OPEN

Automated identification of multiple seizure-related and interictal epileptiform event types in the EEG of mice

Rachel A. Bergstrom[1], Jee Hyun Choi[2,3], Armando Manduca[4], Hee-Sup Shin[2,3], Greg A. Worrell[5] & Charles L. Howe[5,6]

[1]Neurobiology of Disease PhD Program, Mayo Graduate School, Mayo Clinic, Rochester, MN 55905, [2]Center for Neural Science, Korea Institute of Science and Technology, Seoul, Korea, [3]Department of Neuroscience, University of Science and Technology, Daejon, Korea, [4]Department of Physiology and Biomedical Engineering, Mayo Clinic, Rochester, MN 55905, [5]Department of Neurology, Mayo Clinic, Rochester, MN 55905, [6]Department of Neuroscience, Mayo Clinic, Rochester, MN 55905.

Visual scoring of murine EEG signals is time-consuming and subject to low inter-observer reproducibility. The Racine scale for behavioral seizure severity does not provide information about interictal or sub-clinical epileptiform activity. An automated algorithm for murine EEG analysis was developed using total signal variation and wavelet decomposition to identify spike, seizure, and other abnormal signal types in single-channel EEG collected from kainic acid-treated mice. The algorithm was validated on multi-channel EEG collected from γ-butyrolacetone-treated mice experiencing absence seizures. The algorithm identified epileptiform activity with high fidelity compared to visual scoring, correctly classifying spikes and seizures with 99% accuracy and 91% precision. The algorithm correctly identifed a spike-wave discharge focus in an absence-type seizure recorded by 36 cortical electrodes. The algorithm provides a reliable and automated method for quantification of multiple classes of epileptiform activity within the murine EEG and is tunable to a variety of event types and seizure categories.

A nimal studies of epilepsy have relied upon the Racine scale for seizure classification and quantification[1]. However, the Racine scale does not allow for identification of the seizure focus and does not sufficiently measure partial seizure activity that involves areas not relevant to movement and behavior. Likewise, as a discrete measurement based upon subjective interpretation of mouse behavior, the Racine scale has limited sensitivity to small changes in seizure activity associated with therapeutic testing.

Intracranial electroencephalography (EEG) in rodents permits the monitoring of seizure evolution and the localization of seizure foci within the brain. As a continuous measurement, the EEG provides greater sensitivity than behavioral scoring for the detection of small changes in seizure activity. Unfortunately, primary analysis of EEG is typically performed by visual inspection of the signal by expert observers, with emphasis on the frequency, voltage, amplitude, regularity of the waveforms, and spatial range and temporal persistence of the signal events[2]. Visual analysis and scoring of EEG signals is time-consuming and subject to observer bias and error. Indeed, studies of inter-observer reliability among expert scorers show highly variable levels of agreement on scoring of epileptic and arousal-related signal types[3–5].

Automated EEG event detection algorithms overcome the problems of visual inspection. The primary focus of such algorithms has been the detection of either interictal spikes[6,7] or seizure onset[8–11]. A number of computational methods have been used for EEG analysis, including Fourier transforms[12], artificial neural networks[13,14], logistic regression[15], and time-frequency analysis[16]. Line length, or total variation of the signal, was originally proposed for EEG analysis as a computationally efficient method for real-time seizure detection[10]. Line length is defined as the sum of the absolute values of the differences between neighboring data points over a specified time interval. The line length per unit time of a baseline EEG recording remains generally constant while a spike or seizure event transiently increases the line length.

Wavelet transformation isolates different frequency components of the EEG signal and permits analysis of the individual components at a scale-matched resolution that detects transient, inhomogeneous events localized by both time and frequency[17]. The wavelet is an analytical alternative to methods with poor time resolution such as Fourier transformation or poor frequency resolution such as windowed or short-time Fourier transformation. The Daubechies family of wavelets[18] has been used for seizure analysis by several groups[2,11,13,19–22].
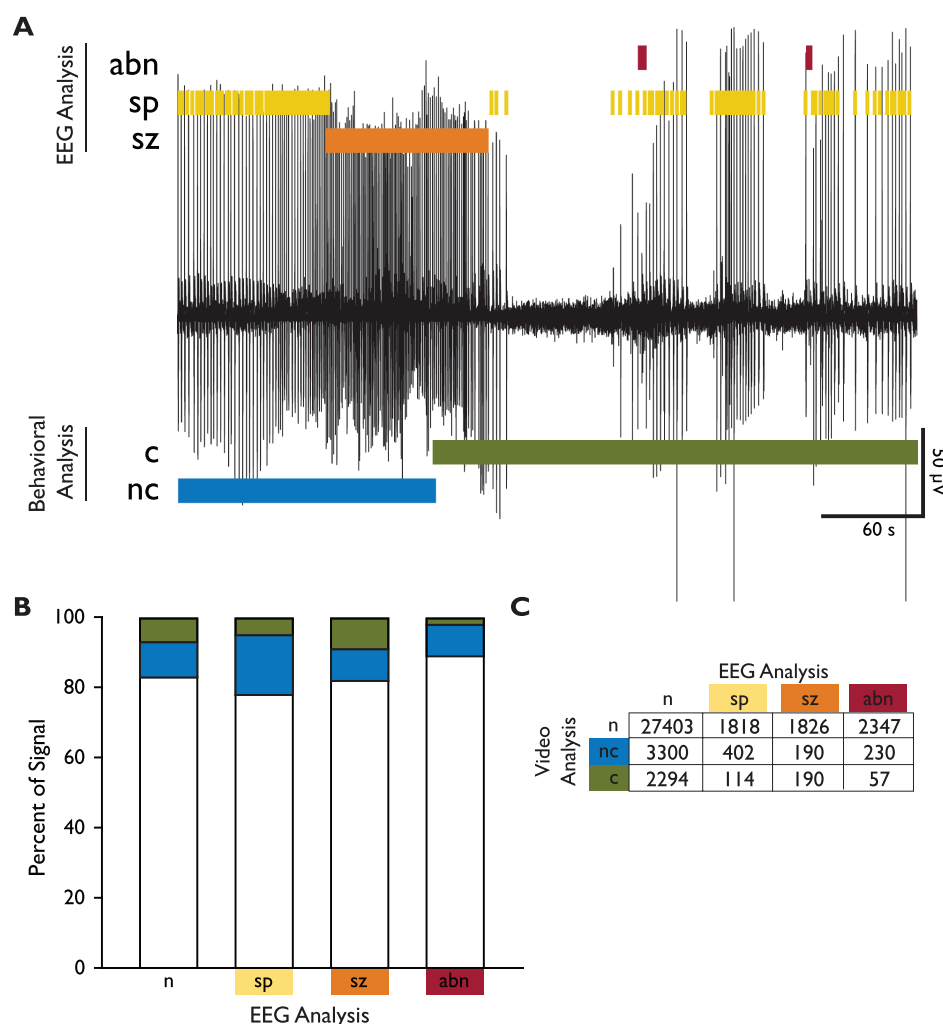
Here we present an automated method for EEG signal analysis in mice that detects and quantifies seizures, spikes, and other abnormal signal events based on total variation following wavelet decomposition of the signal. Our method returns event classification, count, and duration within the timeline of the recording. We also introduce an automated signal cleanup algorithm that removes movement artifacts based on signal amplitude in an empty channel.

