



OPEN

## Early prediction of developing spontaneous activity in cultured neuronal networks

David Cabrera-García<sup>1,5</sup>✉, Davide Warm<sup>1,4</sup>, Pablo de la Fuente<sup>1</sup>,  
M. Teresa Fernández-Sánchez<sup>1</sup>, Antonello Novelli<sup>1,2</sup>✉ & Joaquín M. Villanueva-Balsera<sup>3</sup>✉

Synchronization and bursting activity are intrinsic electrophysiological properties of *in vivo* and *in vitro* neural networks. During early development, cortical cultures exhibit a wide repertoire of synchronous bursting dynamics whose characterization may help to understand the parameters governing the transition from immature to mature networks. Here we used machine learning techniques to characterize and predict the developing spontaneous activity in mouse cortical neurons on microelectrode arrays (MEAs) during the first three weeks *in vitro*. Network activity at three stages of early development was defined by 18 electrophysiological features of spikes, bursts, synchrony, and connectivity. The variability of neuronal network activity during early development was investigated by applying k-means and self-organizing map (SOM) clustering analysis to features of bursts and synchrony. These electrophysiological features were predicted at the third week *in vitro* with high accuracy from those at earlier times using three machine learning models: Multivariate Adaptive Regression Splines, Support Vector Machines, and Random Forest. Our results indicate that initial patterns of electrical activity during the first week *in vitro* may already predetermine the final development of the neuronal network activity. The methodological approach used here may be applied to explore the biological mechanisms underlying the complex dynamics of spontaneous activity in developing neuronal cultures.

Burst activity and synchronization are intrinsic features of neural network development *in vivo* and *in vitro*, important for both neuronal communication and information processing<sup>1–3</sup>. Brain neuronal networks in newborn mammals are immature and present similar electrophysiological properties<sup>4</sup>, characterized by a combination of local synchronous patterns<sup>5</sup>. Moreover, specific firing patterns of synchronization and desynchronization are correlated to cognitive processes and pathophysiological conditions<sup>6</sup>. Understanding the relationship between network dynamics and development is crucial to identify the mechanisms underlying the maturation of neural circuits<sup>7</sup>. Cortical neurons *in vitro* are an effective model to study these mechanisms because neuronal cultures retain some of these network dynamics that can be recorded with the use of microelectrode arrays (MEAs)<sup>8</sup>.

Dissociated neurons on MEAs display sparse spontaneous activity during the first week *in vitro* that shifts to synchronous firing and bursting activity between the second and third week *in vitro*<sup>9–12</sup>. This developmental pattern comprehends a complex range of spiking and bursting dynamics that vary between dishes and cultures<sup>10</sup>. Although variability in burst activity and synchronization is partly explained by differences in experimental conditions<sup>13</sup> and cellular density<sup>10,14</sup>, broad heterogeneity remains even in cortical cultures with standardized conditions<sup>10,15</sup>.

Due to the intrinsic variability of neuronal cultures, standardized mean firing rate (MFR) is the most frequent feature reported for *in vitro* neuropharmacological evaluation<sup>15,16</sup>. However, the complex behavior of neural networks *in vitro* is not completely characterized without considering the synchronization and burstiness of spontaneous events. Multiple tools are available for the analysis of these features of neuronal firing activity<sup>17</sup> and,

<sup>1</sup>Department of Biochemistry and Molecular Biology and University Institute of Biotechnology of Asturias (IUBA), Campus “El Cristo”, University of Oviedo, 33006 Oviedo, Spain. <sup>2</sup>Department of Psychology and University Institute of Biotechnology of Asturias (IUBA), Campus “El Cristo”, University of Oviedo, Institute for Sanitary Research of the Principedom of Asturias (ISPA), 33006 Oviedo, Spain. <sup>3</sup>Project Engineering Area, University of Oviedo, 33004 Oviedo, Spain. <sup>4</sup>Present address: Institute of Physiology, University Medical Center of the Johannes Gutenberg University Mainz, Duesbergweg 6, 55128 Mainz, Germany. <sup>5</sup>Present address: Department of Synapse and Network Development, Netherlands Institute for Neuroscience, 1105 BA Amsterdam, The Netherlands. ✉email: cabreragarcia.david@gmail.com; anovelli@uniovi.es; jmvillanueva@uniovi.es

while gold-standard methods are lacking, some methods provide more robust results. For example, Maximum Interval or logISI are highly consistent burst detectors among different dynamics of spontaneous activity<sup>18</sup>. Likewise, measures of correlation such as the Spike Time Tiling Coefficient (STTC) are efficient in detecting spike train synchrony independently of firing rate<sup>19</sup>. Temporal correlations between the spike trains recorded by each electrode can be also analyzed using graph theory to infer connectivity features of the neural network<sup>20,21</sup>. Thus, the analysis of functional connectivity provides a proxy of the topological formation and organization of neuronal networks on MEAs<sup>21–23</sup>.

Integration of spiking, bursting, synchrony, and connectivity features creates a multidimensional profile of network activity which renders the analysis difficult to tackle with traditional statistical methods. To overcome these limitations, machine learning methods arise as an alternative approach for extracting information from large datasets of neural recordings<sup>24</sup>, as well as for predicting variables<sup>25</sup>. Although the use of machine learning algorithms is widely adopted in other fields of neuroscience<sup>26</sup>, few studies have applied these techniques to the study of neural networks in vitro. Dimensionality reduction and classification methods have been applied to integrate electrophysiological features and to identify developmental stages<sup>27,28</sup>, and unsupervised machine learning methods have been used to cluster burst patterns in mature networks<sup>29</sup>. However, to the best of our knowledge, machine learning has not been previously used to study the broad range of developing network activity in vitro.

While modern machine learning techniques like Deep Neural Networks are powerful methods for neural decoding<sup>30,31</sup>, the use of simpler algorithms provides easier interpretation and fewer overfitting problems<sup>32</sup>. Clustering techniques like Self-Organizing Map (SOM)<sup>33</sup> have been successfully applied to visualize developmental patterns in organogenesis<sup>34</sup>. Furthermore, machine learning regression algorithms have proved to be effective for the prediction of biological data, including Multivariate Adaptive Regression Splines (MARS)<sup>35</sup>, Support Vector Machines (SVM)<sup>36</sup>, and Random Forests<sup>37</sup>. The application and comparison of these methods to investigate whether the development of neuronal networks in vitro is determined by early patterns of electrical activity remains largely unexplored.

Here, we report an integrated methodological approach for the study of developmental patterns of network activity in dissociated cortical neurons during the first three weeks in vitro. We used MEAs to characterize the spontaneous activity of cortical neurons in culture by measuring 18 electrophysiological features of spikes, bursts, synchrony, and connectivity. Clustering analysis allowed us to examine neuronal networks with similar development of bursting and synchronous activity. Then, we successfully used three machine learning models (MARS, SVM, and Random Forest) to predict the levels of electrophysiological features at the third week in vitro, suggesting that the development of network activity is determined by the early electrical activity of neuronal networks. The methodology presented here may help to identify the biological factors determining the maturation of in vitro neural networks.

## Methods

**Primary cortical cultures.** Primary cultures of mouse cortical neurons were obtained from the animal colonies hosted in the Bioterium of the University of Oviedo. The animal procedures used were in accordance with the protocols approved by the Institutional Animal Care and Use Committee of the University of Oviedo. All procedures were also carried out in concordance with the European Communities Council Directive (2010/63/UE), Spanish legislation (RD 53/2013), and ARRIVE guidelines. Primary cultures of cortical neurons were prepared from CD1 mice as previously described<sup>38</sup>, adapting the protocol described for cerebellar neurons<sup>39</sup>. Briefly, the brains from 0 to 2-day postnatal pups were sliced, the cortex was dissected from each slice and dissociated both enzymatically (0.125% papain solution) and mechanically (fire-polished Pasteur pipette) in the presence of DNase I. Cells were resuspended in Neurobasal A growth medium supplemented with 1% B27, 2 mM L-glutamine, and 100 mg/ml gentamicin (NB-B27), and seeded in MEA wells (MultiChannel Systems) previously coated with Poly-L-Lysine (5 µg/ml) to achieve an initial density between 1750 and 3500 cells/mm<sup>2</sup> (generally described as dense cortical cultures<sup>10</sup>). MEAs were incubated at 37 °C with 5% CO<sub>2</sub>, and 25% of the media volume was substituted with fresh NB-B27 every 3 days after the first week.

