

Novel Tools and Methods

Assessment of Spontaneous Neuronal Activity *In Vitro* Using Multi-Well Multi-Electrode Arrays: Implications for Assay Development

Joseph Negri,^{1,2} Vilas Menon,³ and  Tracy L. Young-Pearse¹<https://doi.org/10.1523/ENEURO.0080-19.2019>

¹Ann Romney Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, ²Graduate Program in Biological and Biomedical Sciences, Division of Medical Sciences, Harvard University, Cambridge, MA 02138, and ³Department of Neurology, Columbia University Medical Center, New York, NY 10032

Abstract

Multi-electrode arrays (MEAs) are being more widely used by researchers as an instrument platform for monitoring prolonged, non-destructive recordings of spontaneously firing neurons *in vitro* for applications in modeling Alzheimer's, Parkinson's, schizophrenia, and many other diseases of the human CNS. With the more widespread use of these instruments, there is a need to examine the prior art of studies utilizing MEAs and delineate best practices for data acquisition and analysis to avoid errors in interpretation of the resultant data. Using a dataset of recordings from primary rat (*Rattus norvegicus*) cortical cultures, methods and statistical power for discerning changes in neuronal activity on the array level are examined. Further, a method for unsupervised spike sorting is implemented, allowing for the resolution of action potential incidents down to the single neuron level. Following implementation of spike sorting, the dynamics of firing frequency across populations of individual neurons and networks are examined longitudinally. Finally, the ability to detect a frequency independent phenotype, the change in action potential amplitude, is demonstrated through the use of pore-forming neurotoxin treatments. Taken together, this study provides guidance and tools for users wishing to incorporate multi-well MEA usage into their studies.

Key words: multi-electrode arrays; spike sorting

Significance Statement

Multi-electrode arrays (MEA) are an instrument platform being used by an increasing number of neuroscientists for the purpose of monitoring spontaneous firing of neurons *in vitro* over extended periods of time. Through an analysis of existing literature and empirically generated datasets, this study seeks to establish best practices for the use of these instruments for applications employing neuronal cultures. Elements of experimental design and analysis for assaying firing frequencies across arrays are discussed, with a focus on the use of multi-well MEAs. Additionally, methods for (1) resolving signal to individual neurons through unsupervised spike-sorting, (2) assessing network dynamics, and (3) quantifying changing in action potential amplitudes are reported.

Introduction

Among the conditions posing the greatest unmet medical need are diseases of the CNS. Confounding efforts

toward the development of therapies, these conditions occur within an organ system that is largely inaccessible to direct experimentation and thus studies must be per-

Received March 6, 2019; accepted November 6, 2019; First published January 2, 2020.

The authors declare no competing financial interests.

Author contributions: J.N. and T.L.Y.-P. designed research; J.N. performed research; J.N. analyzed data; J.N. and T.L.Y.-P. wrote the paper; V.M. contributed unpublished reagents/analytic tools.

formed using experimental models. Attempting to study any disease or biological process in model systems always comes with inherent compromises. The system must be sufficiently complex to mimic the multiple facets of the *in vivo* physiology. Yet the system must be sufficiently reductionist to: (1) have clearly defined end-points, (2) be executable within reasonable timeframes, and (3) allow for parallel testing of multiple perturbations. Attempts to model these complex systems has motivated the field of therapeutic discovery to increasingly adopt representative phenotypic assays that: (1) use the most relevant cell type, (2) employ a disease relevant perturbation, (3) and rely on an assay metric that is representative of the disease symptomatology (Vincent et al., 2015).

An assay platform for the study of the CNS and associated disorders must fulfill several criteria. It must be capable of capturing neuronal activity and do so within dense, mixed cultures of neurons and glia. The method should generate a rich dataset to allow for the detection of modest changes in phenotypes, particularly when perturbed at physiologically relevant magnitudes which may result in only small yet meaningful responses. Moreover, the method should be amenable to incorporate human cellular models, drawing on advancements in differentiation efficiency of human induced pluripotent stem cell (iPS) derived neurons and glia. While numerous assay platforms are employed in the field of neuroscience for assaying neuronal behavior, one which is capable of fulfilling all of these criteria are multi-electrode arrays (MEAs).

MEAs are instruments that provide a means of monitoring spontaneous electrophysiological activity within *in vitro* neuronal cultures. MEAs consist of a large number (dozens to hundreds) of planar electrodes embedded in the base of a tissue culture chamber that allow for the parallel detection of local field potentials generated by the spontaneous or evoked firing of neurons. MEAs sample the potential difference across recording and reference electrodes at high rates (10–60 kHz), and action potentials (spikes) are detected when the sampled values deviate substantially from the background potential. Activity within cultures is quantified by the frequency of spiking events, and in some applications the occurrence of synchronized spiking events, or “bursts” (Spira and Hai, 2013; Obien et al., 2014).

By monitoring endogenously generated voltage potentials of firing neurons, these instruments do not rely on fluorescent dyes or proteins and microscopy apparatus

required of calcium-imaging techniques. Additionally, MEAs are able to record simultaneously across upwards of hundreds of channels in a much less labor-intensive manner than conventional single-channel electrophysiological techniques such as patch-clamping or sharp-electrode recording, albeit with a lower degree of spatial resolution. As MEA recordings are non-destructive to the cultures, repeated recordings may be performed for as long as culture integrity can be maintained. For these reasons, MEA use is becoming more common in neuroscience and biomedical research.

Despite the long history and increased use of MEAs, methodologies are still being developed across the field. To date, the most comprehensive methodological review of *in vitro* MEA techniques was performed by Novellino et al. (2011). Published in 2011, that study assessed the use of single-well MEAs, single chamber vessels typically with ~60 recording electrodes. Over the last decade, several multi-well MEAs systems have become commercially available, including those from Axion Biosystems, Alpha Med Scientific, Multichannel Systems, and Maxwell Biosystems. These multi-well MEAs adhere to the form factors of tissue culture plates, with a separate array occupying each well (we will subsequently equate a single well of electrodes as an array). The availability of multi-well MEAs has allowed for changes in experimental design and analysis, with the capability to test more conditions in parallel and use array-level (as opposed to electrode-level) activity metrics.

Accounting for these developments, this study sought to extend best practices for multi-well MEA experiments. Following a review of the literature, commonly reported experimental conditions and spike frequency analysis methods were evaluated through an empirically generated dataset. These data were leveraged to develop a method of cohort assignment for multi-well MEA experiments that accounts for inherent variability in firing frequencies. Further, implementation of unsupervised spike sorting to resolve activity to the level individual neurons (single-units) demonstrates how MEA recording data may be extended for examination of wave form amplitude and functional network phenotypes.

Materials and Methods

Tissue culture

Preparation of MEA plates

We prepared 96-well multi-electrode plates (Axion Biosystems) before the addition of cells by coating with poly-Ornithine (poly-O), laminin (both Sigma-Aldrich), and Matrigel (Corning Life Sciences). Briefly, 1 d before establishment of cultures, a solution of 20 $\mu\text{g/ml}$ poly-O: 5 $\mu\text{g/ml}$ laminin in $1\times$ PBS was added to the MEA plates in a volume of 60 μl per well, and incubated overnight at 37°C. The day of culture plating, the poly-O:laminin solution was aspirated and the plates were washed once with $1\times$ PBS. Matrigel was reconstituted 1:20 in ice-cold DMEM culture media (Gibco; ThermoFisher) without added serum or antibiotics and passed through a 40- μm strainer (BD Falcon). The Matrigel solution was added to

This work is supported by NIH Grants R01AG055909 (to T.L.Y.-P.) and R01MH101148 (to T.L.Y.-P.), the Ellison Foundation, and the Head Foundation. J.N. was supported by the National Science Foundation Graduate Research Fellowship Program.

Acknowledgements: We thank Dr. Richard Pearse for critical reading of this manuscript.

Correspondence should be addressed to Tracy L. Young-Pearse at tyoung@rics.bwh.harvard.edu.

<https://doi.org/10.1523/ENEURO.0080-19.2019>

Copyright © 2020 Negri et al.

This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

the MEA plates in a volume of 60 μl per well, and incubated for a minimum of 1 h at 37°C.

Preparation of cortical cultures

Primary cortical cultures were established from E18 Sprague Dawley rats (*Rattus norvegicus*) using individuals of either sex (Charles River Laboratories, Crl:SD, RRID: RGD_737891). Dams were killed by CO₂ euthanasia under a protocol approved by IACUC of Brigham and Women's Hospital. Dissection of complete cortex from pups was performed in ice-cold HBSS (Gibco) under a dissecting microscope (Zeiss). The dissected cortices were suspended in 0.25% trypsin-EDTA (Gibco) for 10 min at 37°C, the excess trypsin-EDTA solution was then aspirated. The tissue was then triturated in Complete DMEM [DMEM culture media (Gibco) supplement with 5% fetal calf serum (Lonza) and 1 \times penicillin/streptomycin/L-glutamine (Gibco)] with a 10-ml serological pipette before being passed through a 100- μm strainer (BD Falcon). Counts of the cell suspension were taken in triplicate, and the cell suspension was back diluted to 1.5 \times 10⁶ cells/ml in Complete DMEM. The cell suspension was added to the MEA plates in a volume of 50 μl for a plating concentration of 7.5 \times 10⁴ cells/well or 2.4 \times 10⁵ cells/cm². Plates were then placed in a tissue culture incubator (37°C, 95% humidity, 5% CO₂) for 4 h to allow for cells to attach to the culture surface. After the 4-h incubation, 150 μl of BrainPhys Media (1 \times BrainPhys culture media containing SM-1 neuronal supplement; StemCell Technologies) was added to each well. The cultures were maintained in a tissue culture incubator in BrainPhys Media, with semi-weekly half-volume media changes.

Cell viability assay

Cell viability was assessed using the Pierce LDH Cytotoxicity assay kit (ThermoFisher Scientific) according to manufacturer's protocol. Briefly, 50 μl of treated culture media along with 50 μl of detection reagent were combined in a clear 96-well microtitre plate, and incubated at room temperature for 30 min. Following incubation, the stop reagent was added, and optical density (OD) was measured at 490 and 680 nm using a Synergy HT plate reader (BioTek). Percentage viability was calculated by first subtracting the OD₆₈₀ from the OD_{490nm}, then normalizing each observation to the median values of the cell lysis positive control (PC) and the untreated negative control (NC) using the equation:

$$N(x) = \left(1 - \frac{OD_x - \overline{OD_{NC}}}{\overline{OD_{PC}} - \overline{OD_{NC}}}\right) \times 100$$

MEA recording and analysis

All MEA recordings were performed using a Maestro multi-well MEA recorder (Axion Biosystems). During recordings, plates were kept on a heated stage maintained at 37°C and ventilated with a mixture of 5% CO₂:95% air (AirGas) at a rate of 1 cubic-foot/h. to prevent evaporation of liquid within wells by convection and condensation on the underside of the plate lid, the MEA plate was covered with an air-activated oxidizing iron heater (HotHands, Ko-

bayashi) placed on top of an aluminum plate cut to size, for even dispersal of heat. Voltage potentials within wells were simultaneously recorded across 768 channels (eight electrodes per well, 96-well plate) at a sampling frequency of 12.5 kHz using AxIS acquisition software version 2.4 (Axion Biosystems). This sampling rate of 12.5 kHz was chosen as it represents the maximum rate of the Axion Maestro instrument used. Note that sampling rates upwards of 20–50 kHz are achievable with other commercially available multi-well MEA systems. The raw voltage recordings were subjected to a Butterworth filter of 200 Hz to 2.5 kHz, and neuronal firing events (spikes) were detected when the voltage exceeded a “crossing threshold” set at 5.5 SD away from the root mean square (RMS) of the background potential calculated over a 10-ms moving window. All recordings were performed for 30 min unless otherwise specified.

Raw voltage, timestamp, value of crossing threshold for each spike event were extracted from the .spk files of MEA recordings produced AxIS acquisition software, using custom MATLAB scripts (MathWorks) using extractor functions provided with AxIS version 2.4. Following extraction of the raw recording data, all analyses and simulations were performed using the R statistical programming language (R Core Team, 2017).

Mean firing rate (MFR) calculation

To remove spurious spike events arising from by “high-noise” electrodes, an upper limit to the crossing threshold was established by examining the crossing threshold (μV) for all spike events detected and calculated the value corresponding to 3 SD greater than the mean crossing threshold, all events detected at a crossing threshold greater than this upper limit were excluded from the analysis. The MFR (Hz) was calculated as the ratio of the total number of spikes recorded (n), and the duration of recording in seconds, $MFR = n/s$. The log transform of the MFR was calculated as $\log_{10}(n + 1/s)$, to account for instances of $n = 0$, the log of which is undefined.

Treatment group assignment

A pool of active arrays from a multi-well MEAs was established by selecting those arrays that are no >2 SD below the median firing frequency ($\log_{10}\text{Hz}$) of the sample set. A panel of i possible treatment assignments is generated by randomly assigning arrays to treatment groups g , each with n members. For the purposes of this study, $i = 10^4$. For each instance of i , a one-way ANOVA was performed, assessing $\log_{10}\text{Hz}$ as a function of g . The instance of i resulting in the lowest value of the F-statistic, was used as the treatment group assignment.

Neurotoxin treatments

α -Hemolysin (αHL) and tetrodotoxin (TTX) were purchased from Sigma-Aldrich. Both reagents were reconstituted in sterile water, and diluted to the specified concentration in BrainPhys culture media containing SM-1 neuronal supplement (StemCell Technologies); 30-min baseline recordings of mature [more than day *in vitro* (DIV)21] rat primary cortical cultures were performed, and arrays were assigned to cohorts with comparable activity (six to eight replicates per group) using the technique

described above. Cultures were treated with a titration of TTX (0.001–1 μM), and recorded again for 30 min both immediately following treatment and 24 h following treatment. For array-level spike frequency analysis the magnitude of effect of each condition was determined from the coefficients of a linear mix-effect model, while for cluster-level spike frequency analysis, the magnitude of effect of each condition was determined from the coefficients of Γ -distributed generalized linear model (GLM).

Spike sorting

For each spike event, the following metrics were calculated based on the vector of 38 momentary voltage measurements within each instance: maximum voltage (peak), minimum voltage (valley), wave form amplitude (peak-valley range), time interval between peak and valley, area under the curve (AUC), and non-linear energy (NLE). AUC was calculated using the auc function within the R MESS library (Ekström, 2018), NLE was calculated using the equation provided by Kim and Kim (2000):

$$\Psi(x_n) = x_n^2 - (x_{n-1} \times x_{n+1}).$$

The data were aggregated across all recordings in a given experiment, and segmented based on individual electrode. Principal components analysis was performed using the prcomp function within the base R stats library, following log transformation of the metrics as recommended by Venables and Ripley (2002). Mean-shift clustering of spike events was performed based on the values of the first two principal components using the msClustering function within the R MeanShift library (Ciollaro and Wang, 2016), using a Gaussian kernel and bandwidth value of $h = 1.5$.

Network analysis

Spike clusters identified by unsupervised spike sorting analysis were deemed to represent individual neurons. Functional connections between neurons were established based on the spike time tiling coefficient (STTC) calculated between the spike trains of events attributed to each individual. STTC was calculated by the method reported by Cutts and Eglén (2014). Where between two spike trains A and B ,

$$\text{STTC} = \frac{1}{2} \left(\frac{P_A - T_B}{1 - P_A T_B} + \frac{P_B - T_A}{1 - P_B T_A} \right),$$

with a Δt value of 100 ms. Observed STTC values were taken to represent functional connections if they below the 0.05% or above the 99.5% quantile of a null distribution STTC values derived from 1000 permutations of random spike trains with comparable numbers of events as A and B . The network cluster coefficient (\bar{C}) as reported by Watts and Strogatz (1998) was calculated within each array and recording using the using the graph from data frame and transitivity functions within the R igraph library (Csardi and Nepusz, 2006), with isolates treated as zeros.

All spike sorting, permutation, and simulation analyses were performed on the O₂ high performance computing cluster (Research Computing Group, Harvard Medical

School). All scripts for performing these analyses are available at <https://github.com/SubstantiaNegri/meaAnalysis>.

Statistical analyses

Power calculation

A dataset of 30-min recordings of 1272 unique, untreated MEAs was taken to represent spontaneous firing activity $\log_{10}\text{Hz}$ at time (t_0), this was defined as vector A . The correlation coefficient between the firing frequencies of a population MEAs recorded at separate times was estimated to be $\rho = 0.8$, based on the repeated recordings of multiple culture preparations at DIV20 and DIV21. To simulate the expected variance in firing frequencies between recordings, a second vector B was calculated by the formula:

$$B = \rho A + A^\perp \sqrt{1 - \rho^2},$$

where A^\perp represents the residuals of a linear regression between A and a sample of equal length drawn from a random normal distribution. The resulting value of B was examined to confirm that:

$$\text{cor}(\vec{A}, \vec{B}) = \rho,$$

B is then taken to represent spontaneous firing activity $\log_{10}\text{Hz}$ at time (t_1). For each simulated experiment, *control* and *treatment* groups were generated by drawing the paired $\log_{10}\text{Hz}_{t_0}$ and $\log_{10}\text{Hz}_{t_1}$ values for random arrays for sample sizes ranging 3–16. The $\log_{10}\text{Hz}_{t_1}$ values within the treatment group were offset by effect sizes ranging from 0.1 to 2.0 $\log_{10}\text{Hz}$. A total of 5000 iterations were performed for each sample size:effect size pairing, for a total of 1.4×10^6 total simulated treatments. For each iteration, an analysis of covariance (ANCOVA) was performed by fitting an ANOVA model to linear regression for $\log_{10}\text{Hz}$ as a function of group (control/treatment) and time (pre/post) allowing for interaction between the group and time variables. The coefficients for the model were compared using Tukey's test for honest significant difference, and the p value for the comparison of control versus treatment at time t_1 was extracted. The linear regression, fitting of the ANOVA, and Tukey's test were performed using the lm, aov, and TukeyHSD functions within the base R stats library. Power was calculated as the proportion of iterations within each sample size:effect size pairing for which the difference between control and treatment was calculated to have a $p < 0.05$.

Γ -distributed GLMs (Γ -GLMs)

For the distribution of spike cluster frequencies, Γ -GLMs were constructed using the glm function within the base R stats library. Firing frequency, Hz, was modeled as the dependent variable, while treatment and recording were used as interacting categorical independent variables. The inverse link function was applied for translating the model coefficients estimating β , the rate term of the Γ distribution, to the distribution average μ .