Results

Racine scoring does not predict EEG events. KA is a widely-used seizure-induction agent in rodent models of temporal lobe epilepsy (TLE) and triggers a hippocampus-specific injury similar to patterns observed in human TLE patients[23–25]. Behavior was assessed in KA-treated mice by blinded, off-line analysis of video and categorized as non-convulsive or convulsive. Simultaneous EEG recordings were manually scored by expert observers and events were categorized as spikes, seizures, or other abnormal events. Comparison of total behavioral analysis (83% normal, 7% convulsive, 10%

non-convulsive) and manual EEG analysis (82% normal, 8% spikes, 7% seizures, 4% other abnormal) revealed similar scoring (Supplemental Figure 1A). A clear difference between baseline and KA-treated EEG recordings was also identified in all scoring categories using this method (Supplemental Figure 1B). However, when manual EEG and behavioral scores were compared on a second-by-second basis, we found that there was little predictive value between the EEG signal score and the behavior score (Figure 1A). Indeed, the odds ratio for a relationship between video events and EEG events was 1.017 ($X^2 = 0.139$, P = 0.709) indicating only a random interaction between the measurements. Of particular interest is that mice behaved normally for a majority of all EEG events (no Racine scoreable behavior) and convulsive and non-convulsive behaviors could not be used to predict spikes, seizures, or EEG abnormalities at the recording electrode (Figure 1B, 1C). Visual comparison of the EEG signals associated with normal behaviors such as grooming and exploring (Supplemental Figure 1C) and abnormal EEG signals associated with non-convulsive behavior

	EEG Analysis			
Video Analysis	n	sp	sz	abn
n	27403	1818	1826	2347
nc	3300	402	190	230
c	2294	114	190	57

Figure 1 | Behavioral analysis does not adequately predict local EEG activity in the hippocampus. (A) The behavioral and signal analysis scores (colored bars) superimposed upon the corresponding raw EEG signal demonstrate that the behavioral scores are not predictive of the signal at this recording electrode. Of particular interest is the transition from non-convulsive to convulsive behavior during a period of time scored as seizure in the EEG. Additionally, a period of normal EEG is associated with convulsive behavior. Behavior: non-convulsive = blue, convulsive = green; Signal: seizure = orange, spike = yellow, other abnormal = red. (B) The signal and behavioral scores were compared on a second-by-second basis to assess concordance. The horizontal axis represents the EEG category. The vertical axis represents the amount of each particular EEG signal category that is associated with a given behavioral score. The predominant behavioral state associated with all four signal types (normal, spiking, seizure, other abnormal) is normal behavior (white). (C) Contigency matrix of the data shown in panel B. Overall, there is little to no predictive power for behavior (video analysis) on the EEG. n = normal (behavior or EEG), c = convulsive behavior, nc = non-convulsive behavior, sp = spiking signal, sz = seizure signal, abn = other abnormal signal.

(Supplemental Figure 1D) or convulsive behavior (Supplemental Figure 1E) confirmed the variability and overlap of signal types associated with different behaviors.

Automated signal analysis. *Automated EEG signal cleanup.* To determine which portions of the raw, bulk signal were noise, we calculated a representative average signal standard deviation (ThreshA) using an empty channel simultaneously recorded with the EEG. The electrode for the empty channel terminated in the dental cement of the headmount and did not record any brain or muscle activity. Major signal events that deviated from the baseline within this channel were therefore attributed to movement or other noise artifacts at the headmount or in the recording system. Ten, one-minute segments of signal were randomly selected from the empty channel data, and the average standard deviation of the signal within these segments was calculated and used to set the threshold (ThreshA) for cleanup. The standard deviation of the entire raw EEG signal was then computed within non-overlapping 250-msec windows and compared to ThreshA. A 250 msec window was selected based on the observation of many short bursts of noise in the empty channel that lasted 100 to 300 msec and were associated with similar events in the recording channel. A 250 msec window provided optimal removal of the majority of these noise events without excessive loss of otherwise normal signal. Windows with a standard deviation greater than twice ThreshA were set to zero across all channels (Supplemental Figure 2A and 2B). This cleanup method identified and deleted movement artifacts within the recording that might otherwise be identified as epileptiform activity by the algorithm (for example, physical manipulation of the recording wires by the investigator during drug delivery). The average total time removed by this method was small compared to the overall signal length (Supplemental Table 1) and visual inspection of the automatically removed segments confirmed that none were seizure-like.

Determination of baseline line length parameters. To establish analytical parameters, the Daubechies db4 wavelet was used to decompose the cleaned baseline EEG signals into details at four different scales (D1-D4) and an approximation (A4)[2,18]. The A4 approximation was chosen for further analysis because it retained spike and seizure features (Supplemental Figure 2C) across the physiologically relevant frequency band from 0 to 25 Hz[2,13]. Line length was calculated in A4 over a sliding window corresponding to 240 msec (15 data points) using equation (1):

$$LL_w = \sum_{i=1}^{N-1} abs(x_{i+1} - x_i) \qquad (1)$$

where i is the index of the data point within the window, x is the voltage within the signal at index i, abs is absolute value, N is the total number of data points within the window, and w is the window number[10,13]. The median (LL_{med}) and standard deviation (LL_{stdev}) of the line lengths across the entire baseline A4 approximation were calculated and used below to analyze experimental recordings.

