

Characterization of Early Cortical Neural Network Development in Multiwell Microelectrode Array Plates⁺

Ellese Cotterill^{1#} and Diana Hall^{2#}, Kathleen Wallace², William R Mundy², Stephen J Eglon^{1†}, Timothy J Shafer^{2†}.

¹ Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, United Kingdom, and ²Office of Research and Development, National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, Research Triangle Park, NC

Corresponding Author:

Timothy J Shafer, Ph.D.
Integrated Systems Toxicology Division, MD-B105-03
Office of Research and Development
U.S. Environmental Protection Agency
Research Triangle Park, NC 27711

Phone: 919-541-0647

Fax: 919-541-4849

Email: Shafer.tim@epa.gov

⁺Preparation of this document has been funded in part by the U.S. Environmental Protection Agency. This document has been reviewed by the National Health and Environmental Effects Research Laboratory and approved for publication. Approval does not signify that the contents reflect the views of the Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

[#], these authors contributed equally to this study.

[†]These authors co-directed this study.

Acknowledgements

EC was supported by a Wellcome Trust PhD studentship and NIHR Cambridge Biomedical Research Centre studentship.

DH was supported by student services contract #EP-13-D-000108 and by a travelling fellowship from the Company of Biologists.

Abstract

We examined neural network ontogeny using microelectrode array (MEA) recordings made in multi-well MEA plates (mwMEAs) over the first 12 days in vitro (DIV). In primary cortical cultures, action potential spiking activity developed rapidly between DIV 5 and 12. Spiking was sporadic and unorganized at early DIV, and became progressively more organized with time, with bursting parameters, synchrony and network bursting increasing between DIV 5 and 12. We selected 12 features to describe network activity; principal components analysis using these features demonstrated segregation of data by age at both the well and plate levels. Using random forest classifiers and Support Vector Machines, we demonstrated that 4 features (CV of within burst ISI, CV of IBI, network spike rate and burst rate) could predict the age of each well recording with >65% accuracy. When restricting the classification to a binary decision, accuracy improved to as high as 95%. Further, we present a novel resampling approach to determine the number of wells needed for comparing different treatments. Overall, these results demonstrate that network development on mwMEA plates is similar to development in single-well MEAs. The increased throughput of mwMEAs will facilitate screening drugs, chemicals or disease states for effects on neurodevelopment.

Introduction

Microelectrode array (MEA) recordings are a useful tool to study the activity of networks of interconnected neurons, both in vitro and in vivo. In vitro, neural networks on MEAs demonstrate many characteristics of intact neural networks; this includes extracellular recordings of action potentials (“spikes”) and groups of action potentials (“bursts”) simultaneously from multiple points in the network¹. The spontaneous activity in these networks exhibits pharmacological responsiveness and plasticity²⁻⁵. Thus, primary cultures of neural networks on MEAs have been widely utilized to study neurophysiology, neuropharmacology and neurotoxicology (for review, see ⁶). In addition, the ontogeny of network activity on MEAs has been described by numerous different laboratories ^{4,7-10}. Until recently, however, the throughput of MEA devices has been limited, such that it was not possible to study more than a small handful (e.g. 4-6) of networks at a time.

Recently, two manufacturers of MEA devices have introduced multi-well MEA (mwMEA) devices, which allow for recordings to be made from 12-96 wells simultaneously, with 8-64 electrodes per well. The increase in throughput offered by mwMEA devices expands the capabilities of MEA systems, allowing for drug and toxicant screening^{11,12}. Further, when combined with cultures that have undergone genomic manipulation^{13,14} or with patient-derived inducible pluripotent stem cells^{15,16}, mwMEA recordings have been used to describe how network function is affected by disease states. Finally, mwMEAs offer the ability to screen large numbers of chemicals for potential effects on developing networks^{17,18}. Given the significant public concern over the potential role of

chemicals in neurodevelopmental diseases¹⁹, study of chemical effects on the early stages of neural network ontogeny using mwMEAs offer a functional measure for developmental neurotoxicity hazard characterization.

However, in order for such studies to take place, the basic development of activity in networks on mwMEAs needs to be described in detail. In lower throughput MEA systems, neural network development has been demonstrated to transition from low activity at early developmental ages (e.g. the first week in vitro), to one of coordinated bursting, network spikes and synchrony at later time points (e.g. the second week in vitro and beyond). While it is expected that such properties will be retained in multi-well systems, it remains to be demonstrated, and the time-course, variability and other characteristics defined. Further, while single-well MEAs contain ~60 microelectrodes, only 12 well mwMEA plates have an equivalent number (64/well), and the extent to which network properties can be defined with fewer electrodes has not been determined.