Linear mixed effect models

For estimating changes in array-level spike frequency and cluster-level wave form, amplitudes were constructed

Table 1. Considerations for MEA experimental design

Experimental conditions	Analysis parameters
Array surface preparation	Spike detection/event filtering
Culture media	Assay metric
Age of cultures	Array exclusion criteria
Recording duration	Treatment group assignment
Environmental control	Statistical analysis

using the lme function within the R nlme library (Pinheiro et al., 2018). In all cases, the magnitude of effect of each condition was determined by aggregating the data from a minimum of three separate experiments. For array-level spike frequency, $\log_{10}\text{Hz}$ was modeled as the dependent variable; treatment and recording (pre-treatment, post-treatment) were interacting categorical fix-effect variables; and each individual experiment as a random effect. For cluster-level wave form amplitude, $\log_{10}\mu\text{V}$ was modeled as the dependent variable; treatment and recording were interacting categorical fix-effect variables; and spike cluster, electrode, and experiment were nested random-effect variables.

General linear hypothesis tests

General linear hypothesis tests for comparing the predictions of the GLM and LMEs were performed using the `contrMat`, `glht` summary.`glht` functions within the R `multcomp` library (Hothorn et al., 2008). Adjustments to p-values for multiple comparisons was done using the default “single-step” method within `summary.glht`. All figures were generated using the R `ggplot2` and accompanying `ggpubr` libraries (Wickham, 2016; Kassambara, 2018).

Table 2. Comparison of MEA methodology

Reference	Instrument format	Culture model	Surface preparation	Culture media	Age (DIV)	Duration (min)	Detection threshold	Assay metric(s)	Statistical analysis
Varghese et al. (2010)	Single-well	Primary rat (E18)	DETA	NB	12–16	-	-	MFR	<i>t</i> test
Kuperstein et al. (2010)	Single-well	Primary mouse	-	-	8–10	45	5 SD	MFR	<i>t</i> test
McConnell et al. (2012)	Multi-well	Primary rat (P2)	Laminin	NB	12–22	33	8 SD	MFR, active channels	-
Frega et al. (2012)	Single-well	Primary rat (E18)	PDL+laminin	NB	21–28	20	-	MFR, burst rate	Kruskal–Wallis
Bateup et al. (2013a)	Single-well	Primary mouse	PDL+laminin	NB	-	2	9 μV	MFR	<i>t</i> test, ANOVA
Biffi et al. (2013)	Single-well	Primary mouse	PLL	NB	4–35	10	5X baseline	MFR, active channels	Kruskal–Wallis
Vincent et al. (2013)	Single-well	Primary rat (E18)	PEI	-	-	20	3 SD	MFR	Wilcoxon
Bateup et al. (2013b)	Single-well	Primary mouse	PDL+laminin	-	12–19	2	9 μV	MFR	<i>t</i> test, ANOVA
Wainger et al. (2015)	Single-well	Reprogram-fibroblast neurons	PDL+laminin	DMEM/F12	>28	-	10 μV	MFR	Mann–Whitney
Wainger et al. (2014)	Both	hiPS-derived neurons	PDL+laminin	-	>24	-	10 μV , 5.5 SD	MFR	Kruskal–Wallis, linear regression
Illes et al. (2014)	Single-well	Primary mouse	PDL+laminin	DMEM/F12	>21	-	6.2 SD of baseline	MFR	<i>t</i> test, ANOVA
Weir et al. (2014)	Single-well	Primary mouse (P1)	-	aCSF	7–21	-	7SD	MFR, burst rate, etc.	<i>t</i> test, ANOVA
Charkhkar et al. (2015)	Single-well	Primary mouse (E17)	PDL+laminin	DMEM	21	20	5 SD	MFR	<i>t</i> test, ANOVA
Vertkin et al. (2015)	Single-well	Primary mouse (P2)	-	-	14–21	-	-	MFR	<i>t</i> test, ANOVA
Bardy et al. (2015)	Multi-well	hiPS-derived neurons	Laminin	BP	2–21	10	6 SD	MFR	-
Slomowitz et al. (2015)	Single-well	Primary mouse (P2)	-	-	15–22	60	4–5 SD	MFR	-
Liu et al. (2017)	Single-well	Primary mouse	-	-	-	15	6 SD	MFR	<i>t</i> test
Alshawaf et al. (2018)	Single-well	hiESC-derived neurons	-	NB	12–56	5	5 SD	MFR, active channels	Kruskal–Wallis
Feng et al. (2018)	Multi-well	Primary mouse (P1)	PEI+laminin	NB	7–28	30	-	MFR, burst rate, etc.	<i>t</i> test
García-León et al. (2018)	Multi-well	hiPS-derived neurons	PEI+laminin	DMEM/F12	70	5	-	MFR, burst rate, etc.	<i>t</i> test
Black et al. (2018)	Multi-well	Primary adult mouse DRG	PEI+PDL+laminin	DMEM/F12	3–21	30	5.5 SD	MFR, active channels	Mann–Whitney
Nehme et al. (2018)	Multi-well	hiPS-derived neurons	-	NB	7–42	-	5 SD	MFR, burst rate, etc.	Kruskal–Wallis
DeRosa et al. (2018)	Multi-well	hiPS-derived neurons	PEI+laminin	NB	>30	10	5.25 SD	MFR	ANOVA
Sarkar et al. (2018)	Multi-well	hiPS-derived neurons	PO+laminin	DMEM/F12	>21	10	5.5–6 SD	MFR	<i>t</i> test, ANOVA
Russo et al. (2018)	Multi-well	hiPS-derived neurons	PO+laminin	NB	>30	3	5.5 SD	MFR	Mann–Whitney
Blake et al. (2018)	Single-well	Primary adult mouse DRG	PEI+laminin	NB	7	-	5.5 SD	MFR	<i>t</i> test, ANOVA

Surface preparation: DETA, diethylenetriamine; PDL, poly-D-lysine; PLL, poly-L-lysine; PEI, polyethylenimine; PO, polyornithine.

Culture media: NB, neurobasal media; DMEM/F12, Dulbecco's modified eagle medium: nutrient mixture F-12; aCSF, artificial CSF; BP, BrainPhys media.

Vacancies in the table indicate instances where the reagent or technique used was not reported or unclear in the study methods.

Results

Prior art of MEA methodologies

As with all bio-assays, there are several facets to consider in experimental design and analysis. While many are universal, some are specific to the particular mode of detection. Those most relevant to the use of MEAs are detailed in Table 1. To assess existing practices for these elements of MEA experiments, the methods employed in 22 reported studies were reviewed. While not all of the experimental design elements list in Table 1 were explicitly addressed among the reported methods of these studies, the most commonly documented aspects of the methodologies are shown in Table 2.

Calculation of MFR and spike detection threshold

Novellino et al. (2011) surveyed six toxicology laboratories employing MEAs; comparing methods of tissue culture, recording, signal processing and assay metrics across the different research groups. Additionally, the study had each laboratory test clinically used neuropharmacological agents in their assays to assess the reproducibility of the results. Based on the parallel testing of pharmacological agents, Novellino et al., found that MFR, defined as the ratio of total spike events to recording duration in seconds, was the most consistent metric of activity. Concordant with this report, the use of MFR for reporting spike frequency within MEA recordings was the most common assay metric reported by the other studies examined (Table 2).

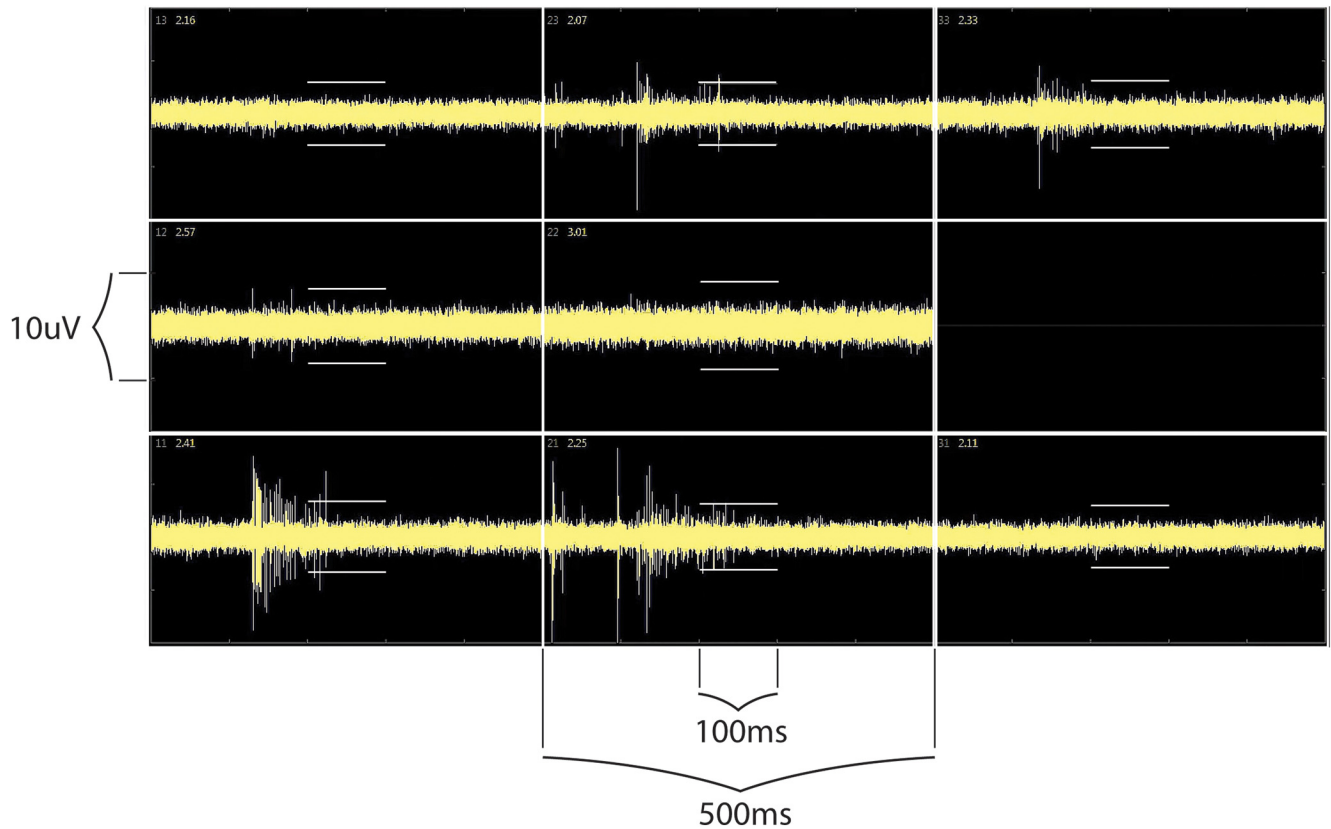


Figure 1. Example of MEA recording traces raw voltage recording of single well from a 96-well MEA containing eight electrodes. Each panel depicts 500 ms of recording data. The crossing thresholds for spike detection representing 5.5 SD the root mean square (RMS) of background voltage indicated by white bars above and below signal traces. Recording captured using AxIS acquisition software, Axion BioSystems.

The calculation of MFR is dependent on the detection of spontaneously generated action potentials, spikes, which occurs when the recorded voltage exceeds a designated threshold, as shown in Figure 1. Among the reported studies, this threshold was either set as an absolute potential value (Bateup et al., 2013a; Wainger et al., 2014) or relative to a number of SDs above the root mean square of the background voltage. In the latter case, the threshold determined by n SD can be dictated by an initial baseline recording (as in Illes et al., 2014). However, a method of dynamic threshold calculation over a moving window (in the range of ~ 10 ms) was more common. The crossing threshold employed by most studies were in the range of 5–6 SD, although others have used crossing thresholds as low as 3 SD (Vincent et al., 2013) or as high as 8 SD (McConnell et al., 2012). While methods are available to detect subthreshold potentials within MEA recording (Henningson and Illes, 2017), the use of large magnitude thresholds is done to insure that the majority of the events recorded represent action potentials.

While any of these approaches for establishing a spike-detection threshold are valid, the dynamically calculated crossing threshold has the value of being extendable for use in a *post hoc* analysis to eliminate spike calls arising from spurious electrical noise. During recording, sporadic electrical noise can cause the background potential to increase several fold, resulting in the erroneous determination of these events as spiking events. This can occur

while using a fixed or dynamically calculated threshold, since while the dynamic threshold will eventually adjust there is latency in doing so. Further, due to the hundreds of channels being recorded for tens of minutes or more, the probability of this occurring is increased, and manual supervision impractical. While this is not discussed in the reviewed literature, to address this issue as part of recordings performed for this study, the values of the crossing threshold at the time of the spike detection for all instances within a recording were examined. Since the background potential is expected to be random normal, all spikes for which the value of the crossing threshold at time of detection exceeded 3 SD of distribution across all crossing thresholds were excluded from analysis. This conservative threshold was chosen as to only exclude events from those channels exhibiting highly deviate background potentials. Following the removal of spike events detected at these aberrantly high crossing thresholds, the MFR was calculated as described above integrating all spike events recorded across all electrodes on an array during a recording session.

Firing frequency from neuronal arrays exhibits log-normal distribution

To assess the distribution of firing frequencies observed across neuronal cultures on MEAs, the MFRs of untreated rat primary cortical cultures were examined (Fig. 2). In Figure 2A, the histogram shows firing frequen-

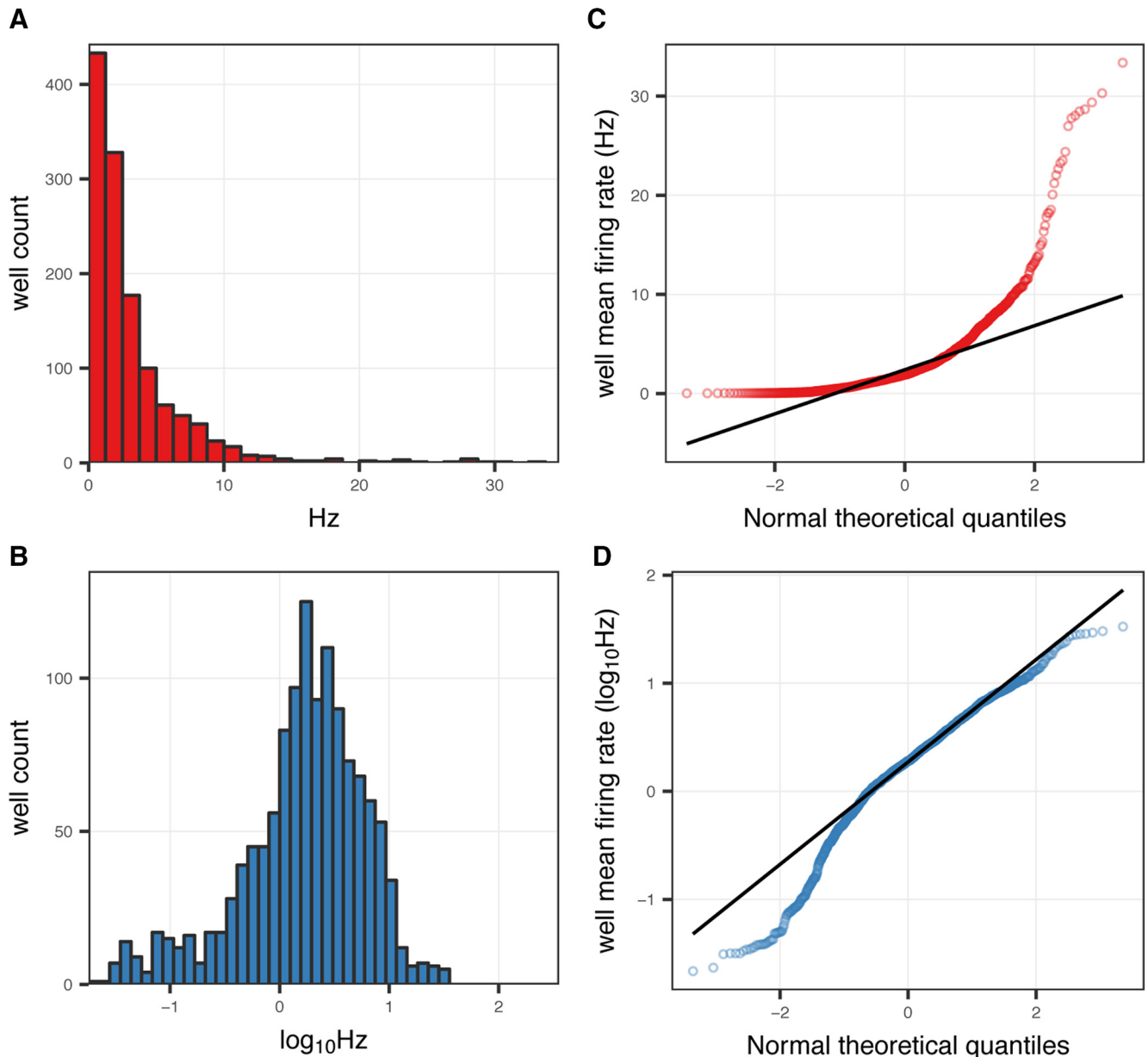


Figure 2. Distribution of MFRs observed in primary cortical cultures. Left, Distribution of firing frequencies (Hz) from 1272 unique arrays across 35, 96-well MEA plates originating from 22 separate culture preparations. Cultures were between DIV19 and DIV35 and recorded for 30 min. MFR shown on a linear (**A**, red) or \log_{10} -transformed (**B**, blue) scale. Arrays exhibiting low firing frequencies, MFR < 0.02 Hz have been removed from this dataset. Quartile-quartile plots comparing firing frequencies on a linear (**C**), log scale (**D**) to a random normal distribution (line).

cies reveals that the distribution is highly skewed on linear scale of Hz. Most studies reporting on MEA experiments make no mention of this highly skewed distribution in firing rates, with the noted exception of the report by Biffi and colleagues that discusses it specifically (Biffi et al., 2013). This observation may have gone unnoticed, due to studies being performed in single-well MEAs having insufficient data for this distribution to be apparent, however a similar highly skewed distribution has been reported for firing frequencies at individual recording electrodes within MEAs (Vincent et al., 2013). This distribution shape is not surprising, given that any frequency based metric is lower bound at zero, and

the only upper limit is the sampling rate at which the observations are being made.

A practical consequence of this pattern of behavior is that the observed distribution array firing frequencies and a theoretical random normal distribution deviate substantially from each other as shown in Figure 2C. Despite this, many studies reporting on spontaneous firing activity of neuronal cultures recorded by MEAs use common parametric statistical tests such as Student's *t* test and ANOVA, which assume a normally distributed dependent variable. Other researchers have presumably observed the asymmetry in the data sets of MEA recordings, and

instead have used non-parametric tests such as Kruskal–Wallis and Mann–Whitney U tests for reporting their results (Table 2).