Bulk signal analysis by comparison to baseline. A thresholding factor, TF_{base}, was derived by iteratively analyzing the baseline EEG recording of an individual mouse. TF_{base} was initially set to 1 and the number of "events" in which the line length of the individual baseline recording varied from the median baseline line length (LL_{med}) was calculated using equation (2):

$$LL_{hits} = \sum_{w=1}^{W} E_w \begin{cases} E = 1, & \dfrac{LL_w - LL_{med}}{TF_{base} \times LL_{stdev}} \geq 1 \\[2mm] E = 0, & \dfrac{LL_w - LL_{med}}{TF_{base} \times LL_{stdev}} < 1 \end{cases} \qquad (2)$$

where E is an event, LL_w is the line length for the wth window within the actual baseline recording, LL_{med} is the median baseline line length calculated above, LL_{stdev} is the standard deviation of the baseline line lengths calculated above, TF_{base} is the threshold factor, W is the total number of windows in the recording, and w is the window number. If LL_{hits} was ≥ 1, TF_{base} was increased by 0.5 and the process was repeated until $LL_{hits} = 0$.
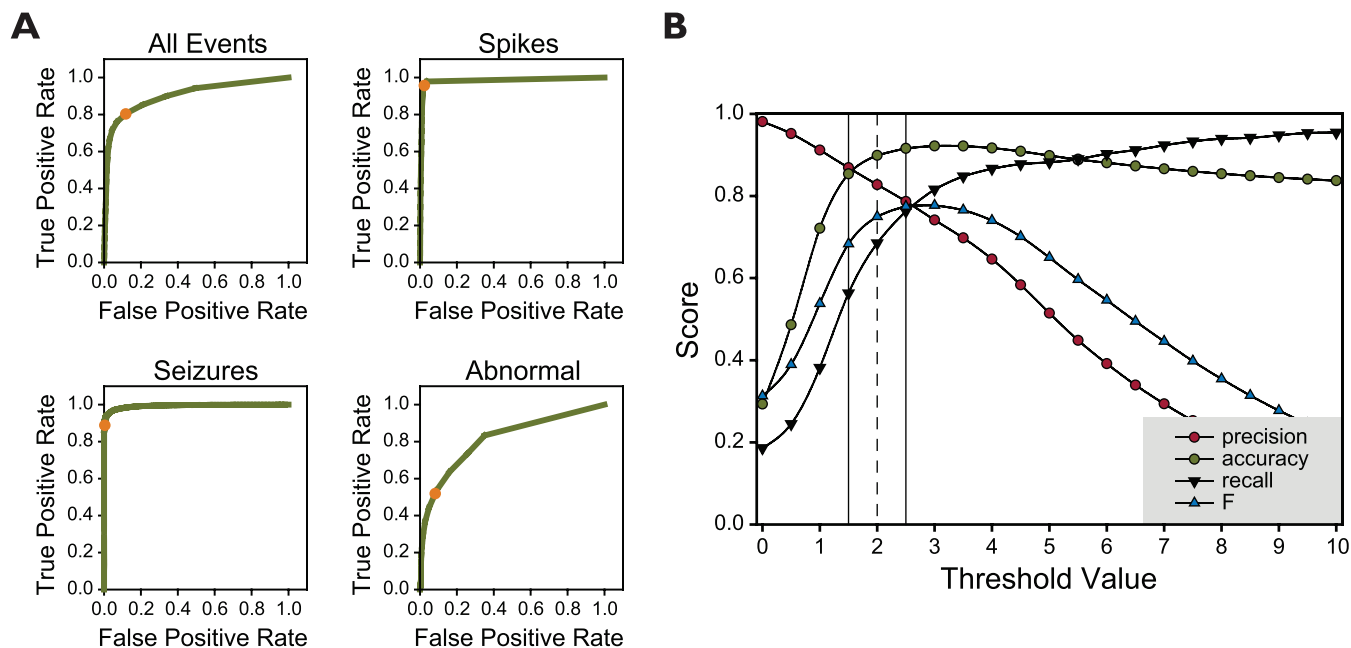
Figure 2 | Overall algorithm performance characteristics. (A) To analyze the performance of the algorithm, area under the curve (AUC) was calculated from the receiver-operator characteristic (ROC) curve of the algorithm-based EEG categorization calls. ROC curves were calculated using a threshold from 0.5 to 10 times the standard deviation of the median line length. The AUC was calculated for all events, spikes only, seizures only, and other abnormal only. (B) The optimal threshold for event identification was determined by ROC curve analysis and determination of trade-off between algorithm performance characteristics. The final threshold used for all experimental analyses (Equation (3), with $TF_{cat} = 2.0$) is noted on the ROC curves in panel A by the orange dot.

Next, the A4 approximation was calculated for each experimental recording for the same mouse and the line length for each 240 msec sliding window across the entire experimental A4 was calculated using equation (1). Using the TF_{base} calculated above, the experimental line lengths (LL_w) were evaluated with equation (2). If LL_{hits} was ≥ 1, the experimental recording was considered different than baseline. We found that 16 of 31 total recordings collected from KA-treated mice were not different from baseline. Manual scoring confirmed the bulk analysis, as these 16 recordings exhibited only 1.0% of the signal as spikes, 0.04% as seizures, and 0.03% as abnormal signal. In contrast, for the 15 recordings identified as different than baseline, manual scoring identified 4.7% of the signal as spikes, 8.5% as seizures, and 4.5% as abnormal signal. Finally, when behavior was simultaneously analyzed by video in ten EEG recording sessions, no convulsive behavior was observed in five sessions that were identified as not different than baseline by bulk analysis. We conclude that the bulk analysis algorithm effectively identifies overall EEG signal patterns as the same as or different than baseline.

Automated categorization of EEG events. Electroencephalographically relevant events were identified in individual sliding windows across the A4 approximation of each experimental recording using equation (3):

$$E_w = \begin{cases} 1, & \dfrac{LL_w - LL_{med}}{TF_{cat} \times LL_{stdev}} \geq 1 \\ 0, & \dfrac{LL_w - LL_{med}}{TF_{cat} \times LL_{stdev}} < 1 \end{cases} \quad (3)$$

where LL_w is the line length for the wth window within the experimental recording, LL_{med} is the median baseline line length calculated above, LL_{stdev} is the standard deviation of the baseline line lengths calculated above, TF_{cat} is the categorization threshold factor (see below), and w is the window number. A vector **E** of binary E_w values from all windows across the recording was constructed. **E** was eroded and dilated using a half-window of 1 (equivalent to 40 ms) to remove isolated hits that were not part of a seizure, spike, or other abnormal event. **E** was restored to the original signal length by four rounds of dyadic upsampling at odd indices. The vector was dilated and eroded again using a half-window of 8 (equivalent to 20 ms) in order to fill in

the events following the dyadic upsampling. These steps produced a final vector of the correct length with discrete events occurring as contiguous sequences of ones separated by non-event sequences of zeros. The start, end, and length of each discrete event in **E** were identified and used to categorize the event based on duration. Any event lasting longer than five seconds was defined as a seizure. **E** was then mapped back to a vector containing the EEG signal amplitudes. Event durations shorter than five seconds with an average amplitude during the event greater than 250 μV were defined as spikes. Any other event patterns were defined as abnormal other.