**MEA recordings.** Extracellular recordings of neuronal spontaneous activity were obtained using standard 60 electrode MEA chips (60MEA200/30iR-Ti-gr), 30 µm electrode diameter spaced by a 200 µm distance, with a MEA1060-Inv-BC amplifier (Multi Channel Systems). Raw analog signals were amplified (bandwidth 1 Hz–3 kHz) and sampled at 25 kHz before being filtered with a 200 Hz high-pass filter (Butterworth second-order). Recordings were performed at 37 °C and MEAs were covered with a MEA-MEM Teflon membrane (ALA Scientific) to maintain CO<sub>2</sub> conditions outside the incubator. After a 3 min stabilization period, 5 min of spontaneous activity was recorded and spikes were extracted from the filtered electrophysiological signal using a threshold method with the MC\_Rack software v.4.6 (Multi Channel Systems, <https://www.multichannelsystems.com/software/mc-rack>). Spikes that crossed a negative threshold set to 5.5 times the SD of the baseline noise were detected and stored in “mcd” files (MC\_Rack). Data files generated by MC\_Rack 4.6 were converted into HDF5 file format using MultiChannel DataManager (Multi Channel Systems, <https://www.multichannelsystems.com/software/multi-channel-datamanager>) and imported to MATLAB 9.8 (The MathWorks Inc, <https://www.mathworks.com>) or Python 3.7 (<https://www.python.org/>) through the relative toolboxes (<https://github.com/multichannelsystems>).

**MEA data analysis.** A dataset of 231 recordings between day in vitro (DIV) 6 and 18 was pre-processed to remove inconsistent and low-quality recordings. We included for analysis MEAs with one recording at DIV 6–8 and at least one additional recording between DIV 9–18 with spiking activity (> 3 spikes/min) in 10 or more channels. MEAs containing more than 6 channels with noise (amplitude larger than 500 µV or non-characteris-

tic extracellular biphasic waveforms) and with a decrease of network spike activity bigger than 50% during the second week in vitro were discarded from analysis. With this acceptance criteria, 141 recordings of 47 independent MEA dishes from 12 cortical cultures were included for analysis.

Four main groups of electrophysiological features were analyzed to characterize the spontaneous activity of cortical neurons in vitro: spikes, bursts, synchrony, and connectivity. Mean values of the features were used for the analysis. See Table S1 for a description of the 18 electrophysiological features included in the analysis. We selected three developmental stages to study the changes occurring during the second week in vitro and, in order to balance our dataset, the 141 recordings were grouped in three DIV intervals: 6–8 ( $n = 50$ ), 9–12 ( $n = 30$ ), and 13–18 ( $n = 61$ ). Spike analysis was performed in Neuroexplorer 5 (Nex technologies, <https://plexon.com/products/neuroexplorer/>) for the number of channels with spikes (Ch. spikes), the mean firing rate (MFR), network spikes/s (Network spikes), and interspike interval (ISI). Bursts were analyzed using the maximum interval method in Neuroexplorer 5 (Nex technologies) to obtain the number of channels with bursts (Ch. bursts), mean bursting rate (MBR), Burst duration, network bursts/min (Network bursts), interburst interval (IBI), percentage of spikes in bursts (Burst % spikes), interspike interval in bursts (Burst ISI), peak frequency in bursts (Burst PeakFreq), and burst surprise detection (Burst surprise). We defined a burst by the following parameters: the maximum initial ISI to start the burst (0.1 s), the maximum ISI to define the burst end (0.2 s), the minimum interval between bursts (0.2 s), the minimum burst duration (0.003 s), and the minimum number of spikes in the burst (3).

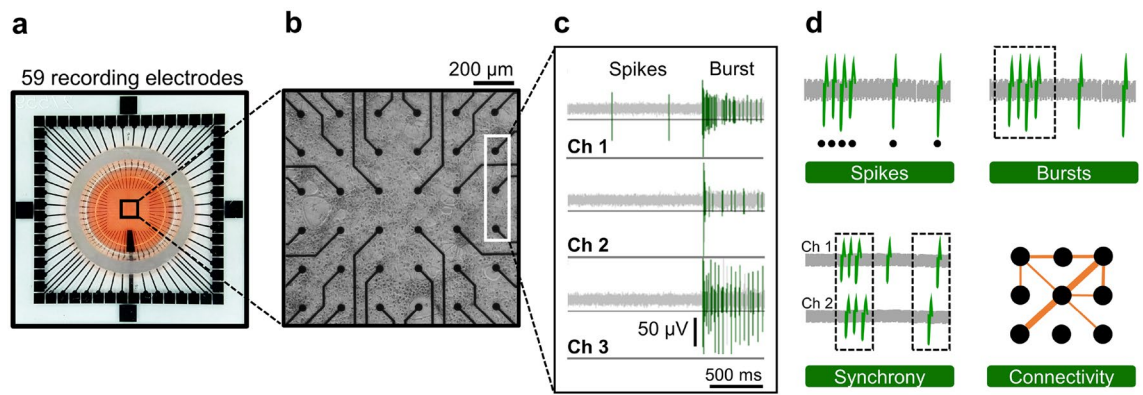
Synchronization between pairs of electrodes was analyzed with the spike time tiling coefficient (STTC) method<sup>19</sup> using a Python repository<sup>40</sup>. Briefly, the STTC is calculated as  $STTC = \frac{1}{2} \left( \frac{P_A - T_B}{1 - P_A T_B} + \frac{P_B - T_A}{1 - P_B T_A} \right)$ , where  $P_A$  is the proportion of spikes in channel A that occur within  $\Delta t$  of a spike from channel B, and  $T_A$  is the proportion of the total recording time in channel A that falls within  $\Delta t$  of a spike from channel A.  $P_B$  and  $T_B$  are calculated similarly. We considered a predefined time window of 100 ms to quantify the correlation between pairs of spikes trains, and the mean value of STTC was calculated from the square matrix. Then, to identify clusters of synchronized neurons, we used the density-based spatial clustering of applications with noise (DBSCAN) algorithm<sup>41</sup> on STTC matrices (DBSCAN STTC) implemented in the Python library Sklearn<sup>42</sup>, using  $\text{eps} = 0.2$  ( $\text{eps}$  is the maximum distance between two points to be considered as neighbors, i.e., highly synchronized pairs of electrodes with  $STTC > 0.8$ ), and 3 as the minimum number of samples to be considered as a core point.

Connectivity parameters were computed and connectivity matrices were generated using the Brain Connectivity Toolbox<sup>43</sup> in MATLAB 9.8 (The MathWorks Inc). First, spike time series were binned in 100 ms time windows for each electrode, and the spike count correlations were computed as the Pearson correlation coefficient for all pairs of electrodes. Only significant correlations ( $P < 0.05$ ) were retained in the correlation matrices. Then, an absolute weight threshold of 0.35<sup>22</sup> was applied to discard spurious connections and to obtain undirected connectivity matrices, in which nodes (electrodes) are linked by binary edges (connections), and the following network parameters were extracted: average node degree (Node degree), clustering coefficient (Clustering coeff), and global efficiency (Efficiency) (see Table S1 for definitions).

**Machine learning.** To examine the distribution of neuronal networks by developmental stage in a low-dimensional projection, we performed principal component analysis (PCA) with the above 18 electrophysiological features using the R package FactoMineR<sup>44</sup>. Values were scaled to unit variance and the 18-dimensional feature vector was projected onto the first two principal components (PCs) to display data points corresponding to DIV intervals (Fig. 2) and clustering analysis (Fig. S3).

To identify developmental clusters of individual neuronal networks (MEA dishes), we applied the SOM technique<sup>33</sup> using the R package Kohonen as previously described<sup>45</sup>. Briefly, SOM is a self-organizing neural network with an input and output layer, such that the output cells are activated as a function of the input data. For the 47 independent MEA dishes used in the study, the mean values of the electrophysiological features Ch. bursts and STTC at each DIV interval were taken as inputs to the SOM network. The network competes to activate a unique output cell by every input, which results in a learning model that categorizes similar cases by activating the same output cell. The output layer is visualized as a two-dimensional map of hexagonal cells, where each cell is associated with the centroid of the input cases that activate the cell. SOM clustering algorithm was applied to centered and scaled datasets using random initialization with 50 iterations and the total number of units in the competition layer was set to 25 ( $5 \times 5$ ). The size of the network topology was estimated using a heuristic function and  $5 \times 5$  was assigned as the first level of abstraction. To cluster the neuronal networks (MEAs) with similar behaviors, the k-means clustering technique was applied to the results obtained with the SOM technique. The number of clusters with the best fit ( $k = 3$ ) was adjusted with the Davies-Bouldin validity index<sup>46</sup>. Mean, median, standard error of the mean (SEM), and interquartile range (IQR) were calculated for the 18 electrophysiological features in each cluster. Additionally, PCA was performed as described above on the data points at DIV 6–8 and DIV 13–18, and clusters defined by k-means and SOM analysis were displayed on the principal components projections.

