CellExplorer: a graphical user interface and standardized pipeline for visualizing and characterizing single neuron features

Peter C. Petersen[1] and György Buzsáki[1, 2, 3]

Affiliations:
[1]Neuroscience Institute and [2]Department of Neurology, Langone Medical Center, New York University, New York, NY 10016, USA
[3]Center for Neural Science, New York University, New York, NY 10003, USA

*Correspondence: gyorgy.buzsaki@nyumc.org

Abstract

The large diversity of neuron types of the brain, with numerous unique electrophysiological characteristics, provides the means by which cortical circuits perform complex operations. To quantify, compare and visualize the functional features of single neurons, we have developed a MATLAB-based framework, CellExplorer, consisting of three components: a processing module for extracting and calculating physiological metrics, a standardized yet flexible data structure, and a powerful graphical interface for fast manual curation and feature exploration. This data mining and discovery tool allows for inspection of dozens of computed features of neurons from large-scale recordings and relate them to those of other neurons in any combination at the speed of mouse clicks. The open source design of the CellExplorer permits the optimization of its functions tested against an ever-growing community-contributed database. CellExplorer will accelerate linking physiological properties of single neurons in the intact brain to genetically identified types.

INTRODUCTION

Discovering novel mechanisms in brain circuits requires high-resolution monitoring of the constituent neurons and understanding the nature of their interactions. Identification and manipulation of different neuron types in the behaving animal is a prerequisite to decipher their role in circuit dynamics and behavior. Yet, currently a large gap exists between neuron classification schemes based on molecular and physiological methods (Kepecs and Fishell, 2014; Klausberger and Somogyi, 2008; McBain and Fisahn, 2001; Roux and Buzsáki, 2015; Rudy et al., 2011).

An assertion of large-scale single unit recording method is that the relationship between neuronal firing and behavioral and cognitive variables can provide insights about the computational role of neurons and neuronal assemblies (Barlow, 1972; Buzsáki, 2004). However, exploiting the power of correlations between neuronal firing and behavioral variables requires multiple-level characterization of single neurons and their interactions along with numerous controls to exclude the contributions of many potential hidden variables. Ideally, neuron features need to be described at multiple levels of complexity. Long-term simultaneous recordings from large numbers of neurons allows building of extensive batteries of neuron properties (Fig. 1). The first level is description of the biophysical and physiological characteristics of single neurons. This step includes waveform features, their physical position relative to recording sites and other units (Csicsvari et al., 2003), interspike intervals statistics and autocorrelograms. These first-level parameters are 'fixed' features and can be used to combine data sets across animals both within and across laboratories for first-order separation of single cells into putative major classes, typically excitatory and inhibitory cells. The second level relates single cells to other neurons, and includes cross-correlations, monosynaptic connections, relationship to multiple oscillatory and irregular local field potentials (LFP) and unit population patterns. Additionally, it can describe long-term firing rates in defined brain states. The third level descriptors of single unit activity, ideally, should include the relationship between its firing patterns and over behavioral correlates, including spontaneous motor patterns and autonomic parameters (McGinley et al., 2015; Steinmetz et al., 2019). Verification and refinement of these properties can be assisted by optogenetic methods, which can relate physiological parameters to genetically identified neuron groups (Boyden et al., 2005; Klausberger and Somogyi, 2008; Rudy et al., 2011; Buzsáki et al., 2015; Roux and Buzsáki, 2015). Antidromic and unit-LFP coupling techniques provide further assignment of single neurons to cortical regions, layers and target projections (Ciocchi et al., 2015; Senzai et al., 2019; Zhang et al., 2013).

This three-level description provides universal features of neuronal activity common to all experimental paradigms and, therefore, is communicable across different experiments and laboratories. In turn, these universal features can be contrasted and compared with higher level correlates, such as learning, memory, decision making, emotions and social interactions. Because these inferred high-level correlates of neuronal activity are often paradigm-specific and may differ across laboratories, the three-level analysis can guard against mistakenly assigning cognitive roles of neuronal spiking that may be explained by measurable overt correlates. Yet, even if all of the above information is available separately, factoring out critical variables and their combinations is possible only when the multitudes of the single neuron characteristics can be compared flexibly.

