A manifold of heterogeneous sleep and wake states across the cortex

Julia H. Wang[1], Sylvain Chauvette[2], Robert Kwapich[1], Sara Soltani[2], Igor Timofeev[2], and Tatiana A. Engel[1,3]

[1]*Cold Spring Harbor Laboratory School of Biological Sciences, Cold Spring Harbor, NY, USA*

[2]*CERVO Brain Research Centre, Québec, QC, Canada*

[3]*Princeton University, Princeton, NJ, USA*

Corresponding author e-mail: tatiana.engel@princeton.edu

Brain states are conventionally divided into wake, slow wave sleep (SWS) and rapid eye movement (REM) sleep based on distinct patterns of neural activity and muscle tone. These canonical brain states are thought to be discrete, temporally sustained, and spatially global. While some evidence of transition states, spatially local, and temporally brief states shows violation of these canonical assumption, no alternative framework exists to comprehensively and systematically quantify the spatiotemporal dynamics of global brain state. Here we provide such a framework to describe the spatiotemporal continuum of varying brain states We used simultaneously recorded multi-day electromyogram (EMG) and local field potentials (LFP) across the cortex to obtain low-dimensional representation of brain states and quantify their spatiotemporal dynamics. We found differences in state expression across the cortex, such as a lack of REM-like activity in the lateral somatosensory cortex, stereotyped patterns of state transitions across the cortex, and evidence of temporally and spatially local microstates.

Animals naturally and continuously transition between dynamic brain states—internal modes of operation which profoundly influence sensation, cognition, action, and memory[1–3]. Brain states are conventionally divided into wake, slow wave sleep (SWS), and rapid eye movement (REM) sleep based on distinct patterns of oscillatory activity in local field potentials (LFPs) or electroencephalogram (EEG), and muscle tone, measured through electromyography (EMG). SWS is characterized by rhythmic slow waves in electrophysiological activity ($1-4$ Hz) and low muscle tone, while REM is marked by theta band activity ($4-8$ Hz) and muscle atonia (Fig. 1A)[2,4]. The wake state consists of task-dependent, desynchronized neural activity dominated by high frequency fluctuations paired with a high level of muscle tone[2,4,5] (Fig. 1A). Alternations between these brain states of sleep and wakefulness are an integral part of normal brain function. The lack of sleep impairs cognitive function during wakefulness[2], and ongoing variations in brain state during wake can affect sensory responses and behavior[6–9]. Furthermore, SWS and REM both have distinct roles in learning and memory: SWS supports the consolidation of short-term

memories into long-term storage and REM plays a role in stabilizing these memories[10].

The three canonical brain states of SWS, REM, and wake are generally thought to be (1) discrete, (2) temporally persistent, and (3) spatially uniform across the cortex; however, each component of this conceptual model has been challenged by experimental evidence. First, within each canonical state, the patterns of neural activity vary continuously and exhibit signatures of distinct substates, such as quiet wake (QW) and active exploration (AE) substates of the wake state[5,11] or light and deep substates of SWS[2]. Instead of being discrete, these substates constitute a continuum that also describes transitions between states[6,11]. Second, the temporal persistence of states is often interrupted by microstates, which include short periods of wake-like activity in sleep (micro-arousal) and vice versa (micro-sleep), and can occur regularly as well as modulate behavior[12–14]. Third, these states may not be uniform across the cortex; instead some cortical areas may exhibit signatures of a different state of sleep or wake than others at the same time[11,15–18], suggesting the existence of local brain states or heterogeneity in the expression of brain states across areas. These observations altogether indicate that the canonical model of brain states is incomplete. However, no alternative framework exists to comprehensively and systematically quantify the spatiotemporal dynamics of global brain state. Developing such a conceptual framework requires large-scale recordings from multiple brain regions simultaneously during the natural variation of sleep and wake states, along with analytical tools to quantify the expression and propagation of brain states across regions.

Extracting a quantitative description of brain-state dynamics from large-scale, multi-day neural recording data is an open challenge. Traditionally, this classification of brain states into canonical REM, SWS, and wake from short segments of EMG and EEG/LFP data, known as sleep scoring, is achieved through expert hand-labeling the amplitude of neural oscillations in various frequencies. Hand-picked thresholds for combinations of these amplitudes are then used to classify states[19]. These methods require expert intuition and are not robust to potential differences in state expression across subjects or brain regions. As a consequence, there is only a 63–90% consensus between expert labels of brain states[19]. Moreover, manual adjustment of thresholds for different subjects and brain areas is laborious and does not scale to multi-day recordings from many subjects and multiple brain regions. To ease these issues, several supervised deep learning methods were developed to automatically classify brain states[20–24]. However, these methods still treat expert-labels as the ground truth for training, even though expert-labels are subject to inter-rater variability and limited to three basic states[25,26]. In addition, supervised learning models are trained to predict state labels from only a specific set of data, and are prone to fail when presented with data of a differing quality or with distributional shift[27]. This can lead to a lack of consistency across models and problems in generalizing to recordings from different subjects, brain areas, or disease states[19,25]. Thus, supervised methods are unsuitable to elucidate deviations from the three canonical brain states.

In contrast, unsupervised learning methods learn from inherent statistical structure of the data

which is unconstrained by existing expert-labels of three canonical brain states. Therefore, these methods hold the potential to account for continuous variation in neural activity necessary to capture substates and transition states and have the flexibility to describe heterogeneity in expression of states across regions. In particular, unsupervised dimensionality reduction methods, such as variational autoencoders (VAEs)[28] have been used extensively to analyze high-dimensional biological data[29–31] and can produce interpretable low-dimensional representations of high-dimensional data[32].

To uncover the continuous spatiotemporal structure of brain states, we recorded LFPs simultaneously from 14 sites across the mouse cortex for several days during normal wake and sleep cycles. We developed a time-dependent VAE to embed these recordings in a latent space representing the continuous variation of brain states. Our model accurately separated the three canonical brain states in the latent space, while also capturing microstates, transition states, and a continuum of substates with reproducible transition dynamics. The model revealed regional heterogeneity of sleep state expression, including a complete absence of REM sleep in frontal-lateral regions and an enhanced expression of active exploration states in frontal-medial regions. Furthermore, coexistence of differing states across regions occurred primarily during global state transitions or as spatially local microstates. Our work shows that brain states are not always discrete, temporally persistent, and spatially uniform and offers a conceptual model of spatiotemporal dynamics of brain states across the cortex.

Results

A continuous embedding of brain states. We analyzed brain states in extensive multi-regional cortical LFP recordings from two mice. We simultaneously recorded EMG and LFP signals on 14 electrodes implanted across the cortex, capturing data continuously during normal sleep-wake cycles over multiple days, multiple 24-hour sessions per mouse (mouse 1: 10 sessions, mouse 2: 12 sessions; Methods, Data collection). The LFP and EMG showed traditional signatures of SWS, REM, and wake states, including high muscle tone in wake, delta waves in SWS, and theta rhythms in REM (Fig. 1A).

As an initial step, we used data from a single LFP electrode in the visual cortex to develop a model that embeds brain states within a local cortical region into a continuous, low-dimensional latent space. We use power spectral information from the LFP signal and average EMG power in 2 s time windows as input features for the model (Fig. 1B). For each time window, we concatenate the average EMG power and LFP power in 30 frequency bands to form a 31-dimensional feature vector. We present this feature vector to a time-neighbor VAE (TN-VAE) model, a VAE variant designed to predict the feature vector at the next point in time, which mitigates overfitting and leads to more robust representations[33]. The encoder in the TN-VAE compresses the high-dimensional feature vector into a point within a two-dimensional latent space. We explicitly set one dimension in this latent space to be the average EMG power, essential for separating wake from sleep states, and the second latent dimension is learned by the

model. The decoder then receives both latent dimensions to predict the feature vector at the next point in time (Fig. 1B).