Prediction of the continuous variables STTC, Ch. bursts, and MFR, was assessed using three robust and optimized machine learning techniques for regression: MARS<sup>47</sup>, SVM<sup>48</sup>, and Random Forests<sup>49</sup> through the R packages Earth, e1071, and randomForest, respectively. We selected these methods, with distinct learning approaches, to determine whether the prediction of the electrophysiological features was affected by the machine learning model. MARS is a flexible and fast nonparametric regression technique used to capture the non-linear relationships between the variables. MARS (R package Earth) was performed with 5 degrees of interaction and 35 model terms. Random Forest combines multiple decision trees and averages the output, and it is considered to be a good model to control overfitting and select robust features for the model applied<sup>49</sup>. Random Forest was performed with 100 trees, and 8 variables were randomly sampled as candidates at each split. SVM is a robust classifier that handles noisy data and identifies non-linear relationships efficiently<sup>36</sup>. The SVM constructs a



**Figure 1.** Electrophysiological characterization of cortical neuron cultures on MEAs. (a) Photograph of a 60 electrode MEA device used in the study. (b) Phase-contrast image of cortical neurons in culture at DIV 14. The distance between electrodes is 200  $\mu\text{m}$ . Each black dot corresponds to one of the recording electrodes. (c) Sample traces of spontaneous electrical activity recorded by three channels. Electrical events, spikes and bursts, that crossed a threshold (horizontal black line) were recorded. (d) Spontaneous activity of cortical neurons was analyzed according to electrophysiological features of spikes, bursts, synchrony, and connectivity. The figure was created using MC\_Rack 4.6 (<https://www.multichannelsystems.com/software/mc-rack>) (c) and Microsoft PowerPoint 365 (<https://www.microsoft.com/powerpoint>) (d).

hyperplane and creates boundaries to classify values in this high-dimensional space. SVM was implemented using a radial basis function kernel  $\sigma = 0.125$  and  $C = 6$ . In all cases, 75% of the dataset was used as a training set and 25% as a test set. Regression models were evaluated by the  $R^2$  coefficient (accuracy) and the root squared mean error (RMSE). For visualization purposes, Regression Error Characteristic (REC) curves<sup>50</sup> were estimated as the relation between the standardized error on the x-axis and the accuracy of the prediction method on the y-axis. Then, we used the SVM model to quantify the relative importance of the features for the predictive machine learning model. We balanced each group of electrophysiological features using two features per group: spikes (Ch. spikes, MFR), bursts (Ch. bursts, MBR), synchrony (STTC, STTC DBSCAN), and connectivity (Clustering coeff, Efficiency). We designed a “leave-one-in” strategy, in which only one group of variables was left and the results achieved by the model calculated, and, conversely, a “leave-one-out” which consisted of removing only one group of variables. To ensure the robustness of the training and test datasets, we applied an oversampling method to the dataset based on the Synthetic Minority Oversampling Technique (SMOTE)<sup>51</sup> using the R package smotefamily. Briefly, SMOTE finds the  $k$  nearest neighbors ( $k = 20$ ), selects a random neighbor, and creates a new synthetic instance between the feature space (in this step the minority class is not separated, and all samples were processed). This technique was applied with a ratio of 2 as desired synthetic instances over the original instances and then we evaluated the performance of the SVM model using fourfold cross-validation with 10 iteration cycles to minimize overfitting. Mean R squared ( $R^2$ ) of the out-fold (not selected for training) was used as the indicator of test fit goodness and error bars were computed using the SEM. All R packages are available at the CRAN repository (<https://cran.r-project.org>).

**Statistical analysis.** Statistical analyses were performed using Prism 8.0 (GraphPad Software, Inc, <https://www.graphpad.com>) and R 4.03 (The R Project for Statistical Computing, <https://www.r-project.org/>). Summary data and graphs were presented as mean  $\pm$  SEM, median and Interquartile range (IQR), or Tukey box plots. Significance levels were defined at  $P < 0.05$ . Additional detailed statistical information is provided in the figure legends.