The distribution of array firing frequencies does form a symmetrical, approximately normal distribution following transformation of the values by \log_{10} , as shown in Figure 2B. Other studies have used \log_{10} transformation in the reporting of MEA data, including Wainger et al. (2014) and Black et al. (2018), who did so to compare firing frequencies of individual electrodes, as well as Slomowitz et al. (2015), to compare the activity of individual neurons following spike sorting. Given the log-normal behavior of the observed frequencies, \log_{10} Hz rather than Hz appears to be a more appropriate metric to describe the spontaneous firing activity of neuronal cultures, especially to enable analysis using parametric methods.

Duration of recording for achieving intra-array reproducibility

Given the broad range of firing activity observed across arrays, extending more than two orders of magnitude, an important consideration for MEA based experiments is the duration of recording to perform. While the inter-array variance is the accumulation of the implicit variability of the biological system, and any variability introduced by the construction of the experimental system, the goal is to minimize sampling error by identifying a duration of recording sufficient to accurately capture the activity across the cultures. Most studies employing MEAs report using recordings on the order of tens of minutes (Table 2), although the duration can vary dramatically with some studies performing recordings as short as 2 min (Bateup et al., 2013a) and others up to 1 h (Slomowitz et al., 2015).

To formally address this question of what is a sufficient duration of recording, the correlation of firing frequencies was examined as a function of recording length in repeated recording sessions. Extended 3-h recordings were taken of mature cortical cultures performed on consecutive days (DIV20, DIV21). From these data, the Pearson's correlation (ρ) was calculated between the distribution of MFRs of arrays on the first day versus the second day, for accumulating 1-min intervals starting from the initiation of recording out to the entirety of the recording. The correlation of firing activity from recordings across sequential days increases dramatically within the first 30 min of recording with a value $\rho > 0.8$ observed in all experiments, yet there are only modest increases in ρ seen comparing longer intervals (Fig. 3). Fitting a linear regression to the data shows that the correlation between recordings is approximated by the natural log of recording duration $\rho \propto \ln(\text{time})$. This suggests that substantially increasing the length of recording beyond 30 min will only marginally improve the correlation, since $\delta / \delta x \ln(x) = 1/x$. Based on these observations, 30-min recordings appear to be an adequate recording interval to measure firing activity within MEAs for three-week rat cortical neuronal cultures. This method of analysis can be employed for any cellular system to determine the most appropriate recording time for the culture conditions to be used in a study using MEAs.

Changes in spontaneous firing across arrays with culture maturity

From the published literature, studies performing MEA experiments using primary cultures of rodent cortical neurons have a range of reported age of cultures: DIV post plating from dissected brain, and the time of analysis ranges from DIV7 to DIV35, although DIV14–DIV28 is most common (Table 2). Conversely, studies that instead used neuronal cultures derived from stem cells have a range of culture ages differing to a greater degree, over 30 d (DeRosa et al., 2018; Russo et al., 2018), including up to DIV70 (García-León et al., 2018). This broad range is expected given the variety of differentiation protocols used and increased interval until these cells become electrically active.

To assess how firing activity changed with maturation, four culture preparations were recorded every 2–3 d between ages of DIV3 and DIV22. As shown in Figure 4A, the cultures are electrically active by the end of the first week, and that activity remains generally consistent over the next two weeks. Of course, as described above, the range of activity observed across the population of arrays is broad. Additionally, this depiction shows all arrays, including the portion from which little electrical activity is observed. These are apparent in the lower portion of the figure where the data appears striated, where differences in a single spike event are seen as step-wise changes.

To determine whether the distribution of firing frequencies differs between cultures, further analysis was performed to examine the variance of firing frequencies with culture maturation (Fig. 4B). ANOVA indicates that the SD of firing frequencies is significantly affected by culture preparation and the age of the culture. Comparison of models indicates that the variance is significantly affected by culture age when controlling for the effect of culture preparation ($p = 0.013$). Examining the effect of culture preparation indicates that culture 3 (green data points) does show significantly lower variance than that of cultures 1, 2, and 4 ($p < 0.001$, $p < 0.001$, and $p = 0.0014$, respectively). However, there is insufficient evidence to conclude interaction between culture preparation and culture age. Put more simply, the rate of change in variance across firing activity with respect to time appears consistent across cultures. The implications these data have for assay development using MEA platforms are that the variance across arrays within a single preparation of primary cortical cultures appear to reach a minimum between DIV7 and DIV14. However, substantial variation across array activity persists during this interval. Using cultures at this age may help mitigate some of the array-to-array variability, but other methods described in this paper such as the bootstrapping simulations for experimental group assignment will still be necessary to fully account for the observed variance.

Assignment of treatment groups for MEA experiments

The variability observed across MEAs presents a challenge for the assignment of treatment groups for experiments. A common practice for biological assays within

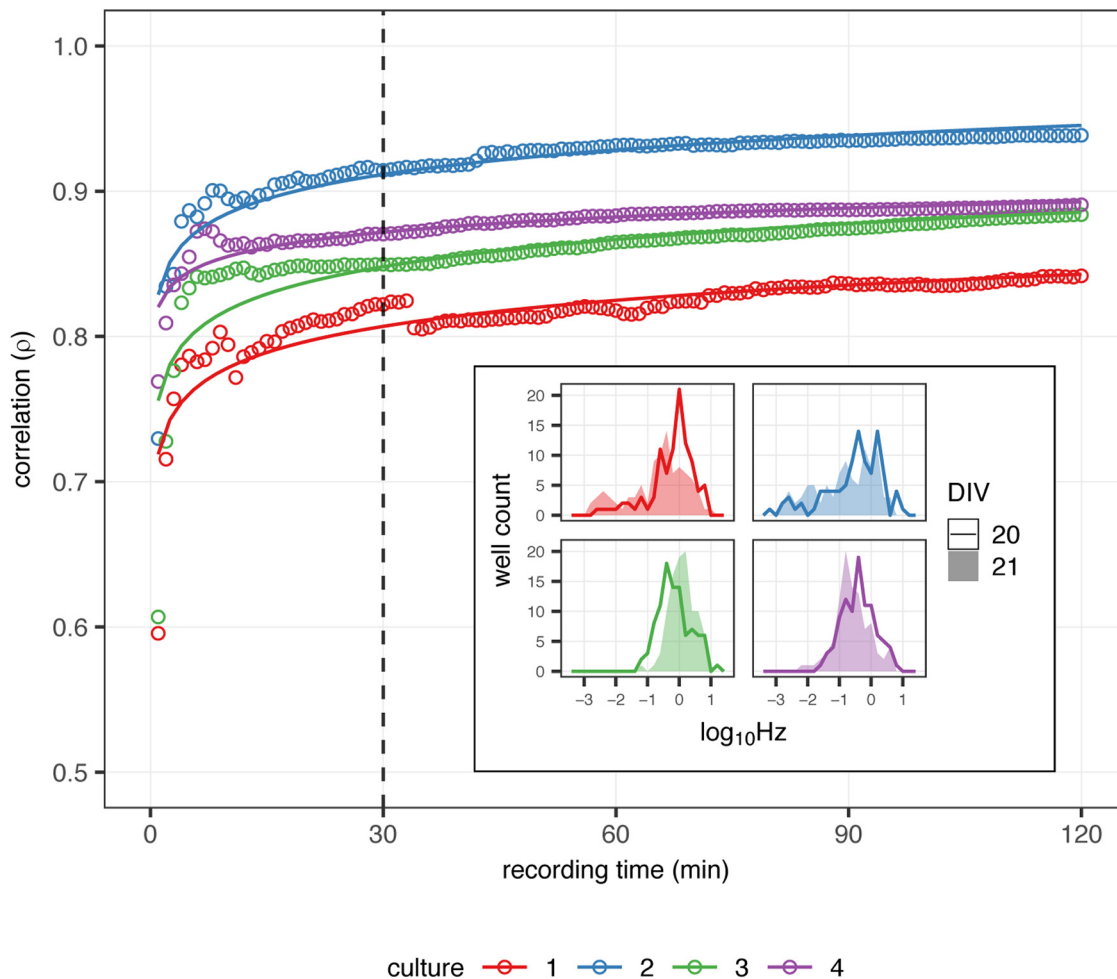


Figure 3. Correlation of firing frequencies in repeated recording as a function of time. Pearson's correlation ρ between firing frequencies of arrays on consecutive days (DIV20 and DIV21) as function of recording time for four culture preparations. Line indicates fit of $\rho \propto \ln(\text{time})$. Initial 2- of 3-h recording shown. Inset, Distribution of array firing frequencies $\log_{10}\text{Hz}$ for 30 min during each recording.

multi-well plates is to use the orientation of the wells in rows and columns to spatially assign treatment groups. However, this practice can be problematic for multi-well MEAs where variability is large and non-uniform. This is demonstrated for a typical multi-well MEA plate shown in Figure 5. Relying on plate rows for the designation of treatment groups can result in significant differences in the firing activity between cohorts even before the initiation of treatment, as shown in Figure 5B, where eight treatment groups ($n = 12$) were designated using the eight rows of the 96-well plate.

While plate-based variability does occur in conventional, colorimetric, fluorescence, or luminescence-based plate assays; in these instances, the cause is usually an external factor, such as liquid handling or excess evaporation along exterior wells, and are typically uniform plate-to-plate. Therefore, these can be addressed by changes to the assay protocol, or *post hoc* application of statistical regression techniques (Clemons et al., 2009). However, unlike in many bioassays data acquisition of MEAs is non-destructive to the cultures, so the varying levels of

activity between arrays can be assessed before initiation of experimental treatments.

The first step in remediating the issue posed by variability across MEAs is to exclude those arrays that exhibit substantially lower levels of firing activity than the population as a whole. These wells are clearly distinguishable as deviating from the otherwise normally distributed population (Fig. 5C). In practice, these low-activity wells are excluded by setting a threshold at 2 SD below the median value of firing frequency across all arrays in the experiment. In this case, seven arrays of 96 are excluded with the remainder closely following a normal distribution (Fig. 5D). Having eliminated these arrays, this now invalidates the approach of using plate rows or columns for the assignment of treatment groups, since the number of available replicates in each group would differ.

While aligning treatment groups to the rows and columns is convenient, it imposes an unnecessary constraint on the positioning of replicates within cohorts. For the original configuration of assigning 96 arrays to eight treatment groups with 12 members each, there are $6.25 \times$

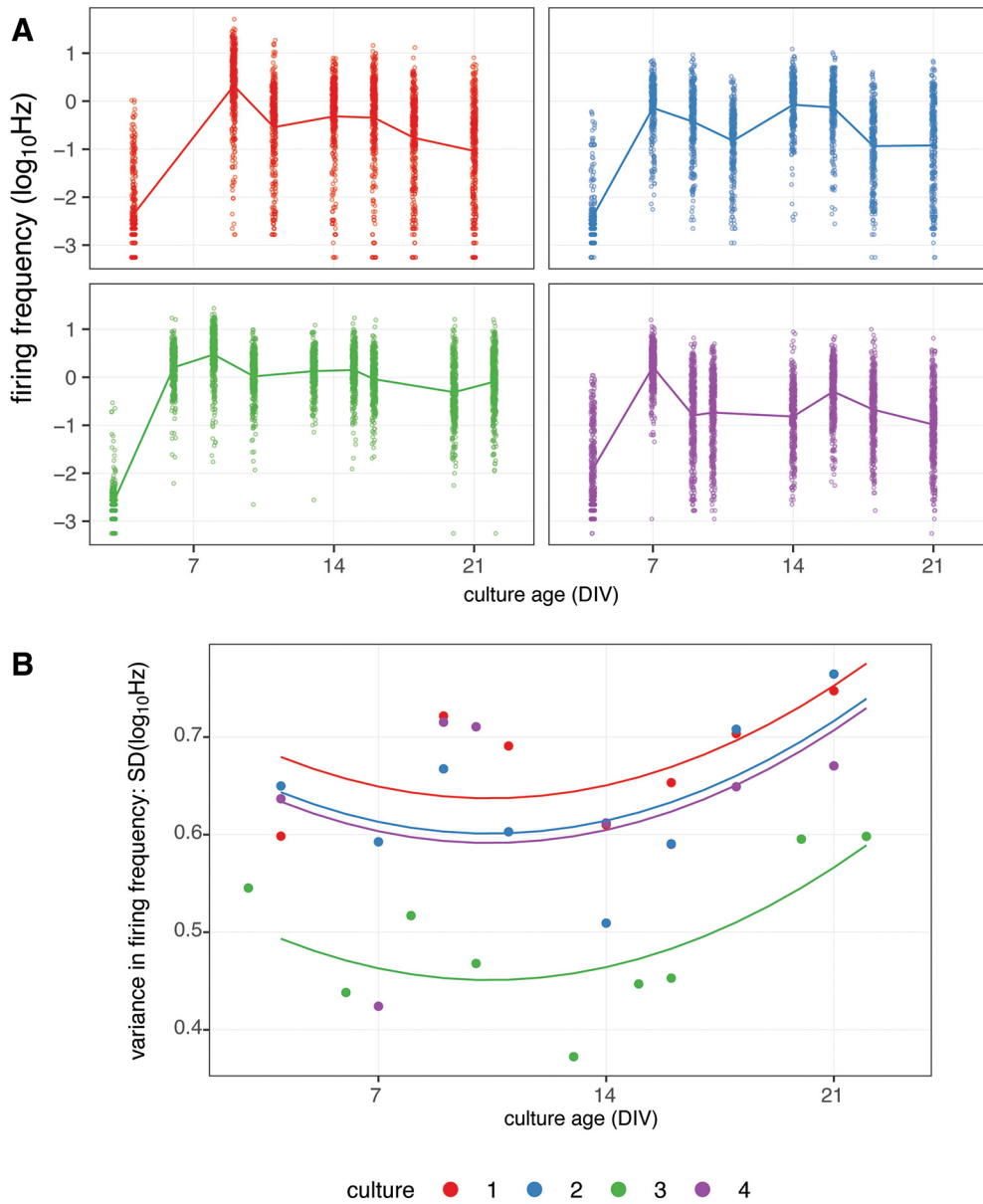


Figure 4. Changes in firing frequency with culture maturation. **A**, Firing frequencies of arrays observed during repeated 30-min recordings over a three-week period (DIV3–DIV22) across four culture preparations. Each point represents an individual array per day; 384 replicate arrays, cultures: 1, 3, and 4; 288 culture: 2. Line indicates mean logHz per recording for each culture. **B**, SD in firing frequencies of arrays observed during repeated 30-min recordings shown in panel **A**. Each line represents the prediction of a linear model for the change in variance as a function of time.

10^{14} potential combinations, of which the option that aligns with the plate rows is only one. When the assay signal is uniform across the plate, one combination of wells is effectively as good as another, so opting for the most convenient makes the most sense. However, with the convenient option removed due to the exclusion of inactive arrays, and adding to it the non-uniform variability across the plate, there is reason to explore the other options to achieve a comparable level of pre-treatment activity between groups.

With the inactive arrays removed, the goal is to find an alternative assignment using the 88 most active of the remaining 89 arrays to generate 8 treatment groups with

11 members each. To do so, we have applied a bootstrapping simulation approach where 10^4 possible treatment assignments are generated, then a one-way ANOVA is applied to each iteration assessing firing frequency ($\log_{10}\text{Hz}$) as a function of group assignment. From these ANOVA results, the iteration that results in the lowest F-statistic, which represents the ratio between group and within group variance, is selected for the assignment of the treatment groups. The results of this approach are shown in [Figure 5E,F](#), where the alternative assignment results have near equal levels of activity across all eight cohorts. Using the described method of removing inactive arrays and determining the optimal assignment through

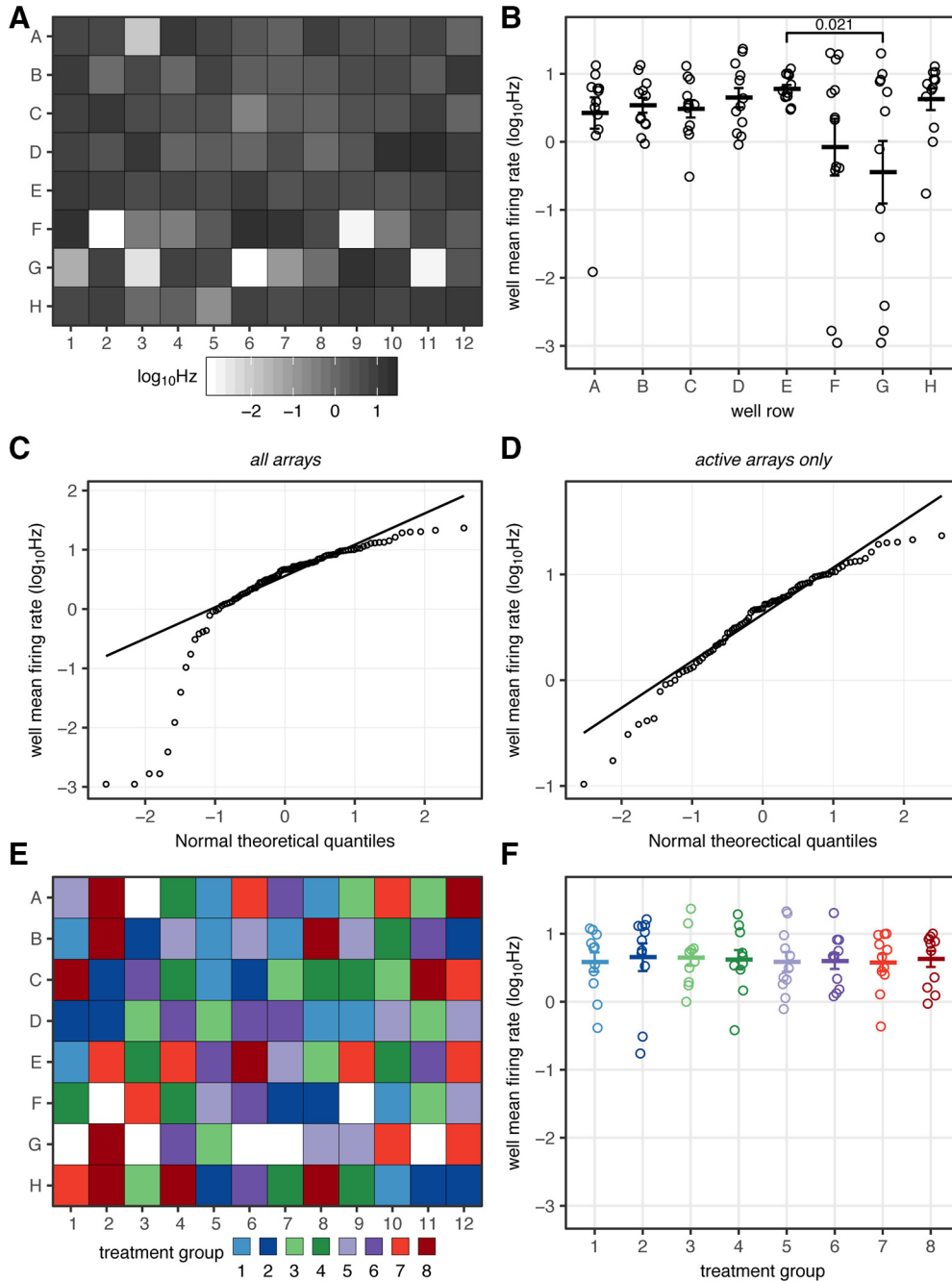


Figure 5. Assignment of treatment groups across multi-well MEAs. **A**, Firing frequencies observed across a single recording of a 96-well MEA. **B**, Firing frequencies across eight cohorts with 12 members (96 total), arising from cohort assignment based on plate row. QQ plots comparing firing frequencies to normal distribution, for all wells (**C**), or following exclusion of wells >2 SD below median (**D**). **E**, Alternative assignment for eight cohorts with 11 members (88 total), with minimal between group variance, following simulation of 10,000 possible assignments. **F**, Firing frequencies within groups from alternative assignment. Data presented as mean \pm SEM.

simulation results in a 78.7% decrease in the variance in firing frequency associated with treatment group assignment ($TotalSS_{sim} = 17.64$ vs $TotalSS_{row} = 82.6$).