To avoid learning spurious features from overfitting to training and validation data [34,35], we selected the optimal model using the neighbor loss, a validation metric promoting temporal smoothness of latent representations[33]. For each dataset, we train many model instances with different initializations and hyperparameter settings (including learning rate, batch size, and regularization weight). We then select the hyperparameter setting and training epoch that yield a model with the minimal neighbor loss, defined as the average normalized distance in the latent space between temporally adjacent data points (Methods: Model architecture). The TN-VAE model, combined with model selection based on the neighbor loss, generates robust latent representations and accurately recovers latent factors on synthetic data[33].

We confirmed the robustness and interpretability of the latent space generated by our model from the LFP data in two ways. First, the model produced consistent latent representations across different initializations and data splits into training and validation sets (Supplementary Fig. 1B). The variability across model instances was smaller than the variability across sessions and mice, reflecting the natural variation in the animals' daily environment and experiences (Supplementary Fig. 1C). Second, the model generalized across sessions and individual animals. When we applied a model trained on data from one mouse to recordings from another mouse, the resulting representations closely matched that of a model trained on the second mouse's data, showing that our model captures consistent, reproducible features (Supplementary Fig. 1D). The consistency of latent space across model instances, animals, and sessions is a prerequisite for interpreting the representations discovered by the model, forming the foundation for our subsequent analyses.

The latent representations derived by our model from a single LFP channel revealed three prominent clusters, corresponding to the canonical wake, SWS, and REM states (Fig. 1C, Supplementary Fig. 2A). To quantify the correspondence between our model and canonical state labels from expert scoring (Methods, Canonical brain states identification), we partitioned latent representations in 3 clusters using a Gaussian Mixture Model (GMM, Methods, Agreement with canonical state labels). The GMM cluster labels showed strong agreement with the canonical state labels (Fig. 1D, Supplementary Fig. 2B; agreement across 24-hour test session, mouse 1: 88.4%, mouse 2: 89.5%), comparable to the coherence observed in inter-expert state scoring[19].

To determine what features our model uses to separate brain states, we examined the distribution of LFP power within classical frequency bands across the latent space. For each 2-second time window, we mapped the average LFP power within four frequency bands (delta $1-40$ Hz, theta$4-8$ Hz, beta $12-30$ Hz, and gamma $30-50$ Hz) to the corresponding location in the latent space (Fig. 1E, Supplementary Fig. 2C). This analysis revealed that our model captures LFP spectral features consistent with the established signatures of canonical states: REM sleep exhibits lower delta power and higher theta

power compared to SWS sleep[4]. Moreover, beta and gamma power also showed systematic variations across the latent space, indicating the presence of additional substates, beyond the canonical definitions. Specifically, beta and gamma power were elevated in the region of the latent space connecting sleep and wake, which may correspond to a drowsy wake substate, known for its high beta and gamma power[4]. Additionally, theta power was enhanced within a subregion of the wake state cluster, which may reflect an active exploration substate, typically associated with elevated theta power[4]. These findings validate that our model captures meaningful electrophysiological features differentiating the three canonical states and enabling the discovery of finer-grained substates.

Substates and dynamics of brain states. The latent space in our model captured continuous, nonlinear variations in LFP power across multiple frequency bands, reflecting natural fluctuations in vigilance states. These representations revealed a more refined architecture of brain states beyond the canonical cycling through wake, SWS, and REM. We therefore hypothesized that different regions within the latent space represent distinct substates, with transition dynamics between them governed by consistent rules.

To test this hypothesis, we used an unsupervised dynamics model to segment latent representations into fine-grained brain states based on their temporal persistence and transition patterns, without relying on expert labels. We fitted the latent embeddings of activity on a single LFP channel with a Hidden Markov Model (HMM), which approximates the data as a sequence of discrete states. At each time step, the sequence can either remain in its current state or stochastically transition to another state, with probabilities depending solely on the current state. Each state maps onto a distinct region within the latent space, modeled as a Gaussian distribution with state-specific mean and covariance. We fitted latent embeddings from all recording sessions to HMMs with varying number of states, ranging from 4 to 15, and selected the optimal number of HMM states using cross-validation (Supplementary Fig. 3, Methods: Substates and dynamics).

An HMM with 8 states was the most parsimonious model, partitioning the latent space into 8 substates of sleep and wake (Fig. 2A). We labeled these substates based on their correspondence with the canonical states and LFP and EMG signatures previously described in the literature[4,11] (Fig. 2A,D, Supplementary Fig. 4A). The canonical wake comprised 4 HMM states: 'active exploration' (AE) with high EMG power and elevated LFP theta power, 'wake' with high EMG power but lower LFP theta power, 'quiet wake' (QW) with intermediate EMG power and desynchronized LFP, and 'drowsy wake' (DW) with yet lower EMG power and increased delta LFP power. The canonical REM corresponded to a single HMM state enriched in LFP theta power at low EMG power. Finally, the canonical SWS encompassed 3 HMM states, which we labeled SWS1, SWS2, SWS3, ranging from lighter to deeper sleep as indicated by progressively lower EMG power.

The transitions between these 8 substates followed sterotypical patterns, with certain transitions occurring more frequently than others (Fig. 2B, Supplementary Fig. 4B). Consistent with previous

literature[11], the brain consistently traversed from wake to SWS before entering REM, and transitions between sleep and wake passed through intermediary quiet or drowsy wake periods. In addition, transitions from sleep to wake and vice versa consistently pass through an intermediary quiet or drowsy wake period. Lighter SWS states transition to deeper SWS states in succession, but deeper SWS only transition to REM, QW, or DW. These stereotypical state transition patterns give insight to how substates of sleep support a continuous traversal of the three basic states.

We next quantified the temporal persistence of brain states by calculating the duration of uninterrupted state sequences across all recording sessions. Histograms of durations per state reveal that most states have a unimodal distribution peaking above 20 seconds ß(Fig. 2C, Supplementary Fig. 4C), consistent with the canonical view of brain states as sustained events. However, DW, QW, and wake have a bimodal distribution, with an additional peak at less than 10 seconds, suggesting the presence of microstates[12] (Fig. 2E), which are defined as short instances of one state within another state. We algorithmically identified microstates as mismatches between a 2-second local state with a surrounding 30-second global state (Fig. 2F, Methods, Microstate identification)[14]. This excludes instances of short-duration states that occur during transitions. We found that these microstates occur rarely, but appear with steretotyped frequency, with some types of microstates more frequent than others (Fig. 2G, Supplementary Fig. 4D). The majority of microstates occur within the two transition regions, reflecting the state duration graphs, and microstate frequency correlates with transition probability (Supplementary Fig. 4E).

Overall, our single-electrode analysis reveal a sophisticated traversal of intermediary substates that make up the continuum of brain state. While states generally persisted for extended periods, they were occasionally interrupted by brief microstates that followed transition patterns similar to full state transitions, suggesting a probing of transition. One caveat was that the substates discovered by the HMM are not always robust across electrodes; specifically, for mouse 2, the HMM reveals different states in the medial somatosensory cortex compared to the visual cortex(Supplementary Fig. 9), indicating regional heterogeneity in brain state expression and dynamics. This observation motivated our subsequent multi-electrode analysis of brain state heterogeneity across the cortex.