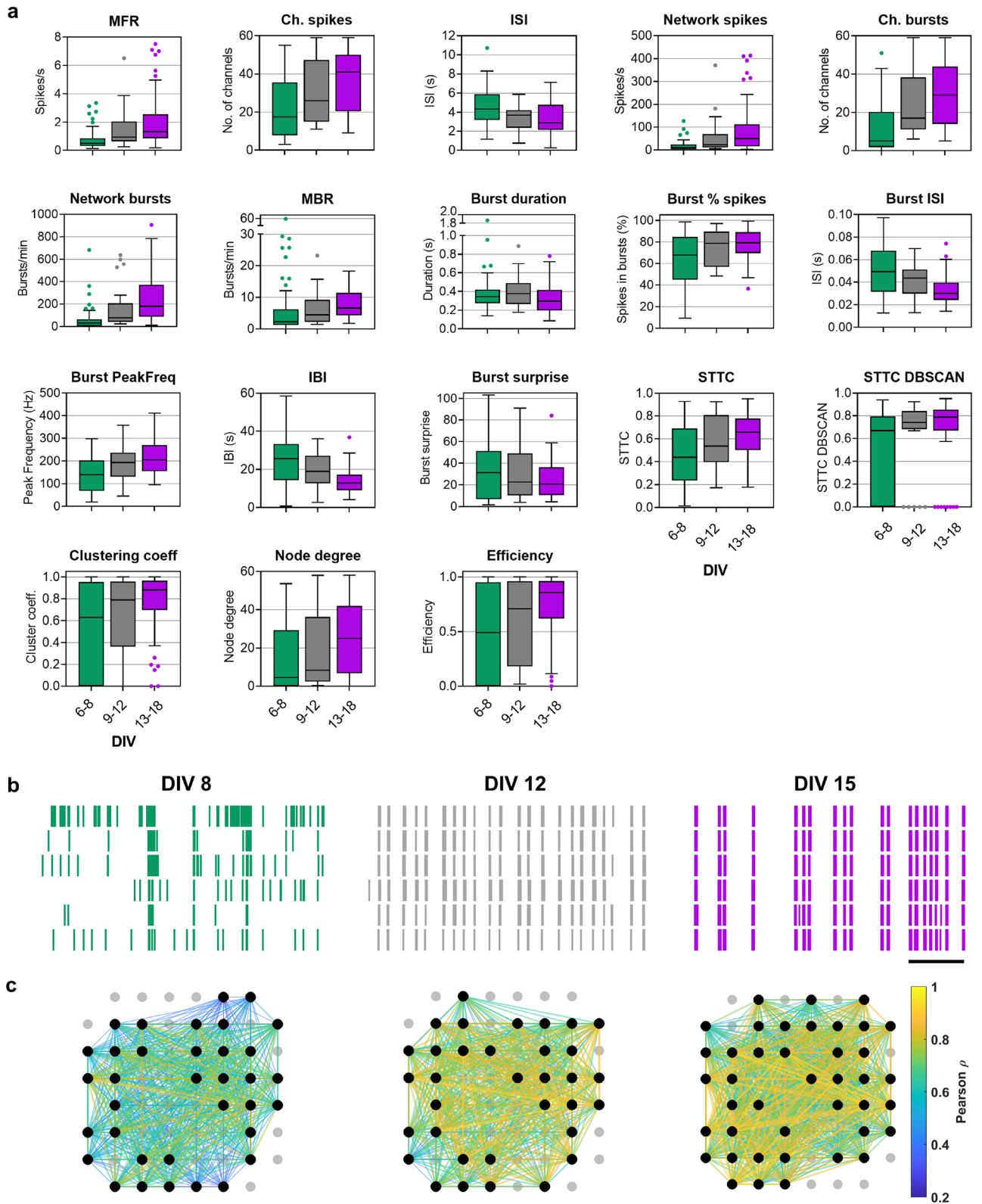
## Results

### Synchronized network activity emerges during early development of cortical neurons in vitro.

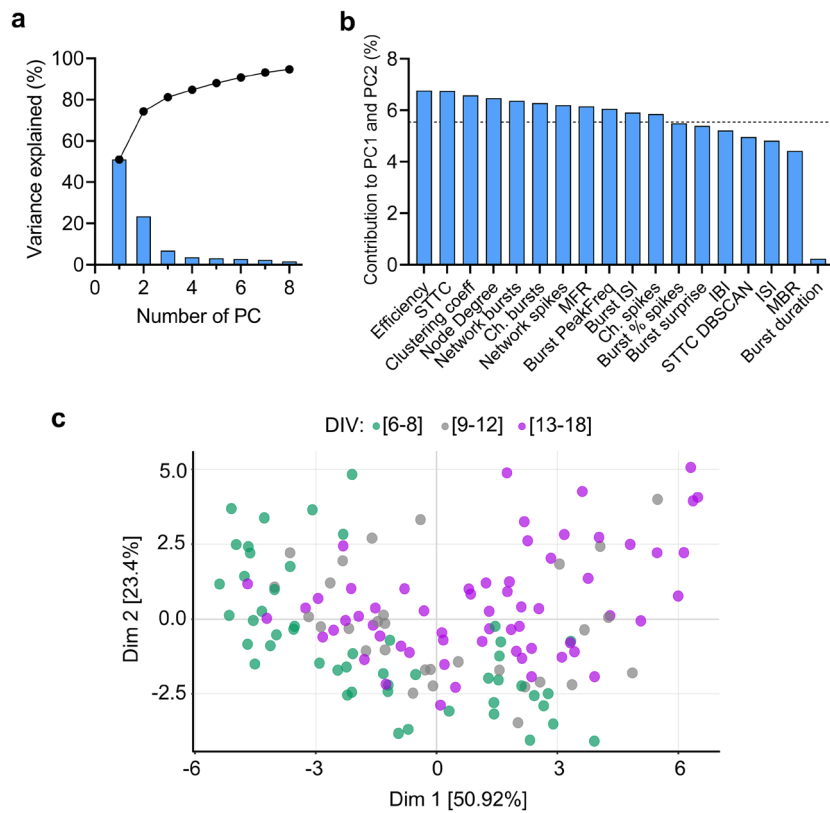
The spontaneous activity of cortical neurons in culture was recorded with MEAs (Fig. 1a, b) during the first three weeks in vitro. The extracellular electrical signals (spikes) crossing a negative voltage threshold were recorded (Fig. 1c) and 18 features of spikes, bursts, synchrony, and connectivity (see Table S1) were analyzed post hoc to characterize the network activity (Fig. 1d).

The values of the 18 electrophysiological features analyzed from MEA recordings were grouped in three DIV intervals (6–8, 9–12, and 13–18) which captured the changes occurring from the first to the third week in vitro (Fig. 2a, Table S2). During the first three weeks in vitro, spiking activity increased from a median MFR of 0.5 [IQR 0.3–0.9] spikes/s at DIV 6–8 to 1.3 [IQR 0.8–2.6] spikes/s at DIV 13–18 while the ISI decreased accordingly (Fig. 2a). The increase in spiking activity was also reflected in a higher number of channels registering spikes at DIV 13–18 (Fig. 2a). Bursts, i.e., intermittent groups of high-frequency spikes, are a key feature of network activity in vitro and were analyzed with the maximum interval method which has a high performance in detecting different types of bursts<sup>18</sup>. A maximum interval of 100 ms was defined to start a burst for consistency with the rest of the measures and within the range of previous studies<sup>18,52,53</sup>. Analysis of bursts showed that the median network burst activity had a sixfold median increase (30 to 180 network bursts/min) from the first





**Figure 2.** Development of spontaneous activity of cortical neurons in culture during the first three weeks in vitro. **(a)** Analysis of 18 electrophysiological features of neuronal network activity at DIV intervals: 6–8 ( $n = 50$ ), 9–12 ( $n = 30$ ), and 13–18 ( $n = 61$ ). The box plots show the median and the interquartile range with Tukey whiskers. Description of features is provided in Table S1 and  $p$ -values for **(a)** are reported in Table S2. **(b)** Raster plots show the spontaneous activity recorded by 6 electrodes from a representative neuronal network at DIV 8, 12, and 15 with an archetypal increase of spontaneous activity and synchrony with days in culture. Vertical lines correspond to spikes. Horizontal scale bar, 20 s. **(c)** Maps of functional connectivity of the same neuronal network and period as in **(b)**. The color of the connections (edges) between electrodes (nodes) represents Pearson's correlation and shows a general increase in connectivity during early development. The figure was created using Graphpad 8.0 (<https://www.graphpad.com>) **(a)**, Neuroexplorer 5 (<https://plexon.com/products/neuroexplorer/>) **(b)**, and Matlab 9.8 (<https://www.mathworks.com/products/matlab.html>) **(c)**.



**Figure 3.** Principal Component Analysis (PCA) of electrophysiological features of neuronal network activity in vitro. **(a)** The graph shows the percentage of variance explained by each principal component (PC) (bars) and the cumulative percentage of variance (dots) by the number of PCs. **(b)** Bar plot displays the contribution (%) of each of the 18 electrophysiological features to the PC1 and PC2. The dotted line represents the significant contribution value (5.55%) for the first two PCs. **(c)** Scatter plot of MEA recordings grouped by DIV intervals in the 2-dimensional PCA based on the 18 electrophysiological features. PC 1 (x-axis) and PC2 (y-axis) captured 74.3% of the total variance and each dot represents a time point recording within each DIV interval: 6–8 (green), 9–12 (gray), and 13–18 (purple) DIV. A gradient from young (DIV 6–8) to mature neuronal networks (DIV 13–18) is captured by the PCA projection: 38.3% of DIV 13–18 recordings are in the positive plane of PC1 and PC2 dimensions while no DIV 6–8 recordings are in the same PC area. The figure was created using Graphpad 8.0 (<https://www.graphpad.com>) (a,b), and R 4.03 (<https://www.r-project.org/>) (c).

to the third week in vitro (Fig. 2a). The percentage of spikes in bursts also increased across development and, consequently, a decrease in ISI in bursts, burst peak frequency, and burst surprise was observed during the same period (Fig. 2a). Whereas the variability and median length of IBI decreased approximately by half during the second week in vitro (25.6 [14.3–33.3] s at DIV 6–8 vs 12.8 [9–17.2] s at DIV 13–18), only a slight decrease in the median burst duration was observed (0.34 s at DIV 6–8 vs 0.30 s at DIV 13–18) (Fig. 2a).

STTC index<sup>19</sup> was used to analyze the level of synchronization of spontaneous activity. The average synchrony obtained with 100 ms bin showed a strong correlation with the percentage of spikes in bursts and within the range of average burst duration (Fig. S1a,b). We observed that levels of STTC and synchronized firing patterns increased from DIV 6–8 to 13–18 (Fig. 2a, b), as well as the number of highly synchronized groups of neurons revealed by applying DBSCAN clustering algorithm to STTC (STTC DBSCAN) results (Fig. 2a). Functional connectivity of neuronal networks on MEAs was analyzed using graph theory<sup>21</sup>, and cross-correlation matrices were calculated over 100 ms bins with an absolute threshold of 0.35 to capture a relevant range of network integration (Fig. S1c), in concordance with previous studies<sup>22</sup>. During the first week in vitro, there was high variability in the levels of Clustering coeff and Efficiency, indicators of network's levels of segregation and integration<sup>43</sup>, respectively (Fig. 2a). The levels of Node degree increased during the second week in vitro (Fig. 2a, c), and high median levels of Clustering coeff (0.88) and Efficiency (0.86) were representative of neuronal networks at DIV 13–18.

### Dimensionality reduction analysis for integrating electrophysiological features of neuronal network activity.

To explore how the 18 electrophysiological features defined the network activity of cortical neurons in each DIV interval, we conducted a PCA as a method to capture the greatest variance in a low-dimensional projection. The lower-dimensional space defined by the first two PC dimensions accounted for 74.32% of the variance (Fig. 3a). The electrophysiological features that better characterized the variance in the first two PCs were related to network features such as Network bursts, Efficiency, and STTC (Fig. 3b). Despite

the lack of clear segmentation between DIV intervals due to the variability within each developmental stage, the first dimension of the PCA captured a left to right gradient from DIV 6–8 to DIV 13–18 (Fig. 3c). A similar distribution of MEA recordings by DIV intervals persisted when the PCA was applied to each cortical culture (Fig. S2a). Moreover, the variability between and within cultures did not seem to depend upon the range of cell densities used in our study (Fig. S2a–c).

**Clustering analysis to explore developmental patterns in neuronal networks in vitro.** We used clustering techniques to explore whether it was possible to characterize individual neuronal networks with similar developing network activity within the range of activity levels found in our dataset. To do this, we applied k-means clustering at the centroids of the SOM<sup>33</sup> on two representative network features of synchrony and bursting activity: STTC and Ch. bursts, respectively. For the 47 MEA dishes used in this study, the matrix of the SOM algorithm showed 25 nodes in a hexagonal grid, in which each node represented neuronal networks with similar network activity development (Fig. S3a, b). Then, the k-means analysis of the SOM revealed 3 clusters of neuronal networks with different patterns of development for the parameters of Ch. bursts (Fig. 4a) and STTC (Fig. 4d). We then quantified the 18 electrophysiological features at DIV 6–8 and DIV 13–18 in the neuronal networks included in each cluster (see Table S3).