The present studies describe the development of networks of mixed primary cortical cultures in 48 well mwMEA plates containing 16 microelectrodes/well. These cultures were prepared from newborn rat cortex and contain excitatory and inhibitory neurons as well as glia^{20,21}. Over the first two weeks in vitro, the neurons extend axons and dendrites²², form synapses²³ and, in single well MEA systems, develop spontaneous network activity¹⁷. The present studies characterized the early ontogeny of activity of these cultures in mwMEAs by describing the firing, bursting, synchrony and network spike properties over the first 12 days in vitro (DIV), as previous studies with this culture model have shown that a significant change in the rates and patterns of activity occur over this time-frame. Further, we sought to evaluate the utility of analysis of multiple features of

network activity as a method to determine the ability of classification approaches to distinguish between cultures under different conditions (e.g. control vs drug/ toxicant treatment, different ages, or genetically-modified vs wild type).

Methods

Experimental Protocol

Cell Culture. All procedures using animals were approved by the National Health and Environmental Effects Laboratory Institutional Animal Use and Care Committee. Primary cultures were prepared from the cortex of 0-24 h old rat pups as described previously^{11,20,21}. Cells were plated (1.5×10^5 cells in a 25 μ l drop of media) onto the surface of 48 well MEA plates (16 electrodes/well) that had been pre-coated with polyethylenimine (PEI) and laminin as previously described¹¹. The resultant cultures contain excitatory and inhibitory neurons and glia (Supplemental Figure 1).

MEA Recordings. Spontaneous network activity was recorded using Axion Biosystems Maestro 768 channel amplifier and Axion Integrated Studios (AxIS) v1.9 (or later) software. The amplifier recorded from all channels simultaneously using a gain of 1200x and a sampling rate of 12.5 kHz/channel. After passing the signal through a Butterworth band-pass filter (300-5000 Hz) online spike detection (threshold = 8x rms noise on each channel) was done with the AxIS adaptive spike detector. On days in vitro (DIV) 5, 7, 9, and 12, plates were placed into the Maestro amplifier and allowed at least 5 min to equilibrate, after which at least 15 min of activity was recorded. As the majority of these plates were used on or around DIV 14 for other experiments, short equilibration and recording times (~30 min total) were selected to minimize the potential impact of repeated

removal of cells from the incubator over the time period during which activity was developing. All recordings were conducted at 37 °C, and since development of activity was being studied, there were no *a priori* thresholds for minimum numbers of active electrodes for inclusion of a well in the data set.

Data

Recordings were made from 656 wells across 16 MEA plates from 15 primary cortical cultures at DIV 5, 7, 9 and 12 for a total of 64 “plate-recordings”. A recording of one plate at DIV 12 was missing from our data set, and one recording at DIV 7 was also excluded from the analysis as its mean firing rate (1.2 Hz) was greater than 2 SD above that of the other DIV 7 plates (mean = 0.6Hz, sd = 0.3). This resulted in a total of 62 plate recordings, with a total of 2976 well recordings, used in our analysis. Activity was usually recorded for 15—30 minutes; only the last 15 minutes of each recording was analysed. Features related to spikes, bursts, network spikes²⁴ and correlations²⁵ were extracted in the R programming environment v. 3.0 using two open source R-packages, SJEMEA and MEADQ, and compiled into a well level data set. Bursts were detected using an implementation of the MaxInterval method by Neuroexplorer, with the following threshold parameters: maximum interspike interval (ISI); 0.25 s, maximum beginning ISI; 0.1 s, minimum interburst interval (IBI); 0.8 s, minimum burst duration; 0.05 s and minimum number of spikes in a burst; 6. Network spikes were identified by dividing the recording period into 3ms bins and determining the number of electrodes in the well that fired at least one spike during each bin; the minimum threshold for a network spike was for spike activity to be present on at least 5 electrodes in a given time bin. Correlation was measured using the spike time tiling coefficient, which was defined as

$$TTC = \frac{1}{2} \left(\frac{r_{12} - r_{21}}{1 - r_{12}r_{21}} + \frac{r_{21} - r_{12}}{1 - r_{21}r_{12}} \right)$$

where r_{12} is the proportion of spikes on electrode 1 that occur within $\pm TTC$ of a spike on electrode 2 and r_{21} is the fraction of the total recording time that lies within $\pm TTC$ of a spike on electrode 1. r_{22} and r_{11} are the equivalent values on electrode 2. Data files generated by Axis were converted into HDF5 file format²⁶; HDF5 files, scripts to generate the features and related R-objects are stored in a public repository

(<http://github.com/sje30/EPAMEadev>). The goal of establishing a public data set is to allow full reproducibility of our analysis and/or to allow novel analyses to be conducted.

Developmental Analysis

Twelve features were chosen to describe the culture activity, which are summarized in Table 1. For all features, the plate value was taken as the median of all non-zero well values on the plate (zero values were ignored).

