Supplemental Information for

Spatiotemporal structure of REM sleep twitching reveals developmental origins of motor synergies

Mark S. Blumberg, Cassandra M. Coleman, Ashlynn I. Gerth, & Bob McMurray

Correspondence to: mark-blumberg@uiowa.edu

Supplemental Inventory

1. Supplemental Figures

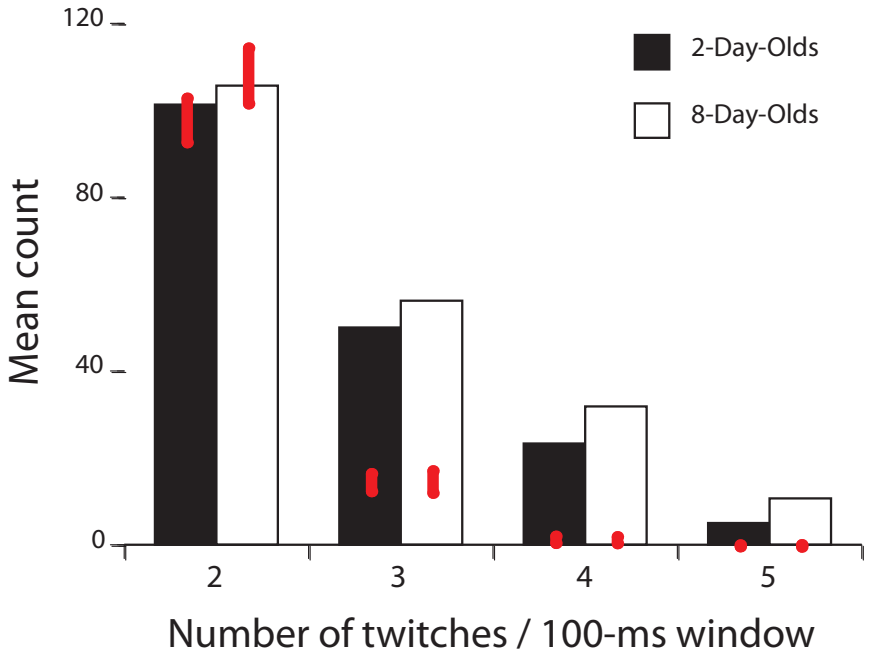

Figure S1. Quantity of multi-joint twitching in relation to chance. Presented are the mean observed counts of 100-ms windows (per pup/litter) containing 2, 3, 4, or 5 twitches in 2- (black bars) and 8-day-old (white bars) rats. Vertical red bars represent the 95% confidence intervals computed using Monte Carlo randomizations (500 iterations). In this Monte Carlo analysis, for each original twitch in the raw dataset we replaced the time with a new time drawn from a uniform distribution (range: 1-20,000 ms), and randomly selected the type of twitch (e.g., right shoulder adduction) with a 1/8 probability. At both ages, 3-, 4-, and 5-twitch events were more likely than expected by chance (*ps* < .002), whereas 2-twitch events were not (2-day-olds: $p = .14$; 8-day-olds: $p = .764$). Events containing 0-1 twitches were excluded from this analysis.

