Please cite this article in press as: Ren et al., A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons, Stem Cell Reports (2026), https://doi.org/10.1016/j.stemcr.2026.102872

STEM CELL REPORTS
Article
ISSCR
OPEN ACCESS

A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons

Xiaoxuan Ren,[1,2,6] Carissa L. Sirois,[3,4,6] Raymond Doudlah,[4] Ethan E. Dayley,[3,4] Natasha M. Méndez-Albelo,[3,4,5] Aviad Hai,[1,2,*] Ari Rosenberg,[4,*] and Xinyu Zhao[3,4,7,*]
[1]Department of Biomedical Engineering, University of Wisconsin-Madison, Madison, WI 53705, USA
[2]Department of Electrical and Computer Engineering, University of Wisconsin - Madison, Madison, WI 53706, USA
[3]Waisman Center, University of Wisconsin-Madison, Madison, WI 53705, USA
[4]Department of Neuroscience, School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI 53705, USA
[5]Molecular Cellular Pharmacology Training Program, University of Wisconsin-Madison, Madison, WI 53705, USA
[6]These authors contributed equally
[7]Lead contact
*Correspondence: ahai@wisc.edu (A.H.), ari.rosenberg@wisc.edu (A.R.), xinyu.zhao@wisc.edu (X.Z.)
https://doi.org/10.1016/j.stemcr.2026.102872

SUMMARY

Neurons derived from human pluripotent stem cells (hPSCs) are valuable models for studying brain development and developing therapies for brain disorders. Evaluating hPSC-derived neurons requires assessing their electrical activity, which can be achieved using multi-electrode arrays (MEAs) for extracellular recordings. Because each electrode channel generally detects activity from multiple neurons, resolving the activity of single neurons requires a process called spike sorting. However, currently available methods were not developed for analyzing data from hPSC-derived neurons and require complex workflows and time-consuming manual intervention. Here, we introduce a semi-automated MEA spike sorting software (SAMS) designed specifically for low-density MEA recordings of cultured neurons. SAMS outperforms commercially available automated spike sorting algorithms in terms of accuracy and greatly reduces computational and human processing time. By providing an accessible, efficient, and integrated platform for spike sorting, SAMS enhances the resolution and utility of MEA in disease modeling and drug development using hPSC-derived neurons.

INTRODUCTION

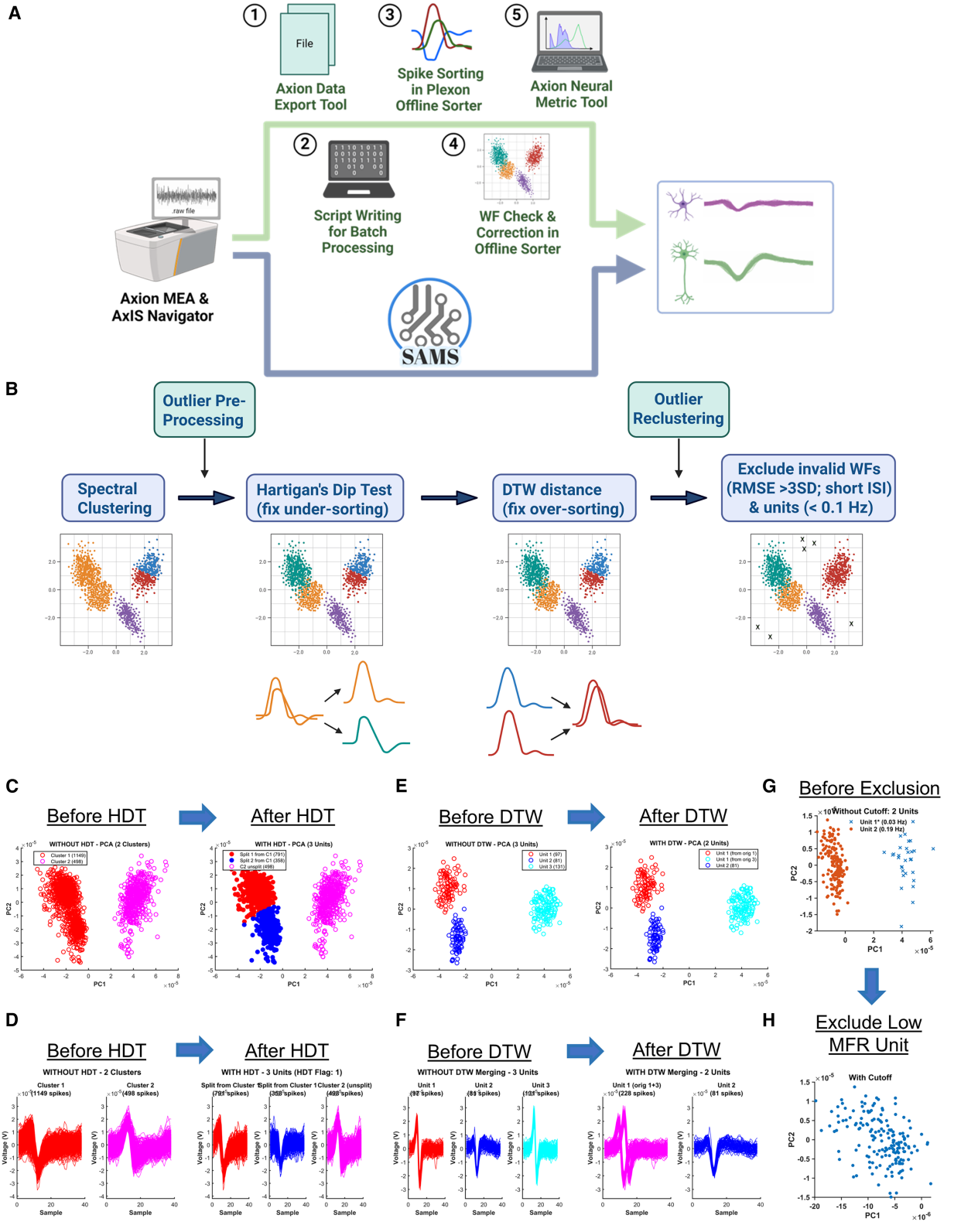
Human pluripotent stem cells (hPSCs) are useful experimental models for defining human-specific biology and disease mechanisms, especially for brain disorders (Niu and Parent, 2020; Sandoval et al., 2024; Zhao and Bhattacharyya, 2018). To evaluate the development and function of hPSC-derived neurons, it is important to assess their electrophysiological properties, particularly their spiking characteristics. Multi-electrode arrays (MEAs) are a widely used tool for simultaneously recording from many neurons *in vivo* (Choi et al., 2021; Guan et al., 2023; Jun et al., 2017; Obaid et al., 2020), *ex vivo* (Zhang et al., 2022), or *in vitro* (Liu et al., 2023; Lv et al., 2023; Spira and Hai, 2013). MEAs detect extracellular field potentials generated by cells proximal to electrodes and have been used to assess the intrinsic properties and network activities of hPSC-derived neurons, allowing for relatively high-throughput assessment of the impact of disease mutations and effects of drug-like compounds (Kiskinis et al., 2014; Liu et al., 2015; Vatine et al., 2017). However, widely used low-density MEA recordings often capture the combined activity of multiple neurons on an individual electrode. To accurately characterize the electrical properties of hPSC-derived neurons, the activities of individual neurons need to be isolated by sorting spikes.

Many spike-sorting pipelines exist for both high-density (electrode spacing <100 μm) and lower-density MEAs (electrode spacing 300–350 μm; Table S1). However, most of these pipelines were created for either *in vivo* recordings or non-human neurons (Carlson and Carin, 2019) and were not optimized for hPSC-derived neurons that have different electrophysiological properties (Cotterill et al., 2016; Hyvärinen et al., 2019; Mayer et al., 2018; Mossink et al., 2021; Napoli and Obeid, 2016). Furthermore, these pipelines require at least some coding skills to run or additional software tools to analyze spike-sorted data. These pipelines thus require time-consuming manual checking and correction of the sorted spikes (McCready et al., 2022; Mehta et al., 2018; Pradeepan et al., 2024; Tyssowski et al., 2019; Vertkin et al., 2015), which limits their application to the analysis of hPSC-derived neurons that contain hundreds of electrodes on a single multi-well plate.

To overcome the challenges of current spike sorting approaches, we developed a semi-automated MEA spike sorting software (SAMS) using low-density MEA data collected from hPSC-derived neurons acquired via the Axion MEA platform. Our goal was to create an open-source end-to-end spike sorting and analysis tool that was user-friendly and accessible to more researchers, especially those without computational expertise. The workflow of SAMS is user interface-based and requires no coding to

Published by Elsevier Inc. on behalf of International Society for Stem Cell Research.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: Ren et al., A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons, Stem Cell Reports (2026), https://doi.org/10.1016/j.stemcr.2026.102872



(legend on next page)

Please cite this article in press as: Ren et al., A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons, Stem Cell Reports (2026), https://doi.org/10.1016/j.stemcr.2026.102872

implement. Importantly, SAMS flags electrodes that need manual verification, thereby reducing the time needed for analysis. We compared the spike sorting performance of SAMS with Offline Sorter (OFS) software (Plexon, Inc.), the current state-of-the-art software recommended for analyzing low-density Axion MEA recordings. SAMS not only outperformed OFS on spike sorting accuracy, but also decreased the amount of manual validation time. We then used SAMS to reanalyze two previously published MEA datasets of hPSC-derived neurons (Guo et al., 2023; Shen et al., 2023). The SAMS pipeline not only confirmed published conclusions but also uncovered additional differences among experimental conditions. Therefore, SAMS enhances the resolution and utility of MEA recordings in disease modeling and drug development using hPSC-derived neurons.

RESULTS

Overview of spike sorting recommendations for the Axion MEA platform

The recommended pipeline for spike sorting of Axion MEA data involves multiple software tools and manual steps (Figure 1A). First, Axion's AxIS Navigator processes the recordings into spike files, which are converted to NeuroExplorer format using Axion Data Export Tool. Spike sorting using OFS is done one file at a time, through either manual sorting or semi-automated sorting, followed by manual inspection and correction of each sorted electrode channel. Finally, the sorted spikes are analyzed using Axion's Neural Metric Tool, and results are exported. This workflow is time-intensive and requires coding for batch processing. Alternative custom pipelines (Table S1) also require coding proficiency for scripting and were not created for low-density recordings from hPSC-derived neurons.

Overview of spike sorting using SAMS

To address the inefficiencies described above, we developed SAMS as an integrated one-stop platform, enabling automated batch processing of MEA spike files directly into analyzed output results (Figure 1A). SAMS provides an intuitive graphical user interface (GUI) to eliminate the need for scripting while integrating manual spike sorting tools. By consolidating multiple manual steps required for other pipelines, SAMS offers a streamlined workflow to accelerate data processing and analysis. The simplified user experience lowers adoption barriers, enabling scientists without specialized computational skills to readily incorporate high-fidelity information on neuronal encodings into their research.

SAMS employs a multi-step spike sorting pipeline for robust classification (Figures 1B and S1). First, detected spike waveforms are projected onto two principal components. Spectral clustering (von Luxburg, 2007) performs initial unit assignments by testing K = 1–5 clusters and selecting the optimal number using the Davies-Bouldin Index. This unsupervised step maximizes between-cluster variance while minimizing within-cluster variance. Second, outlier waveforms exceeding three standard deviations (SDs) are identified and removed to improve cluster quality (Figures S2A–S2E). Third, Hartigan's Dip Test (HDT) (Hartigan and Hartigan, 1985) assesses unit purity by detecting multimodality in the PCA space of each cluster. Clusters failing HDT are split to correct undersorting (Figures 1C, 1D, and S2F–S2K). Fourth, dynamic time warping (DTW) (Cao et al., 2016) resolves oversorting errors by merging units with similar spike waveform templates (Figures 1E, 1F, and S3A–S3H) based on a user-defined threshold (default: 1.5; Figures S3I–S3M). After a second merge pass to verify cluster assignments, a final step removes invalid waveforms: (1) waveforms exceeding three SDs from each unit's mean template (i.e., averaged waveform), (2) waveforms from low-firing-rate units (default: <0.1 Hz) (Mossink et al., 2021) (Figures 1G and 1H), and (3) waveforms with inter-spike intervals (ISIs) falling within the user-defined refractory period (default: 1.5 ms). By integrating quality control metrics at each step—testing cluster integrity, merging probable oversplits, and removing outliers or false positives—SAMS balances the demands of spike sorting accuracy and manual intervention. The multi-tier workflow maximizes the yield of

Figure 1. Comparison of standard spike sorting workflow and SAMS

(A) Top: standard pipeline for spike sorting of Axion MEA data involving multiple software tools and manual steps. Bottom: pipeline for analysis of Axion MEA data using SAMS, which streamlines the processes shown above by directly interfacing with AxIS Navigator output files and performing automatic batch processing.

(B) Schematic of SAMS processing stages: (1) spectral clustering to perform initial waveform clustering, (2) outlier waveform pre-processing, (3) Hartigan's dip test (HDT) to resolve undersorting, (4) DTW distance calculation to resolve oversorting, (5) outlier reclustering, and (6) exclusion of invalid units and waveforms based on spike frequency criteria and short inter-spike interval (ISI), respectively.

(C–H) PCA plots and waveform traces from analyzed MEA data showing examples of electrodes that were corrected by different stages of the SAMS pipeline.

(C and D) Correction of an undersorted electrode (red) using HDT.

(E and F) Correction of an oversorted electrode (red and cyan in the "Before DTW" images) using DTW distance.

(G and H) Removal of a low-firing-rate unit based on the user-defined threshold (here, 0.1 Hz). See also Figures S2 and S3.

high-confidence isolated single units required for revealing biological insights and allows users to customize parameters at each stage of the pipeline to optimize sorting of their data (Table S2). SAMS also returns the parameter settings used for each analysis in the output file to facilitate transparent reporting of methods and to allow users to verify consistent settings between analysis batches.

SAMS provides a user-friendly interface containing key information of output data

To enable users to customize their analyses, SAMS includes panels (Figure 2A) for tuning single-unit burst detection parameters at the single-electrode level (minimum spike count; maximum intra-burst ISI) as well as network burst detection at the multi-electrode level (spike counts, ISIs, and percentage of electrodes participating). While the default parameters are set to be consistent with the defaults applied by AxIS Navigator, which are used by most researchers, the capacity to modify these parameters provides researchers additional flexibility in analyzing their data (Table S2).

SAMS also provides a GUI for manual verification (Figure 2B) and comprehensive output files (Figure 3A). Automated sorting outputs are compiled into a convenient slide deck file (Figure 3B), showing unit assignments, PCA plots, and waveform counts for each electrode. Detailed per-unit metrics are provided in a spreadsheet (Figure 3C), including two lists of flagged electrodes (Figure 3D): the possible multi-unit list (electrodes failing HDT or below the DTW threshold) and the over-excluded unit list (electrodes exceeding the exclusion threshold). Users can reload flagged electrodes into SAMS' manual sorting interface for refinement using click-and-drag selection tools.

SAMS provides consistent performance for spike sorting hPSC-derived neurons

Since the recommended spike sorting tool for data acquired using the Axion platform is Plexon's OFS software (Figure 1A), we benchmarked the performance of SAMS by comparing it to OFS. We first created a "ground truth" spike sorting dataset for comparison (Figures 4A and S4A–S4C) by manually sorting previously published MEA data recorded from dorsal forebrain excitatory neurons derived from one control (CTRL2) and one fragile X syndrome (FXS2) iPSC line (Shen et al., 2023). OFS was used to manually sort waveforms and determine the number of units on each electrode. This ground truth dataset was then used to assess the automated spike sorting performances of SAMS and OFS.

We first evaluated the accuracy of the SAMS pipeline without any manual correction. SAMS correctly sorted more than 70% of electrodes in both CTRL2 and FXS2 conditions (Figure 4B). Among incorrectly sorted electrodes, most were errors of undersorting (17.5%–20.3%), with oversorting being less frequent (6.0%–9.7%). Next, we compared SAMS to the four algorithms offered by OFS. We first ran each algorithm using default settings, then ran them several more times after adjusting one or more parameters. The settings with the highest pooled accuracy are shown in Figures 4C–4F, and accuracy values for each batch across different algorithms are shown in Figures 4G and 4H (see also Figures S4D–S4G; Table S3). Among the OFS algorithms, T-Dist EM achieved the highest accuracy (Figures 4C, 4G, and 4H), with accurate sorting of up to 43.8% of FXS2 and up to 64.5% of CTRL2 electrodes. Although 64.5% is within acceptable accuracy ranges, and in some cases similar to the accuracy achieved by SAMS, the T-Dist EM Scan algorithm exhibited high variability between batches of the same condition (6.9%–88.9% for CTRL2; 1.2%–77.8% for FXS2), while SAMS performed more consistently for all neurons from both FXS2 and CTRL2 conditions (67.4%–94.4% for CTRL2, 62.9%–80.2% for FXS2; Figures 4G and 4H). The other three OFS algorithms performed worse than T-Dist EM Scan (Figures 4D–4F), with none achieving more than 42.9% correct sorting (Figures 4G and 4H). In addition, SAMS consistently achieved lower levels of oversorting than the OFS algorithms (Figures S4D and S4E), while errors of undersorting were similar across all platforms and algorithms (Figures S4F and S4G). Therefore, SAMS, which was developed specifically for sorting hPSC-derived neuron data, consistently achieved accuracy levels comparable to spike sorting methods used to analyze *in vivo* or simulated MEA datasets (Buccino et al., 2020; Mohammadi et al., 2024; Souza et al., 2019; Wang et al., 2023).

Correction of electrodes flagged by SAMS can further improve accuracy

A major challenge in performing spike sorting of MEA data recorded from hPSC-derived neurons is the sheer volume of electrodes to be sorted, because it is necessary to assess not only multiple hPSC lines, but also multiple batches of neurons derived from each hPSC line to minimize both biological and technical variabilities (Anderson et al., 2021). For example, weekly recording of one 96-well plate from Axion for 8 weeks produces data from 6,144 electrodes.

Even though SAMS offers acceptable accuracy levels, users may want to achieve higher accuracy levels than those produced by automatic sorting alone. To help strike a balance between the improved accuracy offered by manual electrode correction after automatic sorting and reduced labor, SAMS returns two lists of electrodes flagged by the algorithm as possibly being incorrectly sorted in each output file (Figure 3D). Manual correction of only flagged electrodes in FXS2 and CTRL2 datasets improved

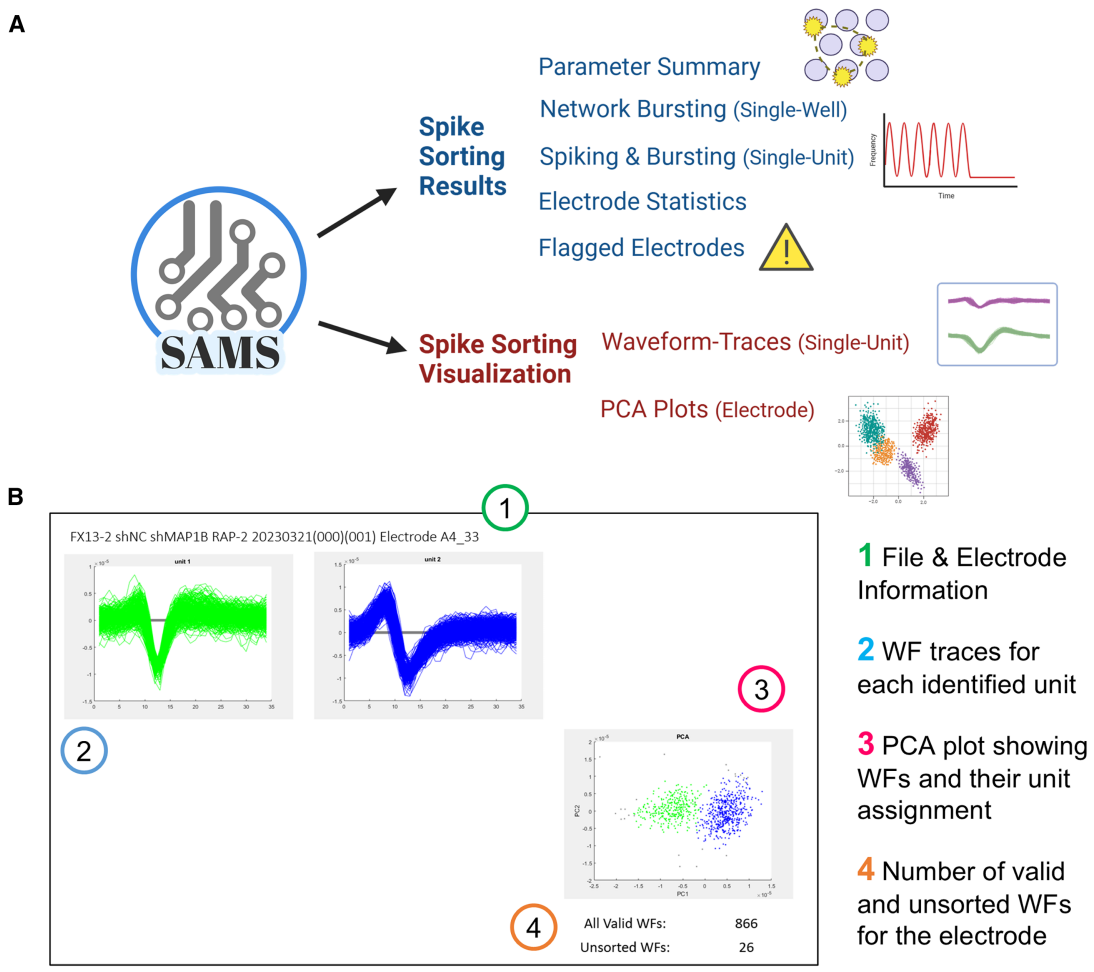


Figure 2. SAMS graphical user interfaces and features

(A) SAMS graphical user interface (GUI) for automatic batch spike sorting, which includes (1) file selection; (2) spike sorting parameter adjustment (Table S2); (3) single-unit burst detection parameters; (4) electrode-level network burst detection parameters; and (5) progress indicator.

(B) SAMS GUI for manual correction of spike sorting results, which includes (1) data file selection; (2–4) parameter adjustments (same as in A); (5) electrode selection; and (6) correction tools. Data shown are for illustrative purposes.

Please cite this article in press as: Ren et al., A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons, Stem Cell Reports (2026), https://doi.org/10.1016/j.stemcr.2026.102872



C

	A	B
1		parameter Value
2	threshold_to_merge	1.5
3	refractoryT	0.0015
4	min_spikes_E	5
5	max_ISI_E	100
6	min_spikes_N	50
7	max_ISI_N	100
8	network_participation_threshold	0.35
9	cutoff_frequency	0.1
10	flag_threshold	0.2
11	std_cutoff	3
12	overlap_threshold	0.7

D

	A	B	C
1	Possible MultiUnit	over-exlcuded unit	
2	A2_11	A6_13	
3	A3_11	A9_33	
4	A4_12	A11_21	
5	A5_33	B6_31	
6	A6_33	B9_31	
7	A7_11	C3_33	
8	A7_13	C7_21	
9	A7_23	C8_12	
10	A7_33	C8_13	

Figure 3. Output files for comprehensive spike sorting

(A) Schematic illustrating SAMS data and outputs.

(B) Electrode-specific spike sorting results are saved in a Powerpoint slide deck. Each slide presents sorting outcomes for an individual electrode (1), displaying color-coded waveforms of defined units (2) alongside corresponding Principal-component analysis (PCA) plot (3)

(legend continued on next page)

the accuracy by 2.0%–8.3% (Figure S5A as compared to Figures 4B and S5B–S5F).

SAMS is highly accurate in analyzing various recording conditions

Longer recordings generate more data points and increase analytical burden, while shorter recordings can affect MEA data consistency (Negri et al., 2020). Therefore, we first assessed whether recording time impacted the performance of SAMS. The human-machine consistency (HMC) remained high (mean: 68.6%–70.8%) and stable across recording durations of up to 15 min for both FXS2 and CTRL2 neurons (Figures 5A and S6A). Figure 5B provides a detailed view of electrode-specific performance across recording durations, revealing patterns of electrodes that were consistently accurately sorted (cyan) compared to our ground truth dataset and those requiring manual correction (red). These results reflect the increase in accurately sorted electrodes indicated by the results in Figure 5A and also show that certain electrodes are difficult to resolve independent of recording duration. Because hPSC-derived neurons continue to mature *in vitro* after plating (Cotterill et al., 2016; Hruska-Plochan et al., 2024; Lemieux et al., 2024; Mossink et al., 2021); we also examined whether the culturing time affected the performance of SAMS (Figure S6F). Here, HMC remained relatively stable (60.7%–100%) between 6 and 8 weeks of culture (Figures 5C and 5D). Therefore, SAMS exhibited high accuracy regardless of the developmental stage of the neural network.

We next asked whether the accuracy of SAMS was related to the number of recorded spikes by plotting the percentage of correctly sorted units (Figure 5E) or electrodes (Figure S6B) as a function of total spike count. The majority of units (87.5%) in the entire dataset (both FXS2 and CTRL2) did not exceed 10,057 spikes over the 15-min recording duration (~11.2 Hz), and accuracy up to this spike count remained at or above 70.2%. We did observe a decrease in accuracy for units and electrodes that had greater than 10,057 spikes, although cells with firing rates at these levels were less common in our dataset. For users aiming to analyze datasets with a larger number of cells with such high firing rates, we strongly suggest that they empirically tune the parameter settings for SAMS to maximize accuracy for that higher range.

Motivated by our findings, and because FXS neurons tend to exhibit higher spike counts than control neurons (Shen et al., 2023), we decided to compare the accuracy of SAMS between the two conditions (Figure S6C). To assess this, Kendall's Tau was selected as a non-parametric measure of ordinal association, and significance was tested for (a) all electrodes, (b) only FXS2, and (c) only CTRL2 conditions (Figure S6D). For all three conditions, there was a significant, monotonic negative association between SAMS performance and the number of spikes. These findings indicate that there is a modest decrease in the performance of SAMS at higher spike counts, although the accuracy achieved at these higher spike counts is still within an acceptable range. We further tested whether SAMS performed differently by condition (FXS2 vs. CTRL2) by using a bootstrapped 95% confidence interval of the proportion of electrodes accurately sorted by SAMS. This confidence interval (−0.0812, 0.0894) contained 0 (no group difference), indicating that there was no significant difference in SAMS performance between FXS2 and CTRL2.