Heterogeneous expression of brain states across the cortex. To investigate whether brain states are consistently expressed across cortical regions, we created a shared latent space by training our model on data aggregated from all 14 electrodes. Each data point represented a 2-second time window of data from a single electrode, but data from all 14 electrodes was used as training data for the model. Given our previous observation that substates and their transition dynamics may be different across electrodes, we used a MultiHMM to identify states. A MultiHMM finds states and emissions that are shared across electrodes, but allows for different transition probabilities between states for each electrode (Methods: Dynamics for 14 electrodes). We again used the elbow in cross-validation negative log-likelihood to determine the optimal number of states as 9 states. These states are the same as in the single electrode

case with an additional SWS substate: REM, DW, QW, wake, AE, SWS1, SWS2, SWS3, and SWS4 (Fig. 3A). Following our single-electrode analysis, we visualized how different frequency bands are distributed across the latent space and corresponded the dominant frequency bands for each substate to determine their state label (Supplementary Fig. 10). We found the states to be consistent between animals, further underscoring that any differences between animals when training on one electrode are due to small differences in electrode placement (Supplementary Fig. 5A).

We found heterogeneity in the embeddings across electrodes, as he distribution of points in the latent space differed across electrodes, challenging the canonical view of spatially uniform brain states (Fig. 3A, Supplemental Fig. 5A). Specifically, the lateral somatosensory region lacks time points in the REM region of the latent space, and electrodes in the frontal region show higher prevalence of AE relative to wake, and more QW than DW. We quantified these differences as the percent of time spent in each state per electrode(Fig. 3B, Supplemental Fig. 5B). We additionally rigorously analyzed the apparent REM-state discrepancy in the lateral somatosensory cortex. We hypothesized that during time periods where other regions exhibit REM activity, the lateral somatosensory area continues to exhibit SWS-like activity. We identified time points of "global REM", where 11 or more electrodes were in REM as determined by expert labels. We then calculated the average spectrogram for medial and lateral somatosensory cortex for each of these time points. While medial regions displayed the characteristic theta-dominated REM signature, lateral somatosensory cortex consistently showed spectral patterns typical of SWS, with prominent delta activity and minimal theta power (Fig. 3C, Supplemental Fig. 5C).

The robust regional heterogeneity in brain state expression, particularly the complete absence of REM-like activity in lateral somatosensory cortex and enhanced expression of active exploration states in frontal-medial regions, provides strong evidence against the canonical view of brain states as spatially uniform across the cortex. Instead, our results demonstrate that brain states manifest with distinct regional signatures. This spatial heterogeneity raises a logical next question: how do these regionally diverse expressions of brain states evolve and interact over time? While our analysis thus far has characterized the static distribution of states across cortical areas, a complete understanding of brain state dynamics requires examining how these states propagate across the cortex during transitions and how regional differences manifest temporally.

Spatiotemporal dynamics of sleep and wake. Next, we extended our multi-electrode analysis extended to understand how this spatial heterogeneity evolves through time. By compiling MultiHMM labels across electrodes, we discovered that less than 70% of time points represent truly global states where all 14 electrodes share the same classification (Fig. 4A, Supp. Fig. 7A; mouse 1: 60.5%, mouse 2: 47.3%), further contradicting the canonical view of spatially global brain states. We went on to define "global states" as periods when at least 11 of 14 electrodes expressed the same state. We found that non-global states are largely less than 30 seconds in duration (Fig. 4B, Supplementary Fig. 7B). The prevalence of spatiotemporal heterogeneity varied significantly across brain states. We found that REM,

AE, wake, QW, and DW states are the most non-global states, while SWS states are largely global (Fig. 4C, Supplementary Fig. 7C). In addition, aggregated states of SWS and wake are global, suggesting spatial heterogeneity occuring within substates of the basic state. (Fig 4D).

We hypothesized transitions between states might also occur heterogeneously across the cortex, originating in one area before propagating to another. Focusing on transitions between the 3 basic states, we found the average trajectory over the latent space for each transition type and each electrode. We visualized these average trajectories by mapping the latent space onto the color wheel, creating a spatiotemporal map of transition (Fig. 4E, Supplementary Fig. 7C). This analysis revealed stereotyped patterns of state propagation across the cortex. Transition from SWS to REM sleep arises first from the medial posterior cortex, whereas transitions to SWS appear more global. Transitions from REM to SWS passes through wake for a few seconds, which has been described in previous literature[36]. These findings reveal the spatiotemporal complexity of brain state transitions demonstrating that state changes often originate in specific cortical regions and propagate non-uniformly.

We extended our analysis to characterize the spatial properties of microstates, finding that they typically occurred locally rather than globally across the cortex. (Supplementary Fig. 8A). The frequency of specific microstate types varied systematically across cortical regions (Supplementary Fig. 8B). We calculated the co-occurence rate of REM in SWS and SWS in REM microstates as the percent of microstate time points shared between two electrodes (Supplementary Fig. 8C). While exact patterns of microstate frequency were not always reproducible across mice, microstate co-occurence did have stereotypical patterns across mice, indicating certain electrodes are consistently more likely to be in a microstate at the same time. Our findings additionally provide new insights into local sleep phenomena, extending previous observations of local slow-waves during wakefulness[16]. In local sleep, SWS-like OFF-states are observed de-synchronized in frontal and parietal regions during wakefulness, increasing with sleep deprivation. In our results, we found that co-existence of DW within QW or wake across different areas corresponded to these previously described local slow-waves, with DW exhibiting higher delta power (Supplementary Fig. 6A). We specifically found that these local DW and QW instances were most represented in frontal and parietal regions (Supp. Fig. 6B). Lastly, we found that these time periods of slow waves in wake lasted no more than 30 seconds, which has not been described previously (Supp. Fig. 6C).

These analyses reveal the complex spatiotemporal organization of brain states, demonstrating that state changes often originate in specific cortical regions and propagate non-uniformly. Furthermore, the prevalence of spatially local microstates and stereotyped patterns of state propagation suggest that brain state transitions follow organized spatial dynamics rather than occurring synchronously across the cortex.

Discussion

Our study provides a comprehensive framework for analyzing brain states as a continuous spatiotemporal phenomenon rather than discrete, persistent, and globally uniform categories. We effectively use the TN-VAE[33] to describe a low-dimensional continuum of brain states. This continuum not only corresponds with previously described states of SWS, REM, and Wake, but also extends their definitions to include variations, substates, microstates, and regional heterogeneity. Our unsupervised approach identified states without reliance on expert labels, allowing for discovery of state variations beyond the traditional classifications.

Three key findings emerge from our analysis. First, our model captures a continuous manifold of brain state with distinct substates and transition dynamics, confirming and extending previous observations by Gervasoni et al. (2004) of brain states as a low-dimensional continuum. However, while earlier approaches relied on linear dimensionality reduction of hand-selected frequency band ratios, our nonlinear transformation provides a more robust and generalizable representation that is consistent across brain regions. Second, we systematically quantified microstates, demonstrating that their occurrence patterns reflect the transition probabilities between substates, suggesting they might represent exploratory state transitions. These microstates appear with stereotyped frequencies and show coordinated spatial organization across the cortex. Third, our multi-electrode analysis revealed substantial regional heterogeneity in state expression, most notably the absence of REM-like activity in lateral somatosensory cortex and increased active exploration states in frontal regions.

Our findings unify and extend previous notions of local sleep. The absence of REM-like activity in lateral somatosensory cortex aligns with Nazari et al.'s[37] observation of persistent slow-waves in these regions during REM sleep. This may result from reduced cholinergic projections from the basal forebrain to these regions rather than simple distance from hippocampal theta. Our study also extends upon existing definitions of f microstates, including micro-arousals and micro-sleeps, [12,13,38,39]. Our finding that the most common types of microstates reflect the transition graph between states; perhaps microstates represent a probing of a transition could complement existing theories indicating that the frequency of microstates may be related to experience or stress[40,41]. In contrast to previous studies[16] that studied these local slow waves in sleep-deprived mice, we show these waves are present in normal sleep-wake cycles and occur more in the frontal regions. It is unclear what the function of heterogeneity of brain states may be. Previous studies have shown that brain regions that are more used during the wake periods may require greater sleep intensity or duration[42,43]. Furthermore, we find that microstates occur heterogeneously, which has been heretofore undescribed. One direction of future study is to further analyze the patterns and frequencies of these spatially heterogeneous microstates.