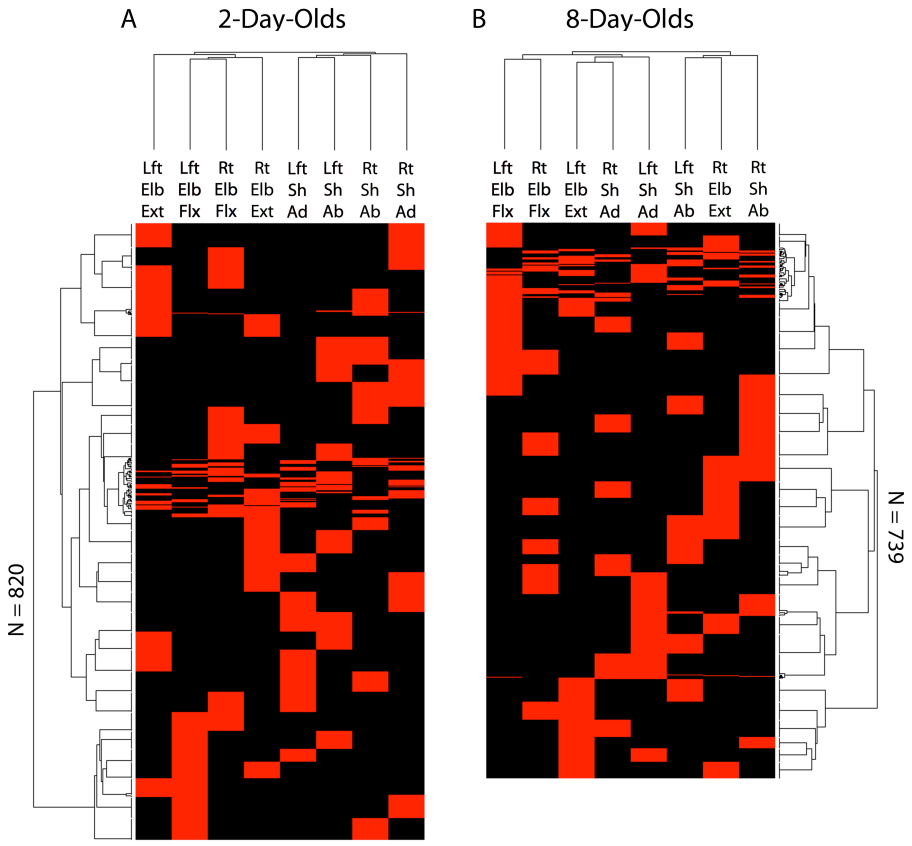

Figure S2. Randomized hierarchical cluster analyses, with seriation, of multi-joint twitching at the shoulders and elbows in (A) 2- and (B) 8-day-old rats. Data were obtained from a single run of the randomization described in Figure S1. Otherwise, these analyses were performed identically to those in Figure 5. The dendrograms at both ages are relatively unstructured (in relation to Figure 5). This is indicated in three ways. First, the large distance of leaves (i.e., individual joints) from the clades (i.e., groupings) in the top dendrograms suggests weak grouping between leaves. Second, the short distance separating clades in the top dendrograms suggests that individual clusters are not highly distinct. Finally, most clades in both the top and rotated dendrograms are simplicifolious (i.e., they comprise a single leaf added to a group, rather than two leaves combining to form a group), suggesting little similarity among individual leaves (i.e., joints in the top dendrograms or events in the rotated dendrograms). Note that there are fewer events (i.e., rows) in these plots than in Figure 5 because, after randomization, there were fewer events comprising more than one twitch and, in particular, there were far fewer events with three or more twitches; see Figure S1). Abbreviations: Rt, right; Lft, left; Sh, shoulder; Elb, elbow; Ad, adduction; Ab, abduction; Flx, flexion; Ext, extension.

2. Supplemental Tables

Table S1. The total number of 2- and 8-day-old subjects used in this study and the number of videos, twitches, and twitches/video obtained.

	Number of subjects/Number of litters	Total number of videos	Total number of twitches	Number of twitches/video
2-day-olds	10/7	35	4966	141.9
8-day-olds	6/6	39	5168	132.5

Table S2. Total number of twitches at each joint in the left and right forelimbs across all 2- and 8-day-old subjects. Ad = adduction; Ab = abduction; Ext = extension; Flx = flexion.

	Left Shoulder		Left Elbow		Left Wrist		Right Shoulder		Right Elbow		Right Wrist	
	Ad	Ab	Ext	Flx	Ext	Flx	Ad	Ab	Ext	Flx	Ext	Flx
2-day-olds	551	363	404	437	461	246	494	359	399	543	481	228
8-day-olds	509	386	473	375	429	316	514	368	458	475	515	350

3. Supplemental Movie

Movie S1. Examples of twitching in an 8-day-old rat. Five clips are shown: (i) 18 seconds of real-time twitching; (ii) a discrete twitch, in slow motion, comprising extension of the left elbow; (iii) a discrete twitch, in slow motion, comprising abduction of the right shoulder; (iv) an example, in slow motion, of a homologous twitch pattern comprising right shoulder adduction followed quickly by left shoulder adduction; and (v) an example, in slow motion, of a complex multi-joint twitch pattern comprising several movements across both forelimbs. The white dots are fluorescent paint for motion tracking of joint movements. All videos were recorded at 250 frames per second.