Finally, we wanted to test the robustness of SAMS by analyzing more diverse MEA recording datasets. Since SAMS was developed using neurons differentiated using a dual SMAD-inhibition protocol (Shen et al., 2023), we next assessed whether SAMS could be used to analyze neurons generated through Neurogenin 2 (*NGN2*)-directed differentiation, as these neurons can exhibit different temporal trajectories in their differentiation and maturation (Hulme et al., 2022; Lin et al., 2021; Rosa et al., 2020). First, we examined the accuracy of SAMS in sorting neurons from iPSCs derived from an ASD patient with 5q13.2 triplication (Guo et al., 2023) Compared to the ground truth data obtained by manual spike sorting in OFS (Figure 4A), SAMS averaged 83.4% accurate sorting of electrodes across the three batches of neuronal differentiation (Figure S6E). We next assessed the performance of SAMS on a single batch of MEA data from *NGN2*-induced neurons recorded from 12 to 23 days *in vitro* by another laboratory (Afshar-Saber et al., 2024b). These iPSCs were derived from a patient with a homozygous *ALDH5A1* mutation (Proband), the patient's mother (Mom), who is a carrier for the mutation, and a CRISPR-corrected isogenic control line (isoCTRL) (Afshar-Saber et al., 2024a). We observed consistent accuracy in spike sorting by SAMS over the duration of the time course for all three conditions, with only one time point from one of the cell lines falling below 75.6% accuracy (Figures 5F and 5G). These accuracy values (65.5%–100%) were also similar to the accuracy values achieved for our datasets (Figures 4B and S6A).

and quantification of valid and invalid waveforms (WFs; 4). This visualization facilitates rapid assessment of automatic sorting quality and aids in identifying potential sorting errors.

(C) Parameter list: summary of parameter settings used for automated spike sorting.

(D) Quality control: list of potentially missorted electrodes flagged by SAMS, enabling targeted manual refinement of sorting results.

Please cite this article in press as: Ren et al., A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons, Stem Cell Reports (2026), https://doi.org/10.1016/j.stemcr.2026.102872

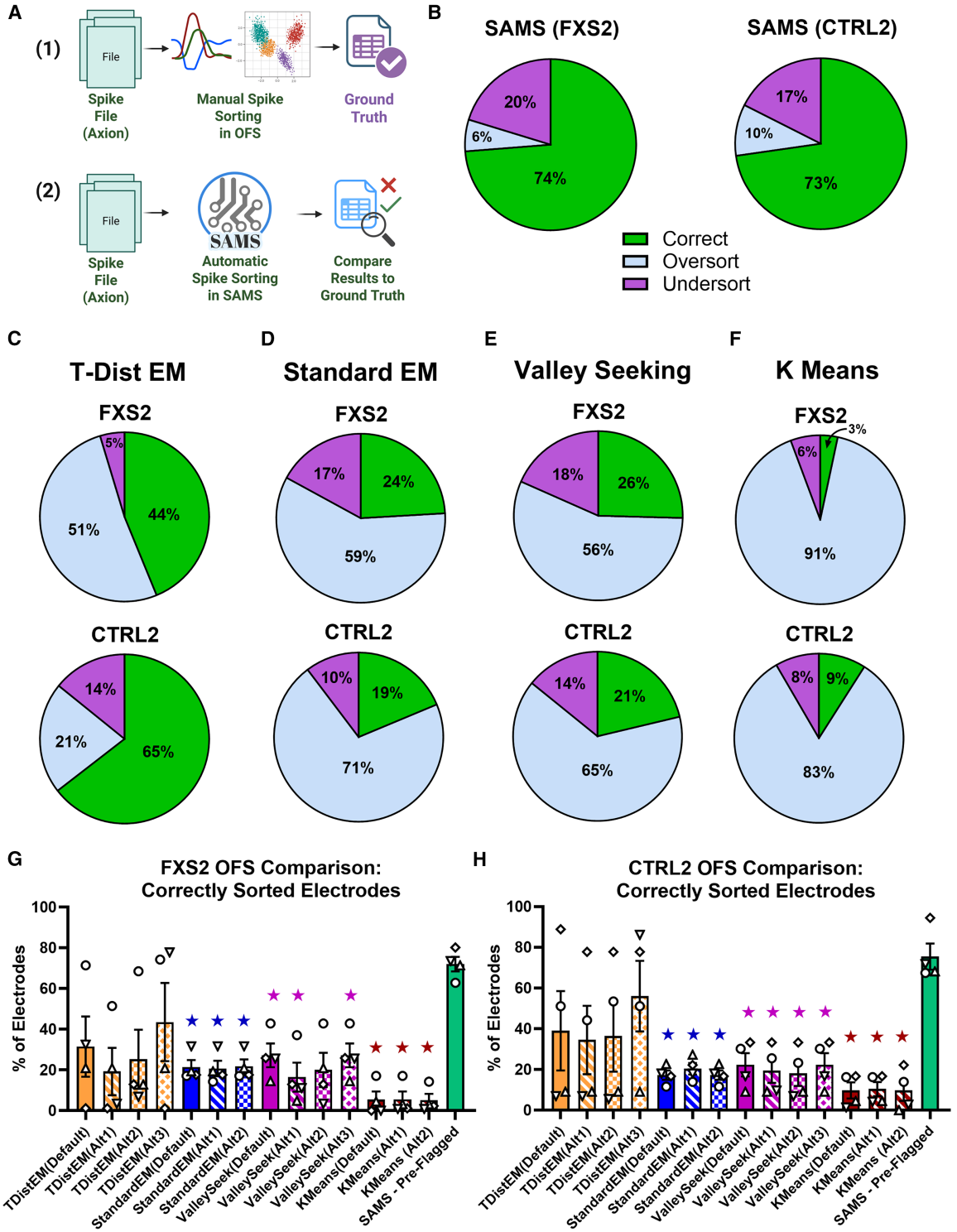


Figure 4. Comparison of SAMS and OFS automatic sorting algorithms

(A) Schematic summarizing (1) ground-truth establishment of MEA data using OFS and (2) comparison of SAMS with ground truth data.

(B) Mean of SAMS without manual correction of any electrodes in FXS2 (left) and CTRL2 (right) neurons.

(C–F) Highest mean accuracy achieved using automated spike sorting algorithms available in OFS software for FXS2 (top) and CTRL2 (bottom) neurons. (C) TDistEM: T-Dist E-M Scan; (D) StandardEM: Standard E-M Scan; (E) ValleySeeking: Valley Seeking Scan; (F) KMeans: K-Means Scan.

(legend continued on next page)

These results collectively indicate that SAMS provides robust and accurate spike sorting across various recording conditions and spike characteristics, with particular strengths in handling lower amplitude signals and moderate spike counts. The algorithm's performance remains consistent across the different maturation stages of neuronal cultures.

Using SAMS to assess electrical activity changes in human neurons with autism-related mutations

To evaluate SAMS as a tool for improving resolution in assessing phenotypes of neurons derived from patients, we next analyzed our published MEA data that had not been subjected to spike sorting. We first examined the activity in neurons that had been differentiated from a pair of FXS (FXS1) and CTRL (CTRL1) hPSCs using *NGN2* induction (Guo et al., 2023). We found that SAMS achieved around 80.5% accuracy for CTRL1 neurons (Figure 6A) and 63.0% accuracy for FXS1 neurons (Figure 6B). We previously reported that FXS1 hPSC-derived neurons exhibit higher mean firing rate (MFR) than CTRL1 neurons, but do not exhibit differences in burst frequency, and that the elevated MFR in FXS1 neurons can be reduced by knocking down *MAP1B* (Figure 6C) (Guo et al., 2023). SAMS analysis confirmed our published results (Figures 6C–6E). Importantly, SAMS resolved an additional difference among experimental conditions: knocking down *MAP1B* in FXS neurons significantly decreased burst frequency at the single-unit level (Figure 6C).

We additionally examined the activity of CTRL hPSC-derived neurons (CTRL3) and hPSC-derived neurons with elevated *MAP1B* (MAP1B-EE) expression differentiated using *NGN2* induction (Guo et al., 2023). We observed significantly increased burst frequency and number of bursts, and no difference in burst duration, which is consistent with our published MEA analysis without spike sorting (Figures S7A–S6C).

We lastly returned to our FXS2 and CTRL2 neurons differentiated using dual SMAD inhibition (Shen et al., 2023). The differences between CTRL2 and FXS2 neurons observed using SAMS, without manual correction, were consistent with the results obtained using standard MEA analysis (Figures 6F–6H; Table S4). SAMS was also able to resolve differences in burst duration and inter-burst interval (Figures S7D–S7I), as well as in other bursting parameters (Table S4), that did not reach statistical significance using standard MEA analysis methods.

We last wanted to confirm whether sorting of spikes into individual units, and the removal of low-activity or outlier waveforms, influenced the electrode level network activity, as SAMS combines the single-unit data for each electrode and then calculates electrode-level network activity parameters from this recombined data. We again observed consistent differences in the network activity between FXS2 and CTRL2 neurons as we did with standard MEA analysis (Figures S7J–S7N; Table S4), including a change in statistical significance for the network burst duration (Figures S7L–S7N). Together, SAMS not only yielded consistent results as the standard MEA analysis but also identified an additional electrophysiological feature that was previously missed.

SAMS allows for separate analysis of bursting and non-bursting neurons from each electrode

Finally, we wanted to utilize SAMS to better understand our neuronal populations. For spiking parameters such as mean firing rate, standard MEA data analysis combines firing rates from neurons that exhibit bursting activity with those neurons that do not. As differences in bursting activity can be due to differences in electrical maturation or other neuronal properties (Hulme et al., 2022; Li et al., 2020; Lin et al., 2021), and bursting differences have been observed in hPSC-derived neuron models of neurodevelopmental and neuropsychiatric disorders (Afshar-Saber et al., 2024b; Amatya et al., 2019; Guo et al., 2023; Que et al., 2021; Shen et al., 2023; Tidball et al., 2020), spike sorting can enable the user to make direct comparisons of neurons that are, for example, more similar in their electrical maturity. Although the increase in MFR in spike-sorted MAP1B-EE neurons did not reach statistical significance, there were significantly higher firing rates in bursting MAP1B-EE neurons compared to bursting CTRL3 neurons (Figure 7A). This finding is particularly interesting given there was no significant difference in the proportion of CTRL and MAP1B-EE neurons that exhibited bursting activity and that 93.2%–99.5% of neurons exhibited bursting activity in each batch of neurons (Figure 7B).

FXS neurons exhibit overall hyperexcitability compared to control neurons, with increased rates of spiking and bursting (Das Sharma et al., 2020; Gildin et al., 2022; Graef et al., 2020; Guo et al., 2023; Liu et al., 2018; Shen et al., 2023). Different genetic or pharmacological interventions to ameliorate these *in vitro* FXS phenotypes decrease overall firing rates but can have varying effects on bursting or network activity (Guo et al., 2023; Shen et al., 2023).

(G and H) Percentage of all electrodes that were accurately sorted by the indicated algorithm in FXS2 (G) and CTRL2 (H) neurons. Data shown are from $n = 4$ independent neuronal differentiations (B1–B4) of $N = 1$ iPSC line per condition. Batches are represented by different symbol shapes. Statistics: (G and H) Data shown as mean (± SEM). Brown-Forsythe & Welch's ANOVA tests with Dunnett's T3 post-hoc correction for multiple comparisons. ★ significant difference compared to SAMS. See also Figures S4D–S4G and Table S3.

Please cite this article in press as: Ren et al., A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons, Stem Cell Reports (2026), https://doi.org/10.1016/j.stemcr.2026.102872

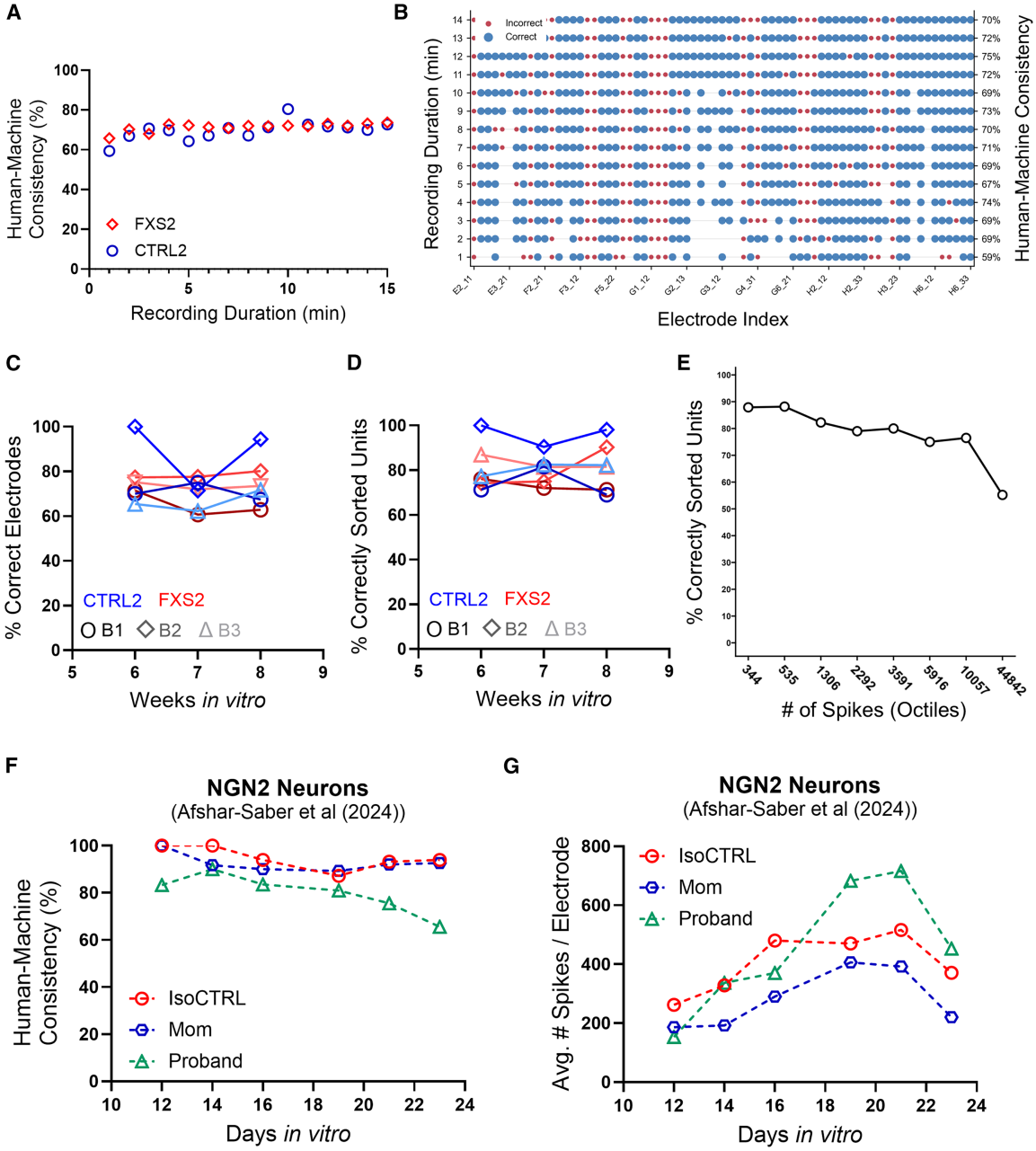


Figure 5. Factors affecting human-machine consistency

(A) Human-machine consistency (HMC; percentage of total electrodes that were accurately sorted) over recording duration length for FXS2 and CTRL2 neurons. Consistency remains above 65% for ≥6-min recording durations. See also Figure S6A.

(B) Electrode-specific HMC (right ordinate) across different recording durations (left ordinate), showing that recording duration does not dramatically impact sorting accuracy on the single-electrode level. Data shown are from CTRL2 neurons at 8 weeks *in vitro*. Cyan: correctly sorted, red: incorrectly sorted.

(C and D) HMC, shown as the percentage of accurately sorted units (C) or electrodes (D), across time of neuronal culturing (6–8 weeks *in vitro*) for FXS2 and CTRL2 neurons. Each batch is represented as a different shape and color. See also Figure S6F.

(E) Percentage of correctly sorted units (both FXS2 and CTRL2) relative to total number of spikes per electrode, split into octiles. See also Figures S6B–S6D. Data in (A and E) are from $n = 4$ independent neuronal differentiations (B1–B4) of $N = 1$ iPSC line per

(legend continued on next page)

Neurons derived from two FXS and two control individuals (FXS1, FXS2, CTRL1, and CTRL2) exhibited increased firing rates as determined using standard MEA analysis (Guo et al., 2023; Shen et al., 2023) or when the single unit activity from both bursting and non-bursting neurons was combined into one group (Figures 6E and 6G). Both sets of FXS neurons that exhibited bursting activity had increased firing rates compared to their respective controls (Figures 7C and 7F). Interestingly, non-bursting FXS1 neurons had similar firing rates as the non-bursting CTRL1 neurons (Figure 7D), while non-bursting FXS2 neurons actually had decreased firing rates compared to CTRL2 (Figure 7G). Additionally, while the proportion of neurons that exhibited bursting activity was higher in both FXS lines, this difference was significant only in FXS2 neurons (Figures 7E and 7H).

Knockdown of the *MAP1B* gene using shRNA partially rescued FXS1 MEA phenotypes both without spike sorting (Guo et al., 2023) and after spike sorting by SAMS (Figure 6D). Interestingly, SAMS revealed that while knockdown of *MAP1B* did not significantly alter the firing rates of bursting FXS1 neurons, there was no longer a statistically significant difference compared to CTRL1 bursting neurons (Figure 7C). Moreover, there was a significant increase in the firing rates of non-bursting FXS1 neurons following *MAP1B* knockdown (Figure 7D) and in the proportion of FXS1 neurons that exhibited any bursting activity (Figure 7E). Together, these results demonstrate the effectiveness of SAMS in enhancing the resolution of MEA data and revealing novel insights when comparing hPSC-derived neurons across conditions.

DISCUSSION

SAMS represents a significant advancement in spike sorting of low-density MEA recordings, particularly hPSC-derived neurons. By integrating multiple algorithms and quality control measures, SAMS addresses several key challenges in the current spike sorting landscape. First, SAMS significantly streamlines the spike sorting workflow. The traditional pipeline, involving multiple software programs and file formats, and extensive manual intervention, is time-consuming and requires specialized coding skills for bulk processing. SAMS consolidates these steps into a single, user-friendly interface, making advanced spike sorting techniques accessible to a broader range of researchers. This democratization of spike sorting technology has the potential to accelerate neuroscience research, particularly *in vitro* disease modeling using hPSC-derived neurons. As the field of neuroscience continues to generate increasingly complex datasets, tools like SAMS will play a crucial role in extracting meaningful insights from neural recordings, ultimately contributing to our understanding of brain function and dysfunction.

The performance of SAMS, which was designed and optimized using hPSC-derived neuron data, was more consistent than the automated sorting algorithms of OFS. This is not surprising given that OFS was designed and optimized for spike sorting of *in vivo* recordings since *in vitro* hPSC-derived neuron recordings differ dramatically from *in vivo* recordings in features such as signal-to-noise ratio. In particular, spike files generated using the recommended default AxIS Navigator settings appear particularly susceptible to oversorting as a result of triggering spike detection with both positive and negative thresholds. Our analysis shows that SAMS achieves relatively high human-machine consistency (mean accuracy: 80.2% (range: 60.7%–100%)) across various recording durations, neuronal culture maturities, batches of neurons, and neurons from different hPSC lines or disease conditions. These levels of accuracy are on par with, or exceed, current-state-of-the-art algorithms that have been developed for spike sorting of MEA data acquired from other platforms or tissue types (Buccino et al., 2020; Mohammadi et al., 2024; Pradeepan et al., 2024; Souza et al., 2019; Wang et al., 2023), and this robustness is crucial for ensuring reliable results across diverse experimental paradigms. The ability of SAMS to maintain performance across different maturation stages of neuronal cultures is particularly valuable for developmental studies, long-term experiments, and repeated recordings of the same neurons over weeks in culture.

Regarding whether this improved performance reflects reduced neuronal overlap rather than algorithmic robustness: low-density arrays do present fewer overlapping waveforms per electrode, which simplifies certain aspects of spike sorting. However, our algorithm was specifically designed as a tool to address the challenges common to these recordings, including similar waveform shapes across units, burst detection, and noise contamination. The consistent performance across multiple culture types, recording conditions, and developmental stages within our dataset suggests that the improvements are not solely attributable to signal sparsity.

condition. Data in (C and D) are mean ± SEM from $n = 3$ independent neuronal differentiations (B1–B4) of $N = 1$ iPSC line per condition.

(F and G) SAMS performance over 11 days *in vitro* (DIV) using an external dataset of MEA recordings from $n = 1$ batch of *NGN2*-induced neurons from a patient with homozygous *ALDH5A1* genetic variant (Proband), as well as from an isogenic CRISPR-engineered heterozygous line (IsoCTRL) and from the patient's mother (Mom).(Afshar-Saber et al., 2024b). See also Figure S6.

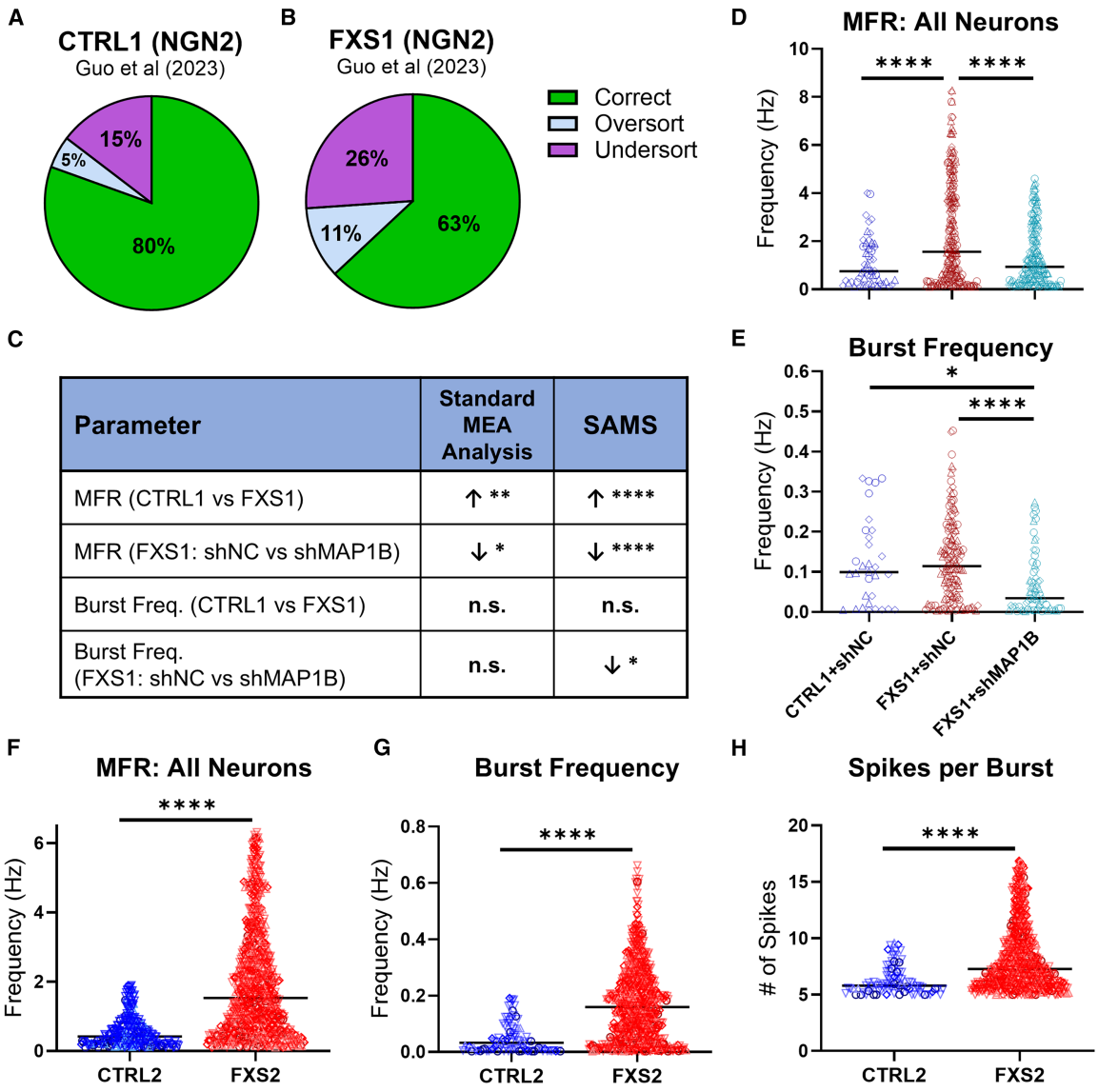


Figure 6. Comparison of SAMS findings to published bulk MEA analysis results

(A and B) Mean accuracy of SAMS in sorting CTRL1 (A) and FXS1 (B) data from (Guo et al., 2023). $n = 3$ independent batches of neurons (B1–B3) from $N = 1$ iPSC line per group.

(C) Summary of results for the two parameters reported in Guo et al. and following analysis of the same data by SAMS.

(D) Mean firing rate (MFR) and (E) burst frequency in CTRL1+shNC, FXS1+shNC, and FXS1+shMAP1B neurons after performing spike sorting with SAMS. $n = 3$ independent batches of neuronal differentiation per condition (D: $n = 56–232$ individual neurons; E: $n = 32–150$ individual neurons).

(F–H) Results from spike sorting of data from (Shen et al., 2023). (F) Mean firing rate, (G) burst frequency, and (H) number of spikes per burst in CTRL2 and FXS2 neurons. $n = 4$ independent batches of neuronal differentiation (B1–B3 and B5). (F) $n = 251$ (CTRL2), 759 (FXS2) neurons; (G) $n = 94$ (CTRL2), 720 (FXS2) neurons; (H) $n = 92$ (CTRL2), 619 (FXS2) neurons. Batches are represented by different symbol shapes. See also Figure S7. Statistics: (D and E) Brown-Forsythe & Welch's ANOVA with Dunnett's T3 correction for multiple comparisons; (F–H) Welch's t test. $^{*}p < 0.05$, $^{****}p < 0.0001$.