PCA

We performed principal component analysis using the R-package FactoMineR²⁷ using all wells and all 12 features. Two PCAs were performed. The first PCA was conducted using data in which a well constituted one observation, while the second PCA was conducted using data in which a plate median constituted an observation. For each PCA, the 12 dimensional feature vector was projected down onto the plane created by the first two principal component dimensions. The purpose of the projection was to visually assess the level of differentiation among the 4 ages. A scree plot was made to describe the cumulative percent of variation explained by the use of additional principal components to describe the data. The scree plots aid in quantifying the extent to which data may be well represented with fewer dimensions.

Classification

Classification was performed to understand whether and to what extent the features chosen above could distinguish between networks with different characteristics (for example, control vs compound-treated). Since this dataset did not contain networks treated with compounds, our classification examined the ability of the chosen features to discriminate between networks of different ages. Two classification techniques, random forests and Support Vector Machines (SVMs), were used to predict the age of each well

based on the twelve features used in our analysis. In some cases, due to the low number of electrodes on a well, lack of bursting or lack of network spike activity, some feature values were missing; this was particularly evident at early DIV. For classification purposes, for those wells with no bursts, the within burst firing rate and burst duration were set to zero. Similarly, the network spike peak and duration were set to zero for all wells that exhibited no network spikes over the recording period. Any wells that had null values for the remaining features, namely correlation, CV of IBI and CV of within burst ISI, were excluded from the classification. This resulted in 370/2976 well recordings, or approximately 12.4% of the total wells, being excluded from the classification.

Initially, classification was performed on the remaining data using a random forest model and all twelve features. The relative importance of each of the features was determined based on the amount they reduced the Gini index. Next, Support Vector Machines were used to examine the classification accuracy obtained by using various subsets of the total twelve features. This four-class classification problem was addressed using the 'one-against-one' approach²⁸, which involved building six binary SVMs, one for each pairwise combination of ages. Each of these SVMs were then used to binarily classify every data point, and the class to which each point was most frequently assigned across the six SVMs was taken as its correct class. A radial kernel was chosen over linear and polynomial kernels for the SVMs because of its superior classification accuracy on our data. The optimal regularization parameter values for the radial kernel were found by conducting a grid search across a range of values using ten-fold cross-validation on the entire data set.

The parameters were chosen as those that maximized the cross-validation accuracy, which were $\eta = \frac{1}{10}$ and $\tau = 10$.

In both types of classification, two thirds of the data were used as a training set and the remaining third used to test the classification accuracy of the model. The classification was repeated one hundred times using random choices of the training and test sets in each iteration, and the classification accuracy averaged over the one hundred repetitions.

Results

Developmental Profile

On DIV 2, only rare, individual spikes were recorded (data not shown). Spontaneous activity in the neural networks arose and could be reliably recorded beginning on DIV 5 (Supplemental Figure 2). Activity as assessed by most of the parameters used here increased with DIV. In particular, not only did spiking increase with time, but the organization of spiking into bursts and correlated activity across the network (Supplemental Figure 2) also increased with DIV. Quantification of the changes in activity over development was achieved using a selection of twelve measures (Figure 1), which were used to describe activity at the level of the entire well or individual electrodes, which were then aggregated into well level values by taking the median.

Spontaneous firing rate

In general, activity increased over development, with the mean firing and burst rates both monotonically rising with increasing DIV (Figure 1A, B).

Bursting activity

A clear increase in bursting activity with increasing culture age was also observed. Although burst duration did not show strong developmentally-related changes (Figure 1C), the fraction of bursting electrodes, within burst firing rate and percentage of spikes occurring within bursts all increased over development (Figure 1D-F). The CV of inter-burst intervals (IBIs) and within burst inter-spike intervals (ISIs) also increased with development, indicating a decrease in the regularity of these features (Figure 1G,H).

Synchronous activity

The synchrony of activity within each individual well on a plate was examined using a feature called network spikes. Network spikes were defined as short time intervals in which the number of active electrodes on the well exceeded a threshold value, and their rate, duration and peak number of active electrodes were quantified for each plate (Figure 1I-K). The frequency of network spikes increased with increasing developmental age. To a lesser extent, an increase in the network spike peak (the maximum number of electrodes active during a network spike out of a possible 16) was also observed across development.

As another measure of network synchrony, we calculated the mean of all pairwise correlation coefficients for all electrodes in a well, using the spike time tiling coefficient²⁵. Correlations strengthened over development, particularly at early ages (Figure 1L).

PCA

A PCA was undertaken to visualize the level of differentiation among the four culture ages. The wells projected onto the first two PC dimensions (Figure 2A) show a stochastic organization starting from the earliest age (red, DIV 5) progressing through to the oldest

age (purple, DIV 12). The progression in age is roughly aligned with the first PC dimension, which accounts for over 50% of the variation (Figure 2B). To quantify the relationship between PC dimension 1 and culture age, PC1 was regressed against culture age. The linear model results showed that PC1 increases with increasing culture age (p-value of slope < 0.001, Supplemental Figure 3). This means that the principal mode of variation corresponds to the difference in ages of the cultures. Moreover, all factor loadings are positive on the first PC dimension, meaning that an increase in PC1 is associated with an increase in all 12 variables. Another salient aspect of the PC projection is that variation appears smaller at earlier ages. Similarly, the projection of the plate medians onto the first two PC dimensions yields a rough segregation by DIV. As in the well-level PCA, at the plate-level DIVs are aligned with the first PC dimension (Figure 2C), revealing a consistent age-related characteristic to the data. A greater percentage of variability is captured by the first PC dimension (67%; Figure 2D) as compared with the well-level PCA, related to the fact that taking the median reduces well-to-well variability. Both PCAs display sufficient visual differentiation between observations by DIV that a more thorough quantification of this separation is warranted through classification techniques.