Several technical advances enabled these discoveries. TN-VAE[33] allowed us to move away from discrete categories of wake and sleep by extracting robust, nonlinear transformations that were consistent

across animals, sessions, and brain regions. In addition, the MultiHMM approach allowed us to characterize shared state definitions while permitting region-specific transition dynamics, revealing the rich spatiotemporal organization of brain states. Our methodology is not limited to brain states of wake and sleep; it can be also applied to understand other notions of brain state such as arousal, attention, disease and abnormal states, and more.

Methods

Data collection. We analyzed an experimental dataset described previously[12]. All experiments were performed in accordance with the guideline of the Canadian Council on Animal Care and approved by the Université Laval Committee on Ethics and Animal Research. Two male mice were implanted with 14 LFP electrodes to a depth of 600 μm from the cortical surface to obtain multisite LFP recordings from the major part of the dorsal cortical surface. Two 75 μm electrodes were inserted into the neck muscle to record the electromyogram (EMG) activity. The recordings were done in standard facilities, 24 hours a day, with a light-dark cycle of 12h/12h. LFP and EMG were continuously recorded with a sampling rate of 1 kHz. We analyzed data from 11-13 sessions per mouse (mouse 1—11 sessions, mouse 2–13 sessions), with each session approximately 24 hours (mouse 1 had one 72 hour session, mouse 2 had one 65 hour session, all other sessions were between 21-26 hours). Only sessions at least 1 week after electrode implantation were used.

Canonical Brain States Identification Previous expert-generated brain state labels were used as a point of comparison for our model. These generation of these labels has been described previously[12]. Fast Fourier Transform (FFT) was applied to 5-second windows of LFP and EMG with a sliding window of 1 second. Delta power was defined as the area between 0.2 and 4 Hz, the theta power as the area between 5 and 9 Hz, and the muscle power as the area between 10 and 300 Hz. A hand-set threshold for muscle power was used for wake detection. Another hand-set threshold of delta power was used for SWS state. The theta/(deltaxmuscle) was used as detector of REM episodes, with hand set thresholds per session and mouse. All other time points ($<$1%) were marked as undefined.

Artifact removal. We noticed recording artifacts in the LFP that arose from mouse movement or blink events. Blinking resulted in 1 second slow oscillation events in the LFP, so to remove such blinking artifacts, we high-pass filtered the raw LFP data at 0.8 Hz. Movement artifacts resulted in abnormally high amplitude LFP, so we also remove any time windows where at any point LFP amplitude is greater than 1.5 mV.

Data preprocessing. First, we normalized raw LFP and EMG recordings through z-score for each session, where mean and standard deviation was calculated separately for each session. Then we performed continuous wavelet transformation using complex Morlet wavelets. The formulation of the complex

Morlet wavelets is as follows:

$$\psi(t) = \frac{1}{\sqrt{\pi B}} \exp^{-\frac{t^2}{B}} \exp^{\mathrm{j}2\pi C t} \tag{1}$$

where B is the normalized bandwidth and C is the normalized center frequency. We used a center frequency of 1.0 and a bandwidth of 1.5 to create a base wavelet, and then this base wavelet is scaled at 30 scales corresponding to frequencies evenly spaced on a logarithmic scale between 1 Hz and 50 Hz. For each time step, a wavelet coefficient is obtained for each frequency as a result of the wavelet transformation. The wavelet coefficients are at the same sample frequency as the original data, 1 k Hz. We then calculated the power spectral density (PSD) through $\log_2(|c_t|^2)$, where c_t is the wavelet coefficient at time t. We downsampled the PSD to 100 Hz and then binned to 2-second windows. We averaged the PSD over time within each window. The 30 PSD coefficients of EMG are integrated to obtain the average EMG power. The 30 frequency bands of LFP are combined with the average EMG power to create a 31-dimensional vector input to our model. We z-scored each feature of our input vector separately, with mean and standard deviation calculated per session.

We created a separate dataset per mouse consisting of all recording sessions for that mouse. After data preprocessing, all sessions but one are concatenated into one dataset consisting of all 2 second time windows. The remaining session is held out as test data.

Model architecture. We pass our 31-dimensional vector consisting of 30 frequency bands of LFP and the average EMG power into a TN-VAE[33]. We minimize the following loss function:

$$\mathscr{L} = -\mathrm{E}_{z\sim q_\phi(z_{t+1}|x_t)}[\log p_\theta(x_{t+1}|z_{t+1})] + \beta \mathrm{KL}(q(z_{t+1}|x_t)\|p(z_{t+1})). \tag{2}$$

Here x_t is the data point at time t and z_t is its latent representation. $p_\theta(x_{t+1}|z_{t+1})$ is the likelihood of x_{t+1} conditioned on z_{t+1} under generative model p_θ, and $q_\phi(z_{t+1}|x_t)$ is the approximated posterior of the future latent state z_{t+1} given the observation x_t.

Our encoder consisted of two 250-dimensional fully connected layers and a 2-dimensional latent representation. The decoder architecture also consists of two 250-dimensional fully connected layers, and ReLU activations were used throughout. We specifically constrain one dimension of this 2-D latent representation to be the average EMG power and learn the other dimension from the TN-VAE. Originally, we allowed both dimensions of the latent space to be learned by the model. When we compared the model with expert-generated canonical labels, we saw disagreements occuring only when the model did not give as high importance to EMG as the experts did. Inspired by experts' use of EMG power as a absolute measure of wakefulness, we set one dimension in our latent space to explictly be EMG power. Thus, we have one dimension that represents wakefulness and another that represents cortical LFP activity. Both dimensions are passed to the decoder during training. These models are coded using Tensorflow and Tensorflow-Probability.

Model selection and validation. Datapoints are randomly split into training and validation sets, where 80% of datapoints are in the training set and 20% are in the validation set. We chose the best model hyperparameters based on the lowest Neighbor Loss (NL)in the validation set[33] . We sampled hyperparameter ranges of $5,000$ to $100,000$ for batch size, $1e-3to1e-5$ for learning. We used a β value of 0.001. For mouse 1 visual probe, we used learning rate 10^{-4} and batch size $10,000$. For mouse 2 visual probe, we used learning rate 10^{-4} and batch size $50,000$. Each model is trained for 1000 epochs, where the model sees the entire training dataset per epoch. In order to further validate that the chosen model is robust, we randomly split and initialized the model 3 more times. The pairwise KL divergence was estimated for the resulting encodings for each model and session by estimating the density of the encodings with 50 bins and then calculating the KL divergence between the densities.

Canonical State Label Agreement A Gaussian Mixture Model was fit to the encodings in the training set per mouse to obtain 3 clusters. The confusion matrix between GMM-generated cluster labels and expert-generated canonical labels was calculated on the held-out test set.

Interpretation of encodings through canonical frequency bands. We calculated the spectral power for each 2-second time window in our dataset. To calculate the power for frequency bands delta ($1-4$ Hz), theta ($4-8$ Hz), beta ($12-30$ Hz), and gamma(>30 Hz)[4], we calculated the spectral power of each 2-second window using Fast Fourier Transform (FFT). We then integrated the spectral power within the corresponding range and divided by the total spectral power. We visualized the spectral power for each band on the encodings by coloring by the z-scored power across the session.