Additionally, this method can be extended to generate treatment groups with comparable activity levels across multiple plates for large experiments, which exceed the capacity of a single plate. It also can be used to generate treatment groups with comparable activity but uneven numbers of replicates for instances where the experiment

incorporates a finite resource, like primary patient material, necessitating an unbalanced design.

Adding it all up: power to detect changes in neuronal firing using MEAs

Ultimately, for a researcher seeking to model changes in spontaneous firing activity within neuronal cultures using MEAs, one of the most practical considerations of experimental design comes down to the decision of the

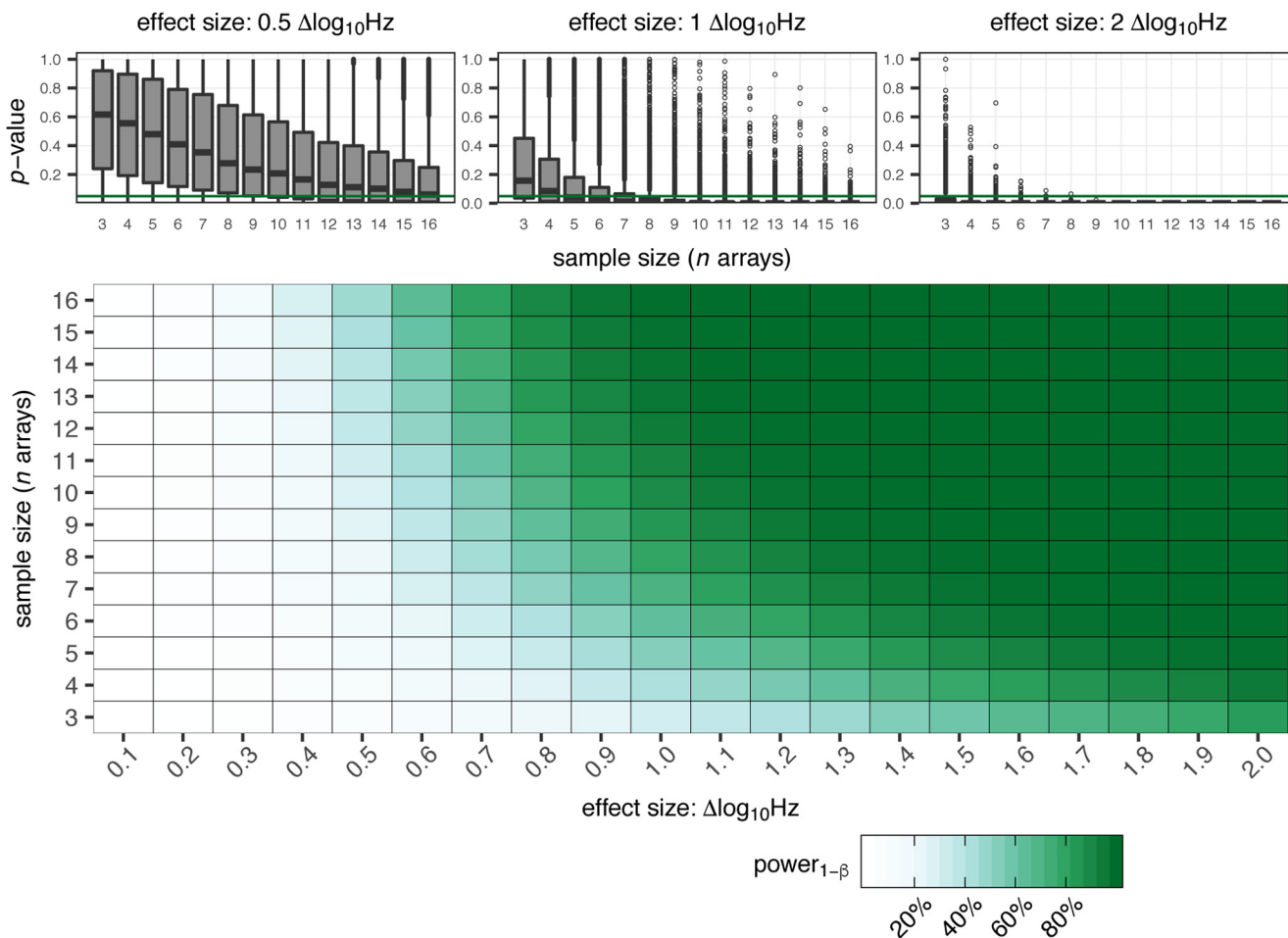


Figure 6. Calculation of statistical power for MEA experiments. Estimates of percentage power_(1-β) based on simulation for a two-condition (control, treatment) repeated measures (pre, post) experiment across sample sizes 3–16 arrays and effect sizes of 0.1–2.0 $\Delta \log_{10}\text{Hz}$. Simulated experiments analyzed by ANCOVA, *p* value extracted from comparison between control and treatment at post measure, 5000 iterations per sample size:effect size pairing. Upper, Distributions of observed *p* values from simulated treatments per sample size for effect size of 0.5, 1.0, and 2.0 $\Delta \log_{10}\text{Hz}$, green line indicates *p* = 0.05.

number of replicates per condition. Having characterized the distribution of firing frequencies observed across a large number of MEAs (Fig. 2) and having established the correlation of activity between repeated recordings (Fig. 3), it is possible to use these factors to estimate the statistical power of MEA experiments.

For demonstration purposes, a two condition (control, treatment), repeated measures (pre, post), experiment is modeled. A bootstrapping approach is applied to simulate experiments using sample sizes ranging from three to 16 replicates per condition, and scenarios in which the treatment resulted in effect sizes of a difference in $\log_{10}\text{Hz}$ ranging from 0.1 to 2.0. The dataset used for the simulation is generated from the firing frequencies of 1272 unique MEAs. The pre-treatment time point $\log_{10}\text{Hz}_{t_0}$, is taken from the existing data, while the post-treatment time point $\log_{10}\text{Hz}_{t_1}$ is randomly generated as to be related to $\log_{10}\text{Hz}_{t_0}$ by a correlation coefficient $\rho = 0.8$, estimated from repeated recordings of MEAs (Fig. 3). The firing frequencies of experimental replicates are pairs of $\log_{10}\text{Hz}_{t_0}$ and $\log_{10}\text{Hz}_{t_0+1}$ with values drawn from random arrays in the population. The $\log_{10}\text{Hz}_{t_0+1}$ values within the

treatment condition are offset by the specified effect size. to estimate the difference between control and treatment at the post time point, and an ANCOVA is performed assigning time as a co-variate and determining the effect of treatment. This method has been shown to provide higher power than other used approaches, such as percentage of baseline or absolute change (Vickers, 2001).

Applying the conventional standard of statistical power $\geq 80\%$ for a robust assay, the results of this simulation show that it is difficult to reliably detect changes below 0.7 $\Delta \log_{10}\text{Hz}$, ~ 5 spikes / sec, even with large numbers of replicates. Based on the results of additional simulations, detecting changes on the order of 0.5 $\Delta \log_{10}\text{Hz}$ would require treatment groups with >40 replicates, and a change of 0.1 would require >600 replicates. to confidently detect changes of 1.0 $\Delta \log_{10}\text{Hz}$ (i.e., a 10-fold difference in spikes / sec), requires treatment groups with eight to nine replicates. Only with large changes in in firing frequency of $\Delta \log_{10}\text{Hz}$ 2.0 or greater (i.e., a 100-fold or more difference in spikes / sec) would it be reasonable to use treatment groups as small as an $n = 3$ (Figure 6). To illustrate the magnitude of these proposed effects, re-

peated experiments with TTX, showed that treating cultures with 10 nM induced decreases of $-1.14 \log_{10}\text{Hz}$ and with 100 nM induced decreases of $-2.24 \log_{10}\text{Hz}$ (data not shown). The reported IC_{50} of TTX-sensitive Na_v channels in rats is 10 nM or less (Catterall et al., 2005). Therefore, this analysis suggests that to have sufficient power to detect changes in spontaneous firing activity comparable to those elicited by treatment with TTX at its IC_{50} concentration requires treatment groups with eight to nine arrays per condition. Of course, it is important to emphasize that these power estimates are based on a simple two condition experiment, so any experiment seeking to examine more conditions would need to allow for more replicates per condition to provide sufficient power for multiple comparisons. This estimate of a requirement of eight to 10 replicates to discern effects on firing frequencies within MEA between treatment conditions of strong perturbations is consistent with replicates used in studies reporting effect of known pharmaceutical compounds or neurodegenerative disease associated mutations (Novellino et al., 2011; Wainger et al., 2015).

Implementation of unsupervised spike sorting for MEA recordings

MEA recordings provide a means of generating rich data sets of electrophysiological activity from a large number of neurons over comparatively long-time intervals of days to weeks. However, resolving the activity of individual neurons within this data are challenging due to the poor spatial resolution of MEA data acquisition since the density of detectors is sparse on a cellular scale, effectively rendering each electrode a point detector. Despite this, it is possible to deconvolute the mixed signals of multiple neurons detected on each recording electrode of a MEA through the implementation of spike sorting analysis in which shape parameters of waveforms arising from a population of neurons are binned by similarity allowing for attribution of individual waveforms to a cell of origin. However, results of spike sorting performed in this fashion come with the caveat that there is no “ground-truth,” since it is not possible to know exactly how many individual neurons are being recorded and signals arising from separate cells may be conflated as coming from the same point of origin or vice versa due to fluctuations in field potentials and propagation of potentials across the array (for review, see Buzsáki et al., 2012; Gibson, 2012). The application of spike sorting analysis to MEA recordings is attractive as it allows for: (1) refinement of spike frequency metrics to be resolved to the level of individual neurons; (2) assessment of frequency-independent phenotypes, such as changes in the amplitude of potentials emitted from firing neurons; (3) determination of network dynamics within cultures through determination of coordinated firing patterns.

A review of commercially available software solutions capable of performing this analysis was performed. One such product available at the time of initiation of this study was Offline Sorter (Plexon) which has been used in the analysis of other MEA studies (Slomowitz et al., 2015; Vertkin et al., 2015). While this software has the capability

to implement similar clustering algorithms to those ultimately employed in the analysis pipeline developed for this study, this software has critical limitations that limit its utility for the analysis of multi-well MEA recording data. First, this software is only capable of running on a single CPU, precluding the ability to take advantage of distributed computing resources available, and thus greatly increasing the processing times of analysis. Second, the software required the data to be analyzed in a recording centric fashion, which greatly diminishes the accuracy of the clustering.

The development criteria for the analysis pipeline were: (1) flexibility to analyze data collected during separate recording sessions; (2) ability to perform clustering analysis in an unsupervised fashion; (3) extendable to parallel processing in a distributed computing environment such as high-performance computing clusters; and (4) able to use open-source software to enable distribution and use by other researchers.

For these analyses, the recording sampling rate was 12.5 kHz, and the spike detection threshold was set at 5.5 SD from the root-mean square of the background voltage potential calculated over a 10-ms moving window. On recording of a voltage above the detection threshold (spike event), 3 ms of recording data were retained, 1 ms preceding and 2 ms following the timing of the threshold crossing event. With a sampling rate of 12.5 kHz, the 3-ms window around each threshold event results in the retention of 38 momentary voltage recordings for each spike event. These 38 momentary voltage recordings are reconstructed into a wave form. From the reconstructed wave form, an array of shape parameters are calculated for each event, including: maximum voltage (peak), minimum voltage (valley), wave form amplitude (peak-valley range), time interval between peak and valley, AUC, and NLE, a parameter describing the “sharpness” around the maxima and minima of the wave form (Kim and Kim, 2000).

The spike sorting process begins by aggregating all of the spike events detected by a given electrode during all recordings of an experiment, as shown in Figure 7A. Based on the wave form parameters of all spike events collected on a given electrode, principal component analysis (PCA) is performed and all spike events were projected along the first two principal components as shown in Figure 7B.

Once projected along principal components, clustering of the waveforms is performed to group those exhibiting the highest degree of similarity. To do so in an unsupervised fashion, the mean-shift algorithm was applied, implemented through use of the MeanShift library within the R programming language (Fukunaga and Hostetler, 1975; Comaniciu and Meer, 2002; Ciollaro and Wang, 2016). The use of mean-shift clustering has particular advantages for the analysis of MEA data over other non-hierarchical clustering techniques such as, *k*-means, since mean-shift does not require an a priori estimate of the number of groups in the dataset. This is an important practical consideration given that the number of cells that will be recorded on a single electrode is unknown and variable. While others have implemented semi-supervised

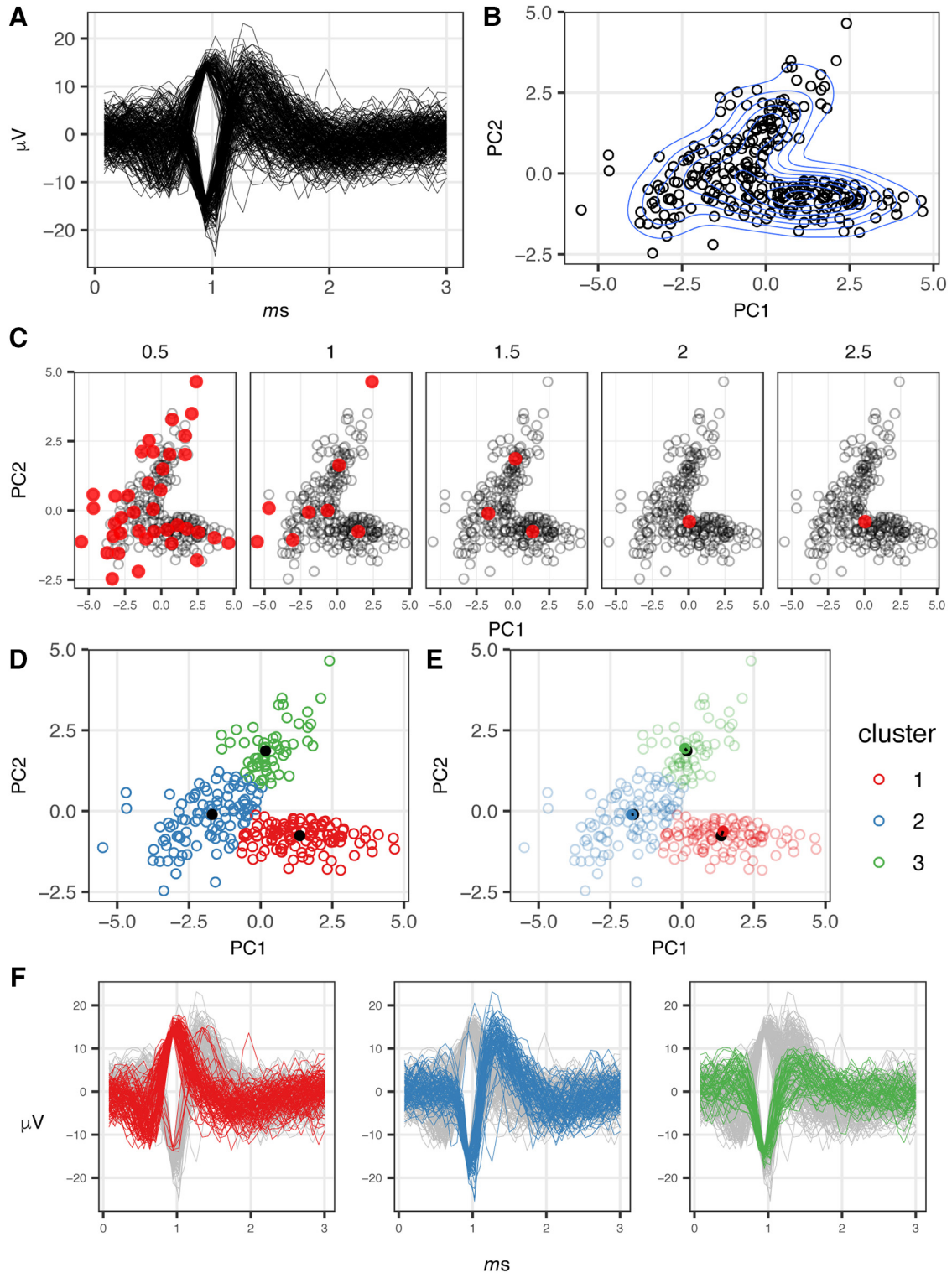


Figure 7. Demonstration of spike sorting. **A**, Waveforms of 269 spike events captured by a single electrode across 7, 30-min recordings. **B**, Projection of 269 spike events along principle first two principal components calculated from wave form shape parameters, blue lines indicate density contours of events. **C**, Varying number of cluster centroids identified by mean-shift clustering as a function of bandwidth value h of KDE. Cluster centroids indicated in red, value of kernel density estimate denoted above each panel. **D**, Spike events colored by cluster membership as determined by mean-shift clustering performed with a KDE $h = 1.5$, cluster centroids are indicated in black. **E**, Mapping of cluster centroids (black circle) to nearest spike (filled circle) event by minimum Euclidean distance. **F**, Waveforms of spike events colored by cluster membership.

clustering methods relying on k -means for the purpose of spike sorting in MEA recordings (Slomowitz et al., 2015), this requires manual intervention to indicate the number of distinct wave form shapes present on each channel. Instead, the mean-shift algorithm operates by applying a density function to the observed data and identifying the local maxima across the distribution of the observations, all observations within the dataset are then mapped or “shifted” to the nearest local maxima. The only variable that needs to be supplied to the algorithm is the bandwidth term (h) to the kernel density estimator (KDE) function to determine the degree of curvature supplied to the density function. For this analysis, a Gaussian density function was applied, while the value of the KDE was determined empirically by applying a range of values to sample clustering of a teach-set of electrodes to establish a value which provided accurate segregation of the waveforms, resulting in the use of $h = 1.5$. The influence of the KDE in the mean-shift algorithm is demonstrated in Figure 7C, in which lower values of h result in insufficient smoothing of the density function, resulting in an over proliferation of local maxima, while higher values of h result in over smoothing of the density function resulting in only a single maximum.