The multi-step approach employed by SAMS—combining spectral clustering, HDT, DTW, and waveform validation—provides a comprehensive tool to help deal with common spike sorting issues by compensating for the limitations of individual spike sorting algorithms. This approach effectively addresses both undersorting and oversorting problems, maintaining a balance between sensitivity and specificity in unit identification. The levels of accuracy achieved

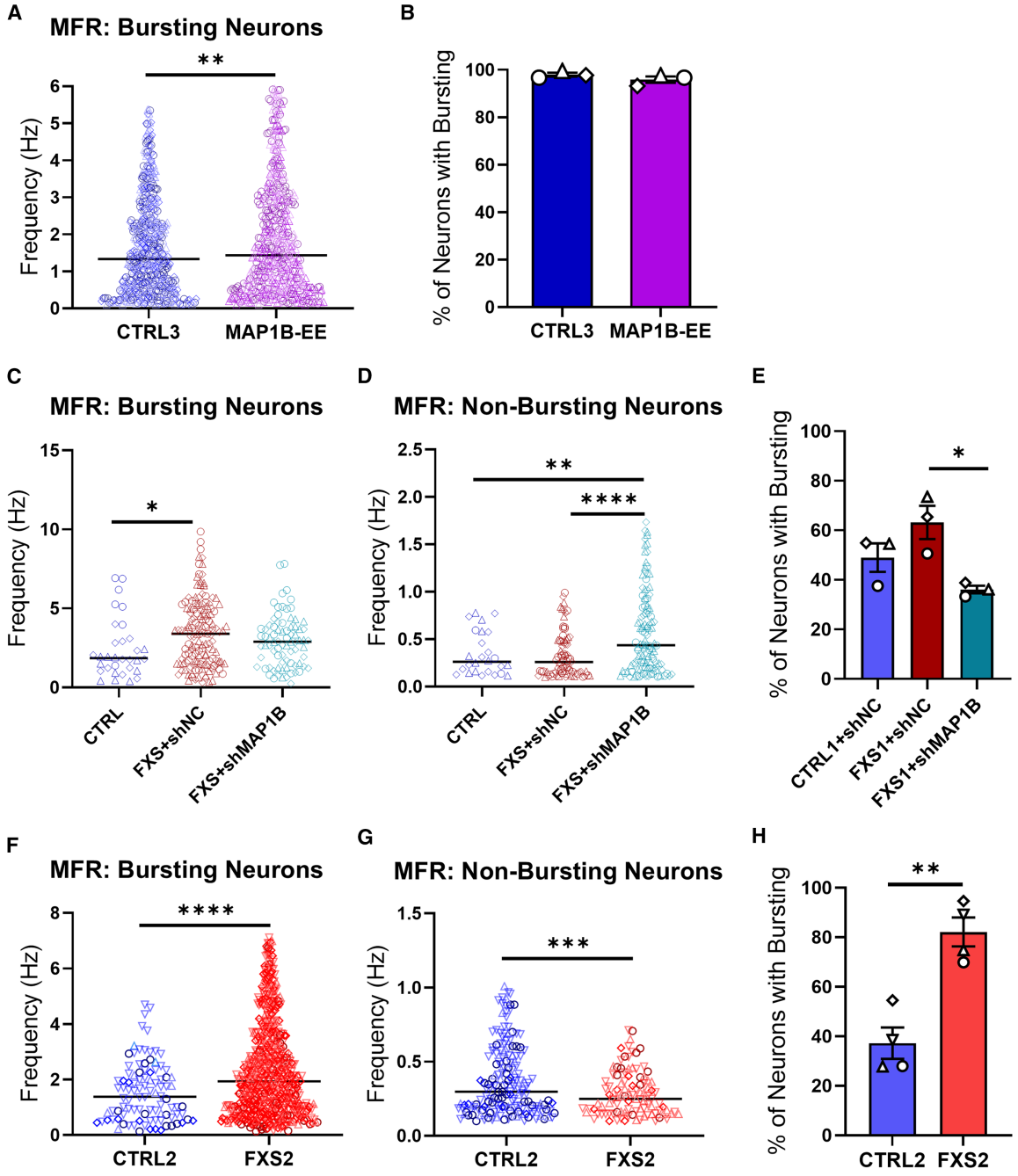


Figure 7. Examination of bursting and non-bursting neuronal populations using SAMS

(A–E) Data from (Guo et al., 2023).

(F–H) Data from (Shen et al., 2023).

(A, C, and F) Mean firing rate in neurons that exhibited bursting activity.

(B, E, and H) Percentage of neurons in each batch that exhibited bursting activity. Data shown as mean (± SEM).

(C and F) Mean firing rate in neurons that did not exhibit bursting activity.

Statistics: (A) Welch's *t* test. *n* = 546 (CTRL3) and 532 (MAP1B-EE) neurons.

(B) Unpaired *t* test. *n* = 3 independent batches of differentiation (B1–B3).

(C) Brown-Forsythe & Welch's ANOVA with Dunnett's T3 correction for multiple comparisons. *n* = 32 (CTRL1), 145 (FXS1+shNC), and 79 (FXS1+shMAP1B) neurons.

(legend continued on next page)

by SAMS, particularly after validation of the Flagged Electrodes lists, is particularly significant in the context of high-content screening using MEA plates in 48- or 96-well plate formats, as these features dramatically reduce the amount of time the user will need to spend checking electrodes before they obtain their final results. The application of SAMS to existing MEA datasets demonstrates its practical utility in various research scenarios. This versatility makes SAMS a valuable tool for a wide range of neuroscience applications, from basic research to disease modeling and drug discovery.

While SAMS was developed for the Axion platform, its core algorithms (HDT, DTW, and spectral clustering) are platform-agnostic. Adaptation to other low-density MEA systems such as Multi Channel Systems would primarily require data interface modifications rather than algorithmic changes and could be implemented in a future version.

Limitations of this study

It is important to note that this first iteration of SAMS, like most other available spike sorting algorithms and platforms, has limitations. SAMS was developed to aid spike sorting, but it is not a solution to all issues that users may encounter. It is essential that users verify and correct any incorrect sorts before drawing conclusions about their data. One limitation is that the algorithm's performance at the default settings appears to decrease with very high spike counts per electrode, suggesting potential room for improvement in handling highly active neurons or high-noise scenarios. Oversorting can occur when spikes are triggered by double thresholding or time shifts, resulting in separate clusters that represent the same neuron. If the waveforms are not similar enough to be detected by DTW as originating from the same unit, they remain split. Undersorting happens when waveforms from different neurons differ only slightly, causing substantial overlap in PCA space, or due to waveform similarities between units or noise contamination that shifts a spike's position in feature space toward a neighboring cluster. It appears as a small cluster hidden within a larger one in the PCA projection, making it difficult to distinguish separate units. As should be done with all new methods, we recommend that users first perform a pilot analysis comparing SAMS to ground truth data so they can determine the settings that work best for their data. We also recommend validation of key findings using complementary methods of measuring single neuron electrical activity, such as patch clamp recording or calcium imaging. All sorts should also be user-verified before drawing conclusions. Future iterations of SAMS may also incorporate machine learning techniques to further refine its performance in these edge cases. Additionally, this current iteration of SAMS cannot identify or define network bursts at the single unit level; therefore, network activity data currently still need to be examined on an electrode-to-electrode basis. This version of SAMS only reports electrode-level network spiking and bursting metrics, as well as the synchrony index; therefore, other measures of synchrony such as area under normalized cross-correlation still need to be acquired through standard bulk MEA analysis. SAMS was developed using MATLAB, a computationally heavy language that can affect performance speed depending on existing hardware. Implementation of SAMS in the future using lower-level languages such as C could increase processing speed. Finally, the data files used for the development of SAMS almost exclusively consisted of cortical excitatory neurons that were cultured for only up to 8 weeks *in vitro*, so analysis of recordings containing more electrically mature neurons, or those containing purely inhibitory or non-cortical neuronal subtypes may produce differences in the overall accuracy. The files used for the development and testing of SAMS shown here only contained data from neurons cultured on 96-well and 48-well MEA plates, which contain 8 and 16 electrodes, respectively, spaced at 350 μm. Since the electrode spacing is only slightly less in the higher-density plates offered by Axion (300 μm), we do not anticipate performance differences if SAMS were to be used on data from these types of plates. Therefore, as SAMS is applied to a broader variety of datasets, additional updates and refinements to the SAMS algorithm can be made to retain its high levels of accuracy. We encourage users to contact us and provide example datasets if they encounter recurring issues with over- or undersorting so that we can continue to further improve the SAMS algorithm and increase its functionality to meet the broader needs of the hPSC-derived neuron community.

(D) Brown-Forsythe & Welch's ANOVA with Dunnett's T3 correction for multiple comparisons. n = 26 (CTRL1), 80 (FXS1+shNC), and 116 (FXS1+shMAP1B) neurons.
(E) Unpaired t test. n = 3 independent batches of differentiation (B1–B3).
(F) Welch's t test. n = 98 (CTRL2) and 664 (FXS2) neurons.
(G) Welch's t test. n = 172 (CTRL2) and 103 (FXS2) neurons.
(H) Unpaired t test. n = 4 independent batches of differentiation (B1–B3 and B5). Batches are represented by different symbol shapes. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.005$, $^{****}p < 0.001$.
See also Figure S7.

Please cite this article in press as: Ren et al., A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons, Stem Cell Reports (2026), https://doi.org/10.1016/j.stemcr.2026.102872

METHODS

Generation of aligned spike files for spike sorting

Raw MEA data files were loaded into AxIS navigator software (Axion Biosystems). "Spontaneous" configuration with the Spike Detector set to "Peak Adaptive Threshold" was selected, while all other settings were set at the default values. The "Peak Adaptive Threshold" setting generates spike files that contain annotated spikes and burst information and aligned waveforms to facilitate spike sorting. The specific values for the Spike Detector settings are in Table S6. Each spike file was converted to the NeuroExplorer (.nex) format using Axion's Data Export Tool, so it was compatible with Plexon's OFS software.

Manual spike sorting of MEA data files to establish ground truth

Prior to assessing the accuracy of SAMS and OFS spike sorting algorithms, the data files to be used for this assessment were first manually analyzed in OFS to determine the number of units on each electrode in each data file. Each .nex file was loaded into OFS to enable viewing of unsorted aligned waveform shapes and waveform PCA data and manual sorting of waveforms into putative units. Detailed manual sorting process is provided in the supplemental methods. Briefly, PCA clustering and the peaks of aligned waveforms were used to define putative units, which were then subdivided into smaller subunits. These subunits were manually inspected to compare their similarity, and similar subunits were remerged into one unit. This process was performed in multiple iterations until a final set of units remained that were distinct from one another based on peak location, trough location, peak-to-trough amplitude, and waveform shape dissimilarity. Units with a firing rate below 0.1 Hz were excluded from the final unit count. All recorded electrodes in each batch of cells were included in the analysis.

SAMS

SAMS was optimized to handle the scale and characteristics of *in vitro* MEA data, where hundreds of electrodes must be analyzed. The key algorithmic components include PCA to project spike waveforms into lower dimension, then spectral clustering to assign initial units by evaluating K = 1–5 clusters. Best fit is chosen based on the Davies-Bouldin Index. To fix undersorting, Hartigan's dip test (HDT) is applied to flag clusters with multimodality in principal component distribution, which are then split accordingly. To fix oversorting, dynamic time warping (DTW) is applied to identify and merge units possessing similar waveform characteristics but temporally shifted. We also have quality control steps: waveforms that deviate more than three SDs from a unit's mean template are removed as outliers, spikes violating the refractory period (default: 1.5 ms) are discarded, and low firing rate units (default: 0.1 Hz) are excluded. SAMS also flags potentially missorted electrodes as possible multi-unit (PMU; failed HDT or low DTW distance) or over-excluded unit (OEU; excessive waveform exclusion) for manual review. Details are provided in the supplemental methods and Table S2 and can also be found on the SAMS GitHub.

SAMS and OFS accuracy assessment

To assess the accuracy of SAMS and OFS algorithms, the number of units identified on each electrode was compared to the number of units identified by manual spike sorting. SAMS was run using the default settings (Table S2). Four different automated spike sorting algorithms available in OFS software were first tested using the default parameters, then each parameter was adjusted individually from the default to determine whether it influenced the algorithm's accuracy and if so, whether the accuracy could be improved. The settings for each parameter tested are provided in Table S3. An electrode was classified as incorrectly "oversorted" when the number of units identified by an algorithm was greater than the number of units identified by manual spike sorting and considered incorrectly "undersorted" when the number of units identified by an algorithm was less than the number of units identified by manual spike sorting.

To more thoroughly assess the accuracy of SAMS, the PCA plots and waveform plots for each electrode were visually inspected to confirm whether waveforms appeared to be assigned to the correct units. Although uncommon, an electrode was classified as incorrectly "missorted" when the number of units identified by SAMS matched the number of units identified by manual spike sorting, but the assignment of waveforms was found to be incorrect upon user inspection. For simplicity of reporting, these units were counted as "undersorted" when calculating the types of errors committed by SAMS or OFS.

To avoid reporting an inflated estimate of SAMS' capabilities, we have reported accuracy as the percentage of electrodes that were completely sorted correctly, unless stated otherwise in the figure legend: an electrode was only considered "correct" if the number of units identified was correct and if the majority of the waveforms (greater than 90%) were assigned to the correct units. Otherwise, the entire electrode was categorized as "incorrect." Further, we also did not exclude any difficult-to-sort electrodes from any of the datasets.

MEA datasets

All MEA datasets used in this study were recorded from hPSC-derived excitatory neurons and were previously

published (Afshar-Saber et al., 2024b; Guo et al., 2023; Shen et al., 2023). For detailed batch information, including the number of active wells and electrodes, see Table S5. For detailed information on iPSC lines and neuronal differentiation protocols, see supplemental methods.

Data analysis and statistics

For detailed statistical information, see supplemental methods. Electrode diagram in Figure 5B was generated using MATLAB. Graphs plotting waveform number against SAMS accuracy were generated using R. Kendall's tau and bootstrapping analyses (Figures 5E, S6B, and S6D) were performed in R (https://github.com/Zhao-Lab-UW/SAMS-R-analyses/tree/main). Graphing and statistical analysis for all other figures was performed using Prism software (GraphPad).

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Xinyu Zhao (xinyu.zhao@wisc.edu).

Materials availability

SAMS (RRID: SCR_027843) can be accessed at https://github.com/Zhao-Lab-UW/SAMS-Semi-Automatic-MEA-Spike-sorting-pipeline-. There was no unique reagent generated in this study.

Data and code availability

The code used for generating the figures and data files used in this study can be accessed at https://github.com/Zhao-Lab-UW/SAMS-Semi-Automatic-MEA-Spike-sorting-pipeline-. SAMS incorporates the following third-party components: AxionFileLoader (Axion BioSystems), Hartigan's dip test (translated by F. Mechler from Hartigan and Hartigan, 1985; Hartigan and Hartigan, 1985), and InterX (NS, 2025).

ACKNOWLEDGMENTS

We thank Drs. Wardiya Afshar-Saber and Mustafa Sahin for kindly sharing their MEA data and members of the Zhao lab for their helpful discussion. This work was supported by the DOD IIRA grant W81XWH-22-1-0621 (to X.Z.); National Institutes of Health (R01MH118827, R01MH116582, R01MH136152, and R01NS138268 to X.Z.; R01EY029438 and R01EY035005 to A.R.; P50HD105353 to Waisman Center; DP2NS122605 to A.H.; and P51OD011106 to A.R.); Retina Research Foundation Edwin and Dorothy Gamewell Professorship, McPherson Eye Research Institute (to A.R.), Kellett Mid-Career Award, Wisconsin Alumni Research Foundation, Jenni and Kyle Professorship, Vilas Distinguished Achievement Professorship, Eagle Autism Foundation, and Simons Foundation Autism Research Initiative (to X.Z.); postdoctoral fellowships from the Wisconsin Stem Cell and Regenerative Medicine Center (SCRMC), Fragile X Research Foundation (FRAXA), and the Autism Science Foundation (to C.L.S.); SciMed scholarships, T32 GM141013, and predoctoral fellowship from the SCRMC (to N.M.M.-A.).

AUTHOR CONTRIBUTIONS

C.L.S. and X.Z. conceived the concept and designed the study. X.R. developed the method with help from A.R., R.C.D., and A.H. C.L.S., X.R., and E.E.D. performed accuracy checking and created figures and tables. N.M.M.-A. created the logo and graphical abstract. X.R., C.L.S., and X.Z. wrote the manuscript. All authors edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2026.102872.

Received: January 20, 2025
Revised: March 1, 2026
Accepted: March 2, 2026

REFERENCES

Afshar-Saber, W., Chen, C., Teaney, N.A., Kim, K., Yang, Z., Gasparoli, F.M., Ebrahimi-Fakhari, D., Buttermore, E.D., Pin-Fang Chen, I., Pearl, P.L., and Sahin, M. (2024). Generation and characterization of six human induced pluripotent stem cell lines (hiPSCs) from three individuals with SSADH Deficiency and CRISPR-corrected isogenic controls. Stem Cell Res. *77*, 103424. https://doi.org/10.1016/j.scr.2024.103424.

Afshar-Saber, W., Teaney, N.A., Winden, K.D., Jumo, H., Shi, X., McGinty, G., Hubbs, J., Chen, C., Tokatly Latzer, I., Gasparoli, F., et al. (2024). ALDH5A1-deficient iPSC-derived excitatory and inhibitory neurons display cell type specific alterations. Neurobiol. Dis. *190*, 106386. https://doi.org/10.1016/j.nbd.2023.106386.

Amatya, D.N., Linker, S.B., Mendes, A.P.D., Santos, R., Erikson, G., Shokhirev, M.N., Zhou, Y., Sharpee, T., Gage, F.H., Marchetto, M. C., and Kim, Y. (2019). Dynamical Electrical Complexity Is Reduced during Neuronal Differentiation in Autism Spectrum Disorder. Stem Cell Rep. *13*, 474–484. https://doi.org/10.1016/j.stemcr.2019.08.001.

Anderson, N.C., Chen, P.F., Meganathan, K., Afshar Saber, W., Petersen, A.J., Bhattacharyya, A., Kroll, K.L., and Sahin, M.; Cross-IDDRC Human Stem Cell Working Group (2021). Balancing serendipity and reproducibility: Pluripotent stem cells as experimental systems for intellectual and developmental disorders. Stem Cell Rep. *16*, 1446–1457. https://doi.org/10.1016/j.stemcr.2021.03.025.

Buccino, A.P., Hurwitz, C.L., Garcia, S., Magland, J., Siegle, J.H., Hurwitz, R., and Hennig, M.H. (2020). SpikeInterface, a unified framework for spike sorting. eLife *9*, e61834. https://doi.org/10.7554/eLife.61834.

Cao, Y., Rakhilin, N., Gordon, P.H., Shen, X., and Kan, E.C. (2016). A real-time spike classification method based on dynamic time warping for extracellular enteric neural recording with large waveform variability. J. Neurosci. Methods *261*, 97–109. https://doi.org/10.1016/j.jneumeth.2015.12.006.

Carlson, D., and Carin, L. (2019). Continuing progress of spike sorting in the era of big data. Curr. Opin. Neurobiol. *55*, 90–96. https://doi.org/10.1016/j.conb.2019.02.007.

Choi, J.S., Lee, H.J., Rajaraman, S., and Kim, D.H. (2021). Recent advances in three-dimensional microelectrode array technologies for in vitro and in vivo cardiac and neuronal interfaces. Biosens. Bioelectron. *171*, 112687. https://doi.org/10.1016/j.bios.2020.112687.

Cotterill, E., Charlesworth, P., Thomas, C.W., Paulsen, O., and Eglen, S.J. (2016). A comparison of computational methods for detecting bursts in neuronal spike trains and their application to human stem cell-derived neuronal networks. J. Neurophysiol. *116*, 306–321. https://doi.org/10.1152/jn.00093.2016.

Das Sharma, S., Pal, R., Reddy, B.K., Selvaraj, B.T., Raj, N., Samaga, K.K., Srinivasan, D.J., Ornelas, L., Sareen, D., Livesey, M.R., et al. (2020). Cortical neurons derived from human pluripotent stem cells lacking FMRP display altered spontaneous firing patterns. Mol. Autism *11*, 52. https://doi.org/10.1186/s13229-020-00351-4.

Gildin, L., Rauti, R., Vardi, O., Kuznitsov-Yanovsky, L., Maoz, B.M., Segal, M., and Ben-Yosef, D. (2022). Impaired Functional Connectivity Underlies Fragile X Syndrome. Int. J. Mol. Sci. *23*, 2048. https://doi.org/10.3390/ijms23042048.

Graef, J.D., Wu, H., Ng, C., Sun, C., Villegas, V., Qadir, D., Jesseman, K., Warren, S.T., Jaenisch, R., Cacace, A., and Wallace, O. (2020). Partial FMRP expression is sufficient to normalize neuronal hyperactivity in Fragile X neurons. Eur. J. Neurosci. *51*, 2143–2157. https://doi.org/10.1111/ejn.14660.

Guan, S., Tian, H., Yang, Y., Liu, M., Ding, J., Wang, J., and Fang, Y. (2023). Self-assembled ultraflexible probes for long-term neural recordings and neuromodulation. Nat. Protoc. *18*, 1712–1744. https://doi.org/10.1038/s41596-023-00824-9.

Guo, Y., Shen, M., Dong, Q., Méndez-Albelo, N.M., Huang, S.X., Sirois, C.L., Le, J., Li, M., Jarzembowski, E.D., Schoeller, K.A., et al. (2023). Elevated levels of FMRP-target MAP1B impair human and mouse neuronal development and mouse social behaviors via autophagy pathway. Nat. Commun. *14*, 3801. https://doi.org/10.1038/s41467-023-39337-0.

Hartigan, J.A., and Hartigan, P.M. (1985). The Dip Test of Unimodality. Ann. Statist. *13*, 70–84.

Hruska-Plochan, M., Wiersma, V.I., Betz, K.M., Mallona, I., Ronchi, S., Maniecka, Z., Hock, E.M., Tantardini, E., Laferriere, F., Sahadevan, S., et al. (2024). A model of human neural networks reveals NPTX2 pathology in ALS and FTLD. Nature *626*, 1073–1083. https://doi.org/10.1038/s41586-024-07042-7.

Hulme, A.J., Maksour, S., St-Clair Glover, M., Miellet, S., and Dottori, M. (2022). Making neurons, made easy: The use of Neurogenin-2 in neuronal differentiation. Stem Cell Rep. *17*, 14–34. https://doi.org/10.1016/j.stemcr.2021.11.015.

Hyvärinen, T., Hyysalo, A., Kapucu, F.E., Aarnos, L., Vinogradov, A., Eglen, S.J., Ylä-Outinen, L., and Narkilahti, S. (2019). Functional characterization of human pluripotent stem cell-derived cortical networks differentiated on laminin-521 substrate: comparison to rat cortical cultures. Sci. Rep. *9*, 17125. https://doi.org/10.1038/s41598-019-53647-8.

Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*, 232–236. https://doi.org/10.1038/nature24636.

Kiskinis, E., Sandoe, J., Williams, L.A., Boulting, G.L., Moccia, R., Wainger, B.J., Han, S., Peng, T., Thams, S., Mikkilineni, S., et al. (2014). Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. Cell Stem Cell *14*, 781–795. https://doi.org/10.1016/j.stem.2014.03.004.

Lemieux, M.R., Freigassner, B., Hanson, J.L., Thathey, Z., Opp, M. R., Hoeffer, C.A., and Link, C.D. (2024). Multielectrode array characterization of human induced pluripotent stem cell derived neurons in co-culture with primary human astrocytes. PLoS One *19*, e0303901. https://doi.org/10.1371/journal.pone.0303901.

Li, M., Shin, J., Risgaard, R.D., Parries, M.J., Wang, J., Chasman, D., Liu, S., Roy, S., Bhattacharyya, A., and Zhao, X. (2020). Identification of FMR1-regulated molecular networks in human neurodevelopment. Genome Res. *30*, 361–374. https://doi.org/10.1101/gr.251405.119.

Lin, H.C., He, Z., Ebert, S., Schörnig, M., Santel, M., Nikolova, M.T., Weigert, A., Hevers, W., Kasri, N.N., Taverna, E., et al. (2021). NGN2 induces diverse neuron types from human pluripotency. Stem Cell Rep. *16*, 2118–2127. https://doi.org/10.1016/j.stemcr.2021.07.006.

Liu, H., Lu, J., Chen, H., Du, Z., Li, X.J., and Zhang, S.C. (2015). Spinal muscular atrophy patient-derived motor neurons exhibit hyperexcitability. Sci. Rep. *5*, 12189. https://doi.org/10.1038/srep12189.

Liu, X.S., Wu, H., Krzisch, M., Wu, X., Graef, J., Muffat, J., Hnisz, D., Li, C.H., Yuan, B., Xu, C., et al. (2018). Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. Cell *172*, 979–992.e6. https://doi.org/10.1016/j.cell.2018.01.012.

Liu, Y., Yao, X., Fan, C., Zhang, G., Luo, X., and Qian, Y. (2023). Microfabrication and lab-on-a-chip devices promotein vitromodeling of neural interfaces for neuroscience researches and preclinical applications. Biofabrication *16*. https://doi.org/10.1088/1758-5090/ad032a.

Lv, S., He, E., Luo, J., Liu, Y., Liang, W., Xu, S., Zhang, K., Yang, Y., Wang, M., Song, Y., et al. (2023). Using Human-Induced Pluripotent Stem Cell Derived Neurons on Microelectrode Arrays to Model Neurological Disease: A Review. Adv. Sci. *10*, e2301828. https://doi.org/10.1002/advs.202301828.

Mayer, M., Arrizabalaga, O., Lieb, F., Ciba, M., Ritter, S., and Thielemann, C. (2018). Electrophysiological investigation of human embryonic stem cell derived neurospheres using a novel spike detection algorithm. Biosens. Bioelectron. *100*, 462–468. https://doi.org/10.1016/j.bios.2017.09.034.

McCready, F.P., Gordillo-Sampedro, S., Pradeepan, K., Martinez-Trujillo, J., and Ellis, J. (2022). Multielectrode Arrays for Functional Phenotyping of Neurons from Induced Pluripotent Stem Cell Models of Neurodevelopmental Disorders. Biology *11*, 316. https://doi.org/10.3390/biology11020316.

Mehta, S.R., Tom, C.M., Wang, Y., Bresee, C., Rushton, D., Mathkar, P.P., Tang, J., and Mattis, V.B. (2018). Human Huntington's Disease iPSC-Derived Cortical Neurons Display Altered Transcriptomics, Morphology, and Maturation. Cell Rep. *25*, 1081–1096.e6. https://doi.org/10.1016/j.celrep.2018.09.076.

Mohammadi, Z., Denman, D.J., Klug, A., and Lei, T.C. (2024). A fully automatic multichannel neural spike sorting algorithm with spike reduction and positional feature. J. Neural. Eng. *21*, 046039. https://doi.org/10.1088/1741-2552/ad647d.

Mossink, B., Verboven, A.H.A., van Hugte, E.J.H., Klein Gunnewiek, T.M., Parodi, G., Linda, K., Schoenmaker, C., Kleefstra, T., Kozicz, T., van Bokhoven, H., et al. (2021). Human neuronal networks on micro-electrode arrays are a highly robust tool to study disease-specific genotype-phenotype correlations in vitro. Stem Cell Rep. *16*, 2182–2196. https://doi.org/10.1016/j.stemcr.2021.07.001.