Classification

Classification techniques were used to determine the degree to which the recordings could be separated into their ages using the features specified above. Firstly, a random forest model was built and used to predict the age of each well, using the twelve features from our analysis. The model was built using two-thirds of the data as a training set and its accuracy determined by using the remaining one-third of the data as a test set. When used to predict

the age of each well from the four possible ages, the accuracy of the random forest model, averaged over one hundred trials, was approximately 72% (compared to the 25% accuracy that could be expected by chance).

From these random forest models, we were also able to determine the relative importance of each of the features in driving the classification (Table 2 “Importance” column). The two most important features were those measuring coefficients of variation, namely the CV of within burst ISI and CV of IBI. In our developmental analysis, these two features both exhibited a monotonically increasing trend with age.

Next, we used Support Vector Machines (SVMs) to quantify the degree to which recordings could be classified correctly by age when only a subset of our features was used. The SVM classifier, built using the same proportion of training and test sets specified above, had a slightly higher level of accuracy, of approximately 73%, compared to the random forest model using all twelve features. Using the ordering of feature importance found above, we were then able to analyze how prediction accuracy varied as the number of features was reduced. Table 2 shows the performance of the SVM as the number of features used in the classification was gradually reduced from twelve, in the bottom row, to just one feature, CV of within burst ISI, in row one. In general we found that the classification accuracy remained high (~70% or higher) as the number of features was reduced. However, four features (burst rate, network spike rate, CV of IBI and CV of within burst ISI) were required to maintain a prediction accuracy $\geq 65\%$ (Table 2, “Accuracy %” column).

We used a similar method to examine the extent to which each pair of ages of arrays could be separated using classification techniques. In this case, rather than using all of the data in

the classification, the SVM classifier was built separately on each pairwise combination of ages. The classifier was most accurate in distinguishing arrays with large differences in age, for example DIV 5 and DIV 12 arrays, for which only the top feature, CV of within burst ISI, was required to achieve almost 92% prediction accuracy (Table 3). Classification performance was poorest for pairs of arrays in which the age difference was low. For example, the prediction accuracy for distinguishing DIV 9 from DIV 12 arrays was only just above chance when using one feature. Using all features improved the ability to distinguish between closely-related ages to ~82-83%, which is well above chance (Table 3).

How many wells are needed?

In our experiments we have used all 48 wells on a plate as replicates of the same experimental condition. This is a conservative way of using the multi-well array, and an alternative, higher throughput, approach might be to use different wells for different experimental conditions. However, there is inevitably a trade-off between the number of experimental conditions tested and the number of replicate recordings of conditions when assigning conditions to wells on a plate.

We therefore sought to investigate how robust our results were if fewer wells were used to form a signature of activity at a given age. Intuitively, we expected that with fewer wells we would get less reliable signatures of activity, and hence poorer classification. Rather than run experiments where fewer wells were used, we simulated the experiments by randomly removing all data for a given number of wells on each of the 16 plates, and then repeated our classification tests to see how well each age could be discriminated. Figure 3 shows that classification accuracy remained above 60% with as few as 4 wells per plate. With 16

wells (1/3 of normal), the classifier accuracy is close to the stable value. For our particular question then of discriminating the four ages, we could get reliable results using $\frac{1}{2}$ (24 wells) or $\frac{1}{3}$ (16 wells) of the data that we generated for each plate.

Discussion

The present results describe the ontogeny of network activity in 48 well MEA plates during the first 12 days in vitro. The results demonstrated a rapid ontogeny of spiking, bursting, synchrony and network spiking activity over this period of time, which is similar to the ontogeny of activity in single well MEAs. Furthermore, these results demonstrate that by considering multiple parameters of network firing, bursting and synchrony properties, principal components analysis and classification methods can be used as reliable predictors of network age at both the plate and well levels. These results demonstrate the neural network ontogeny on mwMEAs offers, relative to single well systems, a high-throughput approach to study network development and its perturbation by drugs, chemicals and disease.