The automated event classifications were compared to manual EEG signal analysis to identify an optimal threshold. Threshold values (TF_{cat}) from 0.5 to 10, increasing by 0.5, were tested using equation (3) and algorithmically identified events were compared to manual calls using receiver-operator characteristic (ROC) curves[26] (Figure 2). The area under the curve (AUC)[27] for the algorithm overall (combined identification of all three event types) was 0.923 (perfect score = 1.0) (Figure 2A). The AUC was also calculated for spikes (0.988) (Figure 2B), seizures (0.987) (Figure 2C), and other abnormal events (0.824) (Figure 2D). Contingency tables (2 × 2) at all threshold values were also constructed by comparing manually identified normal and event calls (pooled spike, seizure, and abnormal) to automated algorithm calls. These tables were then used to determine precision, accuracy, recall, and F characteristics for the algorithm (see further details in the methods section)[26]. Graphing these measurements against threshold value revealed an optimal threshold range from 1.5 to 2.5 (Figure 2E), based on a tradeoff between precision and accuracy ($TF_{cat} = 1.5$) and a tradeoff between precision, recall, and F characteristic ($TF_{cat} = 2.5$). $TF_{cat} = 2.0$ was selected as the optimal threshold for all subsequent event identification.

Mapping **E** (calculated using $TF_{cat} = 2.0$) against the actual EEG recording provided visual confirmation of event calls (Figure 3A, blue-green ticks). Events were clearly differentiated as seizures (Figure 3A, orange), spikes (Figure 3A, yellow), and abnormal other (Figure 3A, red). One-second examples of each identified signal type (baseline, abnormal, spike, and seizure) are shown in Figure 3B, confirming the accuracy of the algorithm. Note that, while both spike and seizure events showed signals with high amplitude, the spikes
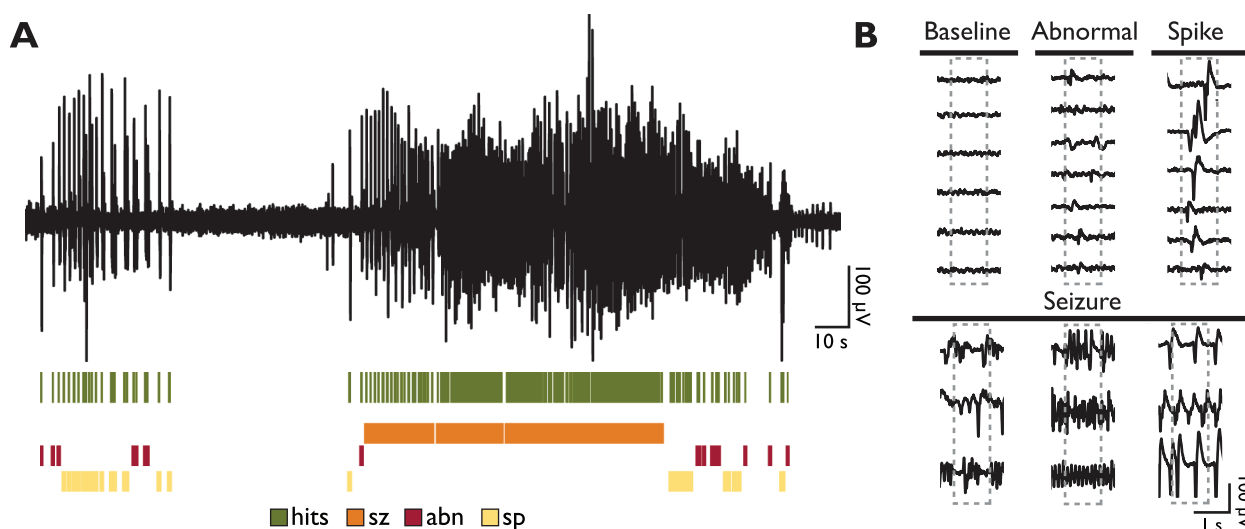
Figure 3 | Automated event identification. (A) The algorithm identifies line lengths greater than the threshold set by the same animal's baseline EEG signal. Events are identified as strings of neighboring hits (blue-green lines) and are sorted according to duration and amplitude. Seizures (sz, orange) are defined as events longer than five seconds. Spikes (sp, yellow) or spike clusters are defined as events shorter than five seconds with a maximum amplitude of more than 250 μV. Other abnormal signal (abn, red) is defined as events shorter than five seconds with a maximum amplitude of less than 250 μV. (B) The algorithm returns three signal score types: spikes, seizures, and other abnormal. Examples are shown for these signal types as well as baseline (normal) signal. Dotted line boxes indicate one second windows.

were isolated, while seizure events generally contained more than one spike-like event per second and persisted longer than the one-second window shown (Figure 3B, gray dotted outline).

Algorithm performance. Finally, the automated results were analyzed on a second-by-second basis and contingency matrices were produced to measure the performance of the algorithm against manual EEG signal analysis (Figure 4). For the first contingency table, spikes, seizures, and abnormal other signals were grouped into "events" versus "normal" EEG signal (Figure 4A). The odds ratio for a relationship between automated and manual scoring was 34.893 ($X^2 = 55214.223$, $P<0.001$), indicating that the algorithm performed exceptionally well. Further analysis of the individual event categories revealed that the predominant contribution to false positives (events identified by the algorithm that were not identified by eye) and false negatives (events identified by eye that were not identified by the algorithm) arose in the abnormal category (Figure 4B and 4C). In contrast, the algorithm correctly identified 77% of spikes and 78% of seizures and recategorized 10% of spikes and 7% of seizures as other events. Overall, the algorithm identified 17% false positive spikes and 10% false positive seizures. In summary, the automated algorithm provided 87% accuracy and 63% precision for all three event categories (sp, sz, abn) and 99% accuracy and 91% precision for spike and seizure classification.

Validation of algorithm on multi-channel SWD recordings. To validate the algorithm under different experimental conditions, we tested the event detection component on multi-channel EEG data acquired via a polyimide-based microelectrode (PBM) extracranial array[28]. The PBM array has been used to identify the cortical foci of absence seizures, identified within the EEG signal as spike-wave discharges (SWDs). The SWD is characterized by 3–5 Hz high-voltage negative spikes followed by a high-voltage positive wave lasting >1 s[29]. This type of seizure is significantly different from the tonic-clonic seizure signal induced by KA treatment, and thus serves as an excellent test for the power and flexibility of our algorithm. The SWD data were collected using a 38-channel PBM array (where channels 8 and 38 were bad, leaving 36 good recording channels) in mice treated with 50 mg/kg γ-butyrolactone (GBL) to induce absence-type seizures. Seizures were identified by visual signal analysis and confirmed by decreased activity in the EMG channel[28].