Napoli, A., and Obeid, I. (2016). Comparative Analysis of Human and Rodent Brain Primary Neuronal Culture Spontaneous Activity Using Micro-Electrode Array Technology. J. Cell. Biochem. *117*, 559–565. https://doi.org/10.1002/jcb.25312.

Negri, J., Menon, V., and Young-Pearse, T.L. (2020). Assessment of Spontaneous Neuronal Activity In Vitro Using Multi-Well Multi-Electrode Arrays: Implications for Assay Development. eNeuro *7*. https://doi.org/10.1523/eneuro.0080-19.2019.

Niu, W., and Parent, J.M. (2020). Modeling genetic epilepsies in a dish. Dev. Dyn. *249*, 56–75. https://doi.org/10.1002/dvdy.79.

NS (2025). Curve Intersections (MATLAB Central File Exchange).

Obaid, A., Hanna, M.E., Wu, Y.W., Kollo, M., Racz, R., Angle, M.R., Müller, J., Brackbill, N., Wray, W., Franke, F., et al. (2020). Massively parallel microwire arrays integrated with CMOS chips for neural recording. Sci. Adv. *6*, eaay2789. https://doi.org/10.1126/sciadv.aay2789.

Pradeepan, K.S., McCready, F.P., Wei, W., Khaki, M., Zhang, W., Salter, M.W., Ellis, J., and Martinez-Trujillo, J. (2024). Calcium-Dependent Hyperexcitability in Human Stem Cell-Derived Rett Syndrome Neuronal Networks. Biol. Psychiatry Glob. Open Sci. *4*, 100290. https://doi.org/10.1016/j.bpsgos.2024.100290.

Que, Z., Olivero-Acosta, M.I., Zhang, J., Eaton, M., Tukker, A.M., Chen, X., Wu, J., Xie, J., Xiao, T., Wettschurack, K., et al. (2021). Hyperexcitability and Pharmacological Responsiveness of Cortical Neurons Derived from Human iPSCs Carrying Epilepsy-Associated Sodium Channel Nav1.2-L1342P Genetic Variant. J. Neurosci. *41*, 10194–10208. https://doi.org/10.1523/jneurosci.0564-21.2021.

Rosa, F., Dhingra, A., Uysal, B., Mendis, G.D.C., Loeffler, H., Elsen, G., Mueller, S., Schwarz, N., Castillo-Lizardo, M., Cuddy, C., et al. (2020). In Vitro Differentiated Human Stem Cell-Derived Neurons Reproduce Synaptic Synchronicity Arising during Neurodevelopment. Stem Cell Rep. *15*, 22–37. https://doi.org/10.1016/j.stemcr.2020.05.015.

Sandoval, S.O., Méndez-Albelo, N.M., Xu, Z., and Zhao, X. (2024). From wings to whiskers to stem cells: why every model matters in fragile X syndrome research. J. Neurodev. Disord. *16*, 30. https://doi.org/10.1186/s11689-024-09545-w.

Shen, M., Sirois, C.L., Guo, Y., Li, M., Dong, Q., Méndez-Albelo, N. M., Gao, Y., Khullar, S., Kissel, L., Sandoval, S.O., et al. (2023). Species-specific FMRP regulation of RACK1 is critical for prenatal cortical development. Neuron *111*, 3988–4005.e11. https://doi.org/10.1016/j.neuron.2023.09.014.

Souza, B.C., Lopes-Dos-Santos, V., Bacelo, J., and Tort, A.B.L. (2019). Spike sorting with Gaussian mixture models. Sci. Rep. *9*, 3627. https://doi.org/10.1038/s41598-019-39986-6.

Spira, M.E., and Hai, A. (2013). Multi-electrode array technologies for neuroscience and cardiology. Nat. Nanotechnol. *8*, 83–94. https://doi.org/10.1038/nnano.2012.265.

Tidball, A.M., Lopez-Santiago, L.F., Yuan, Y., Glenn, T.W., Margolis, J.L., Clayton Walker, J., Kilbane, E.G., Miller, C.A., Martina Bebin, E., Scott Perry, M., et al. (2020). Variant-specific changes in persistent or resurgent sodium current in SCN8A-related epilepsy patient-derived neurons. Brain *143*, 3025–3040. https://doi.org/10.1093/brain/awaa247.

Tyssowski, K.M., Letai, K.C., Rendall, S.D., Tan, C., Nizhnik, A., Kaeser, P.S., and Gray, J.M. (2019). Firing Rate Homeostasis Can Occur in the Absence of Neuronal Activity-Regulated Transcription. J. Neurosci. *39*, 9885–9899. https://doi.org/10.1523/jneurosci.1108-19.2019.

Vatine, G.D., Al-Ahmad, A., Barriga, B.K., Svendsen, S., Salim, A., Garcia, L., Garcia, V.J., Ho, R., Yucer, N., Qian, T., et al. (2017). Modeling Psychomotor Retardation using iPSCs from MCT8-Deficient Patients Indicates a Prominent Role for the Blood-Brain Barrier. Cell Stem Cell *20*, 831–843.e5. https://doi.org/10.1016/j.stem.2017.04.002.

Vertkin, I., Styr, B., Slomowitz, E., Ofir, N., Shapira, I., Berner, D., Fedorova, T., Laviv, T., Barak-Broner, N., Greitzer-Antes, D., et al. (2015). GABAB receptor deficiency causes failure of neuronal homeostasis in hippocampal networks. Proc. Natl. Acad. Sci. USA *112*, E3291–E3299. https://doi.org/10.1073/pnas.1424810112.

von Luxburg, U. (2007). A tutorial on spectral clustering. Stat. Comput. *17*, 395–416. https://doi.org/10.1007/s11222-007-9033-z.

Wang, R., Xu, Y., Zhang, Y., Hu, X., Li, Y., and Zhang, S. (2023). A Fast and Effective Spike Sorting Method Based on Multi-Frequency Composite Waveform Shapes. Brain Sci. *13*, 1156. https://doi.org/10.3390/brainsci13081156.

Zhang, H., Rong, G., Bian, S., and Sawan, M. (2022). Lab-on-Chip Microsystems for Ex Vivo Network of Neurons Studies: A Review. Front. Bioeng. Biotechnol. *10*, 841389. https://doi.org/10.3389/fbioe.2022.841389.

Zhao, X., and Bhattacharyya, A. (2018). Human Models Are Needed for Studying Human Neurodevelopmental Disorders. Am. J. Hum. Genet. *103*, 829–857. https://doi.org/10.1016/j.ajhg.2018.10.009.

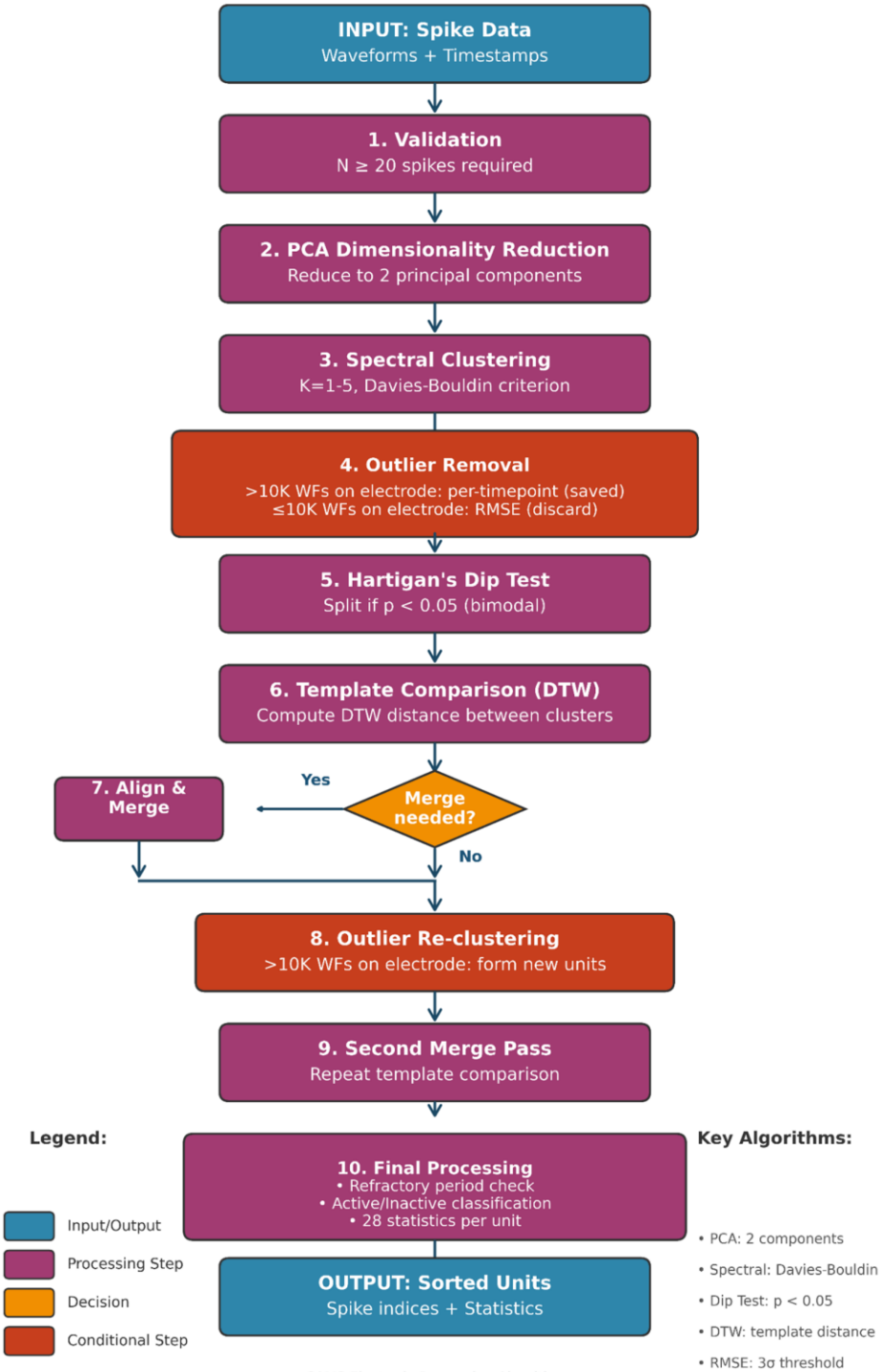
Stem Cell Reports, Volume *21*

Supplemental Information

A semi-automated MEA spike sorting method for high-throughput assessment of cultured neurons

Xiaoxuan Ren, Carissa L. Sirois, Raymond Doudlah, Ethan E. Dayley, Natasha M. Méndez-Albelo, Aviad Hai, Ari Rosenberg, and Xinyu Zhao

SAMS Electrode Processing Algorithm



SAMS Electrode Processing Algorithm

Figure S1. SAMS Workflow

Flowchart outlining the 10 main steps of the SAMS algorithm: (1) Input validation, (2) Principle Component Analysis (PCA), (3) Spectral Clustering, (4) Outlier Pre-Processing based on total number of spikes, (5) Hartigan's Dip Test (HDT) to detect bimodality, (6) Template Comparison using Dynamic Time Warping (DTW) Distance, (7) Alignment and Merging, if needed, based the on results of Step 6, (8) Outlier Re-clustering and (9) Second Merge Pass (repeating of Steps 6 & 7) for electrodes with high total spike numbers, and (10) Final Processing to check for refractory period violations based on inter-spike interval (ISI), to remove low activity units, and to perform burst detection. See **Supplemental Methods** for additional information.

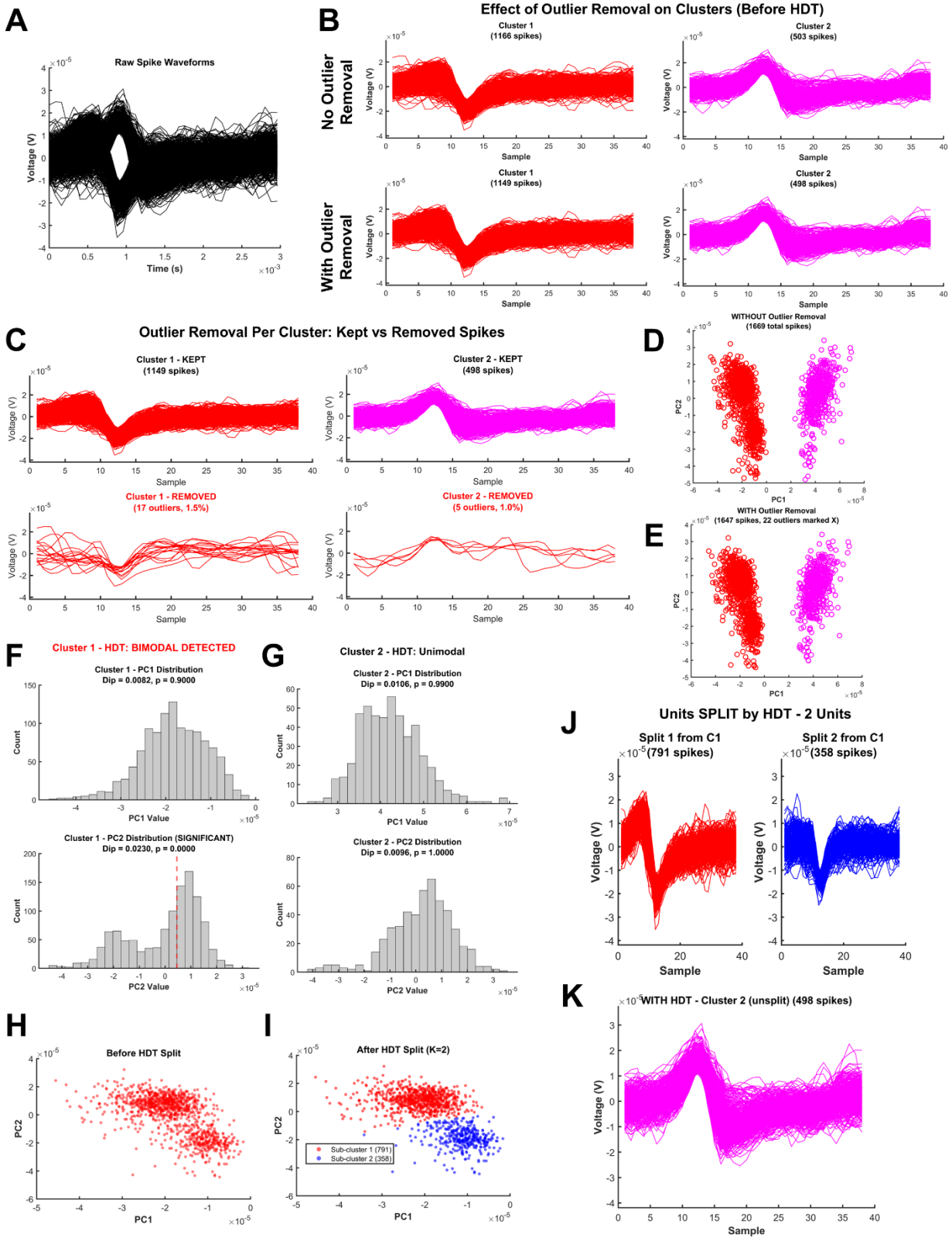

A
Raw Spike Waveforms
Voltage (V)
Time (s)
B
Effect of Outlier Removal on Clusters (Before HDT)
No Outlier Removal
With Outlier Removal
Cluster 1 (1166 spikes)
Cluster 2 (503 spikes)
Cluster 1 (1149 spikes)
Cluster 2 (498 spikes)
Sample
C
Outlier Removal Per Cluster: Kept vs Removed Spikes
Cluster 1 - KEPT (1149 spikes)
Cluster 2 - KEPT (498 spikes)
Cluster 1 - REMOVED (17 outliers, 1.5%)
Cluster 2 - REMOVED (5 outliers, 1.0%)
D
WITHOUT Outlier Removal (1669 total spikes)
E
WITH Outlier Removal (1647 spikes, 22 outliers marked X)
PC1
PC2
F
Cluster 1 - HDT: BIMODAL DETECTED
Cluster 1 - PC1 Distribution
Dip = 0.0082, p = 0.9000
Cluster 1 - PC2 Distribution (SIGNIFICANT)
Dip = 0.0230, p = 0.0000
PC1 Value
PC2 Value
Count
G
Cluster 2 - HDT: Unimodal
Cluster 2 - PC1 Distribution
Dip = 0.0106, p = 0.9900
Cluster 2 - PC2 Distribution
Dip = 0.0096, p = 1.0000
J
Units SPLIT by HDT - 2 Units
Split 1 from C1 (791 spikes)
Split 2 from C1 (358 spikes)
H
Before HDT Split
I
After HDT Split (K=2)
Sub-cluster 1 (791)
Sub-cluster 2 (358)
K
WITH HDT - Cluster 2 (unsplit) (498 spikes)

Figure S2. Hartigan's Dip Test (HDT) to Address Undersorting (Related to Figure 1C,D)
A. Raw spike waveforms for electrode shown in this figure and in **Figure 1C,D. B.** Spike waveforms for Cluster 1 (red) and Cluster 2 (magenta) before (top) and after (bottom) the outlier pre-processing step, which occurs prior to HDT. **C.** Spike waveforms for Clusters 1 and 2 that were kept (top) or removed (bottom) after outlier pre-processing. **D,E.** PCA plots for Clusters 1 and 2 before (**D**) and after (**E**) outlier pre-processing (prior to HDT). **F,G.** HDT performed on PC1 (top) and PC2 (bottom) for Clusters 1 and 2. Only PC2 of Cluster 1 (**F**, bottom) shows significant bimodality. **H,I.** PCA plots of Cluster 1 before (**H**) and after (**I**) splitting into two subclusters (red, blue). **J.** Spike waveforms for the new subclusters split from Cluster 1. **K.** Spike waveforms for Cluster 2, which was not split following HDT.

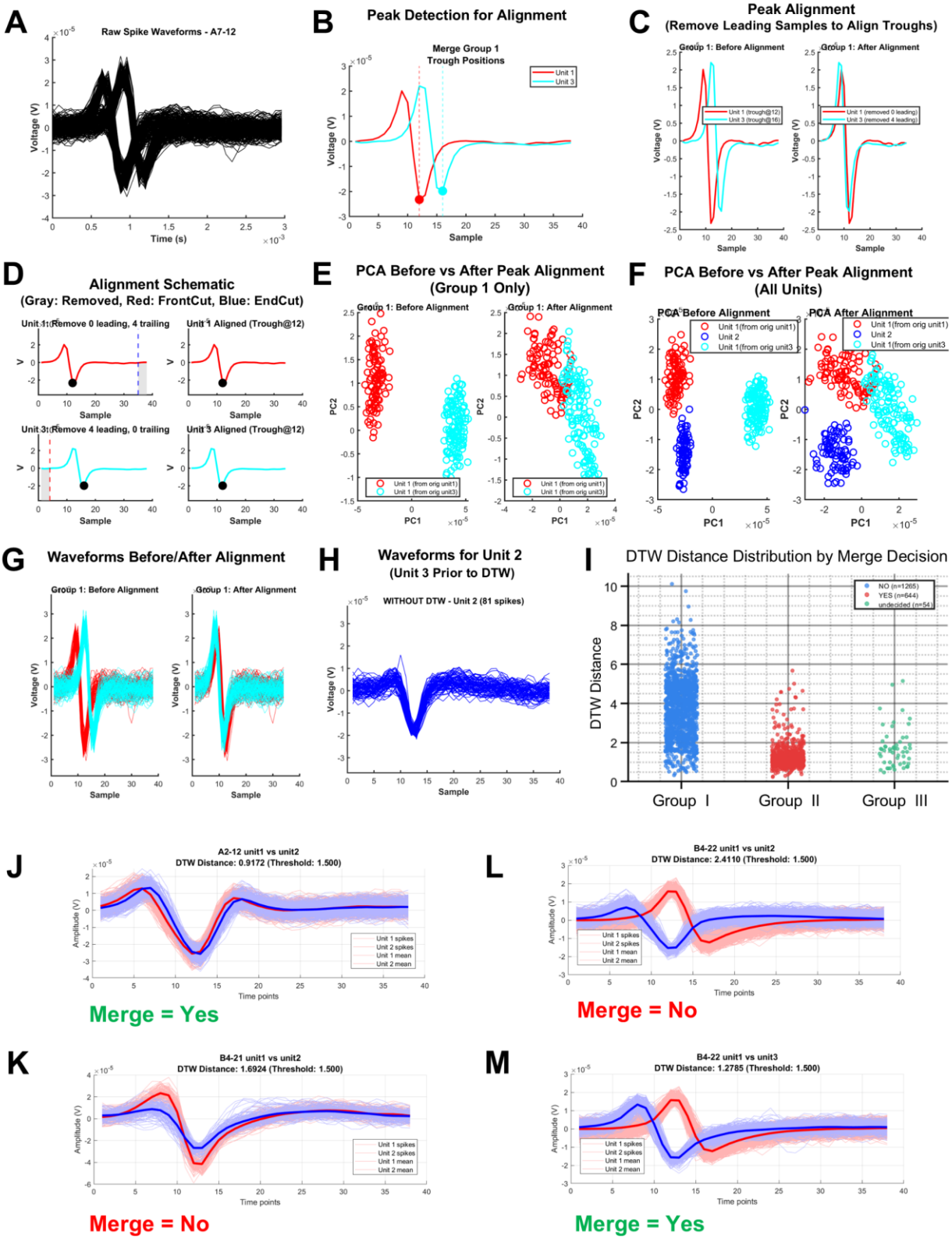

A
Raw Spike Waveforms - A7-12
Voltage (V)
Time (s)
B
Peak Detection for Alignment
Merge Group 1
Trough Positions
Unit 1
Unit 3
Voltage (V)
Sample
C
Peak Alignment
(Remove Leading Samples to Align Troughs)
Group 1: Before Alignment
Group 1: After Alignment
Unit 1 (trough@12)
Unit 3 (trough@16)
Unit 1 (removed 0 leading)
Unit 3 (removed 4 leading)
Voltage (V)
Sample
D
Alignment Schematic
(Gray: Removed, Red: FrontCut, Blue: EndCut)
Unit 1 (Remove 0 leading, 4 trailing
Unit 1 Aligned (Trough@12)
Unit 3 (Remove 4 leading, 0 trailing
Unit 3 Aligned (Trough@12)
V
Sample
E
PCA Before vs After Peak Alignment
(Group 1 Only)
Group 1: Before Alignment
Group 1: After Alignment
Unit 1 (from orig unit1)
Unit 1 (from orig unit3)
PC1
PC2
F
PCA Before vs After Peak Alignment
(All Units)
PCA Before Alignment
PCA After Alignment
Unit 1(from orig unit1)
Unit 2
Unit 1(from orig unit3
PC1
PC2
G
Waveforms Before/After Alignment
Group 1: Before Alignment
Group 1: After Alignment
Voltage (V)
Sample
H
Waveforms for Unit 2
(Unit 3 Prior to DTW)
WITHOUT DTW - Unit 2 (81 spikes)
Voltage (V)
Sample
I
DTW Distance Distribution by Merge Decision
NO (n=1265)
YES (n=644)
undecided (n=54)
DTW Distance
Group I
Group II
Group III
J
A2-12 unit1 vs unit2
DTW Distance: 0.9172 (Threshold: 1.500)
Unit 1 spikes
Unit 2 spikes
Unit 1 mean
Unit 2 mean
Amplitude (V)
Time points
Merge = Yes
L
B4-22 unit1 vs unit2
DTW Distance: 2.4110 (Threshold: 1.500)
Unit 1 spikes
Unit 2 spikes
Unit 1 mean
Unit 2 mean
Amplitude (V)
Time points
Merge = No
K
B4-21 unit1 vs unit2
DTW Distance: 1.6924 (Threshold: 1.500)
Unit 1 spikes
Unit 2 spikes
Unit 1 mean
Unit 2 mean
Amplitude (V)
Time points
Merge = No
M
B4-22 unit1 vs unit3
DTW Distance: 1.2785 (Threshold: 1.500)
Unit 1 spikes
Unit 2 spikes
Unit 1 mean
Unit 2 mean
Amplitude (V)
Time points
Merge = Yes

Figure S3. Dynamic Time Warping (DTW) to Address Oversorting (Related to Figure 1E,F)
A-G. Alignment of waveforms following DTW. **A.** Raw spike waveforms for the electrode that will undergo alignment. The same electrode is also shown in **Figure 1E,F**. **B.** Peak detection for each unit to be aligned. **C,D.** Peak alignment is achieved by removing time points from the start of the traces for one unit to match the earliest peak of the second unit. **E.** PCA plots for Unit 1 (red) and Unit 3 (cyan) before (left) and after (right) peak alignment. **F.** The same data as in **E**, but showing all detected units on the electrode before (left) and after (right) peak alignment. **G.** Spike waveforms for Unit 1 and Unit 3 before (left) and after (right) peak alignment. **H.** Spike waveforms for the Unit 2, which did not undergo alignment. **I.** DTW distances for three groups of sorted spikes. Group I: units for which waveforms were confidently classified as originating from two different cells (DTW distance > 1.5). Group II: units that were classified as originating from the same cell by users but that were over-sorted by SAMS (mean DTW distance ~ 1.5). Group III: units with ambiguous classifications (mean DTW distance ~ 2.0). These cases represent waveforms classified as originating from two different units, but with uncertainty about whether they truly represent distinct units. **J.** Example of an electrode that was misclassified as having two units but that falls below the user-specified DTW distance threshold (here: 1.5). **K.** Example of an electrode that was correctly classified as having two units and that exceeds the user-specified DTW distance threshold. **L,M.** Example of an electrode that contains two units, but that was initially misclassified as having three units. Comparison of Unit 1 and Unit 2 using DTW (**L**) indicates that they are two distinct units (DTW Distance > 1.5), while comparison of Unit 1 and Unit 3 (**M**) indicates that they are waveforms from the same unit (DTW Distance < 1.5).