While the cluster centroids, as shown in Figure 7D, occupy a finite position with respect to the eigenvectors these points may or may not coincide with points representing individual, observed waveforms. Therefore to map the cluster centroid back to the original shape parameters used in deriving the clusters themselves, the wave form closest to the centroid is determined by calculating the minimum Euclidean distance (Fig. 7D,E). The shape parameters of the wave form closest to the centroid can thus be used for evaluating the mean shape differences between clusters.

Spike sorting in a recording-dependent versus recording-independent fashion

Given the longitudinal nature of many MEA-based experiments, it was important to assess the effect of performing the mean-shift clustering in a recording-dependent fashion, in which the clustering is performed several times, once for each set of single recording data, or in a recording-independent fashion, in which the clustering is performed once on the data from all recordings. In the latter case, the meta-data describing the recording origin of each event is maintained, so that the events can still be parsed by recording. The result of these two approaches is shown in Figure 8, in which spikes recorded by a single electrode across multiple recordings were clustered in a recording-dependent fashion (Fig. 8A) or recording-independent fashion (Fig. 8B). The limitation of the recording-dependent approach is exposed as a consequence of the uneven distribution of spike events across recordings. During the fifth recording in which 92 spikes are recorded, the recording-dependent method identifies the same number of clusters, three, as identified in the recording-independent analysis. However, during the third recording in which only 35 spikes are recorded, sporadic events that are otherwise aggregated into cluster #3 (green) in the

recording-independent analysis are oversegregated in the clusters #4 (purple), #5 (dark green), #6 (orange) in the recording-dependent method. The other limitation to the recording-dependent approach is that since the values assigned denoting cluster membership are arbitrary, clusters that occupy the same position in principal component space are assigned to different cluster values in different recordings, presenting an additional challenge of correctly associating clusters hosting similarly shaped waveforms based on the cluster identifier alone.

These data emphasize the importance of performing spike-sorting in an recording-independent manner when examining recording data collected over several sessions to avoid misattribution of firing events to separate clusters that are in fact likely arising from a common point of origin.

Stochastic nature of firing patterns of individual spike clusters

Having implemented spike sorting analysis, it is possible to evaluate the firing frequencies of individual spike clusters, providing a higher resolution of the spontaneous firing than obtained with the array (or well)-level as described above.

The firing frequencies of 237 spike clusters identified within 32 arrays wells across four untreated culture preparations, are shown in Figure 9A. These data show the firing frequencies as average *spikes / sec* (Hz) from 30-min recordings taken on three consecutive days. Across the four experiments, the timing of culture media changes was coordinated with respect to the timing of recordings to minimize differences in firing across experiments as a consequence of the state of media exhaustion.

Examining firing frequencies at the level of individual spike clusters in Figure 9A, the patterns of firing over time appear largely stochastic. First, the firing frequencies, as indicated by increasing values from white \rightarrow black of clusters differ dramatically with respect to each other across clusters within recordings (along rows) and within the same cluster across recordings (down columns), from a minimum of 0 Hz to a maximum of 15.96 Hz. Additionally, while all clusters are observed firing in at least one of the recordings, the number of instances in which clusters are inactive during any one 30-min recording interval in which the culture is observed is substantial. As shown in the cumulative density plot in Figure 9B, the proportion of inactive clusters during a given recording ranges from 24% to 41%. Additionally, while the majority of spike clusters exhibit firing frequencies are <1 Hz, there are a limited number of highly active clusters. The combination of a large proportion of inactive clusters and a limited number of highly active clusters results in a highly skewed distribution of firing frequencies (Fig. 9C, left) analogous to the distribution firing frequencies observed at the array-level (Fig. 2).

The zero-inflated nature of the cluster firing frequencies presents a challenge when aiming to describe differences in distributions of cluster frequencies, such as in response to experimental conditions. The strategy of applying a log transformation to the frequency values as described

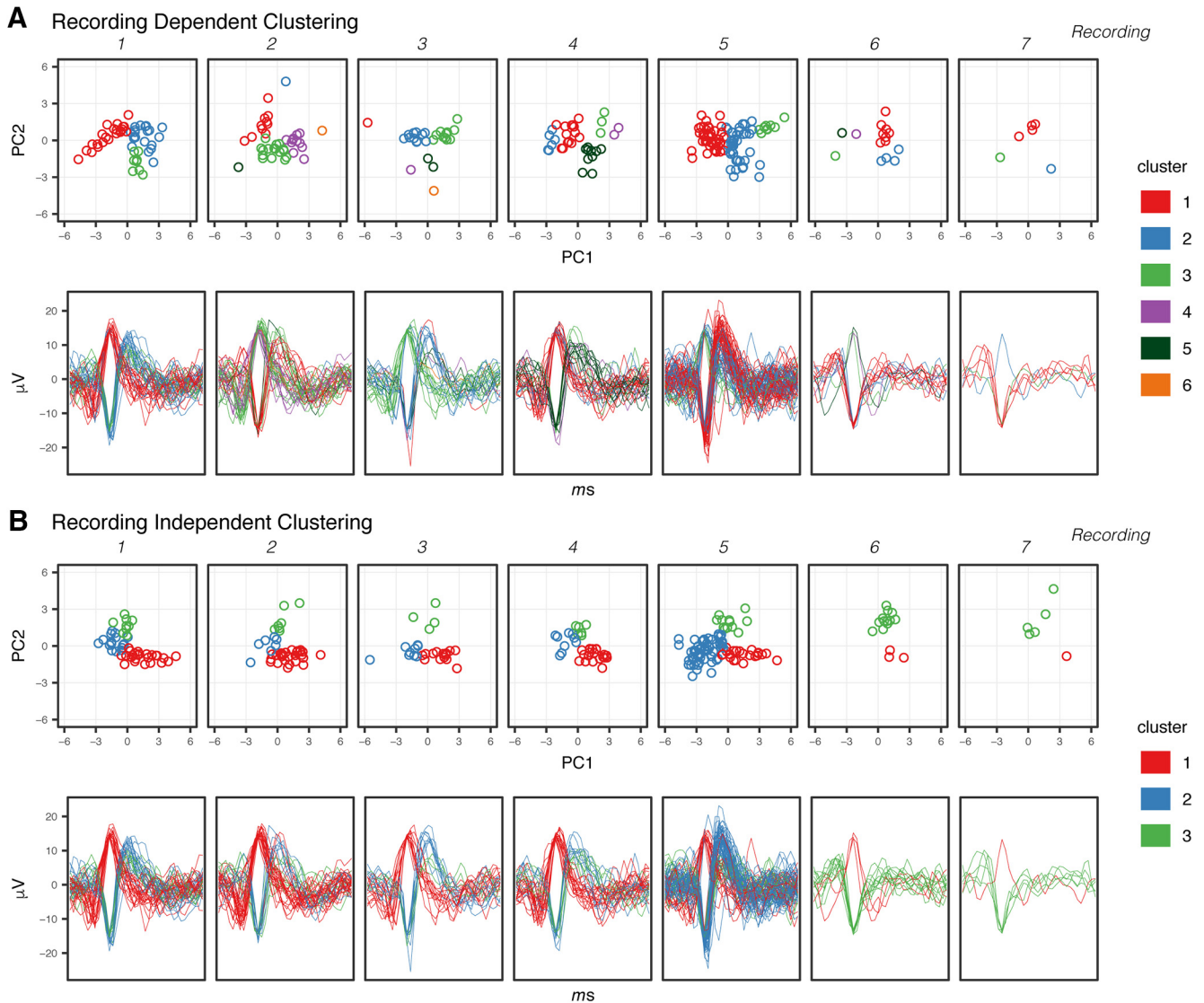


Figure 8. Spike sorting within single recordings versus independent of recording. Result of cluster assignment of 269 spike events recorded from a single electrode across 7, 30-min recordings. Mean-shift clustering performed in recording-dependent fashion on spike events from each recording (**A**) or in recording-independent fashion across pooled spike events from all recordings (**B**).

above for array-level frequency data are not applicable in this instance, since the result of the log transformation is not a symmetrical normal distribution as is achieved in that case, but rather an asymmetrical bimodal distribution (Fig. 9C, right) However, since the values of the firing frequencies can theoretically assume values from 0 → acquisition sampling rate, an alternative approach is to model the distribution of cluster firing frequencies with a Γ distribution. While the distribution of non-transformed array-level firing frequencies could be described using a Γ distribution as well, in practice the log-transformation of array-level firing frequencies is more convenient since most arrays exhibit some basal level of activity across recordings compared to the large proportion of clusters which exhibit no activity during a given recording.

To assess changes in the firing frequencies across distributions of clusters under different experimental con-

ditions, Γ -GLMs can be applied to obtain estimates for the average firing frequency in each condition (McCullagh and Nelder, 1989). This approach is demonstrated in Figure 9D, in which a Γ -GLM was fit to the data model frequency as a function of recording, and the resulting Γ distributions predicted by the model are overlaid on the distribution of clustering firing frequencies observed in each recording. The differences in firing frequencies across recordings are shown in Figure 9E comparing the Γ distributions predicted for each recording and in Figure 9F comparing the differences in average Hz \pm SE between recordings based on the estimates obtained from the fitted model.

Characterizing network behavior of cultures within MEA recordings

Beyond quantifying the distributions of firing frequencies observed across individual neurons (spike clusters),

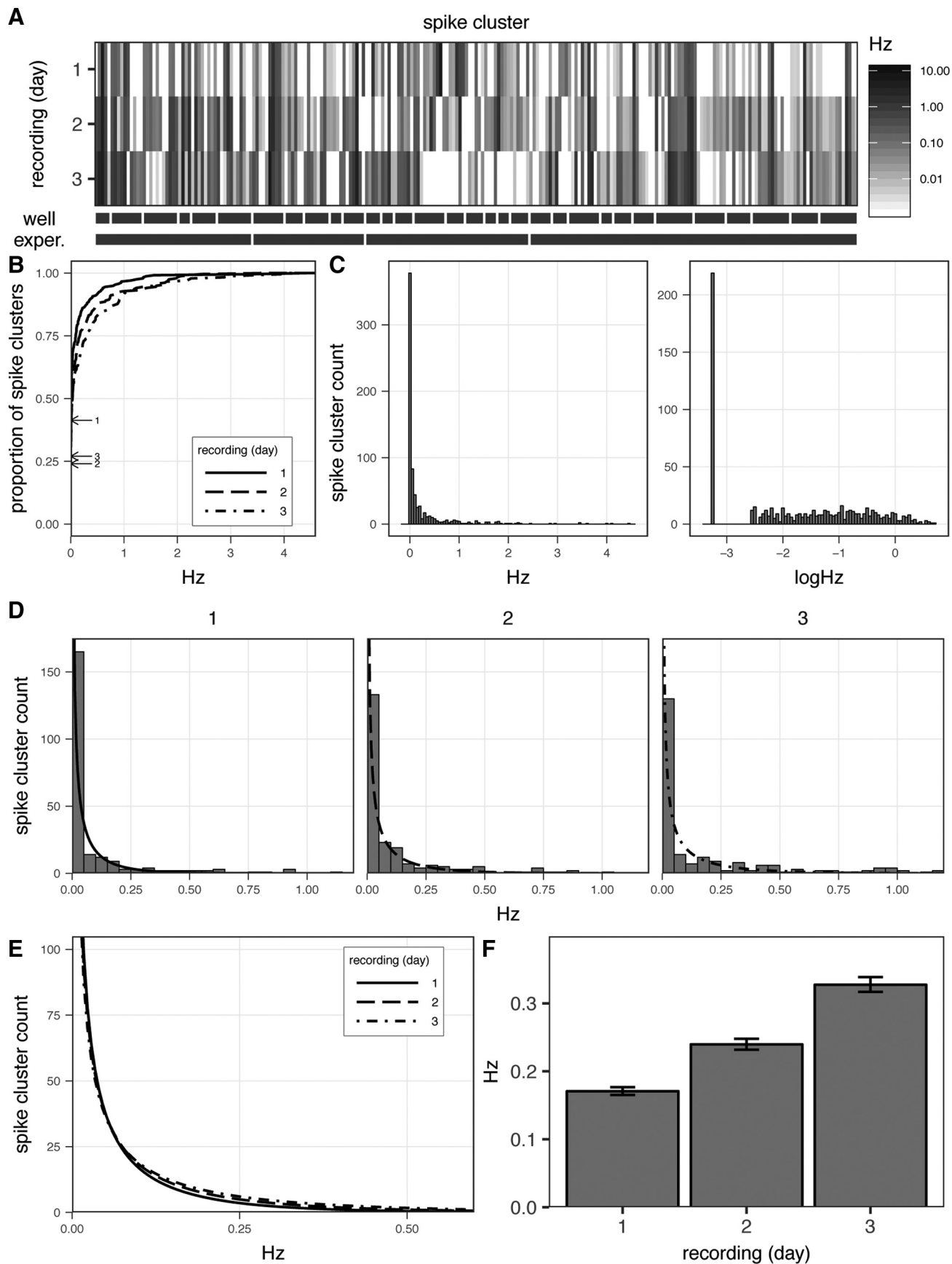


Figure 9. Stochastic firing patterns of spike clusters. **A**, Firing frequencies (Hz) of 237 spike clusters identified in recordings of 156

continued

electrodes in 32 replicate wells across 4 mature, untreated cultures (DIV20–DIV45). Data from 30-min recordings across three consecutive days. Spike clusters within common wells and experiments indicated by hash bars. **B**, Cumulative distribution of spike cluster firing frequencies across recordings, arrows denote y-intercept, indicating proportion of inactive clusters during each recording. **C**, Distribution of firing frequencies Hz across all three recordings on a linear (left) or \log_{10} -transformed scale (right). **D**, Distribution of firing frequencies across each recording overlaid with estimated distribution of firing frequencies obtained from Γ -GLM. **E**, Overlay comparing estimated distribution of firing frequencies from each recording. **F**, Average firing frequency per recording estimated from Γ -GLM, presented as mean \pm SE.

having resolved the recording data to this level it is possible to examine the network dynamics within populations. Compiling all action potentials attributed to an individual neuron (spike cluster) through spike sorting over a given recording interval results in the formation of a spike train for each cell (Fig. 10). Examining the spike trains arising from neurons detected on a single array over an interval of minutes (Fig. 10A) reinforces the variability in firing frequencies described above. However, examining the time course of firing events over a duration of a few seconds reveals synchronized firing patterns (Fig. 10B), a

phenotype of neuronal networks widely reported to be maintained in dissociated cultures (Segev et al., 2001; Van Pelt et al., 2004; Chen et al., 2006). The STTC is a parameter for describing the concordance or discordance of firing events between a pair of spike trains (Cutts and Eglen, 2014). Figure 10C shows the STTC between seven individuals detected within an array over a single 10-min recording. Examining the distribution of over 40,000 STTC values calculated from >14,000 unique pairs of neurons detected from repeated recordings of 387 arrays containing primary cortical cultures, shows that large magnitude,

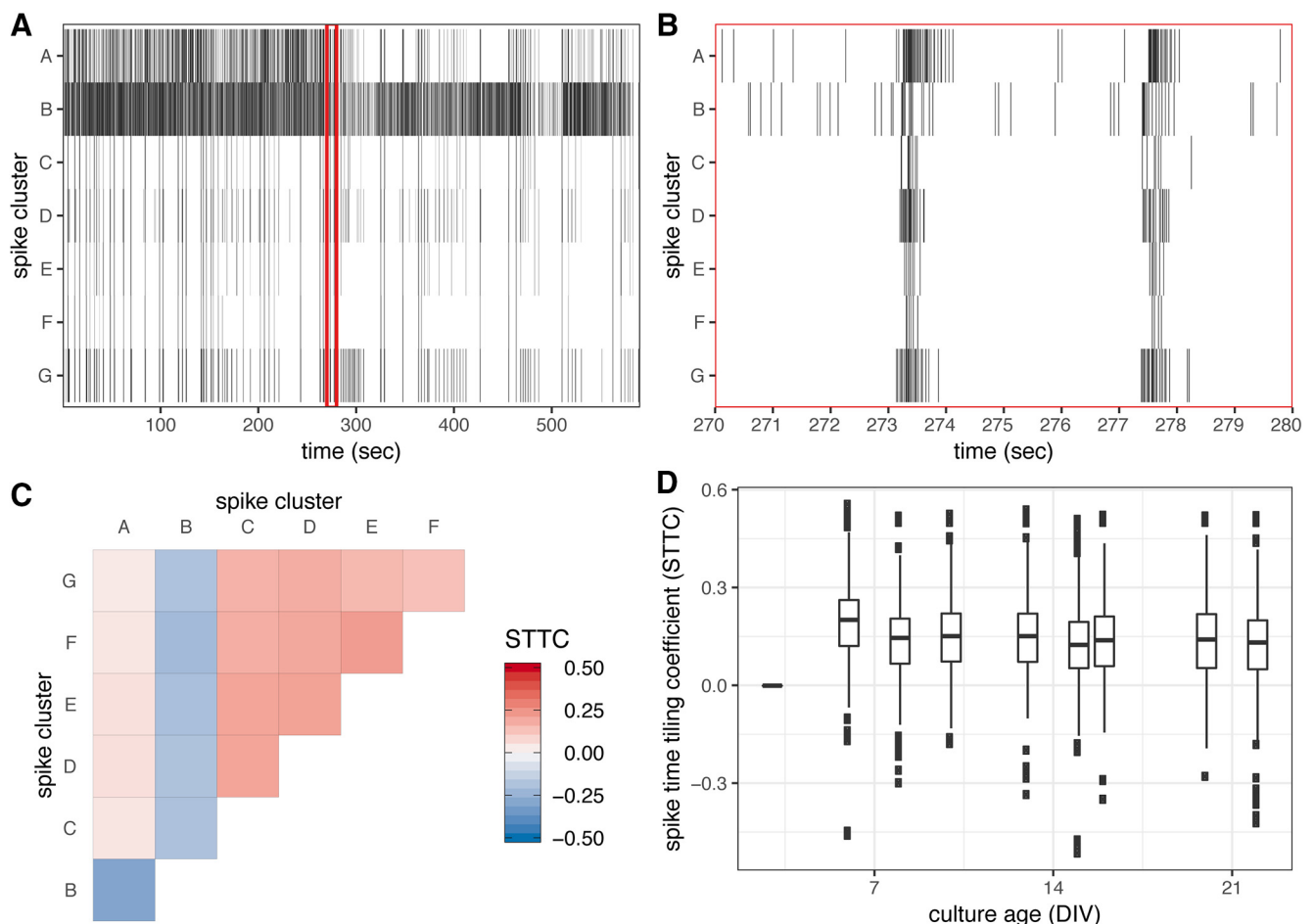


Figure 10. STTC derived from pairwise contemporaneous firing across spike clusters. **A**, Raster plot indicating instances of action potentials attributed to seven individual neurons following unsupervised spike clustering of 10-min recording of DIV8 primary cortical culture. Red box indicates 10-s interval within the recording window. **B**, Instances of action potentials within 10-s interval indicated within the longer 10-min recording depicted in panel **A**. **C**, Value of STTC calculated between spike clusters within a single array across a single 10-min recording. **D**, Distribution of 42,486 STTC values calculated between 14,225 unique spike cluster pairings detected across 387 arrays recorded between DIV3 and DIV22.

positive and negative, STTC values are detected within a week from initiation of culture (Fig. 10D). Moreover, while the distribution of observed STTC values are largely consistent and symmetrical over the observed period, the majority of values are positive indicative of a propensity for in-phase rather than out-of-phase activity relationships in this culture type.