4. Supplemental Experimental Procedures

Experiments were carried out in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals (NIH Publication No. 80-23) and were approved by the Institutional Animal Care and Use Committee of the University of Iowa.

Subjects

Subjects were male and female Sprague-Dawley Norway rats (*Rattus norvegicus)*. A total of 10 P2 rats from seven litters (body weight: 6.1-8.2 g) and six P8 rats from six litters (body weight: 17.2-20.1 g) were used. Litters were culled to eight pups within three days of birth (day of birth = P0). Mothers and their litters were housed and raised in laboratory cages (36 x 27 x 21 cm; Thoren Caging Systems, Hazleton, PA) in the animal colony at the University of Iowa. Food and water were available ad libitum. All animals were maintained on a 12-h light-dark schedule with lights on at 0700 h.

Preparation of subjects for videorecording

On the day of testing and during the lights-on period, a pup with a visible milk band was anesthetized with isoflurane and secured in a supine position in a custom-made silicone mold sized appropriately to the age of the pup. Light restraints were placed over the neck and abdomen to keep the pup in place. An ultraviolet (UV) fluorescent paint was applied at key locations along the two forelimbs and chest (see Figure 1A). After this procedure, which lasted less than 10 min, the pup was transferred to a humidified incubator maintained at thermoneutrality (P2: 35.5˚C; P8: 32˚C) for testing.

Data acquisition

Two high-speed (250 frames/s) digital video cameras (Integrated Design Tools, Tallahassee, FL) with 105 mm micro-Nikkor lenses (Nikon, Melville, NY) were used. These cameras record directly to digital at 1280 x 1024 pixels. Motion Studio software (Integrated Design Tools) was used to synchronize the cameras and record videos.

Recordings began when the pup was cycling between sleep and wakefulness; cycling between states is more rapid at P2 than at P8, but at both ages sleep is the predominant state [1]. Under ultraviolet illumination, multiple 20-s recordings were acquired; this was the maximum duration allowable given camera memory (8 GB) and frame rate. During each 20-s recording period, the experimenter closely monitored the subject and confirmed that only twitches were expressed (if wake behaviors or startles were detected, the video data were not saved). When data were saved, approximately 35 minutes were needed to download the data from the two cameras, after which the next recording began. At the completion of the recording session, each pup was returned to its home cage and the cameras were calibrated for 3-D motion tracking using a calibration fixture and ProAnalyst software (Xcitex, Boston, MA).

For each of the six P8 subjects, 6-8 videos/pup were acquired. For the P2 subjects, however, the range was 2-6 videos/pup; because, in three instances at this age, subjects yielded only two videos (due to fussiness), a littermate was used to provide 2-3 additional videos, thus yielding 4-6 videos from each litter at this age. Thus, the “true” sample size—based on the number of litters rather than the number of pups used—was seven at P2 and six at P8. Although we refer to “pup” in the text (for ease of presentation), in some cases we are referring to a litter.

Data reduction

The protocol for data analysis began with automatic motion tracking of the joints using ProAnalyst 3-D software. Automatic tracking was always supplemented by frame-by-frame confirmation and, when necessary, manual correction. The calibration fixture's coordinates allowed us to pinpoint the location of each fluorescent dot on the subject's body in 3-dimensional space with an accuracy of approximately 0.1 mm.

Based on preliminary analyses, we identified six joint angles or line distances that reliably identified shoulder abduction and adduction, elbow extension and flexion, and wrist extension and flexion, for each of the two forelimbs. These angles and distances were computed using the ProAnalyst software for each of the 5000 video frames from each camera for a given 20-s recording period. Next, the data were imported into Spike2 (Cambridge Electronic Design, Cambridge, UK) as six continuous waveforms representing the six joints across the two forelimbs. To convert these continuous waveforms to discrete twitch-events that indicate movement onset times, we first filtered slow oscillations from the waveforms (time constant = 0.4 s). Next, we calculated the mean baseline quiescent activity for each waveform from multiple time-points across the 20-s recording. The threshold for estimating the onset time for each twitch-event was the standard deviation of this mean multiplied by 15. For quality control, we regularly cross-checked these onset times against video records.