Substates and dynamics. We used an HMM to learn an optimal number of substates and transition probabilities between states. We used the ssm package (www.github.com/lindermanlab/ssm) to train the HMM. For each mouse, we trained a separate HMM on data from all sessions. We trained our model using Gaussian emissions to fit the sequence of points on the latent manifold. In order to determine the number of states, we trained HMMs with varying number of states from 2-14. We performed 3-fold cross validation and calculated the average negative log likelihood per number of states. We chose the the number of states when the negative log likelihood decreased by less than 5% when increasing the number of states. For both mice, we found 8 states as optimal for the visual probe. Transition diagrams were visualized using the pyvis package. Only transitions that accounted for more than 15% of transitions to another state were visualized for clarity reasons. (Note that within state transitions probabilities were 75-95%, so 15% of transitions to another state was at most 4% of all transitions).

Microstate Identification. Microstates were identified by comparing mismatches between and a local state label for a 2-second window and the surrounding 30-sec state label[14]. First, we use a 30-second sliding window and define the mode of the window as the surrounding state. Windows without one state occupying at least 20 seconds in the window are discarded. Then, microstate instances are defined as a mismatch between the 2-second window label and the surrounding window.

Creation of shared embedding across 14 electrodes. We aggregated data from all 14 electrodes to obtain a common embedding space for all electrodes. Each data point is a two second window from a single electrode, but the dataset consists of data from all electrodes. We used the same model architecture as in the single electrode case. We again selected hyperparameters using the NL as a model selection metric. For mouse 1, we trained the model using a learning rate of $1e^{-5}$ and batch size of $50,000$. For mouse 2, we trained the model using a learning rate of learning rate of $1e^{-5}$ and batch size of $10,000$. We again validated the model by ensuring different initializations and dataset splits obtained similar encodings.

Dynamics for 14 electrodes. To understand dynamics and substates for 14 electrodes, we trained a modified HMM, known as a Multi-HMM, that would require the same emissions for different electrodes but allow different initial probability and transition probability matrices for each electrode. We did so by modifying the ssm package with custom written code (www.github.com/engellab/ssm). We again trained a separate Multi-HMM for each mouse, on all sessions per mouse. We again used a cross-validation procedure with 5% cutoff to find the optimal number of states, finding 9 states for both mice.

Average state transitions across the cortex. After obtaining labels from the 9-state HMM for each probe, we mapped each of the 9 states to their corresponding "basic state", SWS, REM, or Wake, and took the most common state across probes as the global state. We then smoothed the global state with a sliding window of 40 seconds to identify transitions between states. This is done because the instantaneous (2-second) state often switches a few times between states before fully transitioning. We take the encodings for each electrode for the 32 seconds around the transition indices and take an average across all instances of that transition type to obtain average encoding pattern for that transition type. We visualize this average encoding pattern by coloring each electrode by its position on the manifold at that point in time, where the color is determined by the angle and distance from the origin of that point.

Acknowledgements

This work was supported by the Cold Spring Harbor Laboratory School of Biological Sciences (J.H.W.), the Jenny and Jeff Kelter Neuroscience Scholarship (J.H.W.), NIH grant RF1DA055666 (T.A.E.), Starr Foundation award I15-0037 (J.H.W. and T.A.E.), and NIH grant RF1NS128901 (J.H.W. and T.A.E).

Author contributions

J.H.W., R.K, S.C., I.T., and T.A.E. designed the research. J.H.W., R.K. and T.A.E. developed the computational analysis framework. S.C., S.S., and I.T. designed the experiments. J.H.W. and R.K. developed the code, performed model training, and analyzed neural recordings data. S.C., S.S. and I.T. performed the experiments and the canonical brain state identification. J.H.W. and T.A.E wrote the paper with input

from S.C. and I.T.

Competing interests

The authors declare no competing interests.

Data availability

Data will be provided upon reasonable request to Igor Timofeev (igor.timofeev@fmed.ulaval.ca).

Code availability

The source code to reproduce the results of this study is freely available on GitHub (https://github.com/engellab/braivest).

References

1. Vyazovskiy, V. V. *et al.* Cortical firing and sleep homeostasis. *Neuron* **63**, 865–878 (2009).

2. McNamara, P. *The Neuroscience of Sleep and Dreams*. Cambridge Fundamentals of Neuroscience in Psychology (Cambridge University Press, New York, NY, US, 2019).

3. Steriade, M. & McCarley, R. W. *Brainstem Control of Wakefulness and Sleep* (Springer US, Boston, MA, 1990).

4. Hernan, A. E. *et al.* Methodological standards and functional correlates of depth in vivo electrophysiological recordings in control rodents. A TASK1-WG3 report of the AES/ILAE Translational Task Force of the ILAE. *Epilepsia* **58 Suppl 4**, 28–39 (2017).

5. Poulet, J. F. A. & Crochet, S. The Cortical States of Wakefulness. *Front. Syst. Neurosci.* **12** (2019).

6. Harris, K. D. & Thiele, A. Cortical state and attention. *Nat. Rev. Neurosci.* **12**, 509–523 (2011).

7. McGinley, M. J. *et al.* Waking State: Rapid Variations Modulate Neural and Behavioral Responses. *Neuron* **87**, 1143–1161 (2015).

8. Engel, T. A. *et al.* Selective modulation of cortical state during spatial attention. *Science* **354**, 1140–1144 (2016).

9. Hulsey, D., Zumwalt, K., Mazzucato, L., McCormick, D. A. & Jaramillo, S. Decision-making dynamics are predicted by arousal and uninstructed movements. *Cell Rep.* **43** (2024).

10. Diekelmann, S. & Born, J. The memory function of sleep. *Nat. Rev. Neurosci.* **11**, 114–126 (2010).

11. Gervasoni, D. *et al.* Global Forebrain Dynamics Predict Rat Behavioral States and Their Transitions. *J. Neurosci.* **24**, 11137–11147 (2004).

12. Soltani, S. *et al.* Sleep–Wake Cycle in Young and Older Mice. *Front. Syst. Neurosci.* **13** (2019).

13. Watson, B. O., Levenstein, D., Greene, J. P., Gelinas, J. N. & Buzsáki, G. Network Homeostasis and State Dynamics of Neocortical Sleep. *Neuron* **90**, 839–852 (2016).

14. Parks, D. F. *et al.* A nonoscillatory, millisecond-scale embedding of brain state provides insight into behavior. *Nature Neuroscience* **27**, 1829–1843 (2024).

15. Funk, C. M., Honjoh, S., Rodriguez, A. V., Cirelli, C. & Tononi, G. Local Slow Waves in Superficial Layers of Primary Cortical Areas during REM Sleep. *Curr. Bio.* **26**, 396–403 (2016).

16. Vyazovskiy, V. V. *et al.* Local sleep in awake rats. *Nature* **472**, 443–447 (2011).

17. Nir, Y. *et al.* Regional Slow Waves and Spindles in Human Sleep. *Neuron* **70**, 153–169 (2011).

18. Bernardi, G. *et al.* Regional Delta Waves In Human Rapid Eye Movement Sleep. *J. Neurosci.* **39**, 2686–2697 (2019).

19. Alsolai, H. *et al.* A Systematic Review of Literature on Automated Sleep Scoring. *IEEE Access* **10**, 79419–79443 (2022).

20. Chambon, S., Thorey, V., Arnal, P. J., Mignot, E. & Gramfort, A. A Deep Learning Architecture to Detect Events in EEG Signals During Sleep. In *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, 1–6 (IEEE, Aalborg, 2018).

21. Gunnarsdottir, K. M. *et al.* A novel sleep stage scoring system: Combining expert-based features with the generalized linear model. *J. Sleep Res.* (2020).