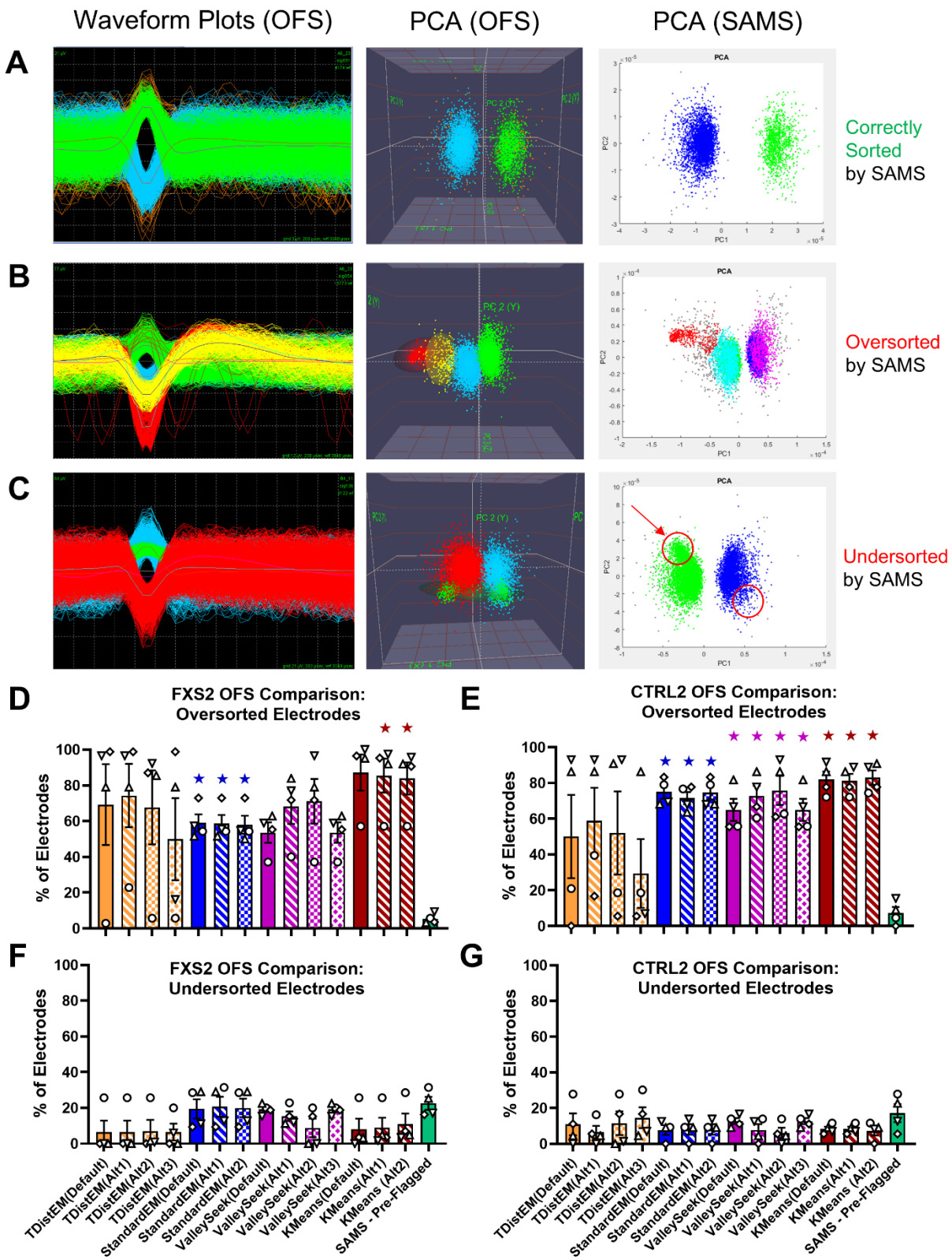

Waveform Plots (OFS)
PCA (OFS)
PCA (SAMS)
A
B
C
Correctly Sorted by SAMS
Oversorted by SAMS
Undersorted by SAMS
PCA
PC1
PC2
D
FXS2 OFS Comparison: Oversorted Electrodes
E
CTRL2 OFS Comparison: Oversorted Electrodes
F
FXS2 OFS Comparison: Undersorted Electrodes
G
CTRL2 OFS Comparison: Undersorted Electrodes
% of Electrodes
TDistEM(Default)
TDistEM(Alt1)
TDistEM(Alt2)
TDistEM(Alt3)
StandardEM(Default)
StandardEM(Alt1)
StandardEM(Alt2)
ValleySeek(Default)
ValleySeek(Alt1)
ValleySeek(Alt2)
ValleySeek(Alt3)
KMeans(Default)
KMeans(Alt1)
KMeans (Alt2)
SAMS - Pre-Flagged

Figure S4. Comparison of the accuracy of SAMS versus automatic spike sorting algorithms in OFS. (Related to Figure 4)

A-C. Examples of ground truth electrodes that were manually sorted in OFS and corresponding PCA plots from OFS and SAMS. **A.** Electrode correctly sorted by SAMS. **B.** Electrode oversorted by SAMS. **C.** Electrode undersorted by SAMS. **D,E.** Percentage of all electrodes that were oversorted by the indicated algorithm in FXS2 (**D**) and CTRL2 (**E**) neurons. **F,G.** Percentage of all electrodes that were undersorted by the indicated algorithm in FXS2 (**F**) and CTRL2 (**G**) neurons. Data shown are the same as in **Figure 4** and **Table S2**, which are from n = 4 independent neuronal differentiations (B1-B4) of N = 1 iPSC line per condition. Batches are represented by different symbol shapes. Statistics: **D,E.** Brown-Forsythe & Welch's ANOVA with Dunnett's T3 post-hoc correction for multiple comparisons; **F,G.** Ordinary one-way ANOVA. ★ indicates a significant difference when compared to SAMS (see Table S2 for detailed statistical information for each comparison).

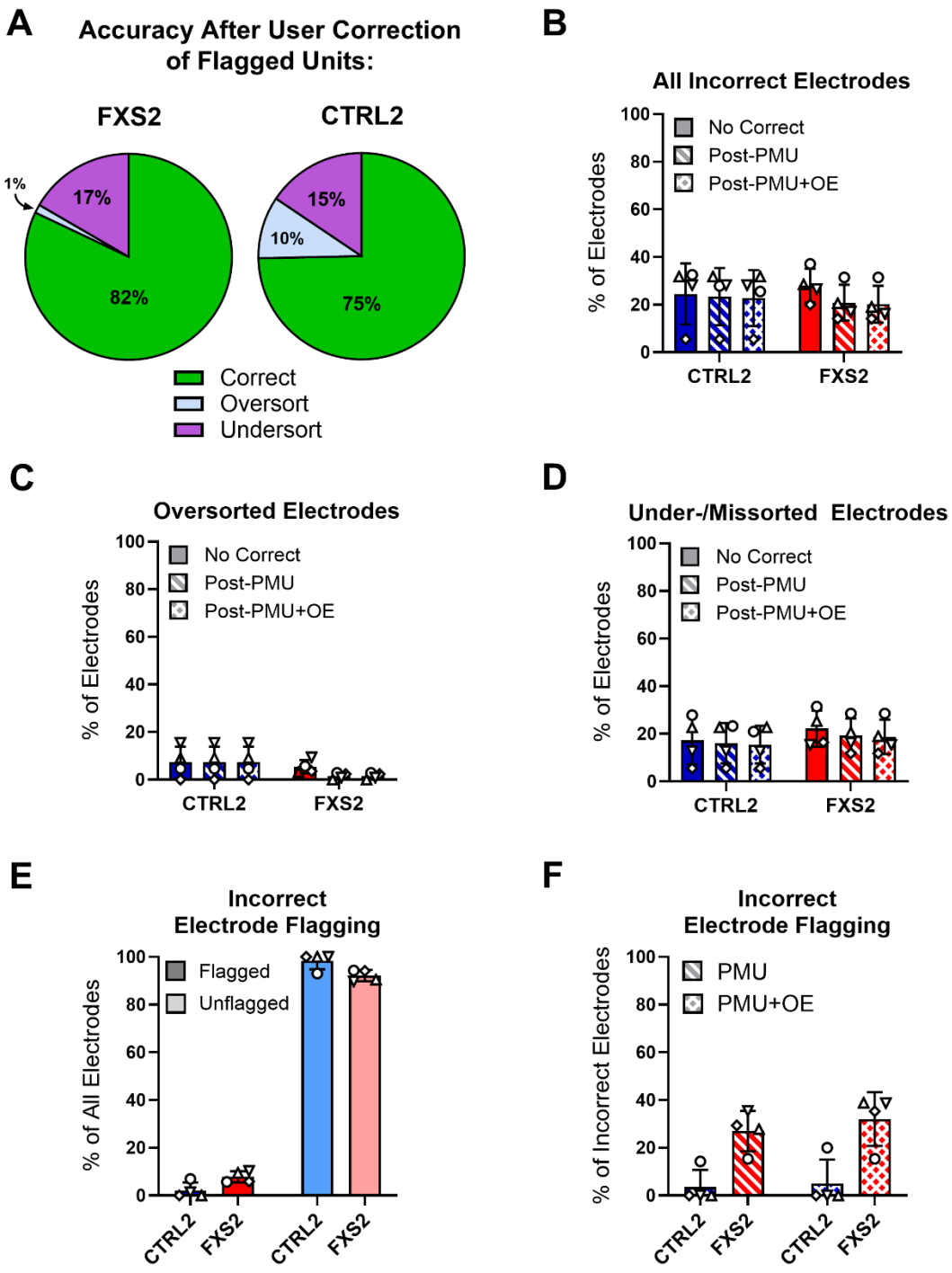

A
Accuracy After User Correction of Flagged Units:
FXS2
CTRL2
1%
17%
82%
15%
10%
75%
Correct
Oversort
Undersort
B
All Incorrect Electrodes
% of Electrodes
No Correct
Post-PMU
Post-PMU+OE
CTRL2
FXS2
C
Oversorted Electrodes
D
Under-/Missorted Electrodes
E
Incorrect Electrode Flagging
% of All Electrodes
Flagged
Unflagged
F
Incorrect Electrode Flagging
% of Incorrect Electrodes
PMU
PMU+OE

Figure S5. Characterization of SAMS Flagging of Potentially Incorrect Sorts

A. The same data as in **Figure 4A**, but after manual correction of any flagged electrodes in FXS2 (left) and CTRL2 (right) neurons. **B.** Percentage of total electrodes that were incorrectly sorted by SAMS in FXS2 and CTRL2 neurons prior to any manual correction by the user (No Correct, solid bars), after correcting incorrect flagged electrodes on only the “Possible Multi-Unit” list (Post-PMU, striped bars), or after correction of incorrect flagged electrodes on both the “Possible Multi-Unit” list or “Over-Excluded Unit” list (Post-PMU+OE, patterned bars). **C.** Same as in **B**, but showing the percentage of total electrodes that were oversorted. **D.** Same as in **B**, but showing the percentage of total electrodes that were undersorted or missorted (when the correct number of units was identified but not all waveforms were correctly assigned as determined by manual inspection). **E.** Percent of total electrodes that were incorrect and that were also flagged by SAMS. **F.** Proportion of incorrect electrodes identified by each flagging type. Data shown are from n = 4 independent neuronal differentiations (B1-B4) of N = 1 iPSC line per condition. Batches are represented by different symbol shapes.

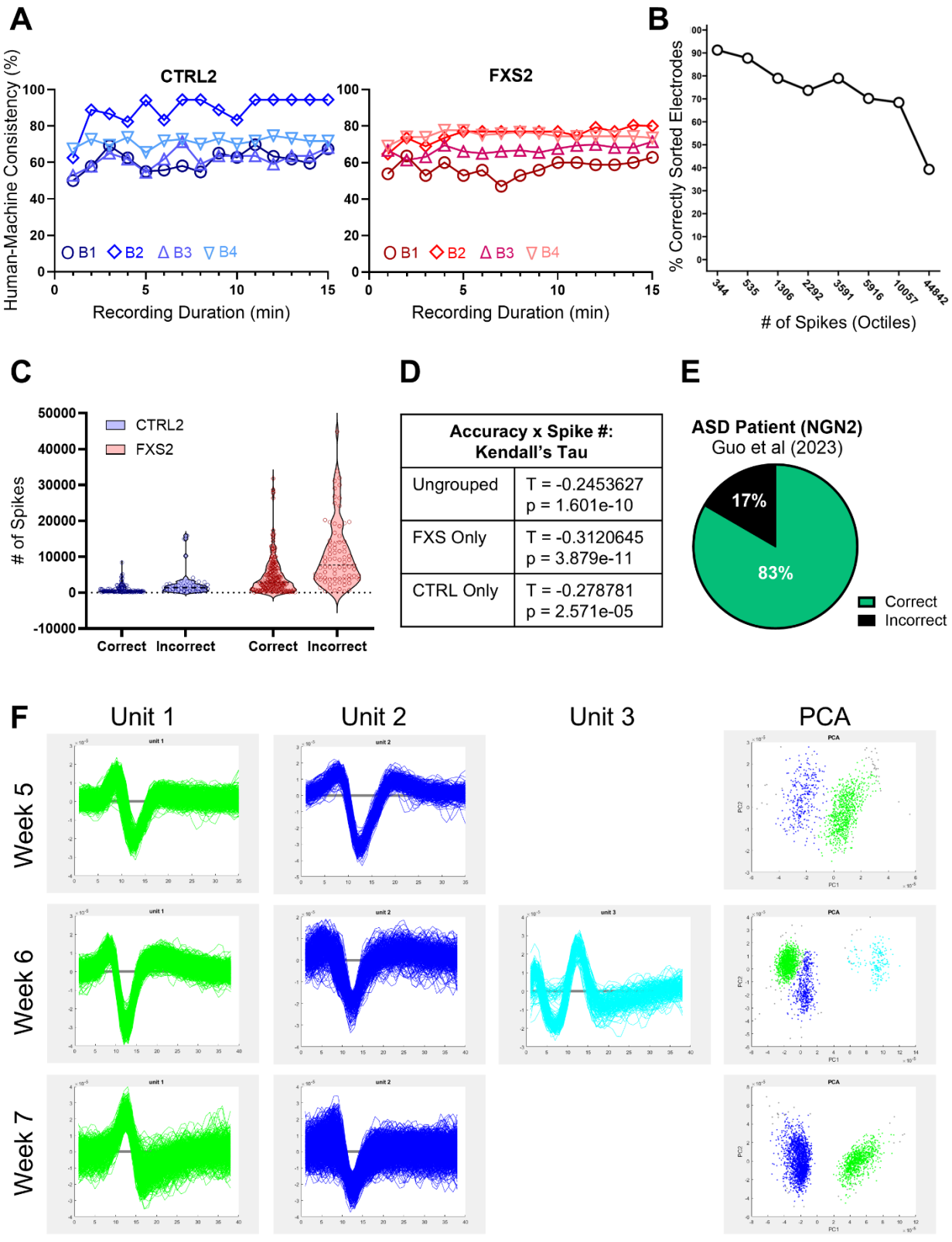

A
CTRL2
FXS2
Human-Machine Consistency (%)
B1 B2 B3 B4
Recording Duration (min)
B
% Correctly Sorted Electrodes
of Spikes (Octiles)
344
535
1306
2292
3591
5916
10057
44842
C
CTRL2
FXS2
of Spikes
Correct
Incorrect
Correct
Incorrect
D
Accuracy x Spike #: Kendall's Tau
Ungrouped
T = -0.2453627
p = 1.601e-10
FXS Only
T = -0.3120645
p = 3.879e-11
CTRL Only
T = -0.278781
p = 2.571e-05
E
ASD Patient (NGN2)
Guo et al (2023)
17%
83%
Correct
Incorrect
F
Unit 1
Unit 2
Unit 3
PCA
Week 5
Week 6
Week 7

Figure S6. Factors affecting human-machine consistency. (Related to Figure 5)
A. The same data as in Figure **5A**, but shown as the mean ± s.e.m. of the 4 individual batches of CTRL2 (left) and FXS2 (right) neurons (B1-B4). **B.** The same data as in **Figure 5E**, but shown as the percentage of correctly sorted electrodes (both FXS2 and CTRL2) relative to the total number of spikes per electrode, split into octiles. **C.** The same data as in **B**, showing the number of spikes on the correctly sorted and incorrectly sorted electrodes in CTRL2 and FXS2 neurons. CTRL2/Correct: n = 113, CTRL2/Incorrect: n = 41, FXS2/Correct: n = 222, FXS2/Incorrect: n = 79. **D.** Kendall's tau was used as a non-parametric measure of ordinal association between SAMS accuracy and number of spikes. Significant monotonic negative associations between performance and number of spikes were seen in all three conditions: ungrouped data (CTRL2 + FXS2), CTRL2 only, or FXS2 only. **E.** SAMS accuracy performance on an additional NGN2-differentiated neuron dataset from an ASD patient with 5q13.2 triplication from Guo et al (2023).[1] Data shown are the average of n = 3 batches of differentiation (B1-B3) at 3 weeks *in vitro*. **F.** Example of SAMS performance over 3 weeks *in vitro* for a single electrode. SAMS was able to correctly characterize changes in the number of active units on this electrode that occurred between recordings (related to **Figure 5C,D**).

A

CTRL3 vs MAP1B-EE		
Parameter	**Std MEA Analysis**	**SAMS**
MFR	↑ *	**n.s.**
# of Bursts	↑ **	↑ ****
Burst Freq	↑ **	↑ ****
Burst Duration	**n.s.**	**n.s.**

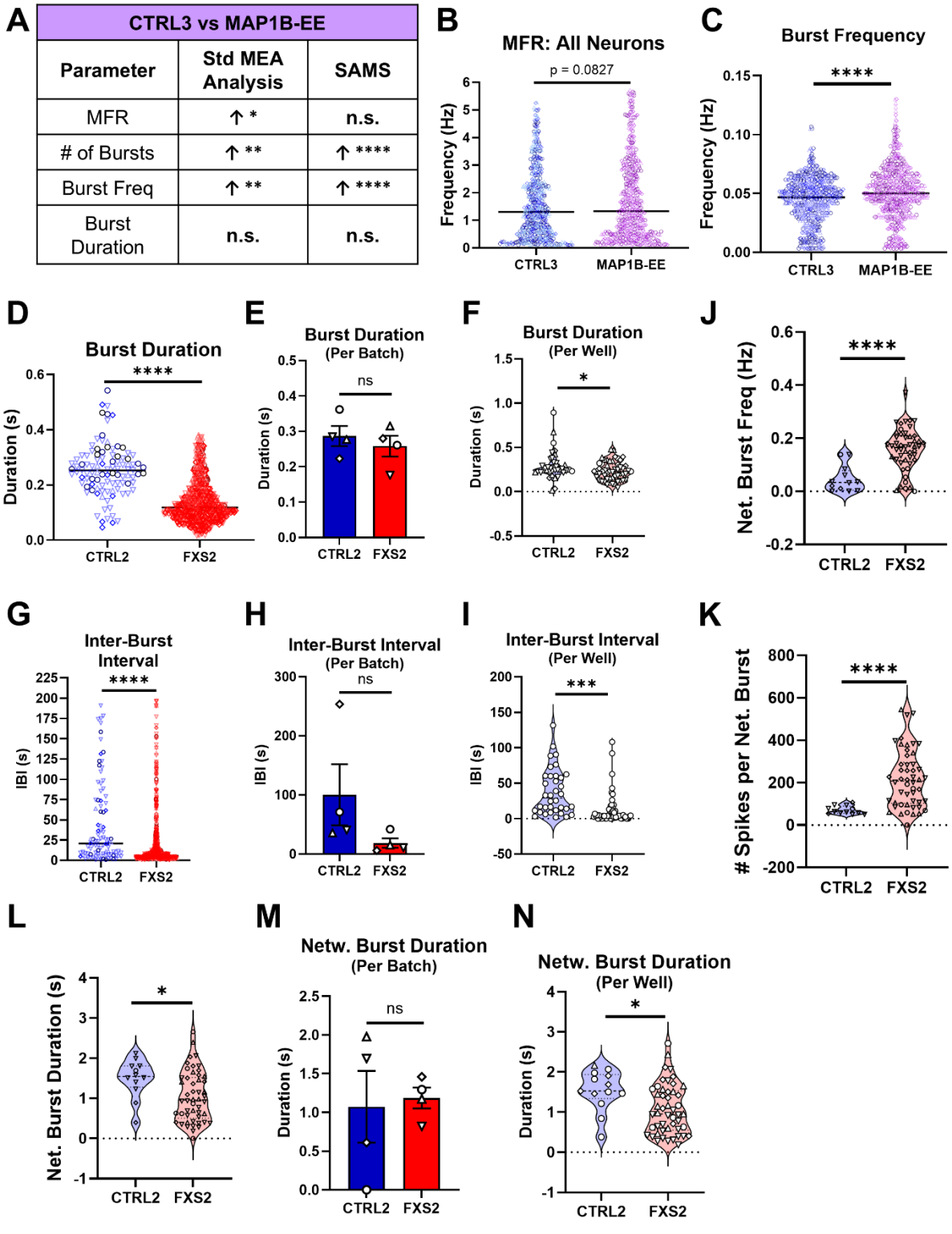


Figure S7. MEA parameters following analysis using SAMS. (Related to Figure 6)
A-C. Analysis of data from Guo et al (2023)[1] comparing control (CTRL3) neurons to isogenic neurons with induced expression of *MAP1B* (MAP1B-EE). **A.** Comparison of results using standard MEA analysis (previously published) and results following spike sorting using SAMS. **B.** Mean firing rate of all neurons identified using spike sorting. **C.** Burst frequency in neurons identified using spike sorting. Statistics: **B**: n = 555 (CTRL3) and 549 (MAP1B-EE) neurons, **C**: n = 577 (CTRL3) and 562 (MAP1B-EE) neurons. 3 independent batches of neuronal differentiation (B1-B3) from N = 1 hESC line per group. Welch's t-test. **D-K.** Analysis of data from Shen et al (2023)[2] comparing control (CTRL2) and Fragile X Syndrome (FXS2) neurons. **D-I.** Examination of two bursting parameters that were inconsistent between SAMS and standard MEA analysis. **D.** Comparison of burst duration in individual units after spike sorting with SAMS shows decreased burst duration in FXS2 neurons. Welch's t-test. n = 103 (CTRL2) and 657 (FXS2) neurons. **E.** Comparison of burst duration using standard MEA analysis, which shows no difference between CTRL2 and FXS2. Unpaired t-test. n = 4 independent batches of neurons (B1-B3,B5). **F.** Comparison of burst duration in individual wells also shows decreased burst duration in FXS, indicating that the inconsistency seen between **E** and **F** is likely an artifact of standard MEA analysis, which is created by averaging all cells, electrodes, and wells together for each batch of cells. Welch's t-test. n = 39 (CTRL2) and 52 (FXS2) individual MEA wells from 4 batches of differentiation (B1-B3, B5). **G.** Comparison of inter-burst interval (IBI) in individual units after spike sorting with SAMS shows decreased IBI in FXS2 neurons. Welch's t-test. n = 96 (CTRL2) and 705 (FXS2) neurons. **H.** Comparison of IBI using standard MEA analysis, which shows no difference between CTRL2 and FXS2. Unpaired t-test. n = 4 independent batches of neurons. **I.** Comparison of IBI in individual wells also shows decreased IBI in FXS, indicating that the inconsistency seen between **G** and **H** is likely an artifact of standard MEA analysis, which is created by averaging all cells, electrodes, and wells together for each batch of cells. Welch's t-test. n = 36 (CTRL2) and 54 (FXS2) individual MEA wells from 4 batches of differentiation. **J,K.** Electrode-level network burst parameters following re-merging of data spike-sorted with SAMS: network burst frequency (**J**) and number of spikes per network burst (**K**). **L-N.** Examination of a network burst parameter that was inconsistent between SAMS and standard MEA analysis. **L.** Comparison of network burst duration in individual units after spike sorting with SAMS shows decreased network burst duration in FXS2 neurons. **M.** Comparison of network burst duration using standard MEA analysis, which shows no difference between CTRL2 and FXS2. Unpaired t-test. n = 4 independent batches of neurons (B1-B3, B5). **N.** Comparison of network burst duration in individual wells also shows decreased burst duration in FXS, indicating that the inconsistency seen between **E** and **F** is likely an artifact of standard MEA analysis, which is created by averaging all cells, electrodes, and wells together for each batch of cells. **J,K,L:** n = 12 (CTRL2) and 50 (FXS2) electrodes, **N:** n = 13 (CTRL2) and 39 (FXS2) individual MEA wells from 4 batches of differentiation (B1-B3, B5). **J,L-N:** unpaired t-test, **K**: Welch's t-test. * $p < 0.05$; *** $p < 0.005$; **** $p < 0.001$. Batches are represented by different symbol shapes. See also **Table S4.**

Table S1. Summary of Available Spike Sorting Programs

Name	MEA Type	Recording Environment	References
Klusta suite	High density	*in vivo*	Rossant (2016)[3]
Mountainsort4	Low or high density	*in vivo*	Chung (2017)[4]
Ironclust	High density	*in vivo*	Jun (2017) *bioRxiv preprint*[5] *[Benchmarked in Magland (2020)]*[6]
WaveClus	Low density	*in vivo*	Quiroga (2004);[7] Chaure (2018)[8]
YASS	High density	*in vivo*	Lee (2020) *bioRxiv preprint*[9]
HTsort	High density	*in vivo*	Chen (2021)[10]
Kilosort4	High density	*in vivo*	Pachitariu (2024)[11]
GEMsort	High density	*in vivo*	Mohammadi (2024)[12]
MixFMM	Low or high density	*in vivo*	Rueda & Rodriguez-Collado (2023)[13]
MultiFq	Low density	*in vivo*	Wang (2023)[14]
Full Binary Pursuit (FBP)	Low or high density	*in vivo*	Hall (2021)[15]
SpyKING CIRCUS	High density	*in vivo, ex vivo*	Yger (2018)[16]
Offline Sorter (Plexon)	Low density	*in vivo, in vitro*	*https://plexon.com/products/offline-sorter/*
HerdingSpikes2	High density	*ex vivo*	Hilgen et al (2017)[17]
[not named]	High density	*ex vivo*	Diggelmann (2018)[18]
[not named]	Low density	*in vitro (rat - cortical)*	Negri (2020)[19]
MEA-ToolBox	Low or high density	*in vitro*	Hu (2022)[20]
µSpikeHunter	High density	*in vitro (rat – cortical; mouse – DRG)*	Heiney (2019)[21]
Mind *in vitro* (MiV)	High density	*in vitro (rat – hippocampal; mouse – mESC-derived MNs)*	Zhang (2024)[22]

Table S2. User-Specified Parameters in SAMS

Parameter	Default Value	Suggested Range *	Description	Refs
Spike Sorting Parameters				
Refractory Period (ms)	1.5	0.5 – 2	Minimum time elapsing between two consecutive spikes to consider them as coming from the same neuron/unit	23-26
Threshold_to_Merge	1.5		DTW Distance threshold for an electrode to be flagged as "Possible Multi-Unit" (See **Fig. S1**)	
Overlap_Threshold	0.7		Controls the strictness of unit similarity detection (higher value: more strict waveform matching; lower value: more lenient merging)	
Flag Threshold	0.2		Threshold for the proportion of excluded waveforms during the invalid waveform exclusion procedure for an electrode to be flagged as "Over-Excluded Unit"	
Cutoff Frequency (Hz)	0.1	0.01 – 0.2	Mean firing rate (MFR) cutoff to exclude low-activity neurons/units	26-29
Standard Deviation Cutoff	3		Number of standard deviations a waveform must exceed from each unit's mean template in ordered to be considered an outlier and therefore excluded from analysis	
Burst Parameters				
Max ISI (ms)	100	80 - 300	Maximum Inter-spike Interval (ISI) allowed within a single burst	1,2,25,30,31
Min # of Spikes	5	4 – 10	Minimum number of spikes allowed within a single burst	24,28,31
Network Bursts				
Min # of Spike	50		Minimum number of spikes required in a network burst	1,2,25,27,30,31
Max ISI (ms)	100	10 - 300	Maximum Inter-spike Interval (ISI) allowed within network bursts	1,2,25,27,29-31
Min Electrodes (%)	35	20-80%	The minimum proportion of total electrodes that a burst must occur on in order to be considered a Network Burst	1,2,22,25,27,28,31

* All parameters have a range of 0 to +∞ in that SAMS will run without errors with any positive value. Suggested Range values have been taken from published *in vitro* MEA and HD-MEA analyses of hPSC-derived neurons.