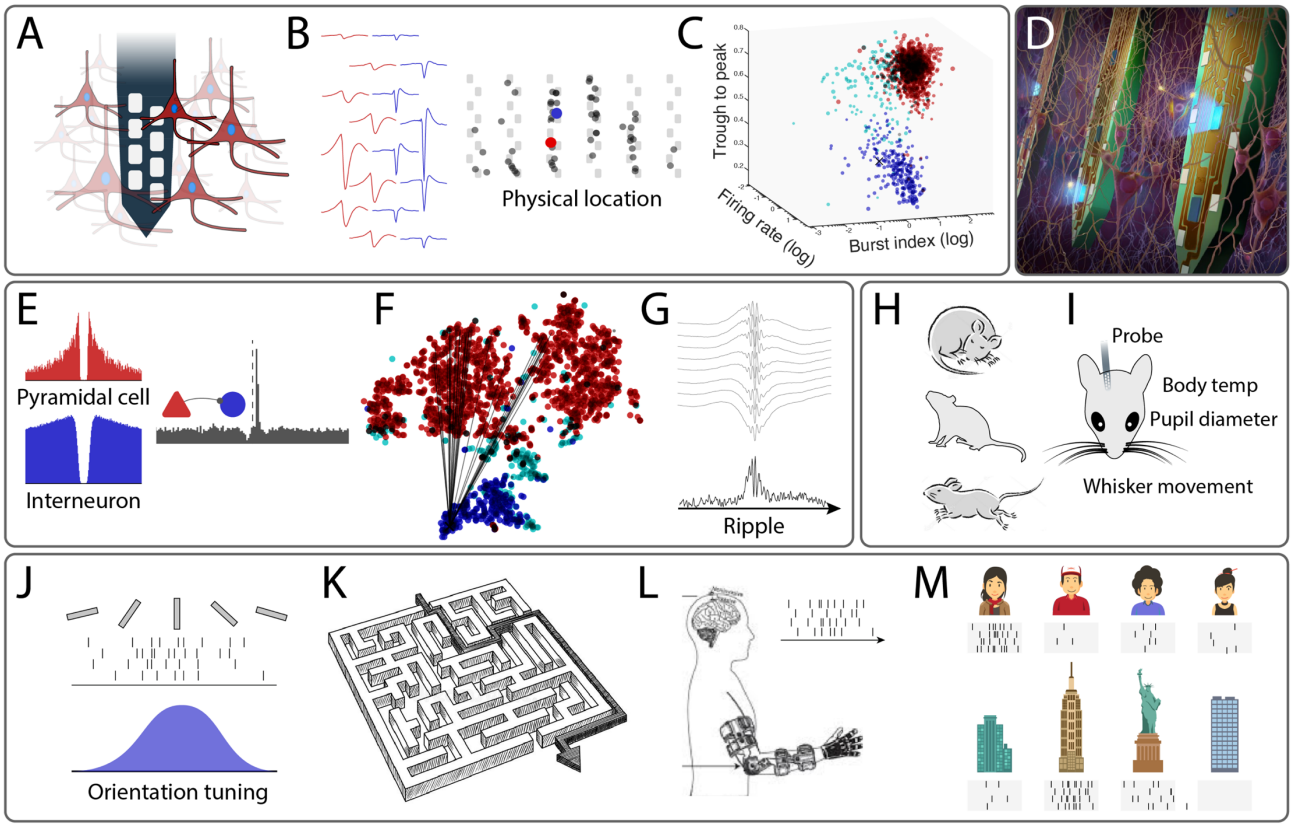


Figure 1: Multifaceted single neuron characterization A. Using high-density silicon probes or multiple tetrodes (shown is a single shank with 8 recording sites), dozens to hundreds of neurons can be recorded simultaneously**. B**, Spikes of putative single neurons are extracted from the recorded traces and assigned to individual neurons through spike sorting algorithms, and assigned to recording sites reflecting the neurons' depth position in the brain (representation shows neurons projected on a silicon probe with 6 shanks). Features of the red neurons are shown in E-F. **C**. Neuron types are separated by first-order physiological parameters. **D**. Optogenetic and other direct methods can ground units to neuron types. **E**. Single neurons are further characterized by their monosynaptic connections to other neurons. **F**. Connection vector of converging pyramidal cells to a single interneuron in a low dimensional representation (t-SNE) from physiological and functional features. **G**. Relating spikes to LFP patterns. **H-I**. Spike pattern correlations with overt behaviors. **J, K**. Spike correlations with experimenter-presented stimuli or situations. **L**, Spike-control of body extensions and robots. **M**. Spike correlates of cognition. **A** to **D**, first-level descriptors. **E-G**, second-level descriptors. **H, I**. third-level descriptors. **J-M**. Spike patterns related to inferred variables.