Clusters 1 to 3 of Ch. bursts showed neuronal networks with developmental patterns from high to low number of channels with burst activity (Fig. 4a) that highly correlated with the changes shown in representative network features of spikes, bursts, synchrony, and connectivity from DIV 6–8 to DIV 13–18 (Fig. 4c). MEAs included in cluster 1 of Ch. bursts were characterized by higher levels of STTC and connectivity at DIV 6–8 as well as the highest increase of network spikes and bursts from DIV 6–8 to DIV 13–18 (Fig. 4c). Conversely, networks with a low number of channels with bursts at DIV 6–8 (cluster 3) evolved into poorly synchronized and integrated networks (Fig. 4b, c).

Clustering based on the synchrony feature STTC included neuronal networks that kept either high (cluster 1) or low (cluster 3) levels of STTC during the three developmental stages, and a group of neuronal networks (cluster 2) with a high increase of synchrony from DIV 6–8 to DIV 13–18 (Fig. 4d). Whereas the changes from DIV 6–8 to DIV 13–18 in the electrophysiological features of neuronal networks included in clusters 1 or 3 of STTC were relatively similar to the equivalent clusters of Ch. bursts (Fig. 4c), neuronal networks included in cluster 2 of STTC were characterized by drastic increases in the levels of STTC and Efficiency from DIV 6–8 to DIV 13–18 (Fig. 4e, f).

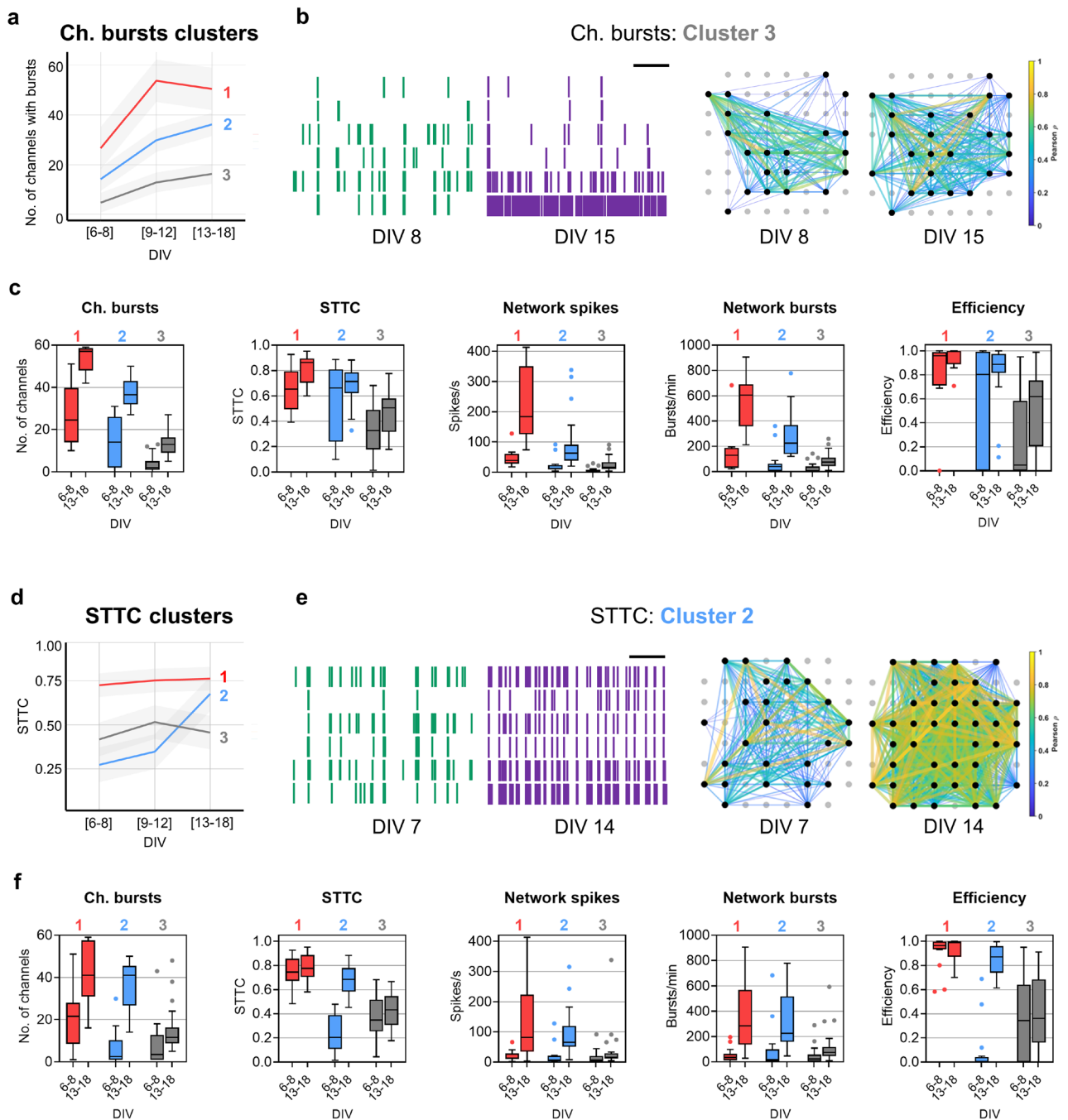
The maximum congruence between clustering methods occurred in neuronal networks included in cluster 3 of STTC and Ch. burst; clusters 1 and 2 of Ch. bursts had a good correlation with cluster 1 of STTC (high levels of STTC at DIV 13–18), but 50% of neuronal networks included in cluster 2 of STTC were also included in cluster 3 of Ch. bursts (Fig. S3c). Additionally, PCA analysis performed on the 18 electrophysiological features of MEA recordings at DIV 6–8 and DIV 13–18 showed that the correlation between electrophysiological features was different at each DIV interval (Fig. S3d), and separation of neuronal networks included in clusters of Ch. bursts (Fig. S3e) and STTC was less clear in the PCA projection (Fig. S3f).

Altogether, these results suggest that clustering methods were effective in characterizing neuronal networks with similar changes in spontaneous activity during early development.

**Prediction of levels of synchrony and bursting activity in cultured cortical neurons using machine learning.** To further investigate the importance of early network activity suggested by the clustering results, we used machine learning techniques to predict the levels of Ch. bursts and STTC at the third week in vitro. These two parameters were compared with the prediction of MFR, one of the most commonly used parameters in MEA studies<sup>15</sup>. The dataset with the 18 electrophysiological features from MEA recordings at DIV 6–8 and DIV 13–18 was divided into training and testing datasets. After training each machine learning model, the performance of the models was evaluated based on the accuracy prediction ( $R^2$ ) of the features at DIV 13–18 in the test dataset (Fig. 5a).