22. Prabhudesai, K. S., Collins, L. M. & Mainsah, B. O. Automated feature learning using deep convolutional auto-encoder neural network for clustering electroencephalograms into sleep stages. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 937–940 (IEEE, San Francisco, CA, USA, 2019).

23. Caldart, C. S. *et al.* Sleep Identification Enabled by Supervised Training Algorithms (SIESTA): An open-source platform for automatic sleep staging of rodent polysomnographic data (2020). BioRxiv preprint at http://biorxiv.org/lookup/doi/10.1101/2020.07.06.186940.

24. Allocca, G. *et al.* Validation of 'Somnivore', a Machine Learning Algorithm for Automated Scoring and Analysis of Polysomnography Data. *Front. Neurosci.* **13** (2019).

25. Fiorillo, L. *et al.* Automated sleep scoring: A review of the latest approaches. *Sleep Med. Rev,* **48**, 101204 (2019).

26. Koch, H., Jennum, P. & Christensen, J. A. E. Automatic sleep classification using adaptive segmentation reveals an increased number of rapid eye movement sleep transitions. *Journal of Sleep Research* **28**, e12780 (2019).

27. Barger, Z., Frye, C. G., Liu, D., Dan, Y. & Bouchard, K. E. Robust, automated sleep scoring by a compact neural network with distributional shift correction. *PLOS One* **14** (2019).

28. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes (2014). ArXiv preprint at http://arxiv.org/abs/1312.6114.

29. Luxem, K. *et al.* Identifying behavioral structure from deep variational embeddings of animal motion. *Commun. Bio.* **5**, 1–15 (2022).

30. Sussillo, D., Jozefowicz, R., Abbott, L. F. & Pandarinath, C. LFADS - Latent Factor Analysis via Dynamical Systems (2016). ArXiv preprint at http://arxiv.org/abs/1608.06315.

31. Higgins, I. *et al.* Unsupervised deep learning identifies semantic disentanglement in single inferotemporal face patch neurons. *Nat. Comm.* **12**, 6456 (2021).

32. Khemakhem, I., Kingma, D., Monti, R. & Hyvarinen, A. Variational Autoencoders and Nonlinear ICA: A Unifying Framework. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, 2207–2217 (PMLR, 2020). ISSN: 2640-3498.

33. Wang, J. H., Tsin, D. & Engel, T. A. Predictive variational autoencoder for learning robust representations of time-series data. In *UniReps: the First Workshop on Unifying Representations in Neural Models* (2023).

34. Bartlett, P. L., Long, P. M., Lugosi, G. & Tsigler, A. Benign overfitting in linear regression. *PNAS* **117**, 30063–30070 (2020).

35. Genkin, M. & Engel, T. A. Moving beyond generalization to accurate interpretation of flexible models. *Nature Machine Intelligence* **2**, 674–683 (2020).

36. Adamantidis, A. R., Zhang, F., Aravanis, A. M., Deisseroth, K. & de Lecea, L. Neural substrates of awakening probed with optogenetic control of hypocretin neurons. *Nature* **450**, 420–424 (2007).

37. Nazari, M. *et al.* Regional variation in cholinergic terminal activity determines the non-uniform occurrence of cortical slow waves during REM sleep in mice. *Cell Rep.* **42**, 112450 (2023).

38. Libet, B., Alberts, W. W., Wright, E. W. & Feinstein, B. Responses of Human Somatosensory Cortex to Stimuli below Threshold for Conscious Sensation. *Science* **158**, 1597–1600 (1967).

39. Edelman, G. M. Naturalizing consciousness: A theoretical framework. *PNAS* **100**, 5520–5524 (2003).

40. Maness, E. B. *et al.* Role of the locus coeruleus and basal forebrain in arousal and attention. *BRB* **188**, 47–58 (2022).

41. Smith, J. *et al.* Regulation of stress-induced sleep fragmentation by preoptic glutamatergic neurons. *Curr. Bio.* **34**, 12–23.e5 (2024).

42. Krueger, J. M., Nguyen, J. T., Dykstra-Aiello, C. J. & Taishi, P. Local sleep. *Sleep Med. Rev.* **43**, 14–21 (2019).

43. Rector, D. M., Schei, J. L., Van Dongen, H. P. A., Belenky, G. & Krueger, J. M. Physiological markers of local sleep. *Eur. J. Neuro.* **29**, 1771–1778 (2009).