Whether testing a specific hypothesis or data mining of ever-growing data sets for discovery, the process can be advanced by fast and user-friendly visualization methods that facilitate efficient

hypothesis testing. Toward this goal, we developed an open source framework, *CellExplorer*, to characterize and classify single cell features from multi-site extracellular recordings. It consists of a pipeline for extracting and calculating physiological features and a powerful graphical interface that allows fast manual curation and feature exploration.

RESULTS

The CellExplorer architecture and operation consist of three main parts: a processing module for feature extraction, a graphical interface for manual curation and exploration, and a standardized data structure (Fig 2). A step-by-step tutorial is available in the supplementary section, and more tutorials available online (Suppl Movie 1). The first step in running the pipeline is defining the data input.

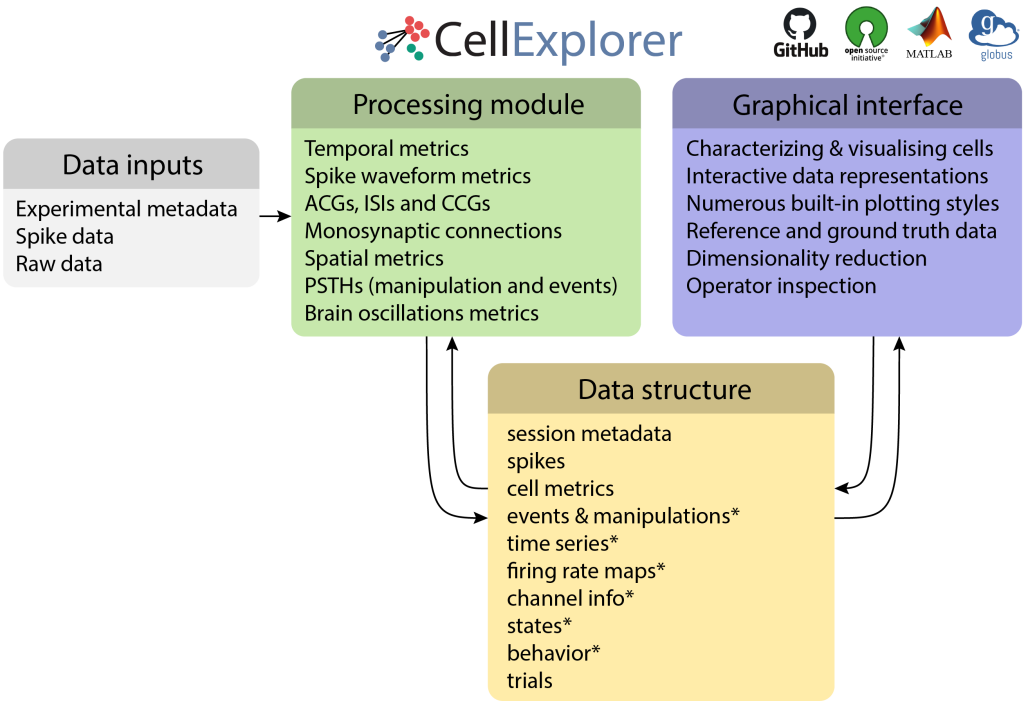


Figure 2. Three-component framework Single yet extensive processing module (green); Standardized yet flexible data structure (yellow); and Graphical interface (purple). Data inputs are compatible with most existing spike sorting algorithms (grey). The data structure joins the Processing module with the Graphical interface (* signifies data containers). CellExplorer is available on GitHub. The software is open source and built in MATLAB. Reference data shared via Globus and a webshare.

Data Input

Before running the pipeline, relevant metadata should be defined describing the spike format, raw data, and experimental metadata (Fig. 2). All experimental metadata are handled in a single MATLAB structure, as part of the data structure, with a GUI for manual entry. The platform supports several spike sorting data formats, including Neurosuite, Phy, KiloSort, SpyKING Circus, MountainSort, IronClust and Wave_Clus (Chung et al., 2017; Hazan et al., 2006; Pachitariu et al., 2016; Quiroga et al., 2004; Schmitzer-Torbert et al., 2005; Yger et al., 2018). The raw data (wide-band) is critical for comparing derived metrics across laboratories, since preprocessed data vary across laboratories and depend on equipment type and filter settings. Yet, for many applications, restricted to examination of

unit–behavior relationship and neuronal spike-interaction analysis, entering metadata and spike time stamps are sufficient.