Previous studies of cortical and hippocampal network ontogeny have demonstrated that activity begins with random, single spiking activity on a single or few channels, and over a period of 2-3 weeks in vitro progresses to bursting activity which becomes more synchronous with time^{4,7-10,29}. This is accompanied over time by the emergence of network bursts. Similar to previous data from our laboratory using single well MEA “chips” where cells were seeded at a high density¹⁷; the ontogeny of spiking and bursting activity occurred rapidly within the first two weeks in vitro, specifically between DIV 5 and 12 in

the present study. Our data are consistent with that of other laboratories who have reported initiation of spiking activity as early as 3-6 DIV^{7,30}. While our data describe the initial stages of the ontogeny of network activity, we did not examine ages beyond DIV 12 in the current study, and cannot rule out that some features examined may continue to “mature” with additional time in culture. Further, the ontogeny of network development in the culture used here does occur earlier than has been reported by other laboratories^{18,31-33}. Factors that may influence these differences include the age at which the cells were isolated (Postnatal Day 0 here vs embryonic day 18 in other studies) as well as the plating density (150K cells here vs 50K in other studies)^{8,11}. This more rapid development of network activity may have some benefits from a screening standpoint, as it could shorten assay times, increase throughput and reduce costs, although ultimately, the screening needs of the user will dictate issues such as culture model and plating density. Regardless of the rate of network ontogeny, the approaches described herein can be applied to the resultant data, and in the present case, it appears that network ontogeny can be reliably predicted by using as few as 3-5 wells/plate by considering all of the parameters.

The use of classification techniques such as random forests and SVMs indicated that the parameters extracted from the spike trains in these experiments could be used to predict reliably the age of the culture from between the four different age categories examined.

While the models performed best when all of the parameters were used to aid classification, several parameters had greater influence on the ability to predict culture age. These included the CV of within burst ISI, the CV of the IBI, mean burst duration, network spike rate and burst rate. When these approaches were used to predict between two

different ages, considering all of the parameters resulted in greater accuracy regardless of age. This indicates that using multiple parameters will provide more robust discrimination of different ages (or perhaps treatments) than relying on a single or a few parameters. The greatest accuracy was achieved when predicting between larger age differences (e.g. DIV 5 vs 12) and likely reflects the relative lack of bursts, network spikes and correlated activity in DIV 5 cultures. This is consistent with the relative lack of connectivity in DIV 5 cultures. Synaptogenesis, as reflected morphometrically by the juxtaposition of pre and post-synaptic markers in this model, begins around DIV 6 and continues through DIV 12 (Harrill et al., 2011). Thus, the initiation of structural evidence of synaptogenesis corresponds nicely with increases in electrophysiological parameters on DIV 7, including those parameters reflecting connectivity (bursts, network spikes and correlated activity).

The present analysis has implications for using mwMEAs for drug development or chemical developmental neurotoxicity screening. Both classification approaches used here provided higher accuracy by including more features. Traditionally, MFR has been widely utilized to describe drug- or chemical-induced alterations in network function^{2,6,11,12,18} as it is easily extracted from the data. However, when possible, determining more features and using them collectively rather than focusing on one or a few features may provide greater sensitivity in detecting effects, as well as possibly facilitating drug or chemical “fingerprinting” approaches³⁴. In addition, the classification approaches used here indicate that age of a network can be reliably determined using 3-8 wells from each plate, indicating that between 6 and 16 different treatment conditions might be possible on a given 48 well

mwMEA plate. We believe our study is the first to assess the important question of how to efficiently use wells on a mwMEA, suggesting that as few as 3-8 wells might suffice to form a reliable pattern of activity. However, this range should be treated with caution: as can be seen from Figure 3, there is an increasing variance with fewer wells, and more importantly, results are likely to differ depending on the size of the effect being measured. Our findings suggest that where there are gross changes in activity patterns, and fewer wells are needed. On the other hand, where changes in activity are more subtle, we would expect more replicates to be required. Our recommendation therefore is that investigators should repeat our sampling approach (Figure 3) to investigate how reducing the number of wells per condition can affect reliability of results.

One factor that influences results from MEAs is the culture to culture variability. Although this was not specifically addressed in the present studies (typically only one plate was available from a given culture), evidence from another data set wherein three plates each were analyzed from several different cultures indicated that culture to culture variability is much greater than plate to plate variability (unpublished data). Thus, obtaining replicate values for different treatments (e.g. concentrations of a drug or chemical) across several different wells and plates from the same culture (e.g. see supplemental material in Wallace et al., 2015³⁵) may be preferable to obtaining replicate values across several different cultures. It is likely that the use of a primary culture model does contribute to the culture to culture differences, as each culture is prepared from different animals. This may in the

future be improved by the use of stem cell derived models, which should, in theory, be more homogeneous.

In conclusion, we have described the early development of neural networks grown in 48 well mwMEA plates and found that it is qualitatively equivalent to development of network activity in single well MEAs. Furthermore, multi-parametric evaluation of the network activity parameters provides an accurate method of classifying networks by age. Together, these results indicate that neural networks cultured on mwMEAs will be a useful tool to study the ontogeny of network activity as well as the potential for drugs, chemicals and diseases to disrupt that activity.