Table S3. Accuracy of SAMS compared to automatic spike sorting algorithms available in OFS software (Related to Figures 4 and S4).

Name	Algorithm	Parameters [a]	Condition	Correct [b]	Oversort [b]	Undersort [b]
TDistEM (Default)	T-Dist E-M Scan	Seed Cluster Pattern: Circular Range of D.O.F. Multiplier: 10 to 30 Step by: 5 Max. Iterations: 5000 Number of Seed Clusters: 5 X Remove Outliers	CTRL	39.02% (± 19.49)	50.09% (± 23.31)	10.89% (± 6.11)
			FXS	31.52% (± 14.79)	69.32% (± 22.59)	6.43% (± 6.43)
TDistEM (Alt1)	T-Dist E-M Scan	Seed Cluster Pattern: Circular **Range of D.O.F. Multiplier: 5 to 25** Step by: 5 Max. Iterations: 5000 Number of Seed Clusters: 5 X Remove Outliers	CTRL	34.50% (± 16.76)	58.91% (± 18.44)	6.6% (± 3.45)
			FXS	19.25% (± 11.60)	74.32% (± 17.72)	6.43% (± 6.43)
TDistEM (Alt2)	T-Dist E-M Scan	Seed Cluster Pattern: Circular Range of D.O.F. Multiplier: 10 to 30 Step by: 5 Max. Iterations: 5000 **Number of Seed Clusters: 4** X Remove Outliers	CTRL	36.48% (± 17.56)	52.03% (± 23.22)	11.49% (± 6.65)
			FXS	25.31% (± 14.49)	67.64% (± 20.69)	7.04% (± 6.23)
TDistEM (Alt3)	T-Dist E-M Scan	Seed Cluster Pattern: Circular Range of D.O.F. Multiplier: 10 to 30 Step by: 5 Max. Iterations: 5000 Number of Seed Clusters: 5 ✔ **Remove Outliers**	CTRL	**56.04% (± 17.33)**	29.37% (± 19.22)	14.6% (± 5.83)
			FXS	**43.56% (± 19.20)**	49.95% (± 22.96)	6.5% (± 4.72)
StdEM (Default)	Standard EM Scan	Seed Cluster Pattern: Circular Beta: 1 Max Iterations: 500	CTRL	17.27% * (± 2.28)	75.01% *** (± 3.69)	7.72% (± 2.69)
			FXS	21.44% ** (± 3.43)	59.03% ** (± 4.78)	19.53% (± 5.36)
StdEM (Alt1)	Standard E-M Scan	**Seed Cluster Pattern: Linear** Beta: 1 Max Iterations: 500	CTRL	**20.38% * (± 2.85)**	71.56% *** (± 3.42)	8.07% (± 2.91)
			FXS	20.73% ** (± 3.78)	58.60% ** (± 4.85)	20.67% (± 5.58)
StdEM (Alt2)	Standard EM Scan	Seed Cluster Pattern: Circular Beta: 1 **Max Iterations: 1000**	CTRL	17.27% * (± 2.28)	74.66% *** (± 3.88)	8.07% (± 2.91)
			FXS	**21.85% ** (± 3.37)**	57.98% * (± 5.1)	19.96% (± 5.15)
ValleySeek (Default)	Valley Seeking Scan	Seed Cluster Pattern: Circular Range of Parzen Multiplier: 0.5 to 1.5 Step by: 0.2 Limit Number of WFs Used to10000 ✔ Assign unsorted to closest unit ✔ Remove Outliers	CTRL	**22.33% * (± 5.71)**	64.97% * (± 6.19)	12.71% (± 1.66)
			FXS	**27.23% * (± 5.85)**	53.58% (± 5.69)	19.2% (± 1.49)
ValleySeek (Alt1)	Valley Seeking Scan	Seed Cluster Pattern: Circular **Range of Parzen Multiplier: 0.3 to 1.3** Step by: 0.2 Limit Number of WFs Used to10000 ✔ Assign unsorted to closest unit ✔ Remove Outliers	CTRL	19.43% * (± 6.00)	72.82% ** (± 5.76)	7.75% (± 3.31)
			FXS	16.51% * (± 7.09)	68.14% (± 9.70)	15.35% (± 2.65)
ValleySeek (Alt2)	Valley Seeking Scan	Seed Cluster Pattern: Circular **Range of Parzen Multiplier: 0.7 to 1.7** Step by: 0.2 Limit Number of WFs Used to10000 ✔ Assign unsorted to closest unit ✔ Remove Outliers	CTRL	18.16% * (± 6.22)	75.83% * (± 8.14)	6.02% (± 2.91)
			FXS	20.05% (± 8.41)	71.13% (± 12.47)	8.82% (± 5.18)

ValleySeek (Alt3)	Valley Seeking Scan	**Seed Cluster Pattern: Linear** Range of Parzen Multiplier: 0.5 to 1.5 Step by: 0.2 Limit Number of WFs Used to10000 ✔ Assign unsorted to closest unit ✔ Remove Outliers	CTRL FXS	**22.33% * (± 5.71)** **27.23% * (± 5.85)**	64.97% * (± 6.19) 53.58% (± 11.39)	12.71% (± 1.66) 19.20% (± 1.49)
K-Means (Default)	K-Means Scan	Seed Cluster Pattern: Circular Range of Units: 2 to 7	CTRL FXS	9.72% ** (± 3.95) 5.42% *** (± 4.05)	81.98% *** (± 4.36) 87.19% (± 10.07)	8.31% (± 1.33) 8.12% (± 5.92)
K-Means (Alt1)	K-Means Scan	Seed Cluster Pattern: Circular **Range of Units: 1 to 6**	CTRL FXS	**10.42% ** (± 3.5)** **5.6% *** (± 3.88)**	81.28% *** (± 3.88) 85.45% * (± 9.51)	8.31% (± 1.33) 8.95% (± 5.62)
K-Means (Alt2)	K-Means Scan	**Seed Cluster Pattern: Linear** **Range of Units: 1 to 5**	CTRL FXS	9.74% ** (± 5.14) 5.18% *** (± 3.08)	83.00% *** (± 4.06) 83.85% * (± 8.97)	7.26% (± 2.52) 10.97% (± 5.91)
SAMS (Pre-Flagged)	SAMS (no user correction of flagged electrodes)	Refractory Period: 1.5 Threshold to Merge: 1.5 Overlap Threshold: 0.7 Flag Threshold: 0.1 Cutoff Frequency: 0.1 Standard Deviation Cutoff: 3	CTRL FXS	**75.47% (± 6.4)** **72.01% (± 3.58)**	7.31% (± 3.3) 5.45% (± 1.43)	17.22% (± 5.01) 22.55% (± 3.61)

[a] Bold font indicates parameter that was changed from the default value
[b] Values shown are the mean percent of the total number of electrodes (± SEM) from n = 4 independent batches of differentiation of FXS2 and CTRL2 neurons (B1-B4; same data used for analyses shown in **Figures 4** and **S4**). Bold font indicates the setting that achieved the highest accuracy.

Statistics: Ordinary one-way ANOVA followed by Tukey's posthoc test (no significant difference in variance between conditions), or Brown-Forsythe and Welch's ANOVA tests followed by Dunnett's T3 posthoc test (significant difference in variance between conditions). Stars indicate a statistically significant difference between SAMS and the indicated algorithm after posthoc testing. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$

Table S4. Comparison of MEA Parameters in Data from Shen, Sirois, Guo, Li et al (2023)[2] Before and After Spike-Sorting Using SAMS

Parameter	Standard MEA Analysis	Results After Spike-Sorting	Per-Well Analysis
Mean Firing Rate	↑ in FXS2 **	↑ in FXS2 ****	
ISI Coefficient of Variation (CoV)	↑ in FXS2 *	↑ in FXS2 ****	
Burst Duration	**n.d.**	**↓ in FXS2 ******	**↓ in FXS2 ***
Number of Spikes per Burst	↑ in FXS2 *	↑ in FXS2 ****	
Mean ISI in Burst	↓ in FXS2 *	↓ in FXS2 ****	
Inter-burst Interval (IBI)	**n.d.**	**↓ in FXS2 ******	**↓ in FXS2 ****
Burst Frequency	↑ in FXS2 *	↑ in FXS2 ****	
Normalized Duration IQR	**n.d.**	**↑ in FXS2 ******	**↑ in FXS2 ******
IBI CoV	n.d.	n.d.	
Burst Percentage	↑ in FXS2 *	↑ in FXS2 ****	
Network Burst Frequency	↑ in FXS2 **	↑ in FXS2 ***	
Network Burst Duration	**n.d.**	**↓ in FXS2 ***	**↓ in FXS2 ***
Number of Spikes per Network Burst	↑ in FXS2 **	↑ in FXS2 ****	
Network Burst Percentage	↑ in FXS2 *	↑ in FXS2 ****	
Network Burst IBI CoV	n.d.	n.d.	
Synchrony Index	↑ in FXS2 *	↑ in FXS2 ****	

Comparison of FXS2 versus CTRL2 at the 7-week time point.
Standard MEA analysis: n = 4 independent differentiations from N = 1 iPSC line per condition.
Results After Spike-Sorting: n = 300 neurons (CTRL2) and 833 neurons (FXS2) identified through spike sorting from 4 independent differentiations, with n = 107 CTRL2 and 723 FXS2 neurons exhibiting bursting activity, and n = 12 CTRL2 and 50 FXS2 electrodes exhibiting network bursting. After removal of outliers: n = 251-759 (Mean Firing Rate), n = 94-720 (bursting parameters), n = 12-50 (network bursting parameters).
Per-Well Analysis was performed for parameters that were inconsistent between Standard MEA Analysis and Results After Spike-Sorting (bold text). n = 30-54 MEA plate wells (bursting parameters), n = 13-49 MEA plate wells (network parameters) from 4 independent differentiations.
Statistics: Unpaired t-test (no significant difference in variance between groups) or Welch's t-test (significant difference in variance between groups). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$

Table S5. MEA Data and Neuron Batch Information

Condition (Batch) [a]	Age	Diff. Protocol	Recording Duration	Plate Format	Active Wells	# of Active Electrodes (% of Total)	Figures	Ref
FXS2 (B1)	7 wk	Dual SMAD	15 min	96w	10 / 12	35 (44%)	4-7, S4-S7	2
FXS2 (B2)	7 wk	Dual SMAD	15 min	96w	11 / 12	86 (98%)	4-7, S4-S7	2
FXS2 (B3)	7 wk	Dual SMAD	15 min	96w	11 / 12	63 (72%)	4-7, S4-S7	2
FXS2 (B4)	7 wk	Dual SMAD	15 min	96w	17 / 18	117 (86%)	4,5,S4-S6	2
FXS2 (B5)	7 wk	Dual SMAD	15 min	96w	24 / 24	185 (96%)	6,7,S7	2
CTRL2 (B1)	7 wk	Dual SMAD	15 min	96w	12 / 12	43 (45%)	4-7, S4-S7	2
CTRL2 (B2)	7 wk	Dual SMAD	15 min	96w	7 / 12	18 (32%)	4-7, S4-S7	2
CTRL2 (B3)	7 wk	Dual SMAD	15 min	96w	8 /12	22 (34%)	4-7, S4-S7	2
CTRL2 (B4)	7 wk	Dual SMAD	15 min	96w	20 / 24	71 (44%)	4,5,S4-S6	2
CTRL2 (B5)	7 wk	Dual SMAD	15 min	96w	23 / 24	102 (55%)	6,7,S7	2
FXS2 (B1)	5 wk	Dual SMAD	15 min	96w	8 / 12	21 (33%)	5C,5D	2
FXS2 (B1)	6 wk	Dual SMAD	15 min	96w	9 / 12	28 (39%)	5C,5D	2
FXS2 (B2)	5 wk	Dual SMAD	15 min	96w	11 / 12	84 (95%)	5C,5D	2
FXS2 (B2)	6 wk	Dual SMAD	15 min	96w	11 / 12	85 (97%)	5C,5D	2
FXS2 (B4)	5 wk	Dual SMAD	15 min	96w	15 / 18	109 (91%)	5C,5D	2
FXS2 (B4)	6 wk	Dual SMAD	15 min	96w	11 / 12	75 (85%)	5C,5D	2
CTRL2 (B1)	5 wk	Dual SMAD	15 min	96w	9 / 12	20 (28%)	5C,5D	2
CTRL2 (B1)	6 wk	Dual SMAD	15 min	96w	9 / 12	24 (33%)	5C,5D	2
CTRL2 (B2)	5 wk	Dual SMAD	15 min	96w	7 / 12	6 (11%)	5C,5D	2
CTRL2 (B2)	6 wk	Dual SMAD	15 min	96w	2 / 12	7 (44%)	5C,5D	2
CTRL2 (B4)	5 wk	Dual SMAD	15 min	96w	19 / 24	55 (36%)	5C,5D	2
CTRL2 (B4)	6 wk	Dual SMAD	15 min	96w	19 / 24	53 (35%)	5C,5D	2
FXS2 (B1)	7 wk	Dual SMAD	1-14 min	96w		26 – 35	5A	2
FXS2 (B2)	7 wk	Dual SMAD	1-14 min	96w		85 – 87	5A	2
FXS2 (B3)	7 wk	Dual SMAD	1-14 min	96w		54 – 60	5A	2
FXS2 (B4)	7 wk	Dual SMAD	1-14 min	96w		113 – 117	5A	2
CTRL2 (B1)	7 wk	Dual SMAD	1-14 min	96w		28 – 40	5A	2
CTRL2 (B2)	7 wk	Dual SMAD	1-14 min	96w		8 – 18	5A	2
CTRL2 (B3)	7 wk	Dual SMAD	1-14 min	96w		17 – 22	5A	2
CTRL2 (B4)	7 wk	Dual SMAD	1-14 min	96w		43 – 71	5A	2
ASD (B1)	3 wk	NGN2 (NPC)	10 min	96w	57 / 60	134 (95%)	S5E	1
ASD (B2)	3 wk	NGN2 (NPC)	10 min	96w	56 / 80	132 (70%)	S5E	1
ASD (B3)	3 wk	NGN2 (NPC)	10 min	96w	51 / 80	99 (64%)	S5E	1
IsoCTRL	12-23 DIV	NGN2 (iPSC)	9 min	48w	5-16 / 16	5 – 44 (6-19%)	5F,5G	32
Mom	12-23 DIV	NGN2 (iPSC)	9 min	48w	5-15 / 16	5 – 28 (6-14%)	5F,5G	32
Proband	12-23 DIV	NGN2 (iPSC)	9 min	48w	5-16 / 16	6 – 145 (8-57%)	5F,5G	32
CTRL3 (B1)	3 wk	NGN2 (NPC)	10 min	96w	16 / 16	106 (83%)	6,7,S7	1
CTRL3 (B2)	3 wk	NGN2 (NPC)	10 min	96w	16 / 16	88 (69%)	6,7,S7	1
CTRL3 (B3)	3 wk	NGN2 (NPC)	10 min	96w	16 / 16	106 (83%)	6,7,S7	1
MAP1B-EE (B1)	3 wk	NGN2 (NPC)	10 min	96w	16 / 16	107 (84%)	6,7,S7	1
MAP1B-EE (B2)	3 wk	NGN2 (NPC)	10 min	96w	16 / 16	107 (84%)	6,7,S7	1
MAP1B-EE (B3)	3 wk	NGN2 (NPC)	10 min	96w	16 / 16	87 (68%)	6,7,S7	1
FXS1 (B1-B3) / shNC	3 wk	NGN2 (NPC)	10 min	96w	24 / 24	41-45 (64-70%)	6,7,S7	1
FXS1 (B1-B3) / shMAP1B	3 wk	NGN2 (NPC)	10 min	96w	24 / 24	37-49 (58-77%)	6,7,S7	1
CTRL1 (B1-B3) / shNC	3 wk	NGN2 (NPC)	10 min	96w	19 / 24	7-19 (18-34%)	6,7,S7	1

[a] B1-B5 indicate independent batches of neuronal differentiation from the same iPSC line or condition

iPSC lines:

FXS1: WC007i-FX13-2

FXS2: WC052i-FX08-25;

CTRL1: WC032i-6007-1

CTRL2, WC031i-5907-6

CTRL3, idCas9A-H9

MAP1B-EE: CTRL3 + doxycycline induced *MAP1B* expression (via dCas9 activator system)

ASD: WC-057-ASD

IsoCTRL: HNDS0005-01 #B2 +/+

Mom: HNDS0005-02 #A

Proband: HNDS0005-01 #B

Table S6. Settings Used for Spike File Generation (AxIS Navigator Software)

Method	
Threshold:	6 x Standard Deviation
Detect Only Crossings:	Yes
Peak Detection:	Maximum Amplitude
Durations	
Pre-Spike:	0.84 ms (11 samples)
Post-Spike:	2.16 ms (27 samples)
Allow Sample Overlap:	No
Coincident Events	
Coincident Event Removal:	Well
Spike Counting	
Interval:	1s (12.5 K samples)

SUPPLEMENTAL METHODS

Manual spike sorting of MEA data files to establish ground truth

First, PCA clustering and the peaks of aligned waveforms were used to define putative units. Next, each putative unit was subdivided into smaller subunits containing 50 to 200 waveforms on average. We then manually inspected these subunits to compare their similarity. Two subunits were considered to be from the same unit based on the waveform amplitude, waveform shape similarity, alignment of peaks and alignment of troughs. Two subunits were considered similar if more than two-thirds of the waveform shapes along their lengths were perceptually indistinguishable or highly overlapping.

Because OFS only allows creation of up to 26 subunits at a time, steps 1-3 were performed in multiple iterations until a final set of units remained that were distinct from one another based on peak location, trough location, peak to trough amplitude, and waveform shape dissimilarity. Next, units and subunits with similar shapes but whose waveforms did not align at their peaks and troughs were also compared to determine whether they were the same unit that was separated by a time shift. Waveforms which had not been assigned to any units during steps 1 through 4 were examined again based on shape, PCA, and degree of overlap to determine whether they belonged to any existing units or whether they belonged to a new unit. Finally, units with a firing rate below a threshold of 0.1 Hz were excluded from the final unit count. All recorded electrodes for each batch of cells were included for analysis.

MEA Datasets

Standard Differentiation Using Dual SMAD-Inhibition

Data from Shen et al (2023) were acquired from neurons differentiated using a standard dual SMAD-inhibition monolayer-based protocol, in which neural progenitor cells (NPCs) were plated for terminal differentiation on MEA plates, then co-cultured with mouse primary astrocytes that were added one week after neuron plating.[2] Neurons were recorded weekly, starting at one week after the addition of astrocytes, for a total of 8 weeks, for 15 minutes in 5% CO_2 at 37°C. The labels “FXS2” and “CTRL2” to refer to data from the WC053i-FX08-25 and WC031i-5907-6 lines, respectively (for more information, refer to Shen et (2023)).

NGN2-induction from NPCs

Data from Guo et al (2023) were acquired from neurons differentiated using an NGN2-based protocol, in which NPCs were co-transduced with lentiviruses containing TetO-NGN2-Neo and reverse tetracycline-controlled transactivator (rtTA) then treated with doxycycline and selected using G418.[1] The resulting cells were plated onto MEA plates, then co-cultured with mouse primary astrocytes that were added one week after neuron plating. Neurons were recorded weekly, starting at one week after the addition of astrocytes, for a total of 6 weeks, for 10 minutes in 5% CO_2 at 37°C. In this paper, we used the labels “FXS1” to refer to the WC007i-FX13-2 line, “CTRL1” to refer to the WC032i-6007-1 line, “CTRL3” to refer to idCas9A-H9 infected with *LV-sgCtrl*, “MAP1B-EE” to refer to idCas9A-H9 infected with *LV-sgMAP1B*, and “ASD Patient” to refers to WC-057-ASD, which was generated from a patient with 5q13.2 triplication (for more information, please refer to Guo et al).

Induced Neurons (iNs)

To further assess the robustness of SAMS, we analyzed MEA data provided by Dr. Mustafa Sahin’s lab.[34] This MEA data was previously recorded from excitatory neurons that were directly induced via NGN2 from the iPSC stage (induced neurons (iNs)) that were co-cultured with hiPSC-derived astrocytes at 12, 14, 16, 19, 21, and 23 days *in vitro* (DIV). In this paper, we use the labels “Proband” to refer to the HNDS0005-01 #B iPSC line, which was derived from a female patient with a homozygous mutation in the *ALDH5A1* gene (c1226G>A;p.Gly409Asp); “Mom” to refer to the HNDS0005-02 #A iPSC

line, which was derived from the mother of the patient who is heterozygous for the same *ALDH5A1* mutation, and "IsoCTRL" to refer to the HNDS0005-01 #B2 +/+ iPSC line, which was generated by CRISPR-mediated correction of the HNDS0005-01 #B iPSC line, allowing it to serve as an isogenic control (for more information, please refer to Afshar-Saber et al (2024b)).[35]

Data Analysis and Statistics

Prior to testing for significant differences, outliers were identified using the ROUT method (Q=1%) and then removed, with the exception of **Figure S7B,** in which Q=1% excluded an excessive amount of data points and so Q=0.2% was used instead. For comparisons between two conditions, unpaired, two-tailed student's t-test (no significant difference in variance between conditions) or two-tailed, unpaired t-test with Welch's correction (significant difference in variance between conditions) was performed. For comparisons between more than two conditions, ordinary one-way ANOVA followed by Tukey's posthoc test (no significant difference in variance between conditions), or Brown-Forsythe and Welch's ANOVA tests followed by Dunnett's T3 posthoc test (significant difference in variance between conditions) was performed.

Statistical comparison between independent batches of neuronal differentiation was performed for the following figures: **Figure 4G,H; S4D-G; S5B-F; S7E,H,M; 7B,E,H**. Statistical comparison of individual wells pooled from the 4 batches of differentiation (where n = the number of wells) of FXS2 and CTRL2 was performed in **Figure S7F,I,N** and where indicated in **Table S4**. Statistical comparison of individual units identified by SAMS pooled from all batches of differentiation for each condition was performed in the following figures: **Figure S6D; Figure 6C-H; Figure S7A-D,G; Figure 7A,C,D,F,G**. Statistical comparison of pooled network activity across individual electrodes was performed in **Figure S7J-L** and where indicated in **Table S4**. The number of wells and electrodes analyzed for all datasets and conditions are listed in Table S5.

Correlation Analysis Between # of Waveforms and SAMS Sorting Accuracy

Since the dataset containing the number of spikes per electrode is heavily right-skewed, Kendall's Tau was selected as a non-parametric measure of ordinal association. Kendall's Tau was calculated and significance was tested for (a) all electrodes, (b) only FXS, and (c) only control conditions.

Bootstrapped Performance Comparison Between Conditions

In addition to the correlation analysis, we also evaluated whether SAMS performs differently by condition. First, we set random seed for reproducibility, then resampled each condition (FXS and CTRL) 10,000 times with replacement. For each resample, we calculated the proportion of each condition which was correctly sorted by SAMS. We then took the difference of the proportions for that resample and stored the value in an array. Once done resampling, we sorted the array and took the 250^{th} and 9750^{th} values. These values are the lower and upper bounds on a confidence interval for the difference in SAMS sorting accuracy between control and FXS conditions. We obtained the confidence interval [-0.068, 0.104] from this procedure, indicating that there is no significant difference in SAMS sorting accuracy between conditions.

SAMS – Principal Component Analysis (PCA) for Feature Extraction

Prior to clustering, spike waveforms are projected into a lower dimensional feature space using principal component analysis (PCA). Each spike waveform is transformed into principal component (PC) scores. The first two to three PCs typically capture >90% of waveform variance and represent the primary distinguishing features between neuronal units, such as amplitude and waveform width or biphasic characteristics. This dimensionality reduction facilitates both computational efficiency and visualization of clustering results.

SAMS – Spectral Clustering for Initial Unit Assignment

Conventional clustering algorithms such as K-means assume spherically distributed clusters and are sensitive to outliers.[36] When applied to spike waveform data in PCA feature space, these assumptions are often violated as neuronal clusters can exhibit elongated, curved, or irregular distributions. These limitations frequently result in misclassification of units, particularly when neurons have overlapping amplitude distributions or when noisy spikes create outliers that skew cluster centers. SAMS implements spectral clustering, which performs clustering on the eigenvectors of the data's affinity matrix. The algorithm solves the following eigenvalue problem:

$(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{v} = \boldsymbol{\lambda D v}$

Where D is the degree matrix, W is the weighted adjacency matrix, v is the eigenvector, and λ is the eigenvalue. The eigenvectors corresponding to the k smallest eigenvalues are used to partition the data into k clusters. This method maximizes between-cluster variance while minimizing within-cluster variance, providing a solid foundation for subsequent refinement steps.

SAMS – Spike Sorting Refinement Procedures

S1. Hartigan's Dip Test for Undersorting Correction

S1.1 Background

Spike sorting algorithms face a fundamental tension between oversorting (splitting one neuron into multiple units) and undersorting (grouping multiple neurons into a single unit). Undersorting tends to happen when waveforms from different neurons have similar shapes but differ in some subtler feature. The sign is a cluster whose distribution in principal component space is multimodal rather than unimodal.

To detect this, SAMS use Hartigan's Dip Test (HDT),[37] a statistical test designed to assess whether a distribution has multiple modes. The test computes a "dip statistic" that measures the maximum difference between the empirical distribution and the best-fitting unimodal distribution. A small p-value indicates that the data are unlikely to have come from a unimodal distribution, suggesting the presence of two or more modes.

The dip statistic D quantifies deviation from unimodality:

$D = \max |F(x) - G(x)|$

Where $\mathrm{F(x)}$ is the empirical distribution function and $\mathrm{G(x)}$ is the unimodal distribution function that minimizes the maximum difference.