The following adjustments were made to the algorithm to identify SWD events. First, the sampling rate for these recordings was 1000 Hz, which necessitated use of the signal approximation at the fifth decomposition (A5) rather than the A4. For this sampling rate, A5 corresponds to a frequency band from 0 to 31.25 Hz, which overlaps with the frequency band used for the KA experiments (400 Hz sampling rate, A4 = 0 to 25 Hz). Similarly, the number of data points within the line length sampling window was adjusted to correspond to approximately 250 ms under the new sampling rate (224 ms, 7 data points at A5). Second, seizures were identified as events lasting longer than one second based on the characteristics of SWD, and events less than 250 ms in duration were discarded as noise. Finally, each channel was analyzed independently, and all thresholds were calculated based on the corresponding baseline for that individual channel.

Visual analysis of the EEG revealed 20 canonical SWD events occuring during a 340 second period. The algorithm identified all 20 SWD events (concordant calls) in 11 of 36 channels and 17 or more of the 20 events were identified in 23 of 36 channels. Of note, the algorithm identified additional SWD events within the EEG from the GBL-treated mouse that were not originally characterized by visual inspection (Figure 5A). None of these events were identified across all channels, suggesting that they were not the result of movement artifacts. The rate of events per 100 seconds (including concordant and additional calls) was compared between baseline and GBL-treated recordings (Figure 5B) and the fold difference in event rate was calculated (Figure 5C). GBL treatment led to an increase in the event rate compared to baseline in channels 1 through 26 and a decrease in event rate in the remaining channels. Mapping electrode location against the fold increase in event rate revealed a probable SWD focus at or near channel four (Figure 5D). The relative difference in event rate decreased as the distance from electrode four increased, and the channels that showed event rate suppression following GBL treatment were located farthest from electrode four (Figure 5D). This conclusion is in agreement with the determination of a frontal-predominant event locus based on visual analysis of the EEG.

Discussion

Here we present an automated algorithm for EEG signal comparison and event identification in mice that is based on signal line length[10] after wavelet decomposition[13]. An ad hoc baseline signal is used to
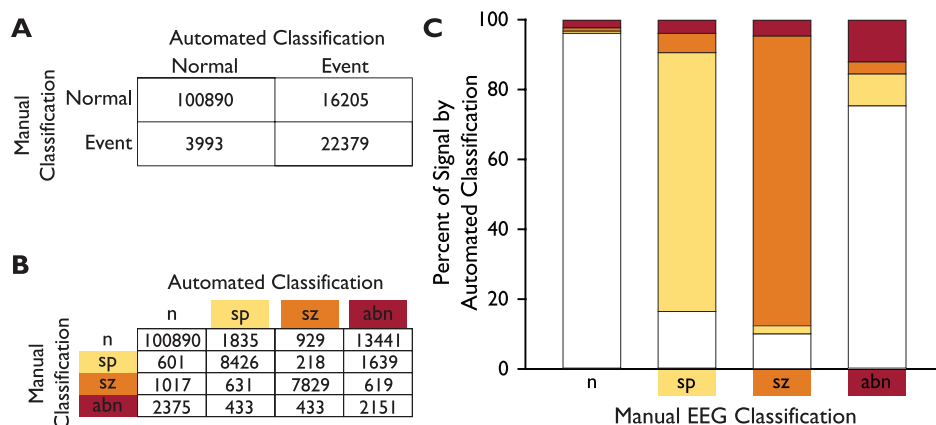
Figure 4 | Algorithmic event categorization performance characteristics. Using a threshold of 2.0 (TF$_{cat}$ in equation (3)), the automated results were compared on a second-by-second basis to manual EEG scoring. (A) Spikes, seizures, and abnormal other calls were combined into "events" versus "normal" EEG in both the algorithm output and the manual analysis and a contingency table was constructed. At this level, the algorithm correctly captured 85% of the events identifed by eye and mis-identified 14% of normal signal as an event. (B) Further analysis of the individual event categories shows that the majority of false positives and false negatives occurred in the abnormal category. (C) Showing the percent concordance graphically reveals that most spikes (77%) and seizures (78%) were captured by the algorithm but a large portion of manually scored abnormal EEG was categorized as normal by the algorithm.
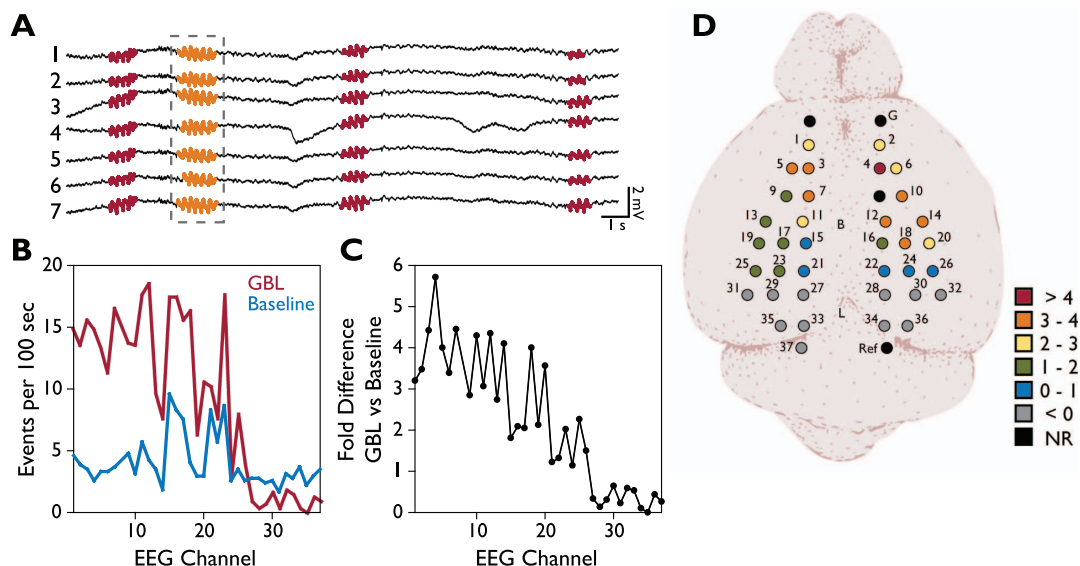
Figure 5 | Validation of event identification using a novel multi-channel EEG data set. Multi-channel EEG was recorded from γ-butyrolacetone-(GBL-) treated mice using a polyimide-based microelectrode array for extracranial multichannel recording. The event detection portion of the algorithm was applied to the recording to identify SWD events. (A) The algorithm identified 20 of 20 visually-identified SWD events, including the sample SWD outlined by the grey dotted line and marked by heavy orange tracing. Additional SWD-like events were also identified by the algorithm (heavy red tracing) in all but one channel. Channels one through seven (of 36 recording channels) are shown. (B) The average event rate/100 seconds was calculated for each channel during baseline (blue) and GBL-treated (red) recordings. (C) The fold change in GBL signal versus baseline at each channel (ratio of GBL event rate to baseline event rate) revealed that the predominant event rate increase occurred in channels 1 through 20. In contrast, channels 27 through 37 experienced a decrease in event rate per 100 seconds relative to baseline, indicating possible suppression or inhibition in this region of the brain in response to GBL-treatment. (D) To identify the brain region of interest, the fold-change in average event rate per 100 seconds was plotted against the location of the associated electrode. The color indicates the fold increase as indicated in the legend (red-orange = high, blue-grey = low). The greatest fold-change was found at electrode 4 and decreased as the relative distance from electrode 4 increased. We conclude that the seizure focus was at or near the position of electrode 4. Of note, electrodes 27 through 37 showed a decrease in events per 100 seconds and were farthest away from electrode 4. NR = non-recording electrode.