References

1. Nam, Y.; Wheeler, B. C. In Vitro Microelectrode Array Technology and Neural Recordings. *Crit. Rev. Biomed. Eng.* **2011**, *39*, 45–61.
2. Gross, G. W.; Harsch, A.; Rhoades, B. K.; et al. Odor, Drug and Toxin Analysis with Neuronal Networks in Vitro: Extracellular Array Recording of Network Responses. *Biosens. Bioelectron.* **1997**, *12*, 373–393.
3. Marom, S.; Shahaf, G. Development, Learning and Memory in Large Random Networks of Cortical Neurons: Lessons beyond Anatomy. *Q. Rev. Biophys.* **2002**, *35*, 63–87.
4. van Pelt, J.; Vajda, I.; Wolters, P. S.; et al. Dynamics and Plasticity in Developing Neuronal Networks in Vitro. In *Development, Dynamics and Pathology of Neuronal Networks: from Molecules to Functional Circuits*; Elsevier, 2005; Vol. Volume 147, pp. 171–188.
5. Chiappalone, M.; Vato, A.; Tedesco, M. B.; et al. Networks of Neurons Coupled to Microelectrode Arrays: A Neuronal Sensory System for Pharmacological Applications. *Biosens. Bioelectron.* **2003**, *18*, 627–634.
6. Johnstone, A. F. M.; Gross, G. W.; Weiss, D. G.; et al. Microelectrode Arrays: A Physiologically Based Neurotoxicity Testing Platform for the 21st Century. *Neurotoxicology* **2010**, *31*, 331–350.
7. Wagenaar, D. A.; Pine, J.; Potter, S. M. An Extremely Rich Repertoire of Bursting Patterns during the Development of Cortical Cultures. *BMC Neurosci.* **2006**, *7*, 11.
8. Chiappalone, M.; Bove, M.; Vato, A.; et al. Dissociated Cortical Networks Show Spontaneously Correlated Activity Patterns during in Vitro Development. *Brain Res.* **2006**, *1093*, 41–53.
9. Illes, S.; Fleischer, W.; Siebler, M.; et al. Development and Pharmacological Modulation of Embryonic Stem Cell-Derived Neuronal Network Activity. *Exp. Neurol.* **2007**, *207*, 171–176.
10. Biffi, E.; Regalia, G.; Menegon, A.; et al. The Influence of Neuronal Density and Maturation on Network Activity of Hippocampal Cell Cultures: A Methodological Study. *PLoS One* **2013**, *8*, e83899.
11. Valdivia, P.; Martin, M.; LeFew, W. R.; et al. Multi-Well Microelectrode Array Recordings Detect Neuroactivity of ToxCast Compounds. *Neurotoxicology* **2014**, *44*, 204–217.
12. Nicolas, J.; Hendriksen, P. J. M.; van Kleef, R. G. D. M.; et al. Detection of Marine Neurotoxins in Food Safety Testing Using a Multielectrode Array. *Mol. Nutr. Food Res.* **2014**, *58*, 2369–2378.
13. MacLaren, E. J.; Charlesworth, P.; Coba, M. P.; et al. Knockdown of Mental Disorder Susceptibility Genes Disrupts Neuronal Network Physiology in Vitro. *Mol. Cell. Neurosci.* **2011**, *47*, 93–99.
14. Charlesworth, P.; Morton, A.; Eglen, S. J.; et al. Canalization of Genetic and Pharmacological Perturbations in Developing Primary Neuronal Activity Patterns. *Neuropharmacology* **2016**, *100*, 47–55.
15. Wainger, B. J.; Kiskinis, E.; Mellin, C.; et al. Intrinsic Membrane Hyperexcitability of Amyotrophic Lateral Sclerosis Patient-Derived Motor Neurons. *Cell Rep.* **2014**, *7*, 1–11.
16. Woodard, C. M.; Campos, B. A.; Kuo, S.-H.; et al. iPSC-Derived Dopamine Neurons Reveal Differences between Monozygotic Twins Discordant for Parkinson's Disease. *Cell Rep.* **2014**, *9*, 1173–1182.