Statistical significance is assessed through bootstrap sampling (500 iterations) against the null hypothesis of unimodality. When either PC1 or PC2 distributions exhibit significant multimodality ($p < 0.05$), the cluster undergoes spectral clustering. To prevent excessive splitting of single units with high waveform variability, sub-clusters are validated using amplitude overlap and temporal separation criteria (see S1.2).Clusters that fail the dip test (i.e. D exceeds a critical value; here, 0.1), indicating multimodality, are split to improve sorting accuracy.

S1.2 Implementation

Clusters containing more than 10 spikes are evaluated for potential undersorting, where multiple neurons may have been incorrectly grouped together. We apply Hartigan's Dip Test to the first two principal component scores of each cluster, using 500 bootstrap samples. Clusters exhibiting significant multimodality in either component ($p < 0.05$) are flagged for splitting.

For flagged clusters, we determine the optimal number of sub-clusters (K = 2 to 5) using the Davies-Bouldin criterion and then apply spectral clustering to divide the cluster accordingly. Each resulting pair of sub-clusters undergoes validation based on two metrics: amplitude overlap, calculated as the proportion of waveform timepoints where the amplitude ranges of the two sub-clusters overlap by more than 70%, and temporal separation, assessed by computing the minimum inter-spike interval between the combined spike trains.

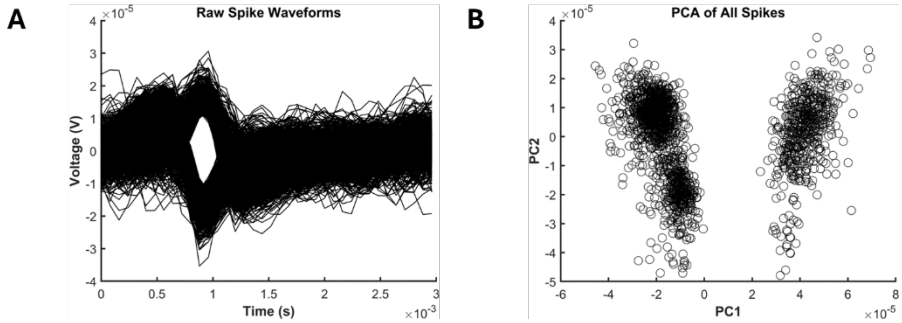
A split is rejected and the sub-clusters are merged back together if more than 70% of timepoints exhibit substantial amplitude overlap and the minimum inter-spike interval exceeds the refractory period. This combination indicates that the sub-clusters share nearly identical waveform morphology and produce no refractory period violations when combined, suggesting they represent the same neuron artificially separated by clustering variability.

If multiple sub-clusters pass validation, the split is accepted and each sub-cluster is retained as a separate unit. If all sub-clusters fail validation and are merged, or if no multimodality is detected initially, the original cluster is preserved as a single unit.

S1.3 Example Electrode

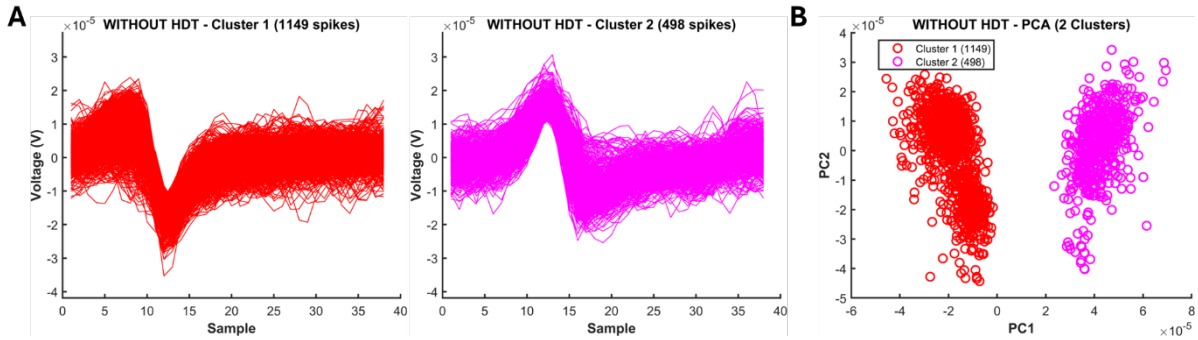
The following example demonstrates the HDT procedure on an electrode where initial clustering identified two units, but one of them actually contained spikes from two distinct neurons. Data are from Shen et al (2023).[2]

SM Figure 1A shows all detected spike waveforms from this electrode. There is substantial variability, with what appears to be at least two distinct waveform shapes present. **SM Figure 1B** shows the corresponding PCA projection. Looking at the left cluster (negative PC1 values), there is a hint of gap in distribution of PC2 that might indicate two overlapping populations.



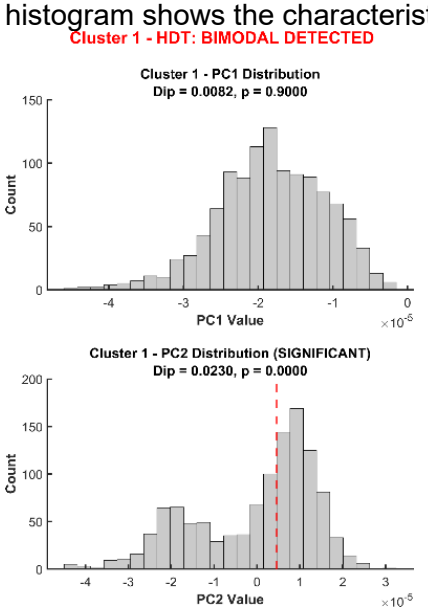
SM Figure 1. A. Raw spike waveforms from the example electrode. **B.** PCA of all spikes. Two main clusters are visible, but the left cluster appears to be consisted by two small clusters along PC2.

Without the HDT correction step, spectral clustering with Davies-Bouldin optimization identifies two clusters: Cluster 1 with 1149 spikes and Cluster 2 with 498 spikes (**SM Figure 2).** Looking at the waveforms, Cluster 1 shows considerable variability—some waveforms have larger amplitudes than others, which might indicate mixing of two neurons. Cluster 2, by contrast, looks fairly homogeneous.



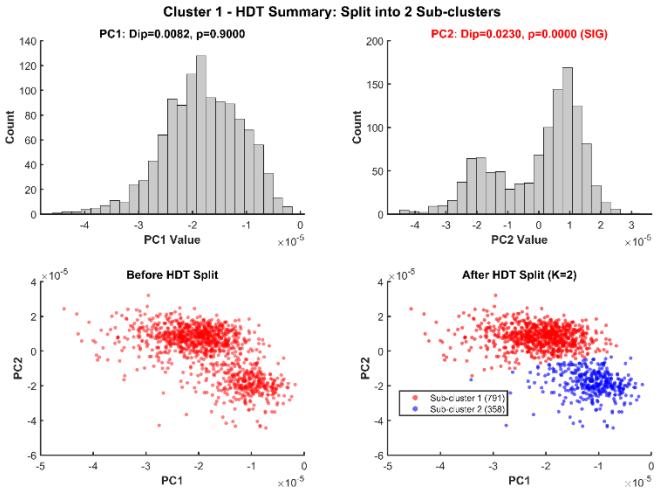
SM Figure 2. A. Sorted waveforms without HDT correction. Cluster 1 (1149 spikes) shows high variability; Cluster 2 (498 spikes) appears more uniform.(The waveforms shown in the figures is reflecting HDT process, removal of outliers of the defined clusters is not shown here). **B.** PCA colored by cluster assignment (without HDT). Cluster 1 (red) spans a large vertical range in PC2.

We ran HDT on both clusters. For Cluster 1, the PC1 distribution yielded a dip statistic of 0.0082 with p = 0.934—clearly unimodal, no evidence of multimodality. But the PC2 distribution told a different story: dip = 0.0230, p = 0.002. This is well below the threshold of 0.05, indicating significant multimodality along the PC2 axis (**SM Figure 3**). The histogram shows the characteristic two-humped shape.



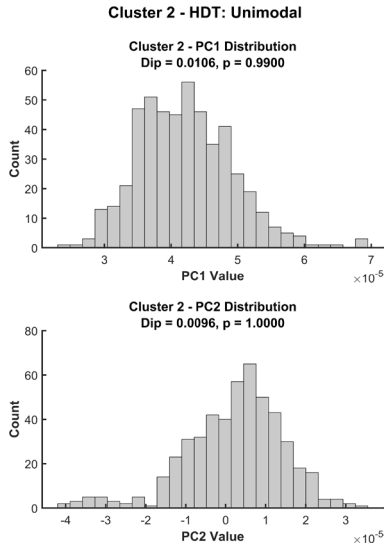
SM Figure 3. Hartigan's Dip Test results for Cluster 1. PC1 (top) is unimodal (p = 0.934). PC2 (bottom) shows significant multimodality (p = 0.002), triggering a split.

SM Figure 4 summarizes the split decision for Cluster 1. The bottom-left panel shows the original cluster in PCA space, and the bottom-right shows the result after forcing K=2 spectral clustering. The split separates the cluster into Sub-cluster 1 (791 spikes, upper region in PC2) and Sub-cluster 2 (358 spikes, lower region in PC2).



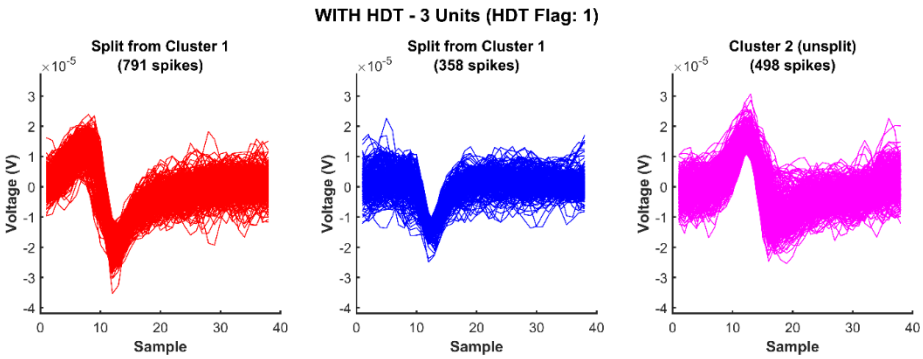
SM Figure 4. Summary of HDT split for Cluster 1. Top row: PC distributions with dip statistics. Bottom row: PCA before (left) and after (right) the split.

For completeness, **SM Figure 5** shows the HDT results for Cluster 2. Both PC1 (dip = 0.0106, p = 0.992) and PC2 (dip = 0.0096, p = 0.998) are solidly unimodal, so this cluster was not split.

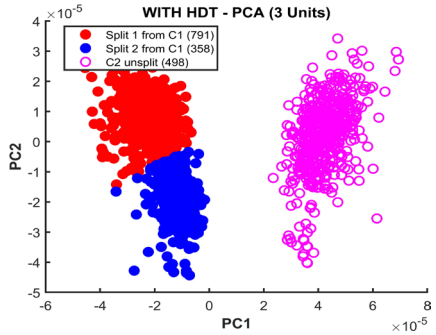


SM Figure 5. Hartigan's Dip Test results for Cluster 2. Both PC1 and PC2 are unimodal ($p > 0.99$), so no split was performed.

After applying HDT, the electrode now has three units instead of two. The original Cluster 1 has been split into two sub-clusters based on the multimodality detected in PC2, while Cluster 2 remains intact (**SM Figures 6–7**).



SM Figure 6. Final sorted waveforms after HDT correction. Three units: Split 1 from Cluster 1 (791 spikes), Split 2 from Cluster 1 (358 spikes), and Cluster 2 unchanged (498 spikes).



SM Figure 7. PCA after HDT correction. The multimodal Cluster 1 has been split into two sub-clusters (red and blue, filled markers), while Cluster 2 (magenta, open markers) remains a single unit.

SM Figure 8 show the individual waveforms for split units after HDT correction. The two sub-clusters from the original Cluster 1 now show more homogeneous waveform

populations. Split 1 (791 spikes) retains the larger-amplitude waveforms, while Split 2 (358 spikes) contains smaller-amplitude spikes.

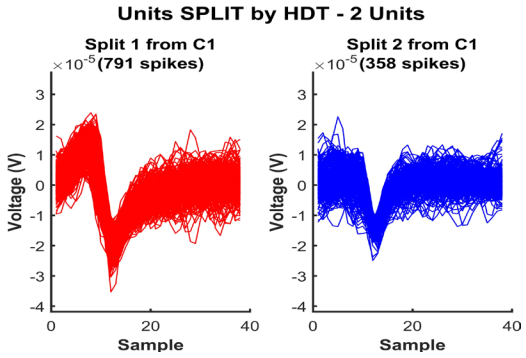


Figure S8. Side-by-side comparison of the two units created by splitting Cluster 1. The amplitude difference between Split 1 (red, 791 spikes) and Split 2 (blue, 358 spikes) is evident.

S2. Dynamic Time Warping-Based Template Merging for Oversorting Correction

S2.1 Background

While HDT addresses undersorting, a separate issue arises from dual-threshold spike detection. Because both positive and negative voltage thresholds can trigger spike extraction, noise in the recording sometimes causes the same neuron's action potentials to be captured with slightly different temporal offsets depending on which threshold was crossed first. The result is multiple populations of spikes that are nearly identical in shape but offset by a few samples in time. If clustering methods solely compare waveforms point-by-point, they will treat these time-shifted spikes as coming from different neurons—a form of oversorting that would inflate unit counts.

We addressed this concern by incorporating Dynamic Time Warping (DTW)[38,39] into the template comparison step. DTW measures similarity between two sequences while allowing for nonlinear stretching and compression along the time axis, making it well-suited for detecting waveforms that differ mainly in temporal alignment rather than shape. However, DTW alone can sometimes be misleading—matching templates that have similar overall energy but genuinely different morphology. To resolve this, we added amplitude-based criteria requiring that the peak and trough values also fall within a specified tolerance.

The DTW distance between two time series X = (x1, ..., xn) and Y = (y1, ..., ym) is calculated as:

$$DTW(X,Y) = \min \sqrt{\sum w_k}$$

Where w_k is the k-th element of a warping path W. The optimal warping path is found using dynamic programming to minimize the cumulative distance between the aligned elements of two waveforms: X and Y. Template waveforms are normalized to [0,1] range before comparison to ensure amplitude-independent shape matching.

S2.2 Template Comparison Procedure

To address oversorting, where a single neuron has been incorrectly divided into multiple clusters, we compare spike waveform templates across clusters to identify candidates for merging.

For each cluster, we first compute a mean waveform template by averaging all spike waveforms assigned to that cluster. All templates are then normalized to a common scale (0–1) using the global minimum and maximum values across the full set of templates, ensuring consistent comparison regardless of amplitude differences between clusters.

We evaluate every pair of templates using two complementary metrics. The first is DTW distance, which quantifies waveform similarity while accommodating minor temporal shifts that may arise from alignment variability. The second is a shape similarity

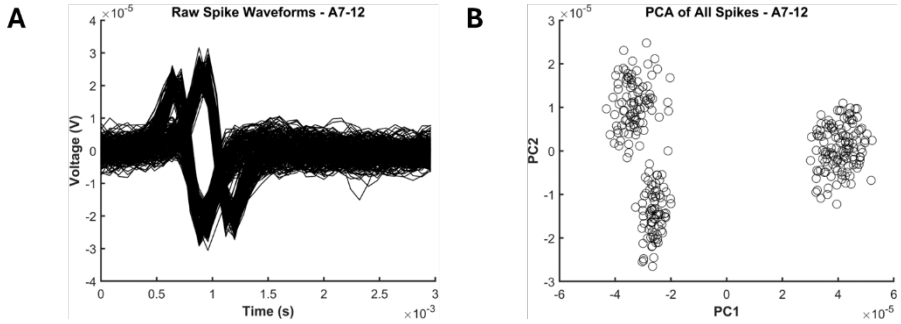
assessment based on waveform features. For each template, we extract the peak and trough amplitudes and calculate the waveform height as the difference between them. Two templates are considered shape-matched if their peak and trough values fall within a tolerance defined as a fraction of the waveform height, allowing for proportional variation in feature magnitudes.

Two clusters are marked for merging when their DTW distance falls below a specified threshold and both peak and trough values satisfy the shape matching criteria. Merge candidates are then grouped using connected component analysis: if cluster A is flagged for merging with cluster B, and cluster B with cluster C, all three clusters are consolidated into a single merged unit. This transitive grouping ensures that clusters representing the same neuron are combined even when identified through indirect pairwise relationships

S2.3 Example Electrode A7-12

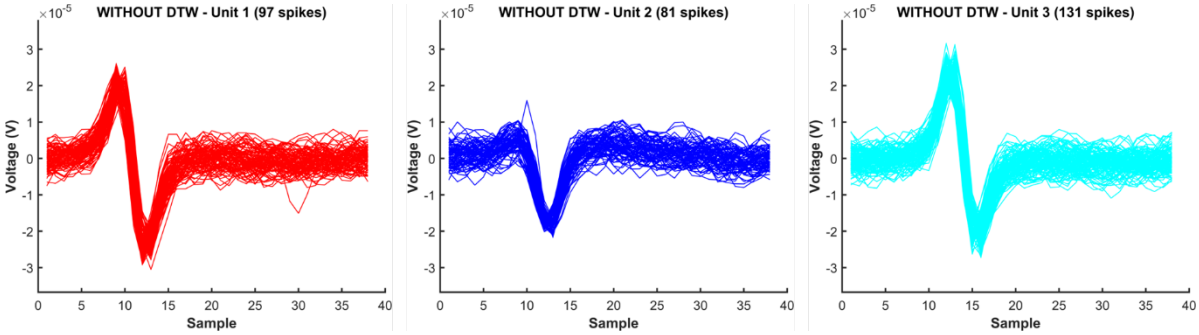
To illustrate the DTW merging procedure, we walk through electrode A7-12, which exhibits the temporal displacement artifact described above. Data are from Shen et al (2023).[2]

SM Figure 9A shows all extracted spike waveforms from this electrode. Even by eye, there is noticeable variation in where the peak occurs, hinting that some waveforms may be time-shifted versions of the same underlying signal. **SM Figure 9B** presents the corresponding PCA projection.

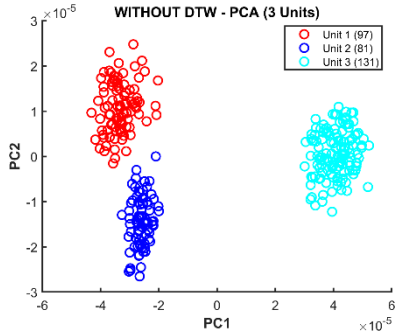


SM Figure 9. A. Raw spike waveforms and **B.** PCA projection of all detected spikes from electrode A7-12.

Without DTW-based template comparison, spectral clustering identifies three separate units: Unit 1 (97 spikes), Unit 2 (81 spikes), and Unit 3 (131 spikes). **SM Figures 10–11** show the sorted waveforms and their PCA distribution. Units 1 and 3 occupy different regions on the PC spaces but have strikingly similar waveform shapes, differing primarily in peak timing. Unit 2, on the right side of the PCA plot, close to unit1, has a smaller amplitude and distinct morphology.

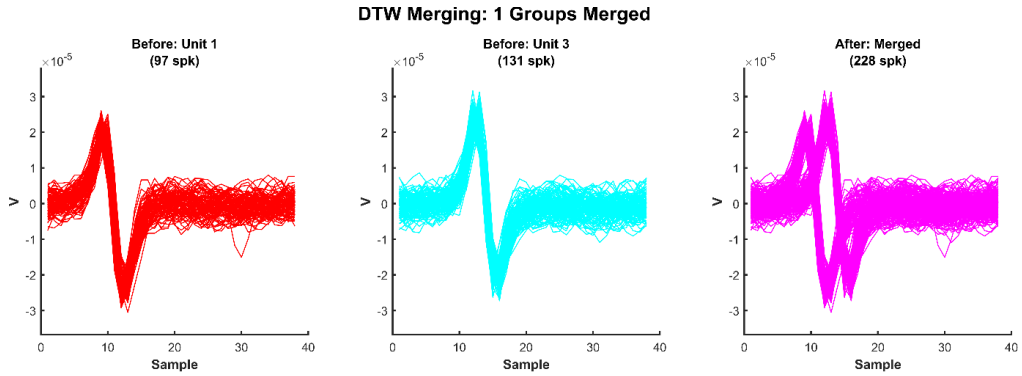


SM Figure S10. Sorted waveforms without DTW merging. Three units are identified.



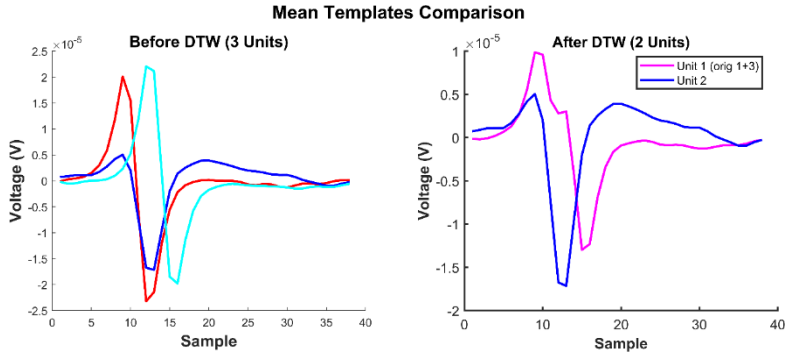
SM Figure S11. PCA colored by cluster assignment (without DTW). Unit 1 (red), Unit 2 (blue), Unit 3 (cyan).

When DTW-based template comparison is applied, the algorithm correctly identifies Units 1 and 3 as belonging to the same neuron. The DTW distance between their templates falls below threshold, and both peak and trough amplitudes match within the 30% tolerance. **SM Figure S12** shows the merge decision: the 97 spikes from Unit 1 and 131 spikes from Unit 3 combine into a single cluster of 228 spikes.

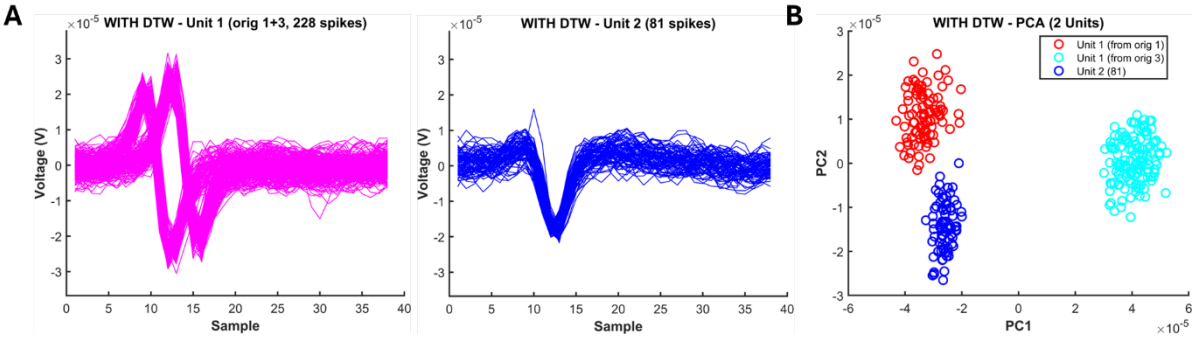


SM Figure S12. DTW merge decision. Before: Unit 1 (97 spikes) and Unit 3 (131 spikes). After: merged unit (228 spikes).

SM Figure S13 compares mean templates before and after merging. The left panel shows the three original templates; note how the red (Unit 1) and magenta (Unit 2) traces are nearly superimposable except for a ~4 sample offset.



SM Figure S13. Mean templates before (left, 3 units) and after (right, 2 units) DTW merging.

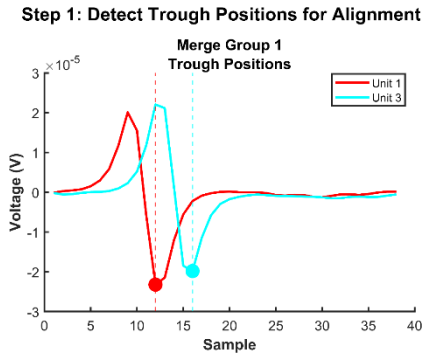


SM Figure S14. A. Final sorted waveforms after DTW merging. Two units remain. **B.** PCA after DTW merging. The merged Unit 1 (red) now spans both of the previously separated left-side clusters.

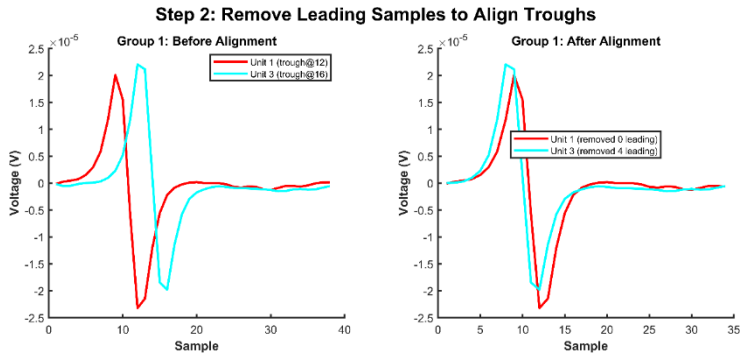
S2.4 Spike Alignment for Visualization

Once merge groups are identified, we align the constituent spikes so that their peaks/troughs coincide. This step is purely for visualization purposes—the cluster assignments have already been determined and are not recalculated. Within each merge group, we identify the cluster whose mean template has the earliest-occurring peak. Spikes from other clusters in the group are then shifted by removing samples from the front of each waveform until their peaks align with the reference. To maintain uniform dimensionality for PCA, we also remove from the end of all waveforms so that every spike ends up with the same length.

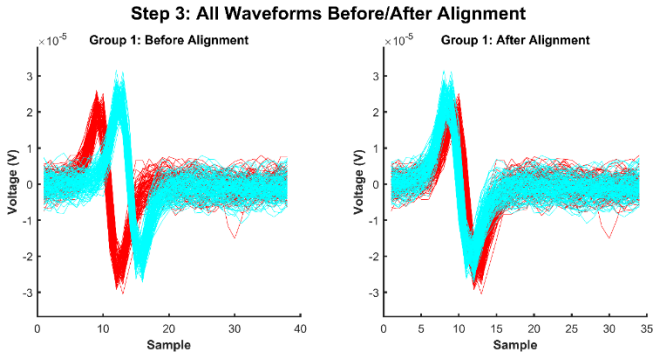
SM Figure S15 shows trough detection for the merge group. Unit 1's template has its trough at sample 12, while Unit 3's occurs at sample 16—a 4-sample lag. Taking Unit 1 as the reference, we remove 4 samples from the front of Unit 3's waveforms. **SM Figure S16** shows templates before and after alignment.