Processing Module

From the input data, the processing module will generate a large battery of spike features corresponding to a 3-level description of neuronal firing and their relationship to experiment-specific behaviors (Figure 2; Suppl. table 1). The processing module is comprised in a single MATLAB script *ProcessCellMetrics.m,* that computes metrics using a modular structure. The first level description provides temporal features, waveform features (filtered and wideband), interspike interval statistics (ISIs) and autocorrelograms (ACGs). Next, the unit parameters are used for initial classification of single cells into broad classes, default into putative pyramidal cells, narrow waveform interneurons and unclassified cells. In experiments with silicon probes, the physical position relative to recording sites are also determined using trilateration (Petersen and Berg, 2016; Csicsvari et al., 2003).

The second level of description relates single neuron spikes to the activity of other neurons and population patterns. These metrics include spike cross-correlograms (CCGs), quantitative identification of putative monosynaptic connections, phase relationships to various oscillations and irregular local field potentials (LFP) and to unit population patterns. Monosynaptic connections, in turn, can be used to physiologically identify putative excitatory and inhibitory interneurons and use this information to refine the primary unit classification (Fig. 1E; Suppl Fig. 4H) (Barthó et al., 2004; English et al., 2017). All parameters can be customized according to the needs of each experimental paradigm (Suppl table 1; petersenpeter.github.io/CellExplorer/datastructure/standard-cell-metrics/).

The third level descriptors are used to assess the relationship between firing patterns of neurons and overt behaviors, including immobility, locomotion and running speed. Level 1-3 descriptors can be then further refined by optogenetic methods, which can relate physiological parameters to genetically identified neuron groups and justify or modify the primary cell type classification (Boyden et al., 2005; Buzsáki et al., 2015; Roux and Buzsáki, 2015). When available, antidromic and unit-LFP coupling techniques provide additional information about single neurons, such as their position in cortical regions, layers and their target projections (Ciocchi et al., 2015; Senzai et al., 2019; Zhang et al., 2013). Because these 3-level descriptors of single unit features are universal, they can be readily compared with similar analyses across laboratories, independent of paradigm-specific features. The Processing Module automatically generates all cell metrics in a standardized fashion.