17. Robinette, B. L.; Harrill, J. A.; Mundy, W. R.; et al. In Vitro Assessment of Developmental Neurotoxicity: Use of Microelectrode Arrays to Measure Functional Changes in Neuronal Network Ontogeny. *Front. Neuroeng.* **2011**, *4*, 1.
18. Hogberg, H. T.; Sobanski, T.; Novellino, A.; et al. Application of Micro-Electrode Arrays (MEAs) as an Emerging Technology for Developmental Neurotoxicity: Evaluation of Domoic Acid-Induced Effects in Primary Cultures of Rat Cortical Neurons. *Neurotoxicology* **2011**, *32*, 158–168.
19. Grandjean, P.; Landrigan, P. J. Neurobehavioural Effects of Developmental Toxicity. *Lancet Neurol.* **2014**, *13*, 330–338.
20. Harrill, J. A.; Robinette, B. L.; Freudenrich, T. M.; et al. Media Formulation Influences Chemical Effects on Neuronal Growth and Morphology. *In Vitro Cell. Dev. Biol. Anim.* **2015**, *51*, 612–629.
21. Björklund, U.; Persson, M.; Rönnbäck, L.; et al. Primary Cultures from Cerebral Cortex and Hippocampus Enriched in Glutamatergic and GABAergic Neurons. *Neurochem. Res.* **2010**, *35*, 1733–1742.
22. Harrill, J. A.; Robinette, B. L.; Freudenrich, T.; et al. Use of High Content Image Analyses to Detect Chemical-Mediated Effects on Neurite Sub-Populations in Primary Rat Cortical Neurons. *Neurotoxicology* **2013**, *34*, 61–73.
23. Harrill, J. A.; Robinette, B. L.; Mundy, W. R. Use of High Content Image Analysis to Detect Chemical-Induced Changes in Synaptogenesis in Vitro. *Toxicol. In Vitro* **2011**, *25*, 368–387.
24. Eytan, D.; Marom, S. Dynamics and Effective Topology Underlying Synchronization in Networks of Cortical Neurons. *Journal of Neuroscience* **2006**, *26*, 8465–8476.
25. Cutts, C. S.; Eglén, S. J. Detecting Pairwise Correlations in Spike Trains: An Objective Comparison of Methods and Application to the Study of Retinal Waves. *Journal of Neuroscience* **2014**, *34*, 14288–14303.
26. Eglén, S. J.; Weeks, M.; Jessop, M.; et al. A Data Repository and Analysis Framework for Spontaneous Neural Activity Recordings in Developing Retina. *Gigascience* **2014**, *3*, 3.
27. Lê, S.; Josse, J.; Husson, F. FactoMineR : An R Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25*, 1–18.
28. Chang, C.-C.; Lin, C.-J. LIBSVM: A Library for Support Vector Machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27:1–27:27.
29. Charlesworth, P.; Cotterill, E.; Morton, A.; et al. Quantitative Differences in Developmental Profiles of Spontaneous Activity in Cortical and Hippocampal Cultures. *Neural Dev.* **2015**, *10*, 1.
30. Kamioka, H.; Maeda, E.; Jimbo, Y.; et al. Spontaneous Periodic Synchronized Bursting during Formation of Mature Patterns of Connections in Cortical Cultures. *Neurosci. Lett.* **1996**, *206*, 109–112.
31. Bikbaev, A.; Frischknecht, R.; Heine, M. Brain Extracellular Matrix Retains Connectivity in Neuronal Networks. *Sci. Rep.* **2015**, *5*, 14527.
32. Bakkum, D. J.; Chao, Z. C.; Potter, S. M. Long-Term Activity-Dependent Plasticity of Action Potential Propagation Delay and Amplitude in Cortical Networks. *PLoS One* **2008**, *3*, e2088.
33. Bakkum, D. J.; Chao, Z. C.; Potter, S. M. Spatio-Temporal Electrical Stimuli Shape Behavior of an Embodied Cortical Network in a Goal-Directed Learning Task. *J. Neural Eng.* **2008**, *5*, 310–323.
34. Mack, C. M.; Lin, B. J.; Turner, J. D.; et al. Burst and Principal Components Analyses of MEA Data for 16 Chemicals Describe at Least Three Effects Classes. *Neurotoxicology* **2014**, *40*, 75–85.
35. Wallace, K.; Strickland, J. D.; Valdivia, P.; et al. A Multiplexed Assay for Determination of Neurotoxicant Effects on Spontaneous Network Activity and Viability from Microelectrode Arrays. *Neurotoxicology* **2015**, *49*, 79–85.

Table 1: Features used in our analysis and a brief description of how they were calculated.

Feature	Description.
Mean firing rate	The mean firing rate (MFR) on each electrode was calculated. The well value was the median value of all active electrodes.
Burst rate	The number of bursts per minute on an electrode was calculated. The well value was the median value from all electrodes that exhibited bursting behaviour.
Burst duration	The mean duration of all bursts on an electrode over the recording period was calculated. The well value was the median value from all electrodes that exhibited bursting behaviour.
Fraction of bursting electrodes	An electrode was classified as bursting if the burst rate on the electrode was at least one per minute. The well value was the number of electrodes classified as bursting as a fraction of the total number of active electrodes on the well.
Within burst firing rate	The ratio of the standard deviation to the mean of the length of all inter- burst intervals (IBI) on an electrode. The well value was the median value from all electrodes that exhibited bursting behaviour.
Percentage of spikes in bursts	The number of spikes on an electrode classified as being within bursts divided by the total number of spikes on the electrode. The well value was the median value from all electrodes that exhibited bursting behaviour.
CV of IBI	The ratio of the standard deviation to the mean of the length of all inter- burst intervals (IBI) on an electrode. The well value was the median value from all electrodes that exhibited bursting behaviour.
CV of within burst ISIs	The ratio of the standard deviation to the mean of the length of all inter- spike intervals (ISI) within bursts on an electrode. The well value was the median value from all electrodes that exhibited bursting b
Network spike rate	The well value was the number of network spikes on the well per minute of the recording period (See Methods section for definition of a Network Spike).
Network spike duration	The duration of a network spike was defined as the length of time during which the number of active electrodes on the well exceeded the threshold value (5). The well value was taken as the median duration of all network spikes on the well during the recording period.
Network spike peak	The maximum number of active electrodes during each network spike. The well value was taken as the median peak value of all network spikes on the well during the recording period.
Mean correlation	The correlation between every pairwise combination of electrodes on a well was calculated using the spike time tiling coefficient (Cutts & Eglen, 2015) with $\Delta t = 50ms$. (See Methods section for definition). The well value was the mean of the pairwise correlations between all distinct electrodes on the well.