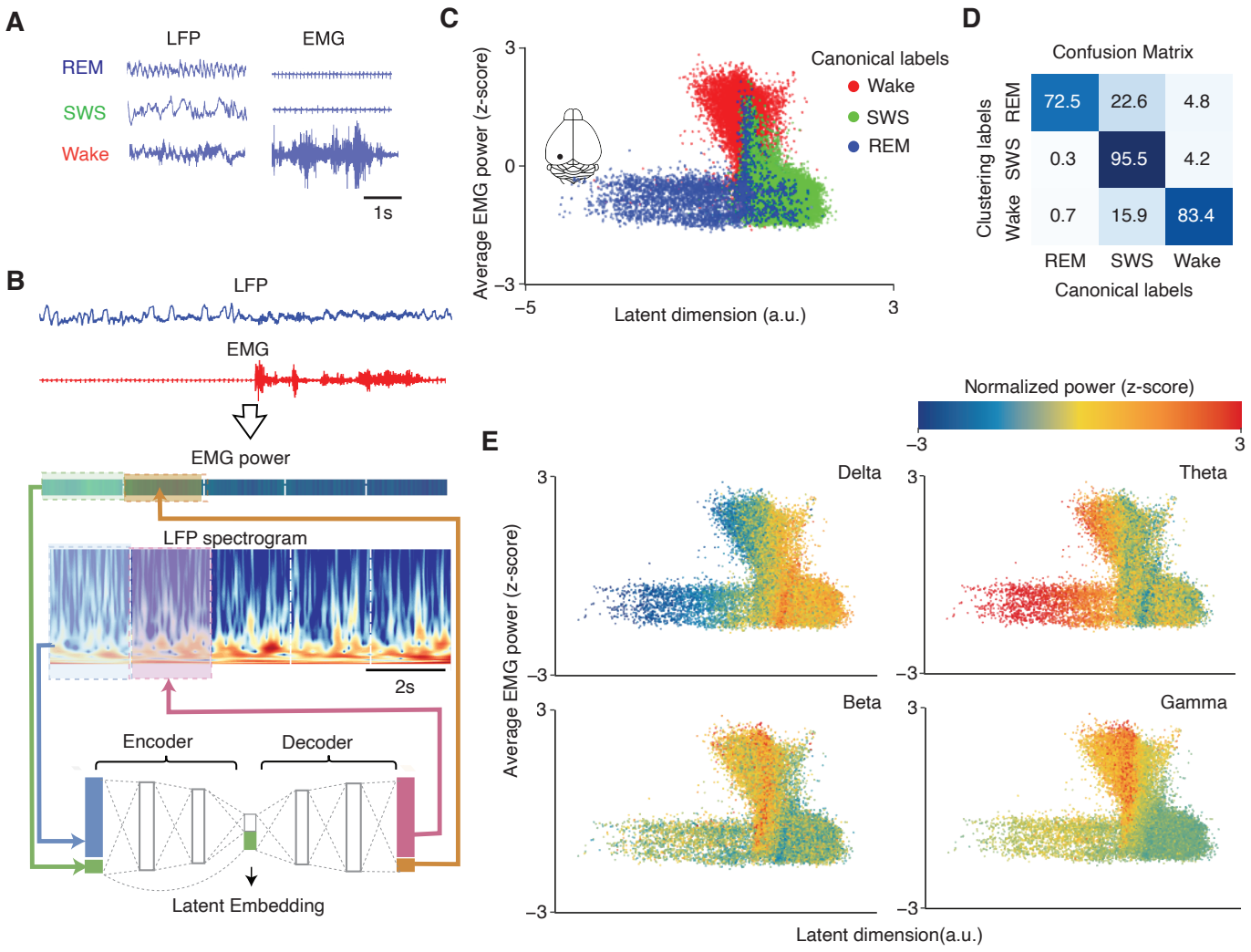


Figure 1. VAE Framework for brain state manifold. **A**, Example LFP and EMG traces recorded during REM, SWS and wake. **B**, Model architecture. For each 2-second window, we compute EMG power and wavelet spectrogram of LFPs, which become inputs to the TN-VAE. The 2-D bottleneck layer of the TN-VAE produces the embeddings for each 2-second window. The average EMG power is explicitly set as one of the latent dimensions. **C**, Two-dimensional latent manifold of the EMG and LFP activity from a single channel in the visual cortex for 24 hours of recording reveals clusters which largely agree with canonical labeling of wake, SWS, and REM states. **D**, Confusion matrix between GMM-generated labels and canonical labels. GMM clustering achieves performance similar to inter-expert agreement. **E**, Latent manifold from (**C**) colored by average power from frequency bands: delta (top left), theta (top right), beta (bottom left), and gamma (bottom right). Each point representing a two second window in time is colored by normalized spectral amplitude ratio for each frequency band divided by total spectral power.

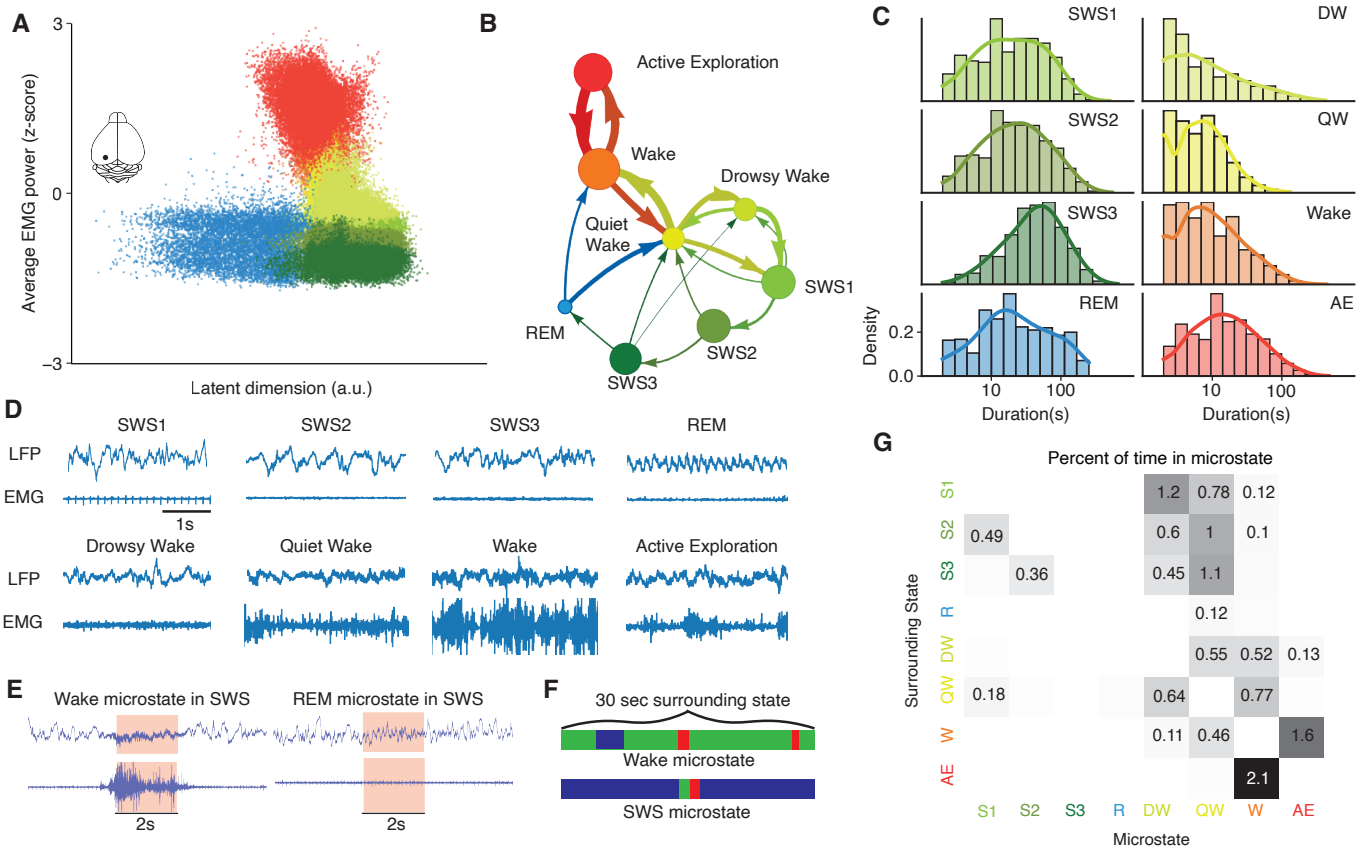


Figure 2. HMM reveals dynamics of brain states and microstates. **A**, Embeddings colored by each of 8 states fit by HMM. **B**, A diagram of most likely transitions. Circle size corresponds to the frequency of the state, and arrow width corresponds to likelihood of transition between states. Transition from any wake state to REM is highly unlikely without first passing through SWS. SWS goes from lighter to deeper before transitioning to REM. **C**, State lengths of each of the 8 states fit by the HMM. **D**, Example raw LFP and EMG traces from each of the 8 discovered states. **E**, Example raw LFP and EMG traces from microstates. **F**, Microstates are identified by mismatches between local state labels for 2 second windows and the surrounding 30-sec state label. **G**, Percent of time spent in microstate out of total time in surrounding state. Color of states corresponds to labels in (**B**).