Two highly trained individuals separately converted the data from waveforms to twitch-events. On a regular basis, three videos from each age group were randomly selected and, for each video, both individuals scored the same joint. Inter-rater reliability

for converting waveforms to onset times was high, with Cohen's Kappa ranging from .83 to .94 (computed using GSEQ software [2]).

Data analysis

Because of the short duration of each individual 20-s recording, most analyses were performed on pooled data at each age. Within Spike2, the pooled data at each age comprised a single datafile denoting records of onset times for each joint movement; breaks between 20-s recordings were marked to prevent inappropriate analyses across sessions. To compute inter-twitch intervals for frequency distributions, twitch onset times were interleaved to produce a single record of all twitches in both forelimbs.

For all inferential statistics, alpha was set at 0.05. All means are presented with their standard error.

Perievent histograms were used to assess pairwise relationships between joint movements. They were computed using the "event correlation" function in Spike2 with one joint movement designated the "target" and the other the "trigger." Histograms were computed using pooled data at each age, and indicate the total number of target events for each 50-ms time bin surrounding the trigger; counts were normalized to percentages in relation to the total number of target twitches within the 500-ms histogram window (250 ms before and after the trigger). To determine statistical significance, we jittered the trigger event data 1000 times within a 250-ms window using PatternJitter [3, 4] and for each jitter constructed 1000 perievent histograms (using a custom-written Matlab program). From this we established a 99% ($p < .01$) criterion, to which we compared each 50-ms bin of the actual data.

To create a windowed dataset, we stepped through the raw data in 75-ms increments. At each time point, all twitches occurring within a 100-ms window were identified (the resulting 25-ms overlap of windows functionally smoothed the data). We chose a 100-ms window based on the inter-twitch interval data (Figure 1C) and the perievent histograms showing that most pairwise twitches occurred within this window (Figures 2 and 3); however, we also examined shorter and longer windows to confirm that our findings were not overly sensitive to this choice. Before analysis, we eliminated windows containing either no twitches or a single twitch. Eliminating these windows amplified the probabilities of multi-twitch events (our primary interest) without changing the relative likelihoods of different types of multi-twitch events. The resulting dataset represented 49.5% of the full set of windows at P2 and 46.4% at P8. The final windowed datasets were composed of 1269 rows at P2 (mean = 181.3 $\pm$ 22.2 rows/pup) and 1242 rows at P8 (mean = 207.0 $\pm$ 21.7 rows/pup). These rows are referred to as "events" below and in the main text.

To examine age-related changes in twitching, we calculated the mean proportion of 100-ms events (in the windowed dataset) that, for example, contained shoulder abductions in the left and right forelimb. These were calculated using the pup/litter as the unit of analysis. Proportions were transformed using the empirical logit and ANOVAs and t tests were performed were performed using SPSS (IBM, Endicott, NY).

Although we examined many different movement categories, only a subset comprising the most clear and significant results are described in the main text.

Hierarchical cluster analysis (HCA) with seriation was performed using PermutMatrix software [5]. The settings used were Euclidean distance and Ward's

Minimum Variance Method. For seriation, the multiple-fragment heuristic was used. Each row of the data was treated as an independent observation for this analysis.

Latent class analysis (LCA) was performed using Latent GOLD software (Statistical Innovations, Belmont, MA). The data at P2 and P8 were analyzed separately and litter was used as the random effect. Cluster convergence occurred for both datasets and the determination of the best model fit was made based by minimizing the value of the Akaike Information Criterion (AIC). We also confirmed that none of the bivariate residuals exceeded a value of 1, indicating that all eight joint movements were independent of each other. The best model fits yielded 28 clusters at P2 and 21 clusters at P8. Each cluster was visualized as a profile plot.