Table 2: Classifier performance at predicting the age of arrays. Features are listed in decreasing order of importance, based on the importance score in column 2, derived from random forest classification and normalized to the top score. The value in each row $n = 1 \dots 12$ of column 3 is the mean percentage of correct classifications using the top n features in the SVM model. For example, row 4 shows that the classifier was 65.0% accurate at predicting age using the top four features.

Feature	Importance	Accuracy %
CV of within burst ISI	1.00	49.2
CV of IBI	0.70	58.3
Network spike rate	0.50	62.0
Burst rate	0.49	65.0
Burst duration	0.44	66.0
% Spikes in bursts	0.39	68.3
Correlation	0.36	69.5
Firing rate	0.35	71.4
Within burst firing rate	0.31	72.7
Bursting electrodes	0.22	73.0
Network spike duration	0.18	73.5
Network spike peak	0.09	73.4

Table 3: Classifier performance at predicting the age of arrays for each pairwise combination of ages (days in vitro). Features are listed in decreasing order of importance, and the value in each row $n = 1 \dots 12$ is the mean percentage of correct classifications using the top n features, as described in Table 2.

Feature	Accuracy %					
	5 vs 7	5 vs 9	5 vs 12	7 vs 9	7 vs 12	9 vs 12
CV of within burst ISI	75.0	87.5	91.8	69.9	78.1	57.4
CV of IBI	77.4	89.5	93.6	76.8	85.5	64.7
Network spike rate	79.3	90.3	95.5	79.6	88.0	68.7
Burst rate	79.3	90.3	95.3	81.0	88.4	72.8
Burst duration	79.7	90.8	95.3	81.0	88.6	74.0
% Spikes in bursts	81.6	91.3	95.6	81.5	90.4	76.4
Correlation	82.2	92.1	95.6	82.3	90.9	77.1
Firing rate	82.3	91.7	95.7	82.3	90.9	80.2
Within burst firing rate	84.2	92.7	96.2	82.4	91.3	81.2
Bursting electrodes	83.6	92.5	96.2	82.6	91.8	81.5
Network spike duration	83.7	92.9	96.8	82.9	92.8	82.0
Network spike peak	82.2	92.5	96.8	82.6	93.0	82.1

Figure Legends

Figure 1: Mean firing (A) and burst rates (B) increase with development. Box plots showing median and interquartile range are shown for $n=16$ plates. (C) Burst duration. (D) Fraction of bursting electrodes. (E) Within Burst Firing rate (F) Percentage of spikes in bursts. (G) Coefficient of variation (CV) of interburst interval (IBI). (H) CV of within burst interspike interval (ISI). (I) Network spike rate. (J) Network spike duration. (K) Network spike peak. (L) Mean pairwise correlation.

Figure 2: (A) Well-level PCA projection of 12-dimensional feature vectors onto PC dimension 1 (x-axis) and 2 (y-axis). Each dot represents a well, colored by day in vitro (DIV) of recording. Rough ordering from youngest (red, DIV 5) to oldest (purple, DIV 12) wells is apparent in change of colors along the positive direction. (B) Scree plot displays % variance explained by the number of PC dimensions. (C) Plate-level PCA projection of plate medians onto PC dimension 1 (x-axis) and 2 (y-axis). As in top, rough ordering of observations by DIV is apparent in the red to purple transition along the x-axis. (D) Scree plot of plate-level PCA. Compared to the well-level PCA scree plot, a larger amount of variation is captured in the first two PC dimensions indicating that taking the plate median reduces variability.

Figure 3: Accuracy of predicting the age of each well by sampling $n \leq 48$ wells on each plate.

Dark blue line shows the mean accuracy, while the vertical lines show the minimum and maximum accuracy over 100 trials with random choices of wells. For $n = 48$, the error bars indicate the small variability in classification due to the partitioning of data into train/test sets. The red dotted line indicates baseline level of performance (25%) for a classifier