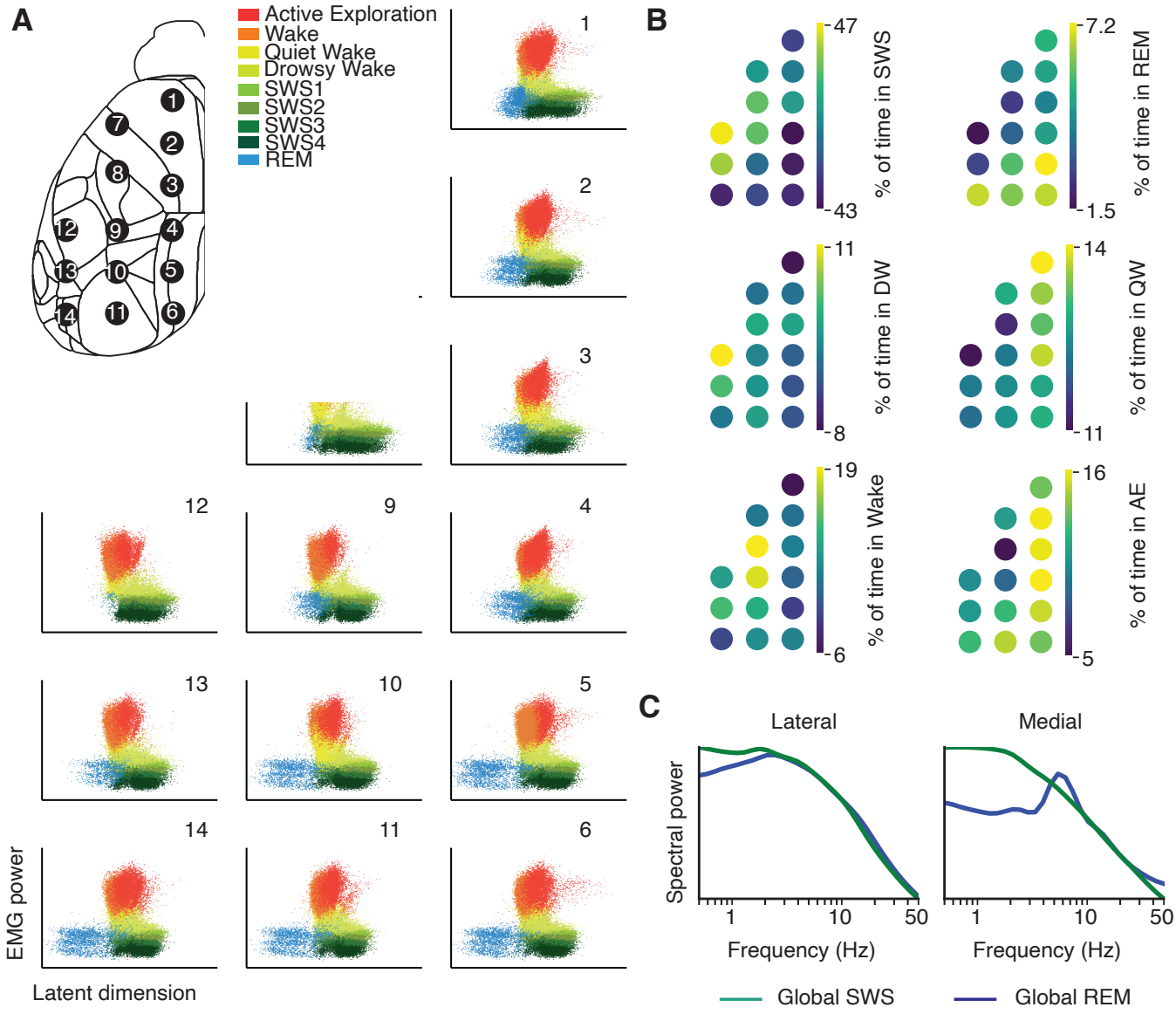


Figure 3. Heteregeneous expression of REM sleep across the cortex. **A**, LFP is recorded from 14 probes simultaneously across the cortex. We fit our model to data compiled from all 14 probes. Embeddings are shown for each separate probe, colored by the state label obtained from the MultiHMM. REM sleep (blue) is missing or partial for some electrodes. **B**, Percent of time spent in SWS (top left), REM (top right), DW (middle left), QW (middle right), wake (bottom left), and AE (bottom right). SWS is combined from SWS1-4. **C**, Periods of global REM and global SWS are defined as time points where greater than 10 electrodes are in that state. For somatosensory 12 (left) and somatosensory 10 (right) electrodes, the average spectrogram for global REM (blue) and global SWS (green) are shown.

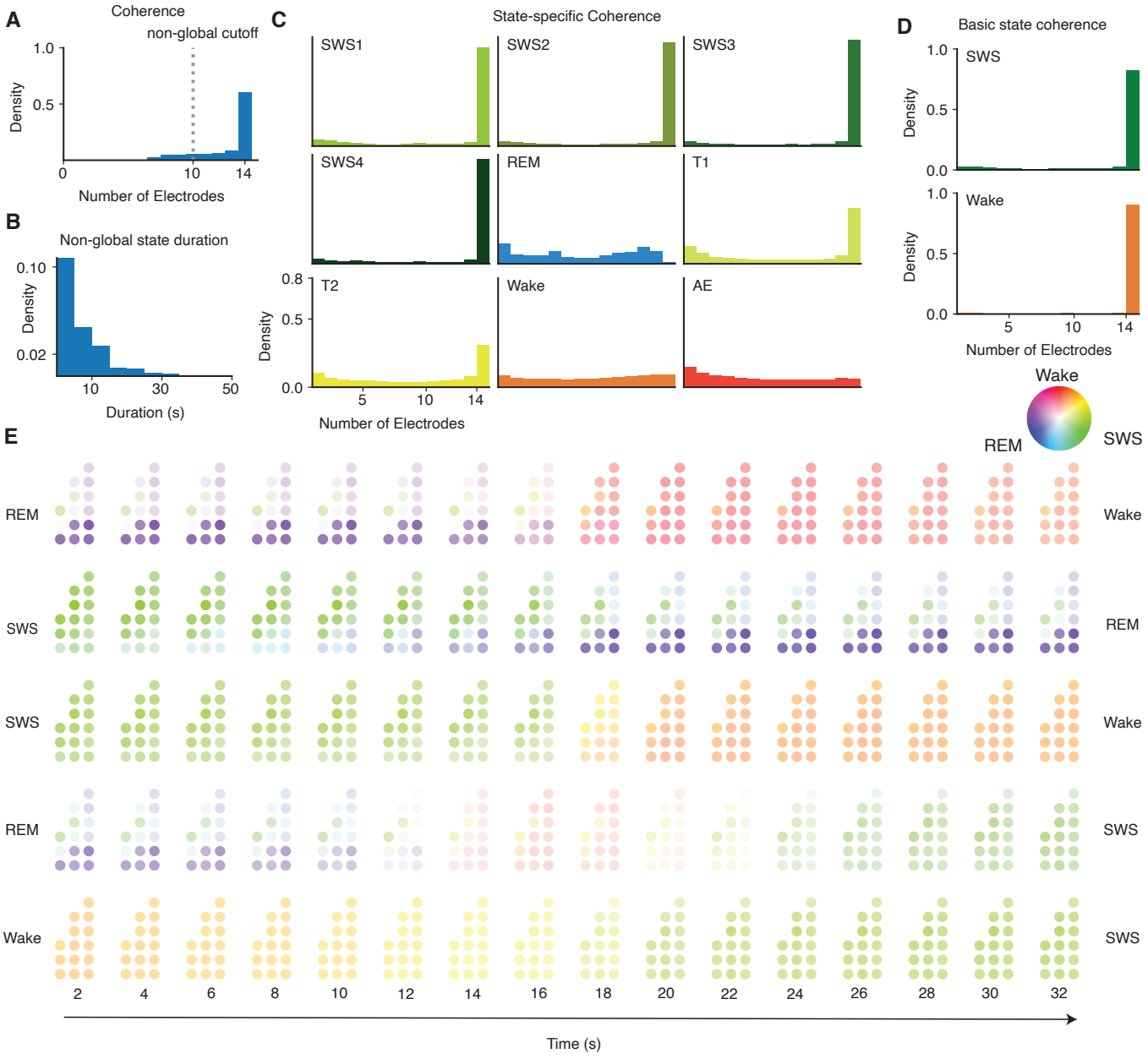


Figure 4. (Caption on following page).

Figure 4. Spatiotemporal dynamics of brain states A, Histogram of coherence of states across 14 electrodes. Coherence is defined as the maximum number of electrodes in the same state (found by the MultiHMM) at any point in time. While most time points are global (14 electrodes in the same state), there are a significant amount of time points with 13 or less electrodes in coherence. **B**, Histogram of the duration of non-global states (states where less than 10 electrodes are in agreement). Most are less than 30 seconds in duration, likely corresponding to micro-states and transition states. **C**, Histogram of number of electrodes (1-14) in each particular state at any point in time divided by the total amount of time any electrode is in that state. REM and wake states show the least amount of time in a global state. **D**, Same as (**C**) but for basic states of SWS (composed of SWS1-4) and wake (composed of QW, DW, wake and AE). While substates of wake are non-global, wake as a whole is global. **E**, Average dynamics for transitions for each transition between basic states of REM, SWS, and wake show spatiotemporal heterogeneity. Colors denote the position on the manifold of that electrode. Each electrode graph represents 2 seconds in time.