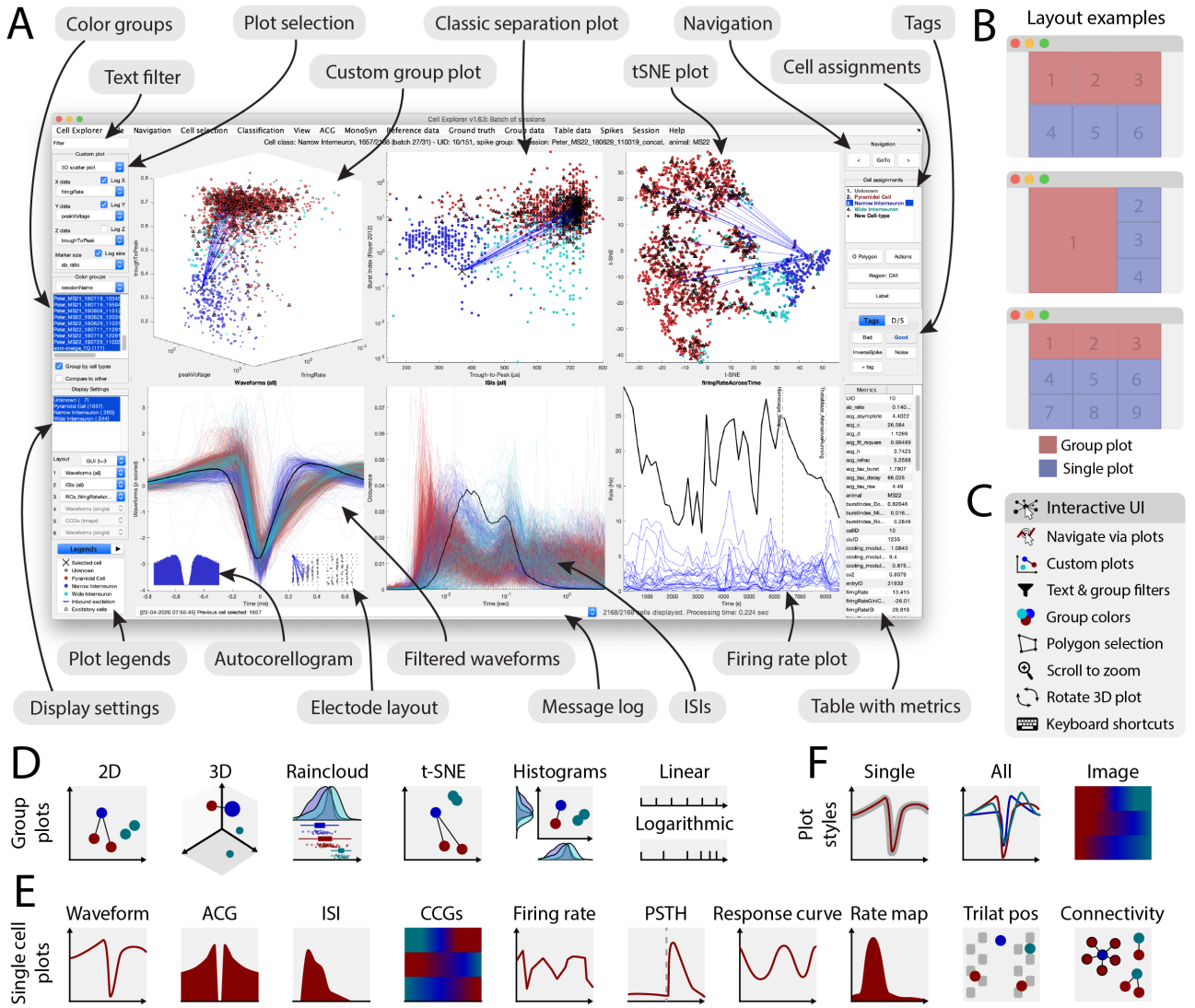


Figure 3: Graphical interface. **A.** The interface consists of 4 to 9 main plots, where the top row is dedicated to population-level representations of the neurons. Other plots are selectable and customizable for individual cells (e.g., single waveforms, ACGs, ISIs, CCGs, PSTHs, response curves, and firing rate maps). The surrounding interface consists of panels placed on either side of the graphs. The left side displays settings and population settings, including a custom plot panel, color group panel, display settings panel, and legends. The right side-panel displays single-cell dimensions, including a navigation panel, cell assignment panel, tags, and a table with metrics. In addition, there are text fields for a custom text filter and a message log. **B.** Layout examples highlighting three configurations with 1-3 group plots and 3-6 single cell plots. **C.** The interface has many interactive elements, including navigation and selection from plots (left mouse click links to selected cell and right mouse click selects the cell from all the plots), visualization of monosynaptic connections, various data plotting styles (more than 30+ unique plots built-in), supports custom plots; plotting filters can be applied by text or selection, keyboard shortcuts, zooming any plot by mouse-scrolling and polygon selection of cells. **D.** Group plotting options: 2D, 3D, raincloud plot, t-SNE, and double histogram. Each dimension can be plotted on linear or logarithmic axes. **E.** Single-cell plot options: waveform, ACG, ISI, firing rate across time, PSTH, response curve, firing rate maps, neuron position triangulation relative to recording sites and monosynaptic connectivity graph. **F.** Most single-cell plots have three representations: individual single cell representation, single cell together with the entire population with absolute amplitude and a normalized image representation (colormap).

Data structure

The data structure (summarized in Fig. 2, and supplementary Fig. 1), is structured into data categories (containers) and MATLAB structures, which functionally separate related data, making them easily interpretable (human-readable) but also makes them machine-readable. The format is derived from

buzcode (introduces the containers and a subset of the structs, github.com/buzsakilab/buzcode), Neurosuite (neurosuite.sourceforge.net), and the Freely Moving Animal (FMA) Toolbox (fmatoolbox.sourceforge.net).