create a threshold for whole signal quantification and event identification. Local line lengths calculated from wavelet-decomposed experimental EEG signals are compared to the baseline-dependent threshold, and line lengths above threshold are summed into events. These events are categorized based on duration and amplitude as seizures, spikes, or other abnormal. This is, to our knowledge, the first use of line length and wavelet decomposition for identification of multiple EEG event types by a single detector.

The detection algorithm is tunable to a variety of EEG features that result in increased line length, and several parameters within the algorithm can be adjusted depending on the event type of interest. In addition to parameters related to differences in sampling rate and event duration as mentioned above, modifications to the decomposition level and line length window size may identify different classes of events, based on the event frequency or magnitude. For example, modification of the wavelet decomposition level from A4 to A5 and adjustment of the sampling window from 15 datapoints to 7 allowed us to accurately detect absence seizures in a unique data set generated by a multi-electrode array with a high sampling frequency. Similarly, a wavelet other than Daubechies db4 can be applied to the signal prior to line length calculations. The Daubechies db4 wavelet was selected for spike and seizure identification based on previous reports of its appropriateness for EEG analysis[2], but a systematic application of other wavelets might reveal additional event clusters or provide further refinement of our algorithm.

Because the digitization of EEG records and the miniaturization of recording equipment allows for chronic and long-term recordings in mice that follow the evolution of seizures and epilepsy, new conclusions are being made regarding the contribution of interictal activity to seizures and epileptogenesis[30–32]. Spikes have been closely associated with epileptic foci and used for guiding surgical decisions

regarding resection[33–36]. Notably, our algorithm is able to monitor the development of both spikes and seizures and can therefore rapidly process signals for the analysis of the impact of spikes on seizures and vice versa over long recordings.

A unique aspect of the algorithm is the inclusion of an empty channel (i.e. one that is not connected to an electrode) for signal cleanup. The automated signal cleanup strategy was originally intended to remove high amplitude electrical noise or movement artifacts. Preliminary visual examination of the signal showed that movement-associated artifacts did not increase with the onset of seizure activity. Instead, the primary contribution to "noise" arose from physical disturbance of the recording wires by the investigator during manipulations of the animal (for example, during drug delivery) or were associated with spontaneous events uncorrelated to mouse movement. Qualitatively, direct comparison of video footage and the raw EEG signal showed that normal seizure-related movements of the instrumented animal did not produce the high-amplitude chaotic noise that is removed via the empty channel. Furthermore, visual inspection of the segments of raw signal removed by the automated cleanup algorithm showed that these removed signals were not seizure-like. Comparison of the amount of signal removed between baseline recordings and KA recordings and visual inspection of the type of signal removed from both groups also revealed no difference, indicating that the automated cleanup is not selectively removing convulsive events in the KA mice. Inadvertent removal of convulsive events would, at worst, underestimate spike and seizure events – but this is not compatible with the performance characteristics described extensively in Figure 4. If there was a systematic loss of convulsive events with biological relevance, the accuracy and precision measurements would have been considerably lower and the odds ratio would not have been significant at $P < 0.001$. Finally, analysis of the empty

channel signal revealed the following characteristics: within non-event domains (i.e. signal segments that are kept by the algorithm) the amplitude ranged from $+4.4$ to -16.2 µV on the empty channel; during the same time domain, the EEG signal ranged from $+77.7$ to -67.0 µV. Furthermore, within the signal segments removed by the algorithm, the emtpy channel signal ranged from $+316.0$ to -788.8 µV, vastly exceeding the normal range of EEG signals (across the EEG signal the range was approximately $+100$ to -150 µV). This indicates that the empty channel is normally very "quiet" but is punctuated by periods of extremely high amplitude noise. In contrast, the true EEG channel does not exhibit such massive amplitude events, perhaps due to overall electronic suppression associated with the presence of a tissue-resident electrode. Regardless, the empty channel algorithm removes any timeframes that may be associated with spurious noise, especially noise in the EEG channel that may be visually indistinguishable from biological events.

In conclusion, we have generated an algorithm for automated analysis of the murine EEG based on the line length feature and wavelet decomposition. Our algorithm will help to reduce the time of analysis for quantification of seizure and spike activity by providing an accurate and reproducible method for processing large EEG data sets automatically.

Methods

Animal use and care. Animal use and care was performed in accordance with guidelines established by the National Institutes of Health and all experiments and procedures were approved by the Mayo Clinic Institutional Animal Care and Use Committee.

Anesthesia and Stereotaxic surgery. Female C57BL/6 mice (Jackson Labs, Bar Harbor, ME) were 40 days old (range 33–47) and weighed 16.8 g (range 14.3–19.2) at the time of surgery. Mice were anesthetized by intraperitoneal injection of ketamine (100 mg/kg), xylazine (10 mg/kg), and acepromazine (3 mg/kg). Upon reaching a surgical plane of anesthesia[37], fur on the head was trimmed and mice were immobilized in a stereotaxic frame (Stoelting, Chicago, IL, USA). Screw electrodes (0.1" length) with 1 mm wire-loops (Pinnacle Technologies, Lawrence, KS) were implanted at the following stereotactic coordinates relative to bregma (caudal, lateral, ventral; in mm): recording electrode (right hippocampus): 2.0, 1.7, 1.0; reference electrode (left somatosensory cortex, hind limb region): 0.5, -1.5, 1.5; ground electrode (preculminate fissure, cerebellum): 5.8, 0.0, 1.0. Electrodes were secured in the skull with dental cement (A–M Systems, Sequim, WA). Standardized six-pin surface-mount adaptors (Pinnacle Technologies) were connected to the electrodes via short lengths (less than 5 mm) of wire-wrapping wire (Radio Shack, Fort Worth, TX). Only one true recording electrode was placed, leaving three unoccupied pins on the headmount. One of these empty channels was used for the automated cleanup algorithm, as described above. Colloidal silver (Ted Pella, Inc., Redding, CA) was applied to the electrode-adaptor connection to ensure electrical conduction. The electrodes and the base of the adaptor were buried in dental cement to stabilize and insulate the connections. The skin was sutured closed around the head mount and triple-antibiotic ointment was applied to the wound to prevent infection. Buprenorphine (0.075 mg/kg) was administered prior to recovery from anesthesia and the morning following surgery. Acetaminophen was added to the water bottle for 48 hours prior to surgery and was maintained in the cage for seven days after surgery. Mice were euthanized by inhalation of isoflurane.