Using STTC to assess coordinated activity between pairs of neurons, this can be extended to infer whether the two cells are functionally connected. Those bona fide connections thereby are considered to represent edges in a network of activity among the neurons. To determine which pairs of neurons are likely to be functionally connected a permutation based analysis was applied similar to that described by Vincent et al. (2013). For each pair of neurons, a null distribution of STTC values was generated by 1000 random permutations of the time stamps of the observed firing events. The empirically observed STTC values were compared to this random distribution, and only those instances in which the observed STTC value was of sufficiently strong correlation (positive or negative) that the likelihood of occurring by chance had a probability of <1% were taken to represent a functional connection between neurons. Using this method, the firing patterns from 1106 arrays representing three primary cultures, each recorded eight to nine times over a three-week period were examined. From this dataset across all recordings, 75,390 STTC values were calculated from 32,414 unique pairs of neurons. Of the 75,390 STTC values calculated from the data, 43,421 (57.6%) were deemed to represent functional connections based on permutation testing. This selection was made with the qualification that these relationships are based on firing patterns and not detection of physical connection.

Specifying these connections, the topography of the functional network between neurons detected on each array during each recording session can be established. To quantify the degree of connectedness within each network, the network cluster coefficient (\bar{C}) is calculated for each array and recording (Watts and Strogatz, 1998). Examining the activity across these 1106 arrays reveals that spontaneous firing activity does not always result in the detection of network activity. As shown in Figure 11, while the vast majority of arrays exhibit spontaneous firing activity emerging within the first week of culture (panel A), network activity as defined by a non-zero (\bar{C}) value is detected in a much lower proportion of arrays. Further, the density of the networks as determined by (\bar{C}) significantly decreases from a peak within the first week at an estimated rate of $-0.008/d$ ($p < 0.001$; Fig. 11C). This decline in network activity is consistent with findings reported by Golshani and colleagues showing that firing patterns of mouse cortical neurons *in vivo* become less synchronized over a period comparable with the age of these *in vitro* cultures (Golshani et al., 2009). However, given the relatively simple topography of the networks detected in these cultures caution should be taken when drawing conclusions based on changes in density parameters. Further, despite this seemingly steady decline in network density with maturation of cultures, the topogra-

phy within individual networks can be seen to change dramatically between recordings as shown in Figure 11D.

Paradigms of spike cluster firing patterns

Given the stochasticity observed in the firing patterns of spike clusters over successive recordings, the firing pattern of any one spike cluster can be thought of as adhering to one of four paradigms: (1) persistent, (2) lost, (3) recovering, or (4) emergent. (1) persistent, pertaining to clusters active during all recordings of interest; (2) lost, pertaining to those that are active during initial recordings but not observed in subsequent recordings, either as the result of chance or the consequence of an experimental perturbation; (3) recovering, pertaining to those that are active during initial recordings, inactive during interim recordings, yet are observed again in subsequent recordings, similarly either as the result of chance or the consequence of an experimental perturbation; and lastly, (4) emergent, pertaining to those clusters that are absent during initial recordings but observed during subsequent recordings. The emergent class harbors the caveat that these clusters may have been absent during initial recordings due to chance, or that the cluster represent waveforms from a cell active during an initial recording but then as a result of experimental perturbation produces waveforms with sufficiently different shape characteristics as to result in the identification of a distinct cluster than those produced originally. In this potential scenario, the analysis would identify two distinct clusters, mutually exclusive with respect to time, in which one is lost and the other is emergent. These classifications of firing patterns are not purely semantic, since there are practical implications for which classes can contribute to the analysis of different phenotypes. For frequency-dependent phenotypes, such as rates of firing, the absence of activity is itself relevant. However, for frequency-independent phenotypes, such as wave form amplitude (discussed below), the phenotype is dependent on activity such that the analysis needs to be restricted to those persistent clusters that present during all recordings when these frequency-independent phenotypes are being assessed.

Examining changes in firing frequency and voltage potential amplitude in response to exogenous treatment

To examine changes in two distinct phenotypes, (1) spontaneous firing frequency and (2) magnitude of voltage potentials, cultures were treated with two well characterized neurotoxins with different mechanisms of action. A pore-forming toxin, α HL derived from *Staphylococcus aureus* (Parker and Feil, 2005) and the Na^+ -channel blocker TTX (Narahashi et al, 1964). A pore-forming toxin was chosen for this analysis since an increase in conductance across the membrane resulting from perforation an ionophore such as α HL would be expected to cause a decrease in wave form amplitude.

Effect of neuronal toxin treatment on cell viability

To determine the maximum tolerated dose of these two toxins, a cell viability assay was performed to assess the

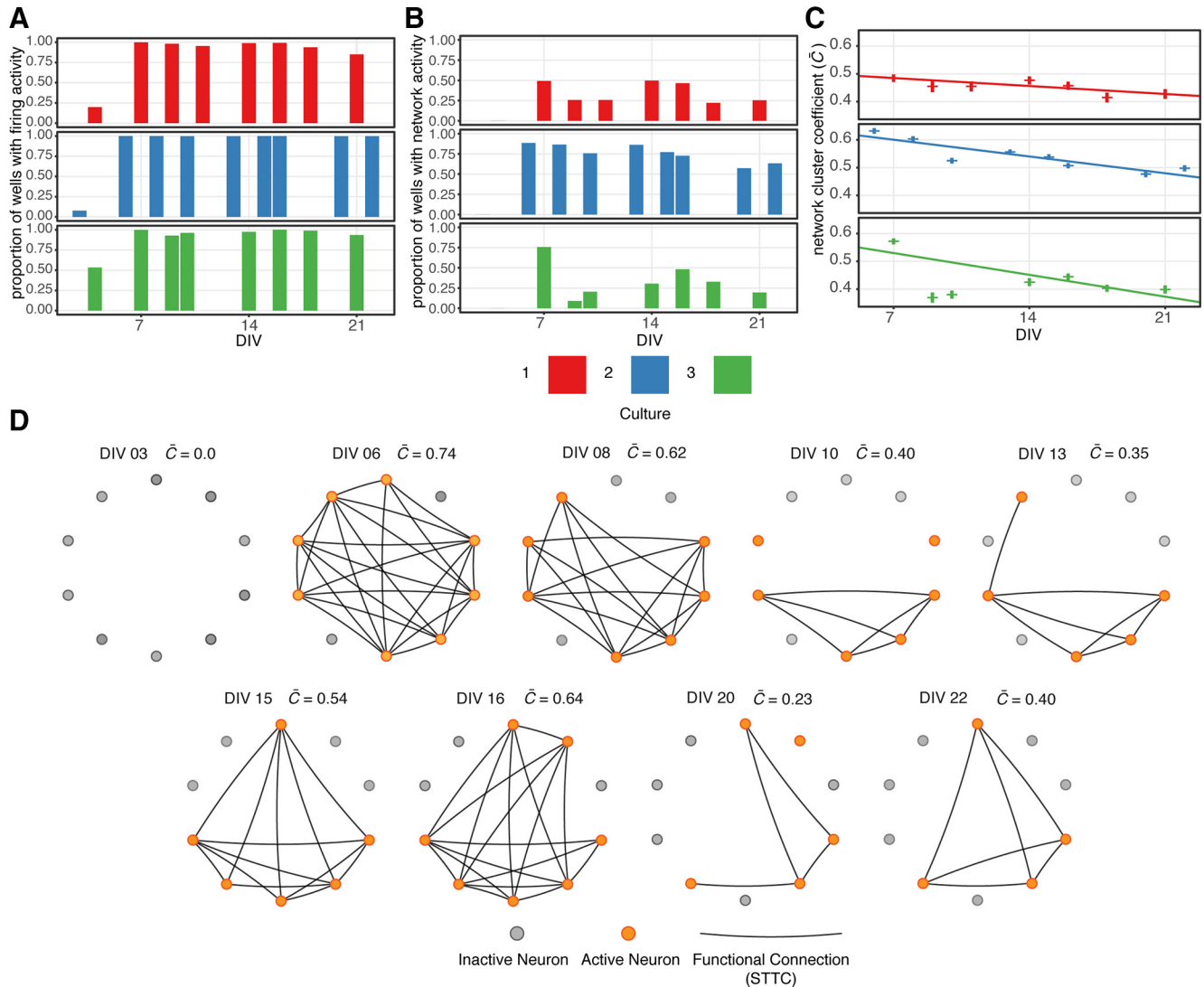


Figure 11. Pattern of network activity within cortical cultures. **A**, Proportion of arrays across three cultures exhibiting spontaneous firing activity recorded between DIV3 and DIV22. **B**, Proportion of arrays across three cultures exhibiting network activity as determined by non-zero, array level average network correlation coefficient. **C**, Change in average network correlation coefficient between DIV3 and DIV22. Bars indicate mean \pm SE, regression fit derived from linear mixed model. **D**, Depiction of network topography between 10 neurons within a single well (array) across nine recordings. Activity state of neuron within recorded indicated by vertex color, lines indicate functional connection between neurons as determined by STTC of significant magnitude.

induction of cell death in cultures following treatment with a titration of each agent. Cortical cultures were treated with titrations of α HL (0.25–16 μ g/ml) and TTX (0.001–1 μ M) for 24 h, after which viability was assessed by detection of lactate dehydrogenase (LDH) released into the culture media. As shown in **Figure 12B**, α HL was tolerated up to 4 μ g/ml causing an \sim 30% reduction in viability, while TTX had negligible effect on viability even at the highest dose tested. Based on these data, α HL was used at a maximum concentration of 4 μ g/ml in all subsequent experiments.

Effect of neuronal toxins on firing frequency

To assess whether these concentrations of TTX and α HL were sufficient to cause an effect on the generation

of action potentials, the frequency of spontaneous firing was assessed in cultures treated with a titration of TTX or α HL at 24 h following treatment, and again at 24 h following the removal of treatment (washout). The data were modeled using a Γ -GLM, estimating the average Hz across the spike clusters within each condition as a function of treatment and recording. Differences in firing frequencies between conditions were assessed using a general linear hypothesis test (Hothorn et al., 2008). As shown in **Figure 12**, treatment with 0.1 μ M TTX (0.0266 ± 0.001 Hz) and 1 μ M TTX ($0.000556 \pm 2e-5$ Hz) as well as 4 μ g/ml α HL (0.05885 ± 0.002 Hz) resulted in a significant decrease in firing frequency compared to treatment with media alone (0.24 ± 0.008 Hz), $p = 0.0148$, $p = 0.0001$,

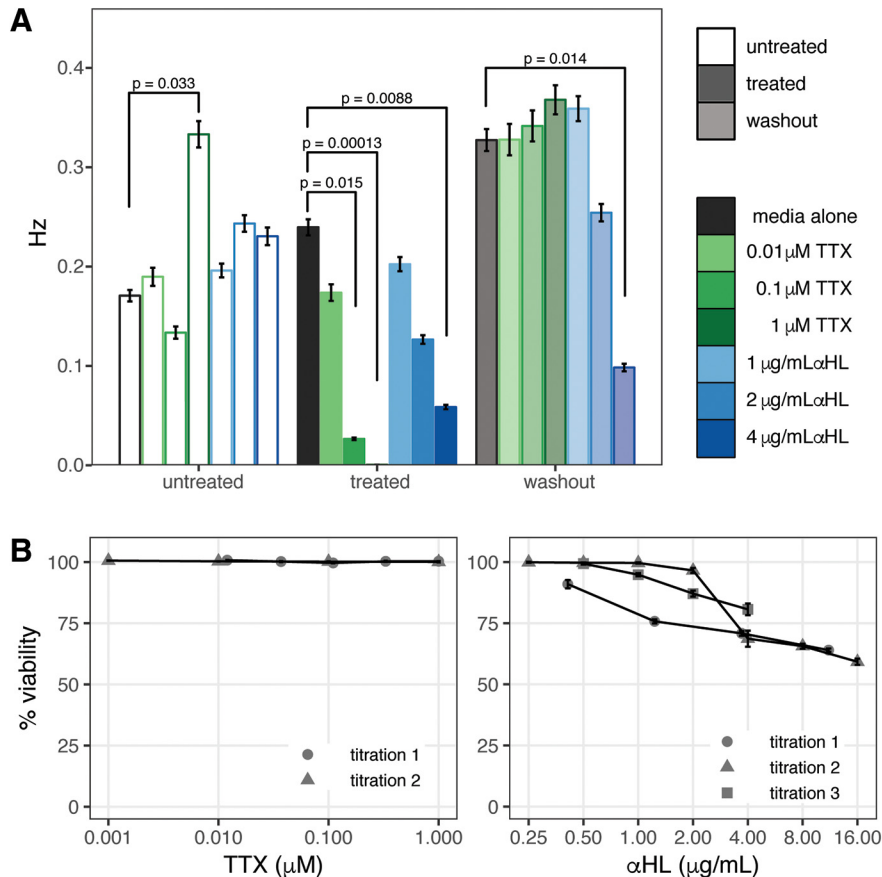


Figure 12. Spike cluster firing frequencies within TTX, αHL-treated cultures. **A**, Firing frequencies (Hz) of spike clusters within cortical cultures treated with titration of TTX (green) or αHL (blue). Frequencies from recordings across three times points: untreated (open bars), 24 h post treatment (filled bars), and washout 24 h following removal of treatment (half-filled bars). Treatment of 22–39 replicate wells per condition across four experiments. Data presented as mean ± SE Hz obtained from estimates of Γ-GLM modeling frequency as function of treatment and recording. Differences in frequencies assessed by general-linear hypothesis test, significance indicated by brackets and associated *p* values. **B**, Percentage viability of cortical cultures following 24-h treatment with TTX (left), or αHL (right) as assessed by LDH release. Data presented as mean ± SE, *n* = 5 per condition.

and *p* = 0.0088, respectively. While 2 μg/ml αHL (0.127 ± 0.004 Hz) also appeared to induce a decrease in firing relative to control, the difference was not significant, *p* = 0.234. Following the washout of treatment, the TTX-treated conditions show a complete restoration of firing activity, returning to levels comparable with that of media alone, while firing within the 4 μg/ml αHL treatment condition was still significantly reduced compared to media alone, *p* = 0.014. While the firing frequency within the 4 μg/ml αHL condition increased following the removal of treatment, to 0.098 ± 0.004 Hz from 0.059 ± 0.002 Hz, suggestive of some recovery, this increase was not significant, *p* = 0.82.

The observed decrease in firing frequency among the 4 μg/ml αHL-treated condition following treatment could be partially attributable to the effect on cell viability at this concentration (Figure 12B), given that the effect was seen both in the presence of the toxin and sustained following its removal. This analysis examined the average firing frequency across all clusters active during each recording, and not specifically the change in firing frequency among cluster that showed persistent activity across re-

cordings. Performing the analysis on the aggregate of all clusters is necessary to capture the effect of a treatment with the magnitude of effect of 1 μM TTX, in which there are no active clusters in the treated condition in which the comparison can be made.

Effect of neuronal toxin on voltage potential amplitude

Having examined changes in firing frequency at the cluster level, the next question was whether it is possible to assess disruptions in membrane integrity by assessing changes in the amplitude of voltage waveforms emitted from the generation of action potentials.

