e7. https://doi.org/10.1016/j.celrep.2018.07.082

Noya, S. B., Colameo, D., Brüning, F., Spinnler, A., Mircsof, D., Opitz, L., Mann, M., Tyagarajan, S. K., Robles, M. S., & Brown, S. A. (2019). The forebrain synaptic transcriptome is organized by clocks but its proteome is driven by sleep. *Science*, *366*, aav2642. https://doi.org/10.1126/science.aav2642

Panagiotou, M., Vyazovskiy, V. V., Meijer, J. H., & Deboer, T. (2017). Differences in electroencephalographic non-rapid-eye movement sleep slow-wave characteristics between young and old mice. *Scientific Reports*, *7*, 43656. https://doi.org/10.1038/srep43656

Rayan, A., Agarwal, A., Samanta, A., Severijnen, E., van der Meij, J., & Genzel, L. (2024). Sleep scoring in rodents: Criteria, automatic approaches and outstanding issues. *The European Journal of Neuroscience*, *59*, 526–553. https://doi.org/10.1111/ejn.15884

Rayan, A., Szabo, A. B., & Genzel, L. (2024). The pros and cons of using automated sleep scoring in sleep research. *Sleep*, *47*, zsad275. https://doi.org/10.1093/sleep/zsad275

Ryan, L. J. (1984). Characterization of cortical spindles in DBA/2 and C57BL/6 inbred mice. *Brain Research Bulletin*, *13*, 549–558. https://doi.org/10.1016/0361-9230(84)90037-6

Simor, P., van der Wijk, G., Nobili, L., & Peigneux, P. (2020). The microstructure of REM sleep: Why phasic and tonic? *Sleep Medicine Reviews*, *52*, 101305. https://doi.org/10.1016/j.smrv.2020.101305

Soltani, S., Chauvette, S., Bukhtiyarova, O., Lina, J.-M., Dubé, J., Seigneur, J., Carrier, J., & Timofeev, I. (2019). Sleep–wake cycle in young and older mice. *Frontiers in Systems Neuroscience*, *13*, 51.

Steriade, M., Amzica, F., & Contreras, D. (1996). Synchronization of fast (30-40 Hz) spontaneous cortical rhythms during brain activation. *The Journal of Neuroscience*, *16*, 392–417. https://doi.org/10.1523/JNEUROSCI.16-01-00392.1996

Sunagawa, G. A., Séi, H., Shimba, S., Urade, Y., & Ueda, H. R. (2013). FASTER: An unsupervised fully automated sleep staging method for mice. *Genes to Cells*, *18*, 502–518. https://doi.org/10.1111/gtc.12053

Tatsuki, F., Sunagawa, G. A., Shi, S., Susaki, E. A., Yukinaga, H., Perrin, D., Sumiyama, K., Ukai-Tadenuma, M., Fujishima, H., Ohno, R.-I., Tone, D., Ode, K. L., Matsumoto, K., & Ueda, H. R. (2016). Involvement of Ca^{2+}-dependent hyperpolarization in sleep duration in mammals. *Neuron*, *90*, 70–85. https://doi.org/10.1016/j.neuron.2016.02.032

Tone, D., Ode, K. L., Zhang, Q., Fujishima, H., Yamada, R. G., Nagashima, Y., Matsumoto, K., Wen, Z., Yoshida, S. Y., Mitani, T. T., Arisato, Y., Ohno, R.-I., Ukai-Tadenuma, M., Yoshida Garçon, J., Kaneko, M., Shi, S., Ukai, H., Miyamichi, K., Okada, T., … Ueda, H. R. (2022). Distinct phosphorylation states of mammalian CaMKIIβ control the induction and maintenance of sleep. *PLoS Biology*, *20*, e3001813. https://doi.org/10.1371/journal.pbio.3001813

Uygun, D. S., Katsuki, F., Bolortuya, Y., Aguilar, D. D., McKenna, J. T., Thankachan, S., McCarley, R. W., Basheer, R., Brown, R. E., Strecker, R. E., & McNally, J. M. (2019). Validation of an automated sleep spindle detection method for mouse electroencephalography. *Sleep*, *42*, zsy218. https://doi.org/10.1093/sleep/zsy218

Van Gelder, R. N., Edgar, D. M., & Dement, W. C. (1991). Real-time automated sleep scoring: Validation of a microcomputer-based system for mice. *Sleep*, *14*, 48–55. https://doi.org/10.1093/sleep/14.1.48

Veasey, S. C., Valladares, O., Fenik, P., Kapfhamer, D., Sanford, L., Benington, J., & Bucan, M. (2000). An automated system for recording and analysis of sleep in mice. *Sleep*, *23*, 1025–1040. https://doi.org/10.1093/sleep/23.8.1c

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Yamabe, M., Horie, K., Shiokawa, H., Funato, H., Yanagisawa, M., & Kitagawa, H. (2019). MC-SleepNet: Large-scale sleep stage scoring in mice by deep neural networks. *Scientific Reports*, *9*, 15793. https://doi.org/10.1038/s41598-019-51269-8

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Yamada, R. G., Matsuzawa, K., Ode, K. L., & Ueda, H. R. (2024). An automated sleep staging tool based on simple statistical features of mice electroencephalography (EEG) and electromyography (EMG) data. *European Journal of Neuroscience*, 1–20. https://doi.org/10.1111/ejn.16465