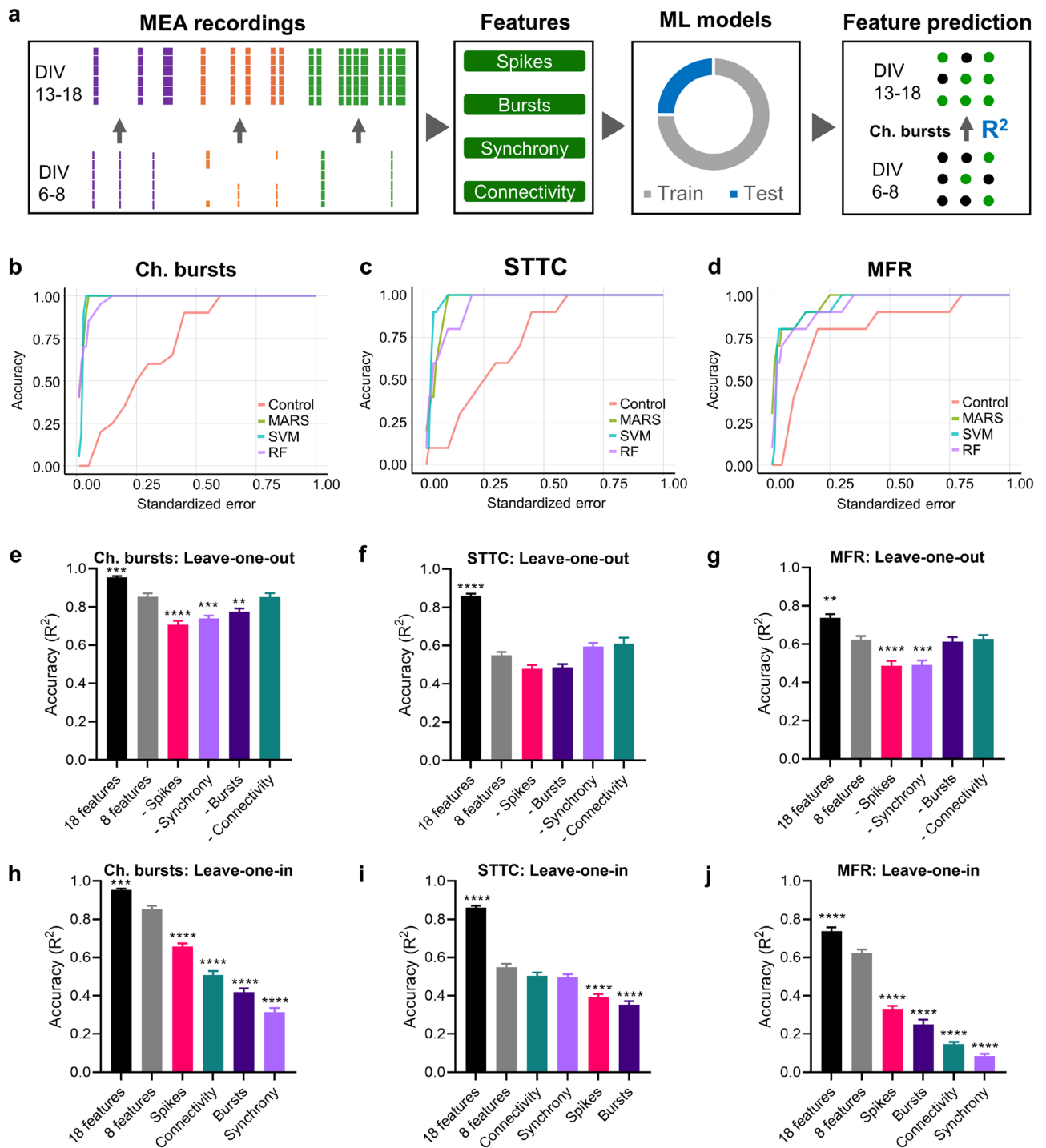
We used three machine learning methods with different learning algorithms: MARS, SVM, and Random Forest. The three models yielded predictive accuracies higher than 0.9 during testing for the features of Ch. bursts and STTC, and higher than 0.8 for MFR (Table S4, and Fig. S4 for the SVM model). Additionally, comparison of model performance with REC curves showed that MARS and SVM models performed slightly better than Random Forest for Ch. bursts and STTC, and similarly for MFR (Fig. 5b–d). In all machine learning models, Ch. spikes and ISI appeared to be the most important electrophysiological features for the prediction of Ch. bursts and MFR, respectively, while the rest of the features had relative importance depending on the model (Table S5). Also, features of functional connectivity had the highest relative importance for the prediction of STTC values in all machine learning models (Table S5). These results show the consistency for the prediction of electrophysiological properties of cortical neurons in culture across machine learning models.

To further identify which electrophysiological features were most informative for the prediction of Ch. bursts, STTC, and MFR, we tested the predictive accuracy of the features in the SVM model using cross-validation leave-one-out and leave-one-in strategies. To balance the number of features, we tested these strategies with 2 parameters in each of the four groups of electrophysiological features: spikes, bursts, synchrony, and connectivity (Fig. 5e–j). In the leave-one-out strategy, one group of features was removed each time, and in the leave-one-in, only one group of variables was left each time. Predictive accuracy with SVM for Ch. bursts was higher than 0.8 when the number of electrophysiological features was reduced from 18 to 8 (Fig. 5e), while accuracy dropped to 0.6 or lower for MFR and STTC, respectively (Fig. 5f, g). Evaluation of feature importance showed that when leaving out only one group of two features, prediction performance for the 3 variables slightly decreased (Fig. 5e–g), suggesting that the rest of the electrophysiological features still retain predictive information. However, when



**Figure 4.** Patterns of developing spontaneous activity in neuronal networks using clustering analysis. **(a,d)** Graphs display three developmental patterns for the features of Ch. bursts **(a)** and STTC identified by k-means clustering of the SOM. The lines represent the means of the observations, labelled by k-means, and the shaded areas represent the 95% confidence intervals. **(b,e)** Examples of raster plots (left panels) and connectivity maps (right panels) of neuronal networks included in either cluster 3 of Ch. bursts **(b)** or cluster 2 of STTC **(e)**. Horizontal scale bar in raster plots, 20 s. Color of the edges between nodes in connectivity maps represents Pearson's correlation. **(c,f)** Changes in Ch. bursts, STTC, Network spikes, Network bursts, and Efficiency between DIV 6–8 and 13–18 in each cluster of Ch. bursts **(c)** and STTC **(f)**. The box plots show the median and the interquartile range with Tukey whiskers. Clusters are indicated on top of the graphs. The figure was created using Graphpad 8.0 (<https://www.graphpad.com>) **(c,f)**, Neuroexplorer 5 (<https://plexon.com/products/neuroexplorer/>), Matlab 9.8 (<https://www.mathworks.com/products/matlab.html>) **(b,e)**, and R 4.03 (<https://www.r-project.org/>) **(a, d)**.





**Figure 5.** Machine learning for the prediction of features of bursts, synchrony, and spikes. (a) Schematics of the machine learning workflow. The 18 electrophysiological features were analyzed from MEA recordings at DIV 6–8 and DIV 13–18 and these features were used to train (gray segment) machine learning (ML) models. Then, accuracy prediction ( $R^2$ ) at DIV 13–18 was evaluated in the test dataset (blue segment) for electrophysiological features (e.g., Ch. bursts). Figure created with Microsoft PowerPoint 365 (<https://www.microsoft.com/powerpoint>). (b–d). ROC curves for control and MARS, SVM, and Random Forest (RF) machine learning models for the prediction of STTC (b), Ch. bursts (c), and MFR (d). ROC curves represent the cumulative error obtained by the machine learning models. (e–j) SVM prediction of Ch. bursts (e–h), STTC (f–i), and MFR (g–j) using leave-one-out (e–g) and leave-one-in (h–j) strategies. We performed fourfold cross-validation (10 iterations) to calculate the mean accuracy of the SVM prediction using 8 electrophysiological features (2 features per group): MFR and Ch. spikes (spikes), MBR and Ch. bursts (bursts), STTC and DBSCAN STTC (synchrony), and Efficiency and Clustering coeff (connectivity). Feature importance was evaluated by removing (leave-one-out) or leaving (leave-one-in) one group of features. Baseline accuracy of cross-validation was calculated using the 18 features. Error bars represent the mean  $\pm$  SEM. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$  vs accuracy obtained with 8 features (one-way ANOVA, followed by Dunnett's post hoc multiple comparison test). The figure was created using Microsoft PowerPoint 365 (<https://www.microsoft.com/powerpoint>) (a), R 4.03 (<https://www.r-project.org/>) (b), and Graphpad 8.0 (<https://www.graphpad.com>) (e–j).

SVM was run on each group of variables individually, only features of spikes (MFR and Ch. spikes) got an accuracy over 0.6 for the prediction of Ch. bursts (Fig. 5h). Synchrony and connectivity features kept similar predictive information for STTC (Fig. 5i), whereas prediction of MFR was not accurately performed (accuracy < 0.4) for any group of features (Fig. 5j). Altogether, the cross-validation analysis confirmed that electrophysiological features at DIV 6–8 were highly correlated with the levels of burst and synchronization at the third week in vitro.

## Discussion

Cortical neurons in vitro exhibit along development a wide variety of firing patterns whose descriptive classification and characterization are challenging<sup>10,54</sup>. In this study, we show that the development of neural network activity in cortical neuron cultures recorded by MEAs can be characterized by combining features of spikes, bursts, synchrony, and connectivity. The use of machine learning techniques allowed us to identify clusters of neuronal networks with similar developing spontaneous activity within the range of synchronized and bursting activity in cultured neurons during early development. Indeed, the levels of Ch. bursts, STTC, and MFR at DIV 13–18 could be successfully predicted based on the activity at DIV 6–8 using three different machine learning models, confirming the correlation between the activity in early developmental stages and features of activity in mature neuronal networks.

Our findings are consistent with the average increase in synchronized spiking and bursting activity during the first 3 weeks in vitro reported in previous studies<sup>9–12</sup>. Increases in peak frequency in bursts, probability of burst events (Burst surprise), and percentage of spikes in bursts from DIV 6–8 to DIV 13–18 are consistent with developmental changes toward a network activity dominated by burst activity. Synchronization (STTC) and highly synchronized hubs of neurons (STTC DBSCAN) progressively appeared during the second week in vitro and this type of changes have been previously shown to be strongly correlated with the strengthening of synapses<sup>9,55</sup>, both glutamatergic and GABAergic<sup>56</sup>. The maturation and consequent rate of spontaneous activity in vitro depends also upon the neuronal density of cultures<sup>10,12,14</sup>. The cortical cultures used in this study can be classified as dense cultures (1750–3500 cells/mm<sup>2</sup>) in which bursting activity can be usually observed after only one week in vitro<sup>10</sup>. Therefore, caution should be taken to extrapolate our results to sparse cortical cultures with different network dynamics during early development.