There is minimal correlation between the rate at which a neuron fires and the detected amplitude of the resulting voltage potential Figure 13A. This lack of correlation stems from the different factors underlying the generation of these properties and the manner in which they are detected in this system. The firing frequency of a neuron, absent any exogenous experimental stimulus, arises from the intrinsic properties of that particular class of neuron and the excitatory and inhibitory inputs it receives from other neurons that have formed synapses on it (Bean,

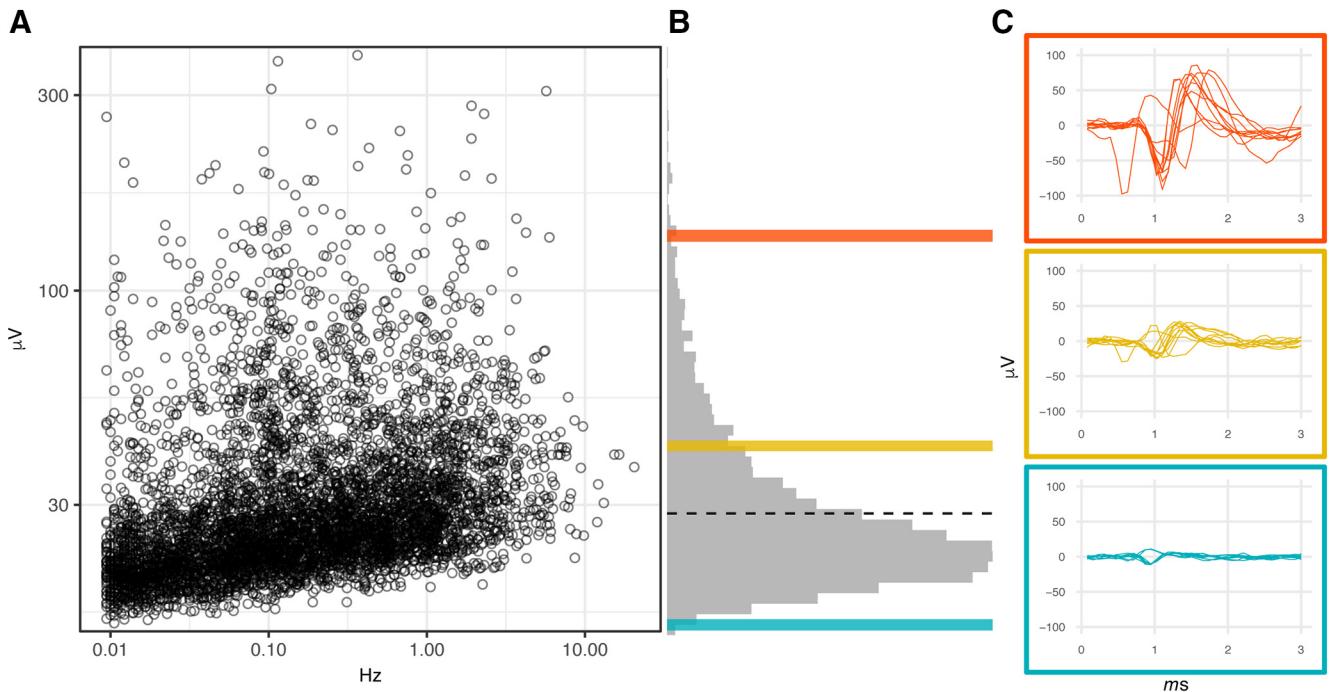


Figure 13. Comparison of firing frequency and wave form amplitude. **A**, Firing frequency (Hz) and wave form amplitude (μV) for 5506 spike clusters within untreated cortical cultures. Wave form amplitude determined by cluster centroid. **B**, Distribution of spike cluster wave form amplitudes, dash line indicating 30- μV threshold used for selecting clusters for assessing changes in wave form amplitude in response to treatment. **C**, Examples of low amplitude (teal), mid amplitude (gold), and high amplitude (persimmon) amplitude waveforms.

2007). Conversely, while the action potentials of all neurons are expected to be of the magnitude of $\sim 70\text{--}100\text{ mV}$ with respect to the charge disparity between the intracellular and extracellular space, the potential that is detected by the extracellular recording electrode of the MEA is dramatically influenced by the physical constraints of the system. If a firing neuron is modeled as a point source of charge, the extracellular potential V_e could be approximated from Coulomb's law, such that:

$$V_e = \frac{I}{4\pi d\sigma},$$

where I is amplitude of a point source of current (A), and d is the distance from the source (meters), and σ is the conductivity (Siemens/meters) of extracellular space (Gold et al., 2007). However, since neurons are better approximate by cylinders whose length is much greater than their transectional radius, the model is refined by the linear source approximation (LSA) described by Holt and Koch (1999). Thereby approximation of the extracellular potential is instead modeled as:

$$V_e = \frac{I}{4\pi\sigma\Delta s} \log \left| \frac{\sqrt{h^2 + r^2} - h}{\sqrt{r^2 - l^2} - l} \right|,$$

where Δs is the length of the cylinder, r is the radial distance from the cylinder, h is the longitudinal distance from the end of the cylinder, and $l = h + \Delta s$ representing the longitudinal distance from the beginning of the cylinder (Holt and Koch, 1999; Gold et al., 2006, 2007). The

practical consequence of this is that the wave form of extracellular voltage potential will vary drastically depending on the spatial orientation with respect to and distance between the firing neuron and the recording electrode, with an amplitude in the range of $\sim 10\text{--}100\ \mu\text{V}$, two to three orders of magnitude below the potential across the cell membrane (Gibson, 2012; Buzsáki et al., 2012). Taken together, these considerations explain the wide range in firing frequencies (Hz, x -axis) and voltage amplitude (μV , y -axis) observed across this population of spike clusters.

Since an increase in conductance across the membrane resulting from perforation by αHL would be expected to cause a decrease in wave form amplitude, the analysis was limited to spike clusters exhibiting a potential of at least 30 μV before treatment to allow for a sufficient dynamic range of response. This threshold is indicated by the dashed line on the histogram in center of Figure 13. Additionally, in an attempt to estimate the magnitude of the change in amplitude on individual cells, the analysis was limited to those persistent clusters that were active during both the untreated and treated recordings.

The estimate of the effect of treatment on wave form amplitude has to account for several factors within the data. First, the estimate of the amplitude of a given neuron during a recording period is derived from the random sample of firing events detected from that neuron, which will exhibit a certain degree of variability. Next, the amplitude across spike clusters varies greatly both before the initiation of treatment, and likely in response to treatment. Additionally, multiple spike clusters can be detected by a single electrode. Each electrode will exhibit a range of

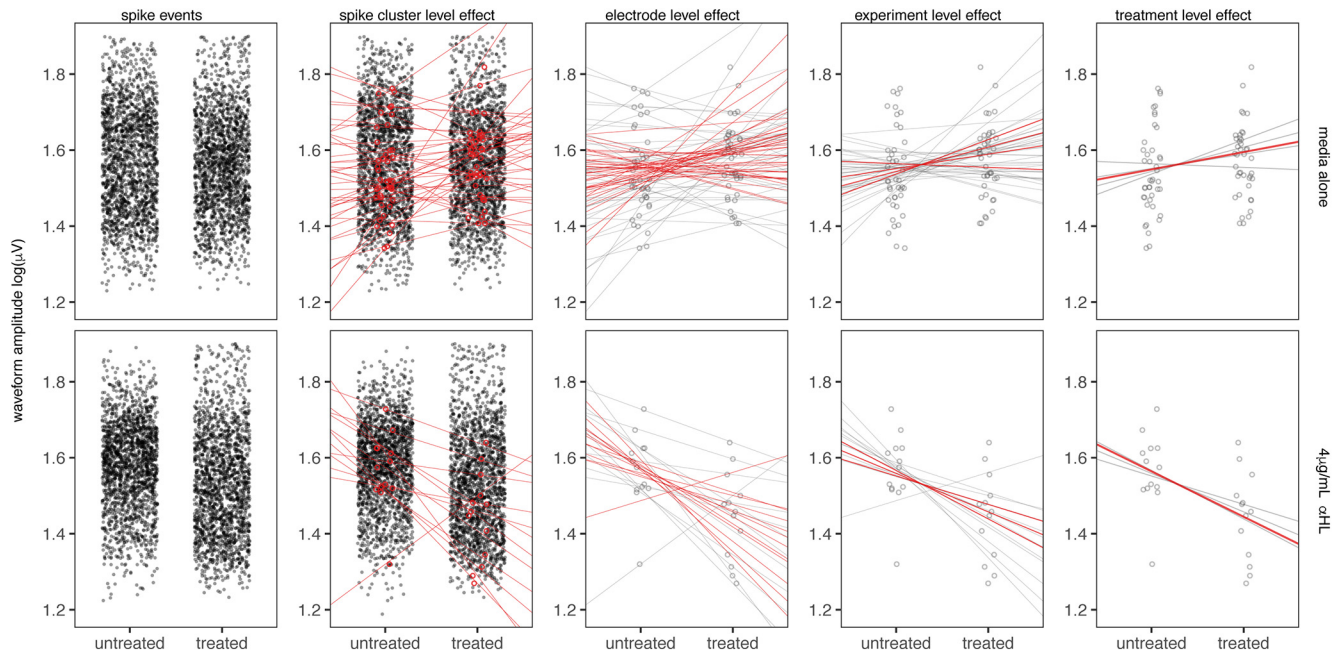


Figure 14. Determining change in wave form amplitude using linear-mixed model. Modeling effect on wave form amplitude in response to treatment media alone (top row) or 4 $\mu\text{g/ml}$ αHL (bottom row) across untreated and treated recordings. Estimates derived from amplitude values of individual waveforms (sample of 2000 shown per condition), and refined at level of spike cluster (open circles), electrodes, experiment, and treatment, indicated above each panel. Estimates at each level shown in red, with estimates of subsidiary level shown in gray.

sensitivity due to either the micro-environment within the culture in which they reside, and/or in the physical circuitry connecting them to the recording device. Finally, these data were collected across several experiments relying on separate preparations of cultures and reagents, which can contribute to additional variability. Therefore, to obtain an estimate for the effect of each treatment on wave form amplitude while accounting for these sources of variability, a linear mixed-effect model (LME) was constructed to predict $\log_{10}\mu\text{V}$ as a function of the fixed effects of treatment and recording (untreated and treated), while controlling for cluster, electrode, and experiment as nested-random effects (Pinheiro and Bates, 2000). The construction of the model is demonstrated in Figure 14 for two treatments, media alone and 4 $\mu\text{g/ml}$ αHL . Based on the raw data of individual waveforms, a series of linear regressions are fit, first based on waveforms within an individual spike cluster, then for all clusters detected by a common electrode, all electrodes within a single experiment, and finally for the effect of treatment across all experiments.

This method was applied to the recording data from treatment of cultures with a titration of TTX and αHL , to determine if an increase in membrane conductance resulting from exposure to a pore-forming toxin would result in a decrease in the amplitude of the voltage waveforms detected. The results in Figure 15 show that treatment with either 2 $\mu\text{g/ml}$ αHL ($-0.068 \pm 0.034 \Delta \log_{10}\mu\text{V}$) or 4 $\mu\text{g/ml}$ αHL ($-0.074 \pm 0.034 \Delta \log_{10}\mu\text{V}$) induced a significant decrease in wave form amplitude compared to treatment with media alone, $p = 0.006$ and $p = 0.003$, respectively. While the media alone condition appeared

have a slight increase in wave form amplitude ($0.046 \pm 0.027 \Delta \log_{10}\mu\text{V}$), this estimate is not significantly different from zero, $p = 0.09$. While treatment with 0.1 μM TTX ($-0.02 \pm 0.04 \Delta \log_{10}\mu\text{V}$) showed a slight decrease in wave form amplitude as well, this difference was not significant, $p = 0.1$. Treatment with 1 μM TTX could not be included in this analysis, since this concentration caused complete inhibition of firing activity.

It is worth noting that the significant effects on wave form amplitude are somewhat modest. The estimated change of $-0.074 \Delta \log_{10}\mu\text{V}$ within the 4 $\mu\text{g/ml}$ αHL -treated cultures translates to a potential difference on average of $-5.74 \mu\text{V}$ after accounting for the baseline amplitude among the spike clusters. That said, due to the attenuation of voltage potentials detected by extracellular electrodes as a result of distance from the neuron and resistance of the extracellular milieu the decrease in potential across the cell membrane is probably a 100 \times greater or more (Buzsáki et al., 2012). Despite this attenuation of signal, the resulting decrease in amplitude is still apparent. Figure 15B shows waveforms from the spike cluster within the 4 $\mu\text{g/ml}$ αHL treatment condition exhibiting the largest change in amplitude ($\Delta 14.2 \mu\text{V}$) between the pre-treatment and post-treatment recordings.

Lastly, is important to emphasize that despite treatment with 4 $\mu\text{g/ml}$ αHL inducing a slight decrease in cell viability (Figure 12B), any such decrease in viability would not be consequential for the effect on wave form amplitude, since these estimates were derived purely from those clusters active during both the untreated and treated recordings.

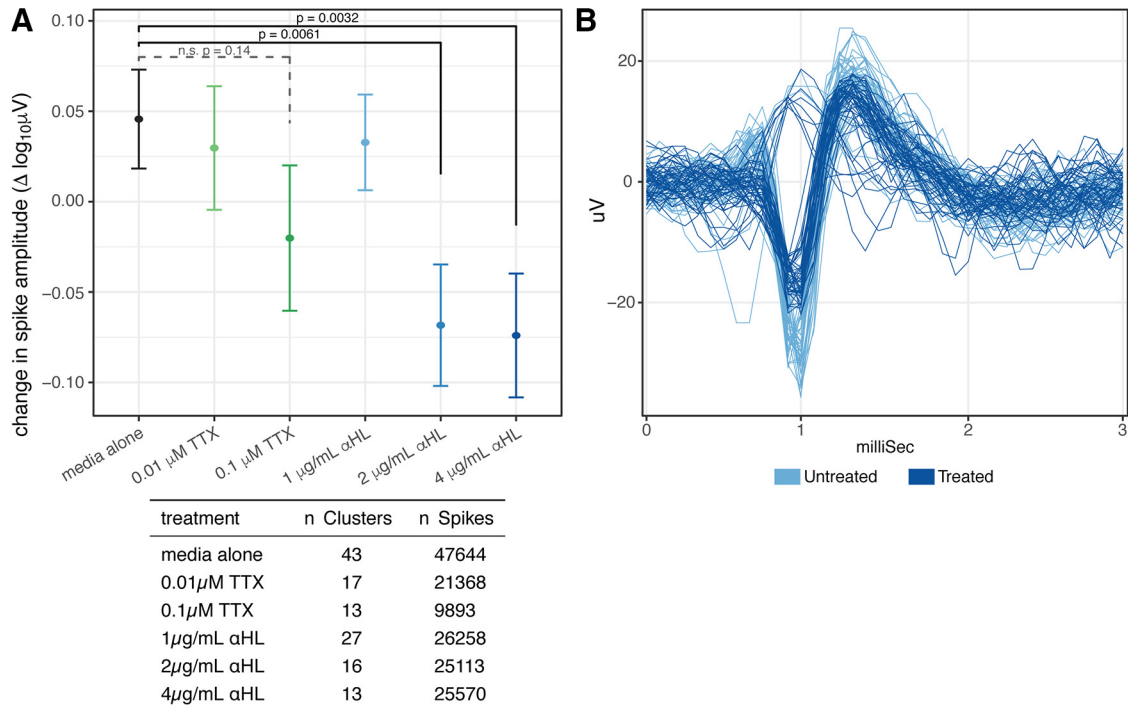


Figure 15. Change in wave form amplitude following treatment with pore-forming toxin. **A**, Estimated of change in spike wave form amplitude ($\Delta \log_{10} \mu\text{V}$) of spike clusters within cortical cultures treated with titration of TTX (green) or αHL (blue), across untreated (0 h) and treated (24 h) recordings. Treatment of 22–39 replicate wells per condition across four experiments. Data presented as mean \pm SE Hz obtained from LME model of wave form amplitude as function of treatment and recording. Differences in wave form amplitude assessed by general-linear hypothesis test, significance indicated by brackets and associated p values. Table indicates number of spike clusters, and spike events available for analysis in each condition. **B**, Comparison of waveforms within spike cluster demonstrating the largest change in amplitude (μV) among the 4 $\mu\text{g/mL}$ αHL treatment condition. Sample of 50 waveforms per recording shown, untreated (light blue) and treated (dark blue).

Discussion

Multi-well, MEAs are a compelling assay platform for neuroscience combining a label-free, functional, electrophysiological read-out with the ability to multiplex experimental conditions. The increased commercial availability of these instruments over the last decade has led to wider adoption and an increase in published reports of MEA-based experiments. However, despite this increased use, there are limited studies delineating rigorous assay development and experimental methods for use of these instruments.

Having generated recording data from a large population of over a thousand MEAs, firing frequencies at the array level were examined, revealing a highly skewed and highly variable distribution. It was shown that application of log-transformation should be used to obtain approximately normal distributions of firing frequencies before statistical analysis, and that implementation of a bootstrapping simulation can be used to accommodate for the broad distribution of firing frequencies to generate treatment with comparable levels of activity. Next, a series of simulated experiments were performed to determine the expected statistical power for a two condition, repeated measure experiment across a range of sample and effect sizes, based on the observed distribution of firing frequencies of arrays and the empirically determined correlation of firing in repeated recordings.

Assessing several methodological aspects of MEA experiments, many of the commonly reported methods appear adequate for estimating firing activity within *in vitro* neuronal cultures, while other elements of MEA experimental design and analysis have either been under-reported in the literature or investigators have employed techniques that may be inappropriate for the task. Specifically, recording duration of 20 min or longer as reported by several studies (Kuperstein et al., 2010; McConnell et al., 2012; Vincent et al., 2013; Slomowitz et al., 2015; Black et al., 2018; Feng et al., 2018), appear sufficiently long to capture the activity in individual arrays with a high degree of reproducibility. Similarly, the common practice of performing experiments with primary cultures that have been aged two to three weeks is reasonable, since the primary rat cortical cultures examined here showed a high degree of spontaneous firing by the end of the first week that plateaued through these time frames. Conversely, the highly skewed distribution of firing frequencies across both individual electrodes and entire arrays has been largely omitted from reports or accounted for in presentation of results, with a limited number of noted exceptions (Biffi et al., 2013; Vincent et al., 2013; Slomowitz et al., 2015; Wainger et al., 2015; Black et al., 2018). Despite this observation, several of the reviewed publications used parametric statistical tests such as Student's t test and ANOVA for assessing differ-

ences in firing frequency between MEAs. This is concerning given that these tests assume a normally distributed dependent variable, a condition that these data fail to meet when examined on a linear scale. Further, some studies used these methods for analyzing longitudinal experiments (Feng et al., 2018; Sarkar et al., 2018), which similarly violate the underlying assumptions of the common forms of these tests which expect measurements to be independent and do not account for the anticipated correlation of values between repeated measurements of the same subject. Rather, ANCOVA and other linear regression based methods that handle time as covariate within the data would be more appropriate (Diggle et al., 2002; Singer and Willett, 2003), as was employed here for the simulation of treatment or as was reported by Wainger et al. (2015).

Despite the increasing number of studies using multi-well MEAs, none of the reports examined described methods for accounting for the wide variability of firing frequencies when assigning treatment groups. Given the ability to easily record MEAs repeatedly, implementation of a simulation-based assignment technique using baseline recording data such as described here seems adequate for the purpose. While the use of complex treatment maps generated by the process may seem impractical for the addition of experimental perturbations, it is certainly feasible for experiments being performed manually, and is trivial for MEA recording systems that are integrated with automated liquid-handlers as is the case with instruments such as the Maestro Apex (Axion Biosystems, Hamilton Robotics).