The two most important structures are the **session** metadata struct and **cell_metrics** struct.
session metadata struct: Contains all session-level experimental metadata. It is a modular structure which makes it flexible and expandable, intuitive and interpretable, and it offers a single structure preventing scattering of metadata. A GUI allows for intuitive manual metadata entry, and a template script can assist in importing experimental metadata.
cell_metrics struct: Modular structure containing all metrics calculated in the processing module. It consists of three types of data-fields for handling the diverse types of data: *numeric double*, *character-cells* and *structs*. Single value metrics are stored in double and character cells for respective numeric and character metrics. Time series (e.g., waveforms), group data (e.g. synaptic connections) and session parameters are stored in predefined struct modules. This structure makes the fields machine readable, including user-defined metrics, providing expandability and flexibility, yet maintaining compatibility with the graphical interface. The single struct allows for processing multiple sessions together in the graphical interface (batch processing) and is convenient for sharing with collaborators and broader scientific community in publications (see Supplementary Section and Supplementary table 1 for a detailed description.

Graphical Interface

The most important component of the framework is the user-friendly Graphical Interface (Fig. 3), which allows for characterization and exploration of all single unit metrics through a rich set of high-quality built-in plots, interaction modes, cell grouping, cross-level pointers, and filters. User-defined numbers of plots can be selected any time and replaced on the screen instantaneously. In the typical layout, the top row displays population-level representations and the bottom row single cell features. Any neuron or multiple neurons can be clicked upon and all other features of the selected neurons are automatically updated in the other plots. For easy navigation and selection, the left mouse click links to selected cell(s) and right mouse click selects the cell(s) from all the plots. These selected groups can be displayed alone or highlighted and superimposed against all data in the same session, multiple sessions or the entire data base. Clusters of neurons of interest can be selected by drawing polygons with the mouse cursor, and the features of the selected groups will be updated in all other display windows. Multiple group selections are also possible for both visualization and statistical comparison.

Flexibility is assisted by self-explanatory side panels, including a custom plot panel, color group panel, display settings panel, and legend (left side) and single-cell dimensions, navigation panel, cell assignment panel, tags, and a table with metrics (right side). A text field is also available for custom text filtering and a message logging. Group plotting options include 2D, 3D, raincloud plot, t-SNE, and double histogram, each plotted either on linear or logarithmic scales.

Examples of the flexible operation of the graphical interface module are illustrated in Fig. 4 and described in more detail in Suppl Movie 1. Here we begin with motives of monosynaptically connected clusters of neurons from the hippocampal CA1 area, as provided by the Processing Module (Fig. 4A). An example sub-network of connected neurons is highlighted in panel B with a selected single neuron (arrow) to be characterized. In turn, selected level 1, 2 and 3 descriptors of the neuron are displayed in panels C to G, respectively. In several panels, the descriptors of the selected neuron are shown against all other neurons. Left mouse clicking any neuron will update all the panels, allowing quick screening and qualitative evaluation of multiple features of each inspected neuron. Neurons of interest can be marked for further quantitative comparisons. Next, level 1-3 descriptions can be compared with paradigm-specific features of the selected neuron(s), such as placed field, trial-by-trial variability of firing patterns, travel direction firing specificity, spike phase precession relative to theta oscillation cycles, and multiple other features predefined by the experimenter. During the data mining process, unexpected features and outliers may be noted, instabilities of neurons ('drifts') can be recognized and artifacts identified. Such experimenter-supervised judgments are also essential for evaluating the quality of quantified data processing and estimating potential single neuron-unique features that might drive population statistics.