Our results indicate that the analysis of features of spikes, bursts, synchrony, and connectivity may help to characterize the developmental stage of neuronal network activity during the first three weeks in culture. By applying the PCA dimensionality reduction technique to these electrophysiological features, we show that a broad separation of neuronal networks at the first and the third week in vitro could be better accomplished by properties of network activity (e.g., number of active channels or connectivity properties) rather than single features of spikes or bursts. Indeed, the PCA analysis suggested that there may be some redundancy in the information encoded by the 18 electrophysiological features since the two first PC dimensions accounted for approx. 75% of the variance and some network features such as Network spikes and bursts seemed to be strongly correlated.

Our findings also indicated that the variability of activity levels observed in neuronal networks within the same developmental stage, also reported in previous studies<sup>10,12,15</sup>, can be efficiently examined when using clustering analysis across developmental stages. The clusters of neuronal networks defined by the k-means analysis of the SOM based on Ch. bursts (i.e., spatial distribution of burst activity) highly correlated with changes in the four groups of electrophysiological features, whereas STTC clustering was less efficient in characterizing neuronal networks with low-intermediate levels of synchronization and divergent development. This might indicate that neuronal networks with highly isolated burst activity at DIV 6–8 may not always develop strong synchronized burst activity. Thus, Ch. bursts may be used as a predictive marker of network development, considering also that bursts are commonly the consequence of cooperative network activity<sup>57</sup>.

Results obtained in the analysis of functional connectivity suggests that the neuronal networks included in the clusters of high Ch. burst and STTC levels (cluster 1 in both clustering methods) already had properties of small-world topology (high levels of segregation and integration) at 7 DIV, whereas other neuronal networks kept their initial random-like topology at the third week in vitro. Previous studies using 60 electrode MEAs have shown that topology may evolve either from random structure<sup>22</sup> or from hubs densely interconnected<sup>23</sup>. However, these divergent results may be related to differences in culture densities<sup>21</sup>. It also remains unclear whether synchronized bursts may be triggered by highly active neurons<sup>58</sup>, or bursts may be instead the result of zones with intermediate activity near the network's boundary as suggested by studies using large-scale MEAs<sup>59</sup>. According to our results, levels of segregation and integration were in general tightly correlated across development. However, as the 60 electrode MEAs cover a partial window of the entire network, the presence of other topologies cannot be totally excluded. Application of clustering techniques to the connectivity analysis of cultured neuronal networks with different cellular densities on high-density electrode arrays may help to elucidate how the emergence of network topology is affected during early development.

Spontaneous activity at early postnatal stages seems to play an essential role in the correct maturation of mammalian cortical networks<sup>5,60</sup> and machine learning techniques allowed us to further explore the correlation between the activity of immature and mature cortical neuronal networks in vitro. Overall, all three machine learning models (MARS, SVM, and Random Forest) used in this study, efficiently captured the nonlinear changes of network activity that occurred during the second week in vitro and were able to predict with high accuracy the values of Ch. bursts, STTC, and MFR. These results suggest that machine learning techniques might be successfully used as internal control for long-term experiments with developing neuronal networks. Analysis of the importance of the different electrophysiological features in the machine learning models showed that the number of channels with bursts can be predicted with good accuracy even with only two features of spikes (Ch. spikes and MFR) whereas prediction of MFR itself had much lower accuracy, suggesting that the fluctuations in firing rate during development might require a continuous dataset for better prediction. Similarly, the

prediction of STTC was less accurate when a lower number of features was used in the SVM model, suggesting that synchronization may require additional features of bursts and connectivity for better characterization. A complementary explanation could be that low and intermediate levels of synchronous activity at DIV 7 may not reflect how synchronization will emerge later in development, as a consequence of recurrent bursting activity<sup>61–63</sup>. In summary, we may argue that, in dense cortical cultures, random spike activity largely distributed across the neuronal network precedes the development of burst activity which may need to reach a certain threshold to become synchronized across the network. Altogether, our clustering and machine learning results suggest that the initial network activity of in vitro neuronal networks may predetermine the consequent developmental trend of network activity.

Due to the somewhat limited dataset and the dependence of network activity on experimental conditions, the interpretation of the results requires caution. Although internal validation of the machine learning models reduces biases and overfitting, further external validation with datasets of similar experimental conditions would reinforce the machine learning results<sup>26</sup>. Moreover, it will be interesting also to explore the influence of parameters such as neuronal survival during the second week in vitro, the balance between excitation and inhibition<sup>64</sup>, or neurite outgrowth<sup>65</sup> on different developmental patterns of network activity. Nevertheless, the methodological approach presented here for the characterization and early prediction of network activity in cultured neurons may be a useful tool to elucidate the biological causes of variability in network dynamics as well as to optimize the use of cortical cultures in models of health and disease.

### Data availability

Any supplementary information, code, and source data are available from the corresponding authors (A.N., D.C.-G., and J.V.B.) upon reasonable request.

Received: 31 May 2021; Accepted: 27 September 2021

Published online: 14 October 2021

### References

- Izhikevich, E. M., Desai, N. S., Walcott, E. C. & Hoppensteadt, F. C. Bursts as a unit of neural information: Selective communication via resonance. *Trends Neurosci.* **26**, 161–167 (2003).
- Salinas, E. & Sejnowski, T. J. Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* **2**, 539–550 (2001).
- Kepecs, A. & Lisman, J. Information encoding and computation with spikes and bursts. *Netw. Comput. Neural Syst.* **14**, 103–118 (2003).
- Khazipov, R. & Luhmann, H. J. Early patterns of electrical activity in the developing cerebral cortex of humans and rodents. *Trends Neurosci.* **29**, 414–418 (2006).
- Luhmann, H. J. *et al.* Spontaneous neuronal activity in developing neocortical networks: From single cells to large-scale interactions. *Front. Neural Circuits* **10**, 1–14 (2016).
- Uhlhaas, P. J. & Singer, W. Neural synchrony in brain disorders: Relevance for cognitive dysfunctions and pathophysiology. *Neuron* **52**, 155–168 (2006).
- Feldt, S., Bonifazi, P. & Cossart, R. Dissecting functional connectivity of neuronal microcircuits: Experimental and theoretical insights. *Trends Neurosci.* **34**, 225–236 (2011).
- Obien, M. E. J., Deligkaris, K., Bullmann, T., Bakkum, D. J. & Frey, U. Revealing neuronal function through microelectrode array recordings. *Front. Neurosci.* **9**, 423 (2015).
- Van Pelt, J., Vajda, I., Wolters, P. S., Corner, M. A. & Ramakers, G. J. A. Dynamics and plasticity in developing neuronal networks in vitro. *Prog. Brain Res.* **147**, 173–188 (2004).
- Wagenaar, D. A., Pine, J. & Potter, S. M. An extremely rich repertoire of bursting patterns during the development of cortical cultures. *BMC Neurosci.* **7**, 1–18 (2006).
- Chiappalone, M., Bove, M., Vato, A., Tedesco, M. & Martinoia, S. Dissociated cortical networks show spontaneously correlated activity patterns during in vitro development. *Brain Res.* **1093**, 41–53 (2006).
- Biffi, E., Regalia, G., Menegon, A., Ferrigno, G. & Pedrocchi, A. The influence of neuronal density and maturation on network activity of hippocampal cell cultures: a methodological study. *PLoS ONE* **8**, e83899 (2013).
- Novellino, A. *et al.* Development of micro-electrode array based tests for neurotoxicity: Assessment of interlaboratory reproducibility with neuroactive chemicals. *Front. Neuroeng.* **4**, 4 (2011).
- Ito, D. *et al.* Minimum neuron density for synchronized bursts in a rat cortical culture on multi-electrode arrays. *Neuroscience* **171**, 50–61 (2010).
- Negri, J., Menon, V. & Young-Pearse, T. L. Assessment of spontaneous neuronal activity In vitro using multi-well multi-electrode arrays: Implications for assay development. *eNeuro* **7**, 1–27 (2020).
- Vassallo, A. *et al.* A multi-laboratory evaluation of microelectrode array-based measurements of neural network activity for acute neurotoxicity testing. *Neurotoxicology* **60**, 280–292 (2017).
- Dastgheyb, R. M., Yoo, S.-W. & Haughey, N. J. MEAnalyzer—A spike train analysis tool for multi electrode arrays. *Neuroinformatics* **18**, 163–179 (2020).
- Cotterill, E., Charlesworth, P., Thomas, C. W., Paulsen, O. & Eglon, S. J. A comparison of computational methods for detecting bursts in neuronal spike trains and their application to human stem cell-derived neuronal networks. *J. Neurophysiol.* **116**, 306–321 (2016).
- Cutts, C. S. & Eglon, X. S. J. Detecting pairwise correlations in spike trains: An objective comparison of methods and application to the study of retinal waves. *J. Neurosci.* **34**, 14288–14303 (2014).
- Bullmore, E. & Sporns, O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **10**, 186–198 (2009).
- Poli, D., Pastore, V. P. & Massobrio, P. Functional connectivity in in vitro neuronal assemblies. *Front. Neural Circuits* **9**, 1–14 (2015).
- Downes, J. H. *et al.* Emergence of a small-world functional network in cultured neurons. *PLoS Comput. Biol.* **8**, e1002522 (2012).
- Schroeter, M. S., Charlesworth, P., Kitzbichler, M. G., Paulsen, O. & Bullmore, E. T. Emergence of rich-club topology and coordinated dynamics in development of hippocampal functional networks In vitro. *J. Neurosci.* **35**, 5459–5470 (2015).
- Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014).
- Glaser, J. I., Benjamin, A. S., Farhoodi, R. & Kording, K. P. The roles of supervised machine learning in systems neuroscience. *Prog. Neurobiol.* **175**, 126–137 (2019).
- Vu, M. A. T. *et al.* A shared vision for machine learning in neuroscience. *J. Neurosci.* **38**, 1601–1607 (2018).