Using simulation methods incorporating data on the variability and reproducibility of MEA recordings to calculate statistical power of experiments showed the magnitude of effects that can be detected with confidence this system. Comparing these simulated results with empirical data from treatment of cultures with the neurotoxin TTX shows that it is possible to detect changes in firing frequency similar to those induced by treatment with TTX at its IC_{50} using treatment groups containing eight to 10 replicates. Conversely, detecting large changes in firing, such as those induced by high doses of TTX ($10 \times IC_{50}$), only requires a few replicates, while detecting more modest changes will require many more replicates. This is an important consideration for experimental design in the different research areas in which MEAs are being used. For toxicology studies in which the effect sizes of pharmacological agents may be larger, a few replicates per condition may suffice, compared to disease modeling studies assessing the functional effects of genetic variants where the effect sizes are expected to be smaller. For these studies, much larger numbers of arrays would be required to discern statistically significant differences.

Further, the high-frequency (>10 kHz) sampling rates of typical MEA instruments can allow for greater resolution of recording data, attributing events to individual cells of origin through the use of spike sorting analysis. Doing so can allow for the data acquired through MEA recordings to be extended beyond aggregate array-level (well-level) activity metrics to more refined phenotypes including

changes in: the spiking frequency of individual cells, the amplitude of action potentials, and the network activity neurons revealed through coordinated firing patterns. Through this work, a spike sorting analysis pipeline was developed implemented in the open source R statistical programming language which provides flexibility to perform spike sorting across data from multiple recordings, and that it can be executed within a parallel computing environment, using multiple CPUs within a high-performance computing cluster. The ability to perform spike sorting across multiple recordings is an important distinction of the pipeline developed here relative to other commercially available software. This is important, given the longitudinal nature of many MEA experiments, and the problems with sorting separately for each time point. Further, the ability to leverage parallel computing is advantageous given the large number of recording channels to be assessed (768 within the instrument used for this study) and the time and computationally intensive nature of the mean-shift algorithm underlying the analysis.

Utilizing this analysis pipeline, untreated cultures examined over multiple days were revealed to exhibit highly stochastic patterns of spontaneous firing among individual cells. This observation provided essential grounding for how to draw inferences about changes in firing patterns across populations of spike clusters, as demonstrated with the use of Γ -GLMs. Additionally, to assess changes in the amplitude of action potentials originating from individual cells, spike sorting analysis was performed on recording data of cultures treated with a pore-forming toxin, α HL, and non-pore forming, channel blocking toxin, TTX. This analysis showed a significant decrease in the voltage potentials within spike clusters of cultures following treatment with α HL, consistent with the expected changes in voltage potentials resulting from perforation of the cellular membrane. This type of analysis may have broader application in the field of toxicology where MEA systems are commonly used. The technique described here for monitoring changes in action potential amplitude is analogous to the methods used for detecting changes in QT intervals in cardiomyocytes recorded using MEAs (Tertoolen et al., 2018). Finally, using these data in conjunction with the STTC method described by Cutts and Eglén (2014) and network cluster coefficient described by Watts and Strogatz (1998), it was shown how the network activity within cultures can be monitored using MEA recordings by assessing the changes in coordinated firing across individual neurons. While previous studies have examined network dynamics within cultures on *in vitro* MEAs, these have typically been carried without first identifying individual cells through spike sorting and instead have examined the temporal correlation of events at the electrode level (Vincent et al., 2013). This obfuscates elements of the network, since activity of multiple cells is likely being aggregated on a single electrode. Incorporating the techniques describe here will facilitate experiments assaying for changes in network dynamics rather than simply aggregated spike frequency. The ability to use network phenotypes as an assay readout will improve as multi-well MEA systems with higher electrode density

become available allowing for better resolution of network activity *in vitro*. This is relevant for modeling disease biology, as the changes in neuronal activity resulting from the pathophysiology of conditions such as Alzheimer's disease are likely to manifest as aberrant firing patterns rather than simply an increase or decrease in the frequency of spiking events (Palop and Mucke, 2010).

References

- Alshawaf AJ, Viventi S, Qiu W, D'Abaco G, Nayagam B, Erlichster M, Chana G, Everall I, Ivanusic J, Skafidas E, Dottori M (2018) Phenotypic and functional characterization of peripheral sensory neurons derived from human embryonic stem cells. *Sci Rep* 8:603.
- Bardy C, van den Hurk M, Eames T, Marchand C, Hernandez RV, Kellogg M, Gorris M, Galet B, Palomares V, Brown J, Bang AG, Mertens J, Böhnke L, Boyer L, Simon S, Gage FH (2015) Neuronal medium that supports basic synaptic functions and activity of human neurons *in vitro*. *Proc Natl Acad Sci USA* 112:E2725–E2734.
- Bateup HS, Deneffrio CL, Johnson CA, Saulnier JL, Sabatini BL (2013a) Temporal dynamics of a homeostatic pathway controlling neural network activity. *Front Mol Neurosci* 6:28.
- Bateup HS, Johnson CA, Deneffrio CL, Saulnier JL, Kornacker K, Sabatini BL (2013b) Excitatory/inhibitory synaptic imbalance leads to hippocampal hyperexcitability in mouse models of tuberosclerosis. *Neuron* 78:510–522.
- Bean BP (2007) The action potential in mammalian central neurons. *Nat Rev Neurosci* 8:451–465.
- Biffi E, Regalia G, Menegon A, Ferrigno G, Pedrocchi A (2013) The influence of neuronal density and maturation on network activity of hippocampal cell cultures: a methodological study. *PLoS One* 8:e83899.
- Black BJ, Atmaramani R, Kumaraju R, Plagens S, Romero-Ortega M, Dussor G, Price TJ, Campbell ZT, Pancrazio JJ (2018) Adult mouse sensory neurons on microelectrode arrays exhibit increased spontaneous and stimulus-evoked activity in the presence of interleukin-6. *J Neurophysiol* 120:1374–1385.
- Blake KJ, Baral P, Voisin T, Lubkin A, Pinho-Ribeiro FA, Adams KL, Roberson DP, Ma YC, Otto M, Woolf CJ, Torres VJ, Chiu IM (2018) *Staphylococcus aureus* produces pain through pore-forming toxins and neuronal TRPV1 that is silenced by QX-314. *Nat Commun* 9:37.
- Buzsáki G, Anastassiou CA, Koch C (2012) The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* 13:407–420.
- Catterall WA, Goldin AL, Waxman SG (2005) International Union of Pharmacology. XLVII. Nomenclature and structure-function relationships of voltage-gated sodium channels. *Pharmacol Rev* 57:397–409.
- Charkhkar H, Meyyappan S, Matveeva E, Moll JR, McHail DG, Peixoto N, Cliff RO, Pancrazio JJ (2015) Amyloid beta modulation of neuronal network activity *in vitro*. *Brain Res* 1629:1–9.
- Chen C, Chen L, Lin Y, Zeng S, Luo Q (2006) The origin of spontaneous synchronized burst in cultured neuronal networks based on multi-electrode arrays. *Biosystems* 85:137–143.
- Ciollaro M, Wang D (2016) MeanShift: clustering via the mean shift algorithm. Available at <https://cran.r-project.org/src/contrib/Archive/MeanShift/>.
- Clemons PA, Tolliday NJ, Wagner BK, eds. (2009) Cell-based assays for high-throughput screening. In: *Methods in molecular biology*, Vol. 486. Totowa, NJ: Humana Press.
- Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal Mach Intell* 24:603–619.
- Csardi G, Nepusz T (2006) The Igraph software package for complex network research. *InterJournal*. Volume Complex Systems. Available at <http://igraph.org>.
- Cutts CS, Eglen SJ (2014) Detecting pairwise correlations in spike trains: an objective comparison of methods and application to the study of retinal waves. *J Neurosci* 34:14288–14303.
- DeRosa BA, El Hokayem J, Artimovich E, Garcia-Serje C, Phillips AW, Van Booven D, Nestor JE, Wang L, Cuccaro ML, Vance JM, Pericak-Vance MA, Cukier HN, Nestor MW, Dykxhoorn DM (2018) Convergent pathways in idiopathic autism revealed by time course transcriptomic analysis of patient-derived neurons. *Sci Rep* 8:8423.
- Diggle PJ, Heagerty P, Liang KY, Zeger ST (2002) Analysis of longitudinal data. Oxford Statistical Science Series. Oxford: Oxford University Press.
- Ekström CT (2018) MESS: miscellaneous esoteric statistical scripts. Available at <https://CRAN.R-project.org/package=MESS>.
- Feng X, Bader BM, Yang F, Segura M, Schultz L, Schröder OHU, Rolfs A, Luo J (2018) Improvement of impaired electrical activity in NPC1 mutant cortical neurons upon DHPG stimulation detected by micro-electrode array. *Brain Res* 1694:87–93.
- Frega M, Pasquale V, Tedesco M, Marcoli M, Contestabile A, Nanni M, Bonzano L, Maura G, Chiappalone M (2012) Cortical cultures coupled to micro-electrode arrays: a novel approach to perform *in vitro* excitotoxicity testing. *Neurotoxicol Teratol* 34:116–127.
- Fukunaga K, Hostetler L (1975) The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans Inf Theory* 21:32–40.
- García-León JA, Cabrera-Socorro A, Eggermont K, Swißen A, Terry J, Fazal R, Nami F, Ordovás L, Quiles A, Lluís F, Serneels L, Wierda K, Sierksma A, Kreir M, Pestana F, Van Damme P, De Strooper B, Thorrez L, Ebner A, Verfaillie CM (2018) Generation of a human induced pluripotent stem cell-based model for tauopathies combining three microtubule-associated protein tau mutations which displays several phenotypes linked to neurodegeneration. *Alzheimers Dement* 14:1261–1280.
- Gibson SP (2012) Neural spike sorting in hardware: from theory to practice. PhD thesis, University of California, Los Angeles.
- Gold C, Henze DA, Koch C, Buzsáki G (2006) On the origin of the extracellular action potential waveform: a modeling study. *J Neurophysiol* 95:3113–3128.
- Gold C, Henze DA, Koch C (2007) Using extracellular action potential recordings to constrain compartmental models. *J Comput Neurosci* 23:39–58.
- Golshani P, Gonçalves JT, Khoshkhou S, Mostany R, Smirnakis S, Portera-Cailliau C (2009) Internally mediated developmental desynchronization of neocortical network activity. *J Neurosci* 29:10890–10899.
- Henningson M, Illes S (2017) Analysis and modeling of subthreshold neural multi-electrode array data by statistical field theory. *Front Comput Neurosci* 11:26.
- Holt GR, Koch C (1999) Electrical interactions via the extracellular potential near cell bodies. *J Comput Neurosci* 6:169–184.
- Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biom J* 50:346–363.
- Illes S, Jakab M, Beyer F, Gelfert R, Couillard-Despres S, Schnitzler A, Ritter M, Aigner L (2014) Intrinsically active and pacemaker neurons in pluripotent stem cell-derived neuronal populations. *Stem Cell Rep* 2:323–336.
- Kassambara A (2018) Ggpubr: 'Ggplot2' based publication ready plots. Available at <https://CRAN.R-project.org/package=ggpubr>.
- Kim KH, Kim SJ (2000) Neural spike sorting under nearly 0-dB signal-to-noise ratio using nonlinear energy operator and artificial neural-network classifier. *IEEE Trans Biomed Eng* 47:1406–1411.
- Kuperstein I, Broersen K, Benilova I, Rozenski J, Jonckheere W, Debulpaep M, Vandersteen A, Segers-Nolten I, Van Der Werf K, Subramaniam V, Braeken D, Callewaert G, Bartic C, D'Hooge R, Martins IC, Rousseau F, Schymkowitz J, De Strooper B (2010) Neurotoxicity of Alzheimer's disease A β peptides is induced by small changes in the A β 42 to A β 40 ratio. *EMBO J* 29:3408–3420.
- Liu DC, Seimetz J, Lee KY, Kalsotra A, Chung HJ, Lu H, Tsai NP (2017) Mdm2 mediates FMRP- and Gp1 mGluR-dependent pro-

- tein translation and neural network activity. *Hum Mol Genet* 26: 3895–3908.
- McConnell ER, McClain MA, Ross J, Lefew WR, Shafer TJ (2012) Evaluation of multi-well microelectrode arrays for neurotoxicity screening using a chemical training set. *Neurotoxicology* 33:1048–1057.
- McCullagh P, Nelder JA (1989) *Generalized linear models*. Chapman and Hall/Crc Monographs on Statistics and Applied Probability Series, Ed 2. Boca Raton: Chapman & Hall.
- Narahashi T, Moore JW, Scott WR (1964) Tetrodotoxin blockage of sodium conductance increase in lobster giant axons. *J Gen Physiol* 47:965–974.
- Nehme R, Zuccaro E, Ghosh SD, Li C, Sherwood JL, Pietilainen O, Barrett LE, Limone F, Worringer KA, Komminen S, Zang Y, Cacchiarelli D, Meissner A, Adolfsen R, Haggarty S, Madison J, Muller M, Arlotta P, Fu Z, Feng G, et al. (2018) Combining NGN2 programming with developmental patterning generates human excitatory neurons with NMDAR-mediated synaptic transmission. *Cell Rep* 23:2509–2523.
- Novellino A, Scelfo B, Palosaari T, Price A, Sobanski T, Shafer TJ, Johnstone AF, Gross GW, Gramowski A, Schroeder O, Jügel K, Chiappalone M, Benfenati F, Martinoia S, Tedesco MT, Defranchi E, D'Angelo P, Whelan M (2011) Development of micro-electrode array based tests for neurotoxicity: assessment of interlaboratory reproducibility with neuroactive chemicals. *Front Neuroeng* 4:4.
- Obien MEJ, Deligkaris K, Bullmann T, Bakkum DJ, Frey U (2014) Revealing neuronal function through microelectrode array recordings. *Front Neurosci* 8:423.
- Palop JJ, Mucke L (2010) Amyloid-beta-induced neuronal dysfunction in Alzheimer's disease: from synapses toward neural networks. *Nat Neurosci* 13:812–818.
- Parker MW, Feil SC (2005) Pore-forming protein toxins: from structure to function. *Prog Biophys Mol Biol* 88:91–142.
- Pinheiro J, Bates D, DebRoy S, Sarkar D; R Core Team (2018) nlme: linear and nonlinear mixed effects models. Available at <https://CRAN.R-project.org/package=nlme>.
- Pinheiro JC, Bates DM (2000) *Mixed-effects models in sand S-PLUS*. In: *Statistics and Computing*. New York, NY: Springer New York.
- R Core Team (2017) *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Russo FB, Freitas BC, Pignatari GC, Fernandes IR, Sebat J, Muotri AR, Beltrão-Braga PCB (2018) Modeling the interplay between neurons and astrocytes in autism using human induced pluripotent stem cells. *Biol Psychiatry* 83:569–578.
- Sarkar A, Mei A, Paquola ACM, Stern S, Bardy C, Klug JR, Kim S, Neshat N, Kim HJ, Ku M, Shokhirev MN, Adamowicz DH, Marchetto MC, Jappelli R, Erwin JA, Padmanabhan K, Shtrahman M, Jin X, Gage FH (2018) Efficient generation of CA3 neurons from human pluripotent stem cells enables modeling of hippocampal connectivity in vitro. *Cell Stem Cell* 22:684–697.e9.
- Segev R, Shapira Y, Benveniste M, Ben-Jacob E (2001) Observations and modeling of synchronized bursting in two-dimensional neural networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 64: 011920.
- Singer JD, Willett JB (2003) *Applied longitudinal data analysis: modeling change and event occurrence*. New York: Oxford University Press.
- Slomowitz E, Styr B, Vertkin I, Milshtein-Parush H, Nelken I, Slutsky M, Slutsky I (2015) Interplay between population firing stability and single neuron dynamics in hippocampal networks. *Elife* 4:e04378.
- Spira ME, Hai A (2013) Multi-electrode array technologies for neuroscience and cardiology. *Nat Nanotechnol* 8:83–94.
- Tertoolen LGJ, Braam SR, van Meer BJ, Passier R, Mummery CL (2018) Interpretation of field potentials measured on a multi electrode array in pharmacological toxicity screening on primary and human pluripotent stem cell-derived cardiomyocytes. *Biochem Biophys Res Commun* 497:1135–1141.
- Van Pelt J, Corner MA, Wolters PS, Rutten WLC, Ramakers GJA (2004) Longterm stability and developmental changes in spontaneous network burst firing patterns in dissociated rat cerebral cortex cell cultures on multielectrode arrays. *Neurosci Lett* 361: 86–89.
- Varghese K, Molnar P, Das M, Bhargava N, Lambert S, Kindy MS, Hickman JJ (2010) A new target for amyloid beta toxicity validated by standard and high-throughput electrophysiology. *PLoS One* 5:e8643.
- Venables WN, Ripley BD (2002) *Modern applied statistics with S. Statistics and Computing*. San Diego: Springer.
- Vertkin I, Styr B, Slomowitz E, Ofir N, Shapira I, Berner D, Fedorova T, Laviv T, Barak-Broner N, Greitzer-Antes D, Gassmann M, Bettler B, Lotan I, Slutsky I (2015) GABAB receptor deficiency causes failure of neuronal homeostasis in hippocampal networks. *Proc Natl Acad Sci USA* 112:E3291–E3299.
- Vickers AJ (2001) The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Methodol* 1:6.
- Vincent F, Loria P, Pregel M, Stanton R, Kitching L, Nocka K, Doyonnas R, Stepan C, Gilbert A, Schroeter T, Peakman MC (2015) Developing predictive assays: the phenotypic screening “rule of 3.” *Sci Transl Med* 7:293ps15.
- Vincent K, Tauskela JS, Mealing GA, Thivierge JP (2013) Altered network communication following a neuroprotective drug treatment. *PLoS One* 8:e54478.
- Wainger BJ, Kiskinis E, Mellin C, Wiskow O, Han SSW, Sandoe J, Perez NP, Williams LA, Lee S, Boulting G, Berry JD, Brown RH Jr, Cudkowicz ME, Bean BP, Eggen K, Woolf CJ (2014) Intrinsic membrane hyperexcitability of amyotrophic lateral sclerosis patient-derived motor neurons. *Cell Rep* 7:1–11.
- Wainger BJ, Buttermore ED, Oliveira JT, Mellin C, Lee S, Saber WA, Wang AJ, Ichida JK, Chiu IM, Barrett L, Huebner EA, Bilgin C, Tsujimoto N, Brenneis C, Kapur K, Rubin LL, Eggen K, Woolf CJ (2015) Modeling pain in vitro using nociceptor neurons reprogrammed from fibroblasts. *Nat Neurosci* 18:17–24.
- Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* 393:440–442.
- Weir K, Blanquie O, Kilb W, Luhmann HJ, Sinning A (2014) Comparison of spike parameters from optically identified GABAergic and glutamatergic neurons in sparse cortical cultures. *Front Cell Neurosci* 8:460.
- Wickham H (2016) *Ggplot2: elegant graphics for data analysis*. New York: Springer.