To test whether the clusters produced by LCA were simply due to the relative independent frequencies of the twitch movements, we used a Monte Carlo method to determine the likelihood of a particular cluster appearing by chance. To do this, we shuffled the assignment of twitches to event times within each 20-s video segment. This conservative approach maintained the relative frequencies of the different twitches (e.g., in the shuffled dataset elbow flexion occurred just as frequently, but at different times) and the temporal structure of the datasets (e.g., if twitches tended to come in sets of three within a 100-ms window, that continued to be the case), but the particular limbs involved were now variable. Shuffling was performed 150 times and LCA was performed on each of the shuffled datasets. Then, for each LCA twitch pattern (or cluster) identified from the original dataset, we determined whether that pattern could have arisen by chance by assessing its likelihood against all of the twitch patterns that LCA detected in the 150 shuffled datasets. (Clusters were matched between the original

dataset and the clusters from the LCA analyses of each of the 150 Monte Carlo runs by converting each cluster, in either set, to a discrete series of joints based on an upper and lower threshold. A range of thresholds was used to ensure that our results were robust.) Even using this conservative method, nearly all of the clusters that LCA identified were unlikely to have arisen by chance ($p < .05$). We conclude that LCA properly identified clusters.

In order to examine developmental changes in the clusters of twitches, we needed to match P2 and P8 clusters. Thus, we tested the similarity of each of the 28 clusters at P2 against each of the 21 clusters at P8, and vice versa, using the eight probability values that comprise each profile plot as the basis of similarity. Two similarity rules were used. First, we used Euclidean distance. Second, since Euclidean distance assumes a linear scaling that may not be appropriate for probabilities, we also used a probabilistic rule (Equation 1) in which similarity was the probability of the same twitch being present (in each cluster) added to the probability that the same twitch was absent from both:

$$Similarity_{x \leftrightarrow y} = \sum_{t \in twtiches} p^t{}_x p^t{}_y + \left(1 - p^t{}_x\right)\left(1 - p^t{}_y\right) \quad (1)$$

Here, $p^t{}_x$ represents the probability that twitch t (e.g., elbow flexion) participated in cluster x (one of the clusters from the P2 analysis) or cluster y ($p^t{}_y$, from the P8 analysis). Combined, this computes the probability that both twitches occurred ($p^t{}_x p^t{}_y$) or that neither did ($[1-p^t{}_x][1-p^t{}_y]$), summed over all eight pairs of corresponding twitches within a given pair of clusters.

A P2 and P8 cluster were deemed a match only if they were the best matches on both similarity rules exclusively and reciprocally (i.e., the P2 cluster's closest match was

the P8 cluster, and the P8 cluster's closest match was the P2 cluster).

Regression analyses of the LCA clusters were performed using SPSS. We computed Shannon's Entropy, *E,* for each cluster by first normalizing the probabilities associated with each twitch (for a given cluster) by dividing each probability by the sum of the probabilities. (Note that in these profile probabilities, each likelihood represents the independent probability of a specific joint movement given the cluster. The LCA coefficients for a given cluster represent eight binomial distributions, not a single multinomial distribution.) We next calculated *E* using Equation 2:

$$E_x = -\sum_n q^n{}_x \log_2\left(q^n{}_x\right) \quad (2)$$

Here, E_x is the entropy of cluster *x*; and $q^n{}_x$ is the normalized probability of twitch *n*, in cluster *x,* output from the LCA (after normalization).

Cluster frequency was obtained from Latent Gold LCA software and was log-transformed prior to analysis.

5. Supplemental References

1. Blumberg, M. S., Seelke, A. M. H., Lowen, S. B., and Karlsson, K. Æ. (2005). Dynamics of sleep-wake cyclicity in developing rats. Proc. Natl. Acad. Sci. USA *102*, 14860–14864.

2. Bakeman, R., and Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences* (Cambridge: Cambridge University Press).

3. Harrison, M. T., and Geman, S. (2009). A rate and history-preserving resampling algorithm for neural spike trains. Neural Comput. *21*, 1244–1258.

4. Amarasingham, A., Harrison, M. T., Hatsopoulos, N. G., and Geman, S. (2012). Conditional modeling and the jitter method of spike resampling. J. Neurophysiol. *107*, 517–531.

5. Caraux, G., and Pinloche, S. (2005). PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics *21*, 1280–1281.