27. Cotterill, E. *et al.* Characterization of early cortical neural network development in multiwell microelectrode array plates. *J. Biomol. Screen.* **21**, 510–519 (2016).
28. Charlesworth, P., Cotterill, E., Morton, A., Grant, S. G. N. & Eglén, S. J. Quantitative differences in developmental profiles of spontaneous activity in cortical and hippocampal cultures. *Neural Dev.* **10**, 1–10 (2015).
29. Madhavan, R., Chao, Z. C. & Potter, S. M. Plasticity of recurring spatiotemporal activity patterns in cortical networks. *Phys. Biol.* **4**, 181–193 (2007).
30. Glaser, J. I. *et al.* Machine learning for neural decoding. *eNeuro* **7**, 1–16 (2020).
31. Barrett, D. G., Morcos, A. S. & Macke, J. H. Analyzing biological and artificial neural networks: Challenges with opportunities for synergy?. *Curr. Opin. Neurobiol.* **55**, 55–64 (2019).
32. Chicco, D. T. quick tips for machine learning in computational biology. *BioData Min.* **10**, 35 (2017).
33. Kohonen, T. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69 (1982).
34. Tsigelny, I. F. *et al.* Analysis of metagene portraits reveals distinct transitions during kidney organogenesis. *Sci. Signal.* **1**, ra16–ra16 (2008).
35. Das, D., Banerjee, N. & Zhang, M. Q. Interacting models of cooperative gene regulation. *Proc. Natl. Acad. Sci. USA.* **101**, 16234 LP–16239 (2004).
36. Modak, S., Mehta, S., Sehgal, D. & Valadi, J. Application of support vector machines in viral biology. In *Global Virology III: Virology in the 21st Century* (eds Shapshak, P. *et al.*) 361–403 (Springer International Publishing, 2019).
37. Riddick, G. *et al.* Predicting in vitro drug sensitivity using Random Forests. *Bioinformatics* **27**, 220–224 (2011).
38. Pérez-Gómez, A. *et al.* Selective enhancement of domoic acid toxicity in primary cultures of cerebellar granule cells by lowering extracellular Na<sup>+</sup> concentration. *Toxicol. Sci.* **161**, 1–12 (2017).
39. Novelli, A., Reilly, J. A., Lysko, P. G. & Henneberry, R. C. Glutamate becomes neurotoxic via the N-methyl-D-aspartate receptor when intracellular energy levels are reduced. *Brain Res.* **451**, 205–212 (1988).
40. Eisenman, L. N., Emmett, C. M., Mohan, J., Zorumski, C. F. & Mennerick, S. Quantification of bursting and synchrony in cultured hippocampal neurons. *J. Neurophysiol.* **114**, 1059–1071 (2015).
41. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* 226–231 (AAAI Press, 1996).
42. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
43. Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *Neuroimage* **52**, 1059–1069 (2010).
44. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, (2008).
45. Wehrens, R. & Kruisselbrink, J. Flexible self-organizing maps in kohonen 3.0. *J. Stat. Softw.* **87**, (2018).
46. Vesanto, J. & Alhoniemi, E. Clustering of the self-organizing map. *IEEE Trans. Neural Netw.* **11**, 586–600 (2000).
47. Friedman, J. H. Multivariate adaptive regression splines. *Ann. Stat.* **19**, 1–67 (1991).
48. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
49. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
50. Bi, J. & Bennett, K. P. Regression error characteristic curves. In *Proceedings of the AIII 20th International Conference on Machine Learning* 8 (2003).
51. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
52. Gullo, F., Maffezzoli, A., Dossi, E. & Wanke, E. Short-latency cross- and autocorrelation identify clusters of interacting cortical neurons recorded from multi-electrode array. *J. Neurosci. Methods* **181**, 186–198 (2009).
53. Brown, J. P. *et al.* Evaluation of a microelectrode array-based assay for neural network ontogeny using training set chemicals. *Toxicol. Sci.* **154**, 126–139 (2016).
54. Stegenga, J., Le Feber, J., Marani, E. & Rutten, W. L. C. Analysis of cultured neuronal networks using intraburst firing characteristics. *IEEE Trans. Biomed. Eng.* **55**, 1382–1390 (2008).
55. Ichikawa, M., Muramoto, K., Kobayashi, K., Kawahara, M. & Kuroda, Y. Formation and maturation of synapses in primary cultures of rat cerebral cortical cells: An electron microscopic study. *Neurosci. Res.* **16**, 95–103 (1993).
56. Ito, D., Komatsu, T. & Gohara, K. Measurement of saturation processes in glutamatergic and GABAergic synapse densities during long-term development of cultured rat cortical networks. *Brain Res.* **1534**, 22–32 (2013).
57. Vardi, R., Goldental, A., Sardi, S., Sheinin, A. & Kanter, I. Simultaneous multi-patch-clamp and extracellular-array recordings: Single neuron reflects network activity. *Sci. Rep.* **6**, 1–9 (2016).
58. Gritsun, T. A., Le Feber, J., Stegenga, J. & Rutten, W. L. C. Network bursts in cortical cultures are best simulated using pacemaker neurons and adaptive synapses. *Biol. Cybern.* **102**, 293–310 (2010).
59. Okujeni, S. & Egert, U. Inhomogeneities in network structure and excitability govern initiation and propagation of spontaneous burst activity. *Front. Neurosci.* **13**, 1–14 (2019).
60. Egorov, A. V. & Draguhn, A. Development of coherent neuronal activity patterns in mammalian cortical networks: Common principles and local heterogeneity. *Mech. Dev.* **130**, 412–423 (2013).
61. Chiappalone, M., Vato, A., Berdondini, L., Koudelka-Hep, M. & Martinoia, S. Network dynamics and synchronous activity in cultured cortical neurons. *Int. J. Neural Syst.* **17**, 87–103 (2007).
62. Maeda, E., Robinson, H. & Kawana, A. The mechanisms of generation and propagation of synchronized bursting in developing networks of cortical neurons. *J. Neurosci.* **15**, 6834–6845 (1995).
63. Wagenaar, D. A., Nadasdy, Z. & Potter, S. M. Persistent dynamic attractors in activity patterns of cultured neuronal networks. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **73**, 051907 (2006).
64. Teppola, H., Acimović, J. & Linne, M. L. Unique features of network bursts emerge from the complex interplay of excitatory and inhibitory receptors in rat neocortical networks. *Front. Cell. Neurosci.* **13**, 1–22 (2019).
65. Okujeni, S. & Egert, U. Self-organization of modular network architecture by activity-dependent neuronal migration and outgrowth. *Elife* **8**, 1–29 (2019).

## Acknowledgements

The authors would like to thank Davide Bassetti for helpful discussion and comments.

## Author contributions

A.N., D.C.-G., J.V.-B. and M.T. F.-S. conceptualized and designed the study. D.C.-G. and D.W. performed the experiments and acquired the data. D.C.-G., D.W., J.V.-B., and P.F. analyzed the data and prepared the figures. All authors contributed to data interpretation. D.C.G. prepared the first draft of the manuscript and all authors edited, commented, and approved the final manuscript.



## Funding

This work was supported by MINECO-18-CTQ2017-86994-R project from the Spanish Ministry of Economy and Competitiveness (MINECO).

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-99538-9>.

**Correspondence** and requests for materials should be addressed to D.C.-G., A.N. or J.M.V.-B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021