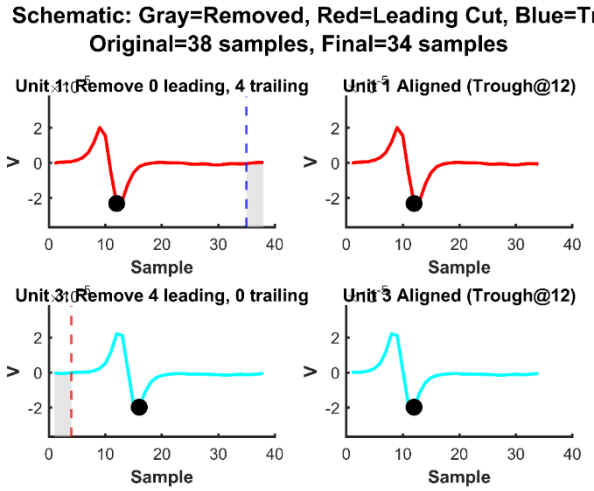
SM Figure S15. Peak detection for alignment. Dashed lines mark the trough positions: sample 12 for Unit 1 (red) and sample 16 for Unit 3 (cyan).



SM Figure S16. Mean templates before (left) and after (right) alignment. Peaks now coincide.

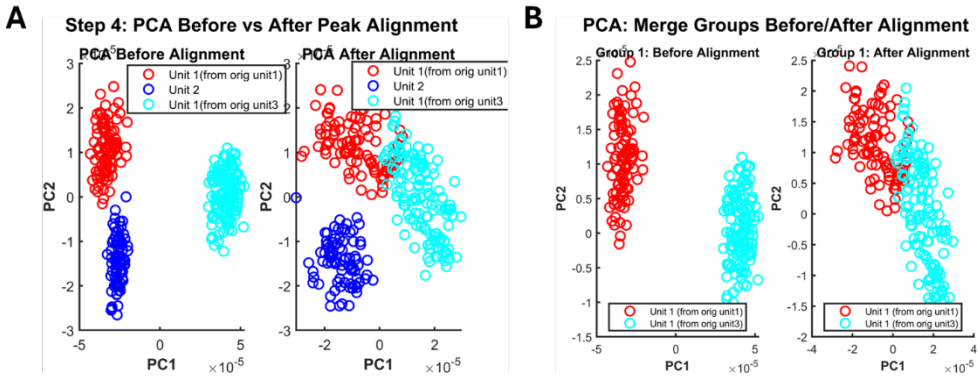


SM Figure 17. All waveforms in the merge group before (left) and after (right) alignment.



SM Figure S18. Alignment schematic showing trimmed regions. Original waveforms were 38 samples; after alignment, all are 34 samples.

The alignment step improves visualization by removing the artificial separation caused by temporal offset. **SM Figure 19** compares the PCA before and after alignment for all three original units. Before alignment, original Units 1 (red) and 3 (cyan) form distinct clusters despite belonging to the same neuron. After alignment, they overlap substantially, confirming that the original separation was driven by timing differences rather than genuine waveform dissimilarity.

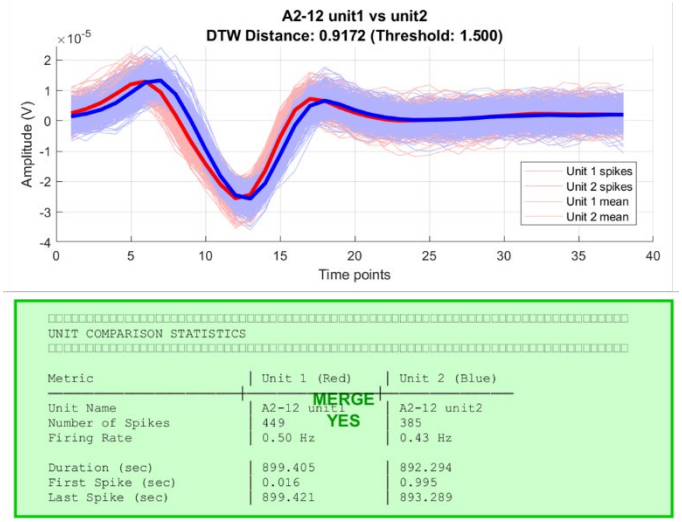


SM Figure 19 A. PCA of all units before (left) and after (right) spike alignment. **B.** PCA of the merge group only. Before alignment (left), the two clusters are clearly separated. After alignment (right), they largely overlap.

S2.5 Additional Merge Decision Examples

To further illustrate the DTW merge decision process, **SM Figures 20–23** present additional electrode examples showing both merge and no-merge outcomes. Each panel

displays the waveform comparison with individual spikes (thin lines) and mean templates (thick lines), along with unit statistics. Green borders indicate merged units (DTW distance below threshold); red borders indicate units kept separate.

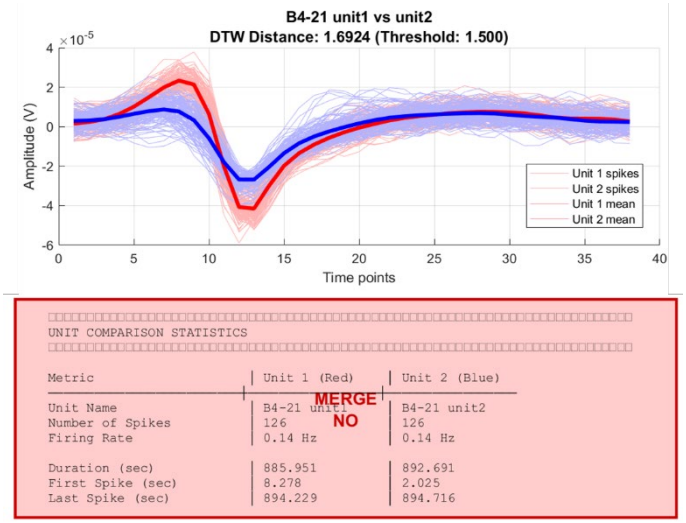


UNIT COMPARISON STATISTICS

Metric	Unit 1 (Red)	Unit 2 (Blue)
Unit Name	A2-12 unit1	A2-12 unit2
Number of Spikes	449	385
Firing Rate	0.50 Hz	0.43 Hz
Duration (sec)	899.405	892.294
First Spike (sec)	0.016	0.995
Last Spike (sec)	899.421	893.289

MERGE
YES

SM Figure 20. A2-12 unit1 vs unit2. DTW distance 0.9172 (below threshold 1.5). The highly similar waveform shapes result in a MERGE decision.

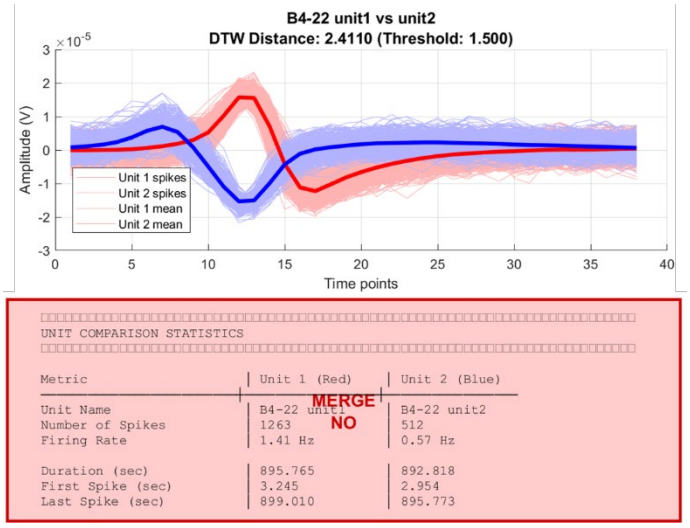


UNIT COMPARISON STATISTICS

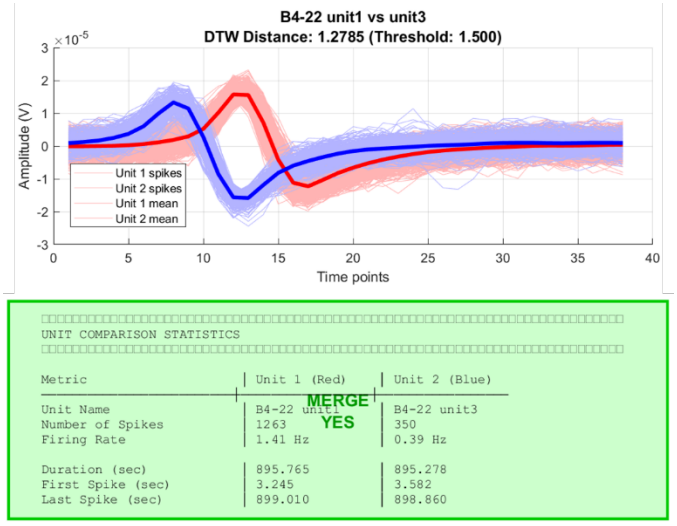
Metric	Unit 1 (Red)	Unit 2 (Blue)
Unit Name	B4-21 unit1	B4-21 unit2
Number of Spikes	126	126
Firing Rate	0.14 Hz	0.14 Hz
Duration (sec)	885.951	892.691
First Spike (sec)	8.278	2.025
Last Spike (sec)	894.229	894.716

MERGE
NO

SM Figure 21. B4-21 unit1 vs unit2. DTW distance 1.6924 (above threshold 1.5). Despite similar overall shape, the amplitude difference results in NO MERGE.

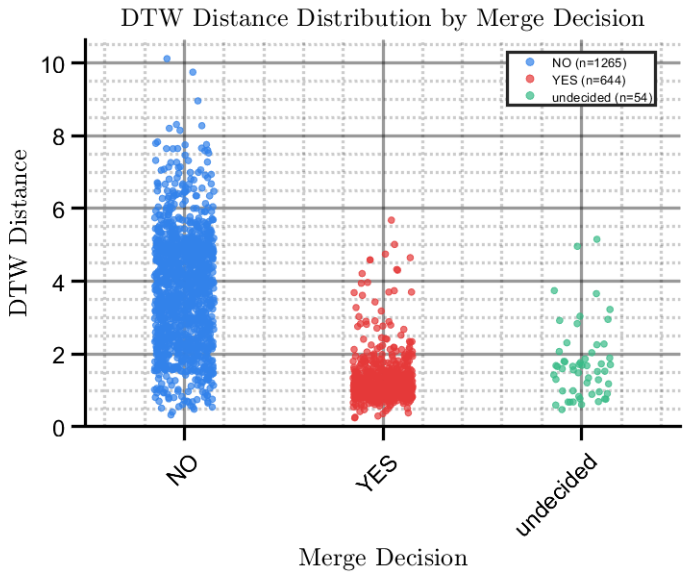


SM Figure 22. B4-22 unit1 vs unit2. DTW distance 2.4110 (above threshold 1.5). Distinct waveform morphologies with different peak timing result in NO MERGE.



SM Figure 23. B4-22 unit1 vs unit3. DTW distance 1.2785 (below threshold 1.5). Similar waveform shapes lead to a MERGE decision.

SM Figure 24 summarizes DTW distances across all unit pair comparisons, showing clear separation between merge decisions (clustered at low DTW values) and no-merge decisions (spanning higher values).



SM Figure 24. DTW distance distribution by merge decision. Blue (n=1265): NO merge; red (n=644): YES merge; green (n=54): undecided.

S3. Parameter Summary

SM Tables 1&2 summarize the parameters used in the HDT and DTW procedures, respectively. These values were selected based on empirical testing across recordings from multiple wells and plates.

SM Table 1. Parameters for Hartigan's Dip Test under-sorting correction.

Parameter	Value
HDT significance threshold	$p < 0.05$
Number of bootstrap replicates	500
Optimal K range	2-5
Overlap threshold for rejecting split	70%
Principal components tested	PC1 and PC2

SM Table 2. Parameters for DTW-based template merging and spike alignment.

Parameter	Value
DTW distance threshold	1.5
Amplitude tolerance	30% of min($height_1$, $height_2$)
Maximum alignment trim	10 samples
Minimum waveform length after trim	20 samples

S4. Summary

The two refinement procedures described here—HDT for under-sorting and DTW for over-sorting—address complementary failure modes in spike sorting. HDT catches cases where initial clustering groups together spikes from multiple neurons by testing for multimodality in the PC distributions and splitting when warranted. DTW catches cases where temporal misalignment (typically from dual-threshold detection) causes a single neuron's spikes to be scattered across multiple clusters. Together, these corrections help ensure that final unit counts more accurately reflect the true number of neurons contributing to the recorded signal.

SAMS – Quality Control and Validation

A. Outlier Detection

Within each sorted unit, waveform consistency is assessed by calculating root mean square error (RMSE) relative to the mean template. Waveforms exceeding three standard deviations are excluded as outliers:

$Outlier\ if: |\ RMSE\_i - \mu\_RMSE\ | > 3\sigma\_RMSE$

This approach removes contaminating spikes from noise, distant neurons, or movement artifacts while preserving natural waveform variability.

B. Refractory Period Violation

Physiological constraints are enforced by removing spikes violating the absolute refractory period (default: 1.5 ms). Inter-spike intervals are calculated for each unit, and violations are eliminated to ensure biological plausibility of sorted units.

C. Firing Rate Threshold

SAMS discards units with firing rates (fr) below a threshold θ: $fr < \theta$, specified by the user (here θ = 0.1 Hz) and have an inter-spike interval (ISI) below a threshold (here: 1.5 ms). This conservative approach helps maintain the integrity of the sorted data for further analyses (see **Figure 1G,H**).

SAMS – Criteria for Flagging of Potentially Missorted Units

SAMS implements two categories of flagged electrodes: Possible Multi-Unit (PMU) and Over-Excluded Unit (OEU). These flags help identify potential issues in spike sorting for further review and can help the user prioritize which electrodes to correct after automatic sorting.

A. Possible Multi-Unit (PMU) Flagging

An electrode may be flagged as PMU under two conditions:

- Hartigan's Dip Test Failure: If an electrode fails the Hartigan's dip test, indicating a single unit may actually be two units, we examine the overlap of the waveform ranges. If the overlap exceeds 70%, the original unit is not split, but the electrode is flagged as PMU.
- Dynamic Time Warping (DTW) Distance: We calculate the DTW distance between aligned elements of two waveforms. Units with DTW distances below a specified threshold (default 1.5) are flagged as PMU. This threshold was determined through expert analysis of thousands of electrodes across multiple files, where a DTW distance around 1.5 typically indicated over-classification (see **Fig. S3I**).

B. Over-Excluded Unit (OEU) Flagging

An electrode is flagged as OEU if the total number of excluded waveforms during the invalid waveform exclusion procedure exceeds a certain threshold (default: 10% of total waveforms).

These flags direct users to the subset of electrodes (typically 10-20%) most likely to benefit from manual curation, enabling practical validation of large-scale MEA experiments. Both the DTW distance threshold for PMU and the exclusion threshold for OEU can be adjusted by users in the SAMS interface to suit specific experimental needs (**Figure 2**).

Supplemental References

1. Guo, Y., Shen, M., Dong, Q., Méndez-Albelo, N.M., Huang, S.X., Sirois, C.L., Le, J., Li, M., Jarzembowski, E.D., Schoeller, K.A., et al. (2023). Elevated levels of FMRP-target MAP1B impair human and mouse neuronal development and mouse social behaviors via autophagy pathway. Nat Commun 14, 3801. 10.1038/s41467-023-39337-0.
2. Shen, M., Sirois, C.L., Guo, Y., Li, M., Dong, Q., Méndez-Albelo, N.M., Gao, Y., Khullar, S., Kissel, L., Sandoval, S.O., et al. (2023). Species-specific FMRP regulation of RACK1 is critical for prenatal cortical development. Neuron 111, 3988-4005.e3911. 10.1016/j.neuron.2023.09.014.
3. Rossant C, Kadir SN, Goodman DFM, Schulman J, *et al.* Spike sorting for large, dense electrode arrays. Nat Neurosci. 2016 Apr;19(4):634-641. doi: 10.1038/nn.4268. Epub 2016 Mar 14. PMID: 26974951; PMCID: PMC4817237.
4. Chung JE, Magland JF, Barnett AH, Tolosa VM, *et al*. A Fully Automated Approach to Spike Sorting. Neuron. 2017 Sep 13;95(6):1381-1394.e6. doi: 10.1016/j.neuron.2017.08.030. PMID: 28910621; PMCID: PMC5743236.
5. Jun JJ, Mitelut C, Lai C, Gratiy SL, *et al.* Real-time spike sorting platform for high-density extracellular probes with ground-truth validation and drift correction. bioRxiv 101030; doi: https://doi.org/10.1101/101030
6. Magland J, Jun JJ, Lovero E, Morley AJ, *et al.* SpikeForest, reproducible web-facing ground-truth validation of automated neural spike sorters. Elife. 2020 May 19;9:e55167. doi: 10.7554/eLife.55167. PMID: 32427564; PMCID: PMC7237210.
7. Quiroga RQ, Nadasdy Z, Ben-Shaul Y. Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. Neural Comput. 2004 Aug;16(8):1661-87. doi: 10.1162/089976604774201631. PMID: 15228749.
8. Chaure FJ, Rey HG, Quian Quiroga R. A novel and fully automatic spike-sorting implementation with variable number of features. J Neurophysiol. 2018 Oct 1;120(4):1859-1871. doi: 10.1152/jn.00339.2018. Epub 2018 Jul 11. PMID: 29995603; PMCID: PMC6230803.
9. Lee J, Mitelut C, Shokri H, Kinsella I, *et al*. YASS: Yet Another Spike Sorter applied to large-scale multi-electrode array recordings in primate retina. bioRxiv 2020.03.18.997924; doi: https://doi.org/10.1101/2020.03.18.997924
10. Chen K, Jiang Y, Wu Z, Zheng N, *et al*. HTsort: Enabling Fast and Accurate Spike Sorting on Multi-Electrode Arrays. Front Comput Neurosci. 2021 Jun 21;15:657151. doi: 10.3389/fncom.2021.657151. PMID: 34234663; PMCID: PMC8255361.
11. Pachitariu M, Sridhar S, Pennington J, Stringer C. Spike sorting with Kilosort4. Nat Methods. 2024 May;21(5):914-921. doi: 10.1038/s41592-024-02232-7. Epub 2024 Apr 8. PMID: 38589517; PMCID: PMC11093732.
12. Mohammadi Z, Denman DJ, Klug A, Lei TC. A fully automatic multichannel neural spike sorting algorithm with spike reduction and positional feature. J Neural Eng. 2024 Aug 5;21(4):046039. doi: 10.1088/1741-2552/ad647d. PMID: 39019065; PMCID: PMC11298775.
13. Rueda C, Rodríguez-Collado A. Functional clustering of neuronal signals with FMM mixture models. Heliyon. 2023 Oct 10;9(10):e20639. doi: 10.1016/j.heliyon.2023.e20639. PMID: 37867904; PMCID: PMC10589779.
14. Wang R, Xu Y, Zhang Y, Hu X, *et al*. A Fast and Effective Spike Sorting Method Based on Multi-Frequency Composite Waveform Shapes. Brain Sci. 2023 Aug 2;13(8):1156. doi: 10.3390/brainsci13081156. PMID: 37626512; PMCID: PMC10452241.

15. Hall NJ, Herzfeld DJ, Lisberger SG. Evaluation and resolution of many challenges of neural spike sorting: a new sorter. J Neurophysiol. 2021 Dec 1;126(6):2065-2090. doi: 10.1152/jn.00047.2021. Epub 2021 Nov 17. PMID: 34788137; PMCID: PMC8715051.
16. Yger P, Spampinato GL, Esposito E, Lefebvre B, *et al*. A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. Elife. 2018 Mar 20;7:e34518. doi: 10.7554/eLife.34518. PMID: 29557782; PMCID: PMC5897014.
17. Hilgen G, Sorbaro M, Pirmoradian S, Muthmann JO, *et al*. Unsupervised Spike Sorting for Large-Scale, High-Density Multielectrode Arrays. Cell Rep. 2017 Mar 7;18(10):2521-2532. doi: 10.1016/j.celrep.2017.02.038. PMID: 28273464.
18. Diggelmann R, Fiscella M, Hierlemann A, Franke F. Automatic spike sorting for high-density microelectrode arrays. J Neurophysiol. 2018 Dec 1;120(6):3155-3171. doi: 10.1152/jn.00803.2017. Epub 2018 Sep 12. PMID: 30207864; PMCID: PMC6314465.
19. Negri J, Menon V, Young-Pearse TL. Assessment of Spontaneous Neuronal Activity In Vitro Using Multi-Well Multi-Electrode Arrays: Implications for Assay Development. eNeuro. 2020 Jan 24;7(1):ENEURO.0080-19.2019. doi: 10.1523/ENEURO.0080-19.2019. PMID: 31896559; PMCID: PMC6984810.
20. Hu M, Frega M, Tolner EA, van den Maagdenberg AMJM, *et al*. MEA-ToolBox: an Open Source Toolbox for Standardized Analysis of Multi-Electrode Array Data. Neuroinformatics. 2022 Oct;20(4):1077-1092. doi: 10.1007/s12021-022-09591-6. Epub 2022 Jun 9. PMID: 35680724; PMCID: PMC9588481.
21. Heiney K, Mateus JC, Lopes CDF, Neto E, *et al*. µSpikeHunter: An advanced computational tool for the analysis of neuronal communication and action potential propagation in microfluidic platforms. Sci Rep. 2019 Apr 8;9(1):5777. doi: 10.1038/s41598-019-42148-3. PMID: 30962522; PMCID: PMC6453950.
22. Zhang X, Dou Z, Kim SH, Upadhyay G, *et al*. Mind In Vitro Platforms: Versatile, Scalable, Robust, and Open Solutions to Interfacing with Living Neurons. Adv Sci (Weinh). 2024 Mar;11(11):e2306826. doi: 10.1002/advs.202306826. Epub 2023 Dec 31. PMID: 38161217; PMCID: PMC10953569.
23. Zanini G, Parodi G, Chiappalone M, Martinoia S. Investigating the reliability of the evoked response in human iPSCs-derived neuronal networks coupled to micro-electrode arrays. APL Bioeng. 2023 Dec 20;7(4):046121. doi: 10.1063/5.0174227. PMID: 38130601; PMCID: PMC10735322.
24. Parodi G, Brofiga M, Pastore VP, Chiappalone M, *et al*. Deepening the role of excitation/inhibition balance in human iPSCs-derived neuronal networks coupled to MEAs during long-term development. J Neural Eng. 2023 Sep 19;20(5). doi: 10.1088/1741-2552/acf78b. PMID: 37678214.
25. Lemieux MR, Freigassner B, Hanson JL, Thathey Z, *et al*. Multielectrode array characterization of human induced pluripotent stem cell derived neurons in co-culture with primary human astrocytes. PLoS One. 2024 Jun 25;19(6):e0303901. doi: 10.1371/journal.pone.0303901. PMID: 38917115; PMCID: PMC11198861.
26. Hruska-Plochan M, Wiersma VI, Betz KM, Mallona I, *et al*. A model of human neural networks reveals NPTX2 pathology in ALS and FTLD. Nature. 2024 Feb;626(8001):1073-1083. doi: 10.1038/s41586-024-07042-7. Epub 2024 Feb 14. PMID: 38355792; PMCID: PMC10901740.
27. Mossink B, Verboven AHA, van Hugte EJH, Klein Gunnewiek TM, *et al*. Human neuronal networks on micro-electrode arrays are a highly robust tool to study disease-specific genotype-phenotype correlations in vitro. Stem Cell Reports. 2021 Sep 14;16(9):2182-2196.

doi: 10.1016/j.stemcr.2021.07.001. Epub 2021 Jul 29. PMID: 34329594; PMCID: PMC8452490.
28. Mossink B, van Rhijn JR, Wang S, Linda K, *et al*. Cadherin-13 is a critical regulator of GABAergic modulation in human stem-cell-derived neuronal networks. Mol Psychiatry. 2022 Jan;27(1):1-18. doi: 10.1038/s41380-021-01117-x. Epub 2021 May 10. PMID: 33972691; PMCID: PMC8960401.
29. Hornauer P, Prack G, Anastasi N, Ronchi S, *et al*. DeePhys: A machine learning-assisted platform for electrophysiological phenotyping of human neuronal networks. Stem Cell Reports. 2024 Feb 13;19(2):285-298. doi: 10.1016/j.stemcr.2023.12.008. Epub 2024 Jan 25. PMID: 38278155; PMCID: PMC10874850.
30. Cotterill E, Charlesworth P, Thomas CW, Paulsen O, *et al*. A comparison of computational methods for detecting bursts in neuronal spike trains and their application to human stem cell-derived neuronal networks. J Neurophysiol. 2016 Aug 1;116(2):306-21. doi: 10.1152/jn.00093.2016. Epub 2016 Apr 20. PMID: 27098024; PMCID: PMC4969396.
31. Muzzi L, Di Lisa D, Arnaldi P, Aprile D, *et al*. Rapid generation of functional engineered 3D human neuronal assemblies: network dynamics evaluated by micro-electrodes arrays. J Neural Eng. 2021 Dec 10;18(6). doi: 10.1088/1741-2552/ac3e02. PMID: 34844234.
32. Afshar-Saber, W., Teaney, N.A., Winden, K.D., Jumo, H., Shi, X., McGinty, G., Hubbs, J., Chen, C., Tokatly Latzer, I., Gasparoli, F., et al. (2024). ALDH5A1-deficient iPSC-derived excitatory and inhibitory neurons display cell type specific alterations. Neurobiol Dis *190*, 106386. 10.1016/j.nbd.2023.106386.
33. Afshar-Saber, W., Chen, C., Teaney, N.A., Kim, K., Yang, Z., Gasparoli, F.M., Ebrahimi-Fakhari, D., Buttermore, E.D., Pin-Fang Chen, I., Pearl, P.L., and Sahin, M. (2024). Generation and characterization of six human induced pluripotent stem cell lines (hiPSCs) from three individuals with SSADH Deficiency and CRISPR-corrected isogenic controls. Stem Cell Res 77, 103424. 10.1016/j.scr.2024.103424.
34. Raykov, Y.P., Boukouvalas, A., Baig, F., and Little, M.A. (2016). What to Do When K-Means Clustering Fails: A Simple yet Principled Alternative Algorithm. PLoS One 11, e0162259. 10.1371/journal.pone.0162259.
35. Hartigan, J.A., and Hartigan, P.M. (1985). The Dip Test of Unimodality. The Annals of Statistics *13*, 70-84, 15.
36. Cao, Y., Rakhilin, N., Gordon, P.H., Shen, X., and Kan, E.C. (2016). A real-time spike classification method based on dynamic time warping for extracellular enteric neural recording with large waveform variability. J Neurosci Methods *261*, 97-109. 10.1016/j.jneumeth.2015.12.006.
37. Wang, R., Xu, Y., Zhang, Y., Hu, X., Li, Y., and Zhang, S. (2023). A Fast and Effective Spike Sorting Method Based on Multi-Frequency Composite Waveform Shapes. Brain Sci 13. 10.3390/brainsci13081156.