EEG and video recording and manual seizure scoring. EEG and video recordings were initiated seven to 14 days after surgery. The EEG recording apparatus consisted of a $10\times$ pre-amplifier configured to record two EEG channels, a low-torque swivel, and a data acquisition and conditioning system connected to a PC running PAL-8200 EEG software (Pinnacle Technologies). Data were sampled at 400 Hz and 14-bit resolution on each channel, with 1.0 Hz high pass filtering and 100 Hz low pass filtering. Custom nine inch-diameter plastic cylindrical recording cages were produced by the Engineering Department of the Mayo Clinic (Rochester, MN). Simultaneous video was recorded using high resolution color cameras and VistaPro 6 server software (Lorex Technology, Inc., Markham, Ont). EEG signals were collected in European Data Format (.edf) and were converted to text files using the Prana Software Suite (Phi Tools, France). At least one session of one hour of video and EEG was collected prior to experimental manipulation. These were used as the baseline control for signal and behavioral analysis. Videos were observed by a blinded observer for scoring according to Racine[1] and Borges[38] with the following modifications: stages 1 and 2 (including mouth and facial movements, head nodding, and immobility) were classified as "non-convulsive," and stages 3, 4, 5, and 6 (including forelimb clonus, rearing, rearing and falling, loss of posture, or jumping) were classified as "convulsive." Visual inspection of the EEG signal was carried out in EEGLab[39]. Signal was characterized as normal (similar in amplitude and frequency to the average baseline recording for each specific mouse), spiking (transient, isolated bursts of

high-amplitude activity at less than one event per second), seizure (greater than five seconds of sustained, high-frequency and high-amplitude, high frequency, or high amplitude activity), or other abnormal (signal that appeared significantly and identifiably different in frequency or amplitude from the average baseline recording for that specific mouse but did not fall into the category of spike or seizure). For the manual analyses, simultaneous video and EEG recordings were edited to remove artifacts related to animal manipulations (injections) prior to examination in approximately one-hour segments. A total of 31 recordings collected from 7 mice were used for the primary signal analyses and algorithm development. Of these recordings, ten were baseline (no treatment) and 21 were collected following injection of kainic acid. For the explicit side-by-side comparison of video analysis and raw EEG, a subset of ten video recordings (collected simultaneously with EEG) from two mice was used. Three of these video recordings were baseline and seven were collected following injection of kainic acid.

Seizure induction by kainic acid. To induce seizures, 1 mg/mL (-)- (α)-Kainic Acid (KA) (Cayman Chemical, Ann Arbor, MI) in PBS was injected into the peritoneum at 10 µg/g body weight at 60 minute intervals (up to five doses) until Racine stage 4, 5, or 6 seizures were observed[1,38]. Mice were euthanized after two hours of status epilepticus or if seizures were not observed after five KA doses.

Algorithm. The algorithm and all algorithm-related data analysis was written and performed in the Matlab (Mathworks, Natick, MA) computing environment. The annotated code can be found within the supplemental materials. The algorithm consists of four blocks applied to each individual animal: 1) automated signal clean-up; 2) determination of baseline line length parameters; 3) bulk signal analysis to categorize experimental recordings as equal to or different than baseline; 4) event identification, categorization, and quantification in the experimental recordings. The data from these calculations were compiled into a Matlab structure, including the file name, recording length, individual event identification with time index and duration, total event count, total event duration, and bulk signal score (0 = not different from baseline), to provide a summary file of the EEG signal analysis.

Algorithm performance analysis. The automated algorithm was objectively compared to manual EEG signal analysis using the receiver-operator characteristic (ROC)[26], a curve-fitting strategy that determines the balance of false positives and true positives. The more steeply the curve rises to a true positive rate of 1.0, the better the performance of the algorithm under analysis. An algorithm that performs at chance would rise at a 45° angle (50:50 odds of a correct call). A quantitative measure of performance arises by measuring the area under the ROC curve (AUC)[27]. A perfectly performing algorithm would show 1.0 true positive rate and 0.0 false positive rate, yielding an area of 1.0 under the ROC curve. A randomly performing algorithm would yield an AUC of 0.5. Further performance analysis is provided by constructing a 2×2 contingency matrix of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). From this, precision is calculated as the ratio of TP to TP+FP; recall (or sensitivity) is the ratio of TP to TP+FN; accuracy is the ratio of TP+TN to TP+FP+FN+TN; and the F characteristic is 2 divided by $(\text{precision})^{-1}+(\text{recall})^{-1}$. Precision is interpreted as the optimal identification of true positive events, relative to total positive calls. Accuracy is interpreted as the optimal identification of both true positives and true negatives, relative to the total calls made. Hence, at low thresholds, precision is high because most true postives are identified, but accuracy is low because the number of true negatives is greatly underestimated. As the threshold increases, precision is lost but accuracy improves as the balance of true positive and true negative calls reaches an optimal relationship. At peak accuracy, the algorithm is correctly identifying the most true positives and true negatives and the least false positives and false negatives. Increasing the threshold further decays both precision and accuracy because true positives are missed.

1. Racine, R. J. Modification of seizure activity by electrical stimulation. II. Motor seizure. *Electroencephalogr Clin Neurophysiol* **32**, 281–294 (1972).
2. Adeli, H., Zhou, Z. & Dadmehr, N. Analysis of EEG records in an epileptic patient using wavelet transform. *J Neurosci Methods* **123**, 69–87 (2003).
3. Abend, N. S. *et al.* Interobserver reproducibility of electroencephalogram interpretation in critically ill children. *J Clin Neurophysiol* **28**, 15–19 (2011).
4. Drinnan, M. J., Murray, A., Griffiths, C. J. & Gibson, G. J. Interobserver variability in recognizing arousal in respiratory sleep disorders. *Am J Respir Crit Care Med* **158**, 358–362 (1998).
5. Gerber, P. A. *et al.* Interobserver agreement in the interpretation of EEG patterns in critically ill adults. *J Clin Neurophysiol* **25**, 241–249 (2008).
6. Coutin-Churchman, P. E. *et al.* Quantification and localization of EEG interictal spike activity in patients with surgically removed epileptogenic foci. *Clin Neurophysiol* **123**, 471–485 (2012).
7. Ko, C.-W. & Chung, H. Automatic spike detection via an artificial neural network using raw EEG data: effects of data preparation and implications in the limitations of online recognition. *Clin Neurophysiol* **111**, 477–481 (2000).
8. Buteneers, P. *et al.* Automatic detection of epileptic seizures on the intra-cranial electroencephalogram of rats using reservoir computing. *Artif Intell Med* **53**, 215–223 (2011).
9. Chiu, A. W. L., Daniel, S., Khosravani, H., Carlen, P. L. & Bardakjian, B. L. Prediction of Seizure Onset in an In-Vitro Hippocampal Slice Model of Epilepsy

Using Gaussian-Based and Wavelet-Based Artificial Neural Networks. *Ann Biomed Eng* **33**, 798–810 (2005).

10. Esteller, R., Echauz, J., Tcheng, T., Litt, B. & Pless, B. in *Proceedings of the 23rd Annual International Conference of the IEEE.* **2**, 1707–1710 (Engineering in Medicine and Biology Society, 2001).

11. Osorio, I., Frei, M. G. & Wilkinson, S. B. Real-time automated detection and quantitative analysis of seizures and short-term prediction of clinical onset. *Epilepsia* **39**, 615–627 (1998).

12. Gotman, J. in *Current Practice of Clinical Electroencephalography* (eds Daly, D. D. & Pedley, T. A.) 51–83 (Raven Press, 1990).

13. Guo, L., Rivero, D., Dorado, J., Rabunal, J. R. & Pazos, A. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *J Neurosci Methods* **191**, 101–109 (2010).

14. Nigam, V. P. & Graupe, D. A neural-network-based detection of epilepsy. *Neurol Res* **26**, 55–60 (2004).

15. Subasi, A. & Erçelebi, E. Classification of EEG signals using neural network and logistic regression. *Comput Methods Programs Biomed* **78**, 87–99 (2005).

16. Guerrero-Mosquera, C., Malanda Trigueros, A., Iriarte Franco, J. & Navia-Vázquez, Á. New feature extraction approach for epileptic EEG signal detection using time-frequency distributions. *Med Biol Eng Comput* **48**, 321–330 (2010).

17. Pavlov, A. N., Makarov, V. A., Mosekilde, E. & Sosnovtseva, O. V. Application of wavelet-based tools to study the dynamics of biological processes. *Brief Bioinform* **7**, 375–389 (2006).

18. Daubechies, I. *Ten Lectures on Wavelets.* 53–105 (SIAM, 1992).

19. Gabor, A. J., Leach, R. R. & Dowla, F. U. Automated seizure detection using a self-organizing neural network. *Electroencephalogr Clin Neurophysiol* **99**, 257–266 (1996).

20. Shoeb, A. *et al.* in *Proceedings of the 23rd Annual International Conference of the IEEE.* **1**, 419–422 (Engineering in Medicine and Biology Society, 2004).

21. Saab, M. E. & Gotman, J. A system to detect the onset of epileptic seizures in scalp EEG. *Clin Neurophysiol* **116**, 427–442 (2005).

22. Khan, Y. U. & Gotman, J. Wavelet based automatic seizure detection in intracerebral electroencephalogram. *Clin Neurophysiol* **114**, 898–908 (2003).

23. Ben-Ari, Y. Limbic seizure and brain damage produced by kainic acid: mechanisms and relevance to human temporal lobe epilepsy. *Neuroscience* **14**, 375–403 (1985).

24. Ben-Ari, Y. & Cossart, R. Kainate, a double agent that generates seizures: two decades of progress. *Trends Neurosci* **23**, 580–587 (2000).

25. Nadler, J. V. Minireview. Kainic acid as a tool for the study of temporal lobe epilepsy. *Life Sci* **29**, 2031–2042 (1981).

26. Fawcett, T. ROC graphs: notes and practical considerations for researchers. *Pattern Recognit Lett* **27**, 882–891 (2010).

27. Ling, C. X., Huang, J. & Zhang, H. in *Proceedings of the 18th International Joint Conference of Artificial Intelligence.* **1**, 329–341 (2003).

28. Choi, J. H., Koch, K. P., Poppendieck, W., Lee, M. & Shin, H.-S. High resolution electroencephalography in freely moving mice. *J Neurophysiol* **104**, 1825–1834 (2010).

29. van Luijtelaar, E. L. & Coenen, A. M. Two types of electrocortical paroxysms in an inbred strain of rats. *Neurosci Lett* **70**, 393–397 (1986).

30. Barkmeier, D. T. *et al.* Electrical, molecular and behavioral effects of interictal spiking in the rat. *Neurobiol Dis* (2012).

31. Staley, K. J., White, A. & Dudek, F. E. Interictal spikes: harbingers or causes of epilepsy? *Neurosci Lett* **497**, 247–250 (2011).

32. White, A. *et al.* EEG spike activity precedes epilepsy after kainate-induced status epilepticus. *Epilepsia* **51**, 371–383 (2010).

33. Cascino, G. D. *et al.* Long-term follow-up of stereotactic lesionectomy in partial epilepsy: predictive factors and electroencephalographic results. *Epilepsia* **33**, 639–644 (1992).

34. Krendl, R., Lurger, S. & Baumgartner, C. Absolute spike frequency predicts surgical outcome in TLE with unilateral hippocampal atrophy. *Neurology* **71**, 413–418 (2008).

35. Miller, J. W. & Gotman, J. The meaning of interictal spikes in temporal lobe epilepsy: should we count them? *Neurology* **71**, 392–393 (2008).

36. Rosati, A. *et al.* Intractable temporal lobe epilepsy with rare spikes is less severe than with frequent spikes. *Neurology* **60**, 1290–1295 (2003).

37. Buitrago, S., Martin, T. E., Tetens-Woodring, J., Belicha-Villanueva, A. & Wilding, G. E. Safety and efficacy of various combinations of injectable anesthetics in BALB/c mice. *J Am Assoc Lab Anim Sci* **47**, 11–17 (2008).

38. Borges, K. *et al.* Neuronal and glial pathological changes during epileptogenesis in the mouse pilocarpine model. *Exp Neurol* **182**, 21–34 (2003).

39. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* **134**, 9–21 (2004).

Acknowledgements

Author contributions

R.A.B. performed the experiments, designed the algorithm, analyzed the data, and wrote the manuscript. J.H.C. performed the multi-channel recordings. A.M. designed the algorithm. H.S.S. designed the multi-channel recording experiments. G.A.W. designed the algorithm. C.L.H. conceived the experiments, designed the algorithm, analyzed the data, and wrote the manuscript. All authors reviewed the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

How to cite this article: Bergstrom, R.A. *et al.* Automated identification of multiple seizure-related and interictal epileptiform event types in the EEG of mice. *Sci. Rep.* **3**, 1483; DOI:10.1038/srep01483 (2013).