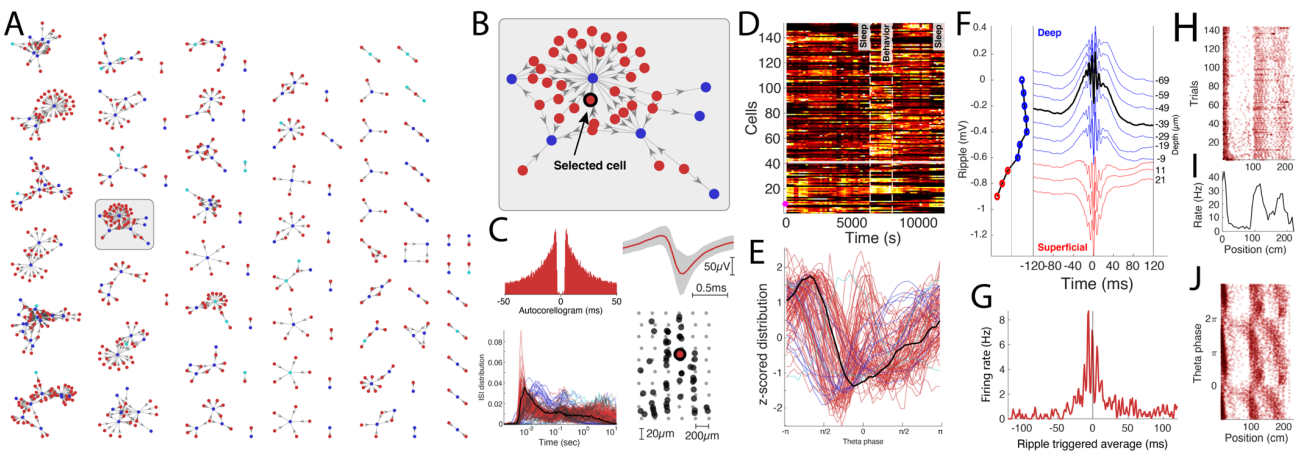


Figure 4. Data exploration example. A. Connectivity graph with monosynaptic modules found across multiple datasets. Cells are color coded by their putative cell types (pyramidal cells in red, narrow interneurons in blue and wide interneurons in cyan). **B.** Highlighted monosynaptic module with single pyramidal cell highlighted (arrow). **C.** First level descriptors: Auto-correlogram, average waveform (top row), the ISI distribution across population and the physical location of the cells

relative to the multi-shank silicon probe. **D.** Firing rate across time for the population, each cell is normalized to its peak rate. The session consists of three behavioral epochs: pre-behavior sleep, behavior (track running) and post-behavior sleep (boundaries shown with dashed lines). **E.** Theta phase distribution for all neurons recorded in the same session (red, pyramidal cells; blue, interneurons) during locomotion with the selected cell highlighted (black line). **F.** Average ripple across multiple sites of a shank. The site of the selected neuron is highlighted (dashed black line). The polarity of the average sharp wave is used to determine the position of the cell relative to the pyramidal layer in CA1. **G.** Ripple wave-triggered PSTH for the selected cell aligned to the ripple peak. **H.** Trial-wise raster for the selected cell across in a maze. **I.** The average firing rate across trials. **J.** Spike raster showing the theta phase relationship to the spatial location of the animal. Each of the three place fields shows phase precession.

Value of large inter-laboratory data sets

While progress in discovery science often depends on investigator-unique approach to novel insights, standardization of data processing and screening is essential in fields where 'big data' generation is not realistic in small and middle-size laboratories. This applies to the current effort to quantitatively relate physiology-based and genetically classified cell 'types' (Klausberger and Somogyi, 2008; McBain and Fisahn, 2001; Rudy et al., 2011). In each experiment, typically only one or limited number of neuron types can be identified. Yet, combining data sets from numerous experiments and different laboratories can generate physiological descriptors grounded by optogenetics and other 'ground truth' data.

Fig. 5 illustrates the feasibility and utility of this approach. Level 1-3 descriptors of neurons recorded from the same brain region and layer can be combined from multiple experiments and contrasted to data quality of units recorded in a single session. An ever-growing data set allows for more reliable modality separation and characterization of neuron types. For example, the initial divisions of neurons into putative pyramidal cells, interneurons and unclassified cells can be further refined by quantifying monosynaptic connections, increasing confidence of pyramidal cell–interneuron separation as well as identifying subsets of the unclassified group as interneurons (Mizuseki et al., 2011; Petersen and Buzsáki, 2020; Peyrache et al., 2015; Stark et al., 2013). Combining extracellular spiking with intracellular recordings can further help determine cell the excitatory or inhibitory identity of neurons (Radosevic et al., 2019). Single neurons identified by opto-tagging or other direct means (Ciocchi et al., 2015; Klausberger and Somogyi, 2008; Royer et al., 2012; Senzai et al., 2019; Stark et al., 2012; Zhang et al., 2013; Roux and Buzsáki, 2015). can be used to examine and contrast level 1-3 features of initially classified neurons into further types. In turn, such 'ground truth' data may offer further clues for including unrecognized brain region and layer-distinct physiological features of single neurons for more effective classification (Senzai and Buzsáki, 2017). An expected outcome is that growing data sets containing ground truth-verified neurons will allow that in future experiments multiple neuron types recording in the same animal can be reliably identified by physiological descriptors only.

Data sets obtained from different brain areas, different electrode types and drug conditions can be compared. t-Distributed Stochastic Neighbor Embedding (t-SNE) plots can highlight inconsistences and differences across recording sessions, identify important regional and layer-specific differences and alert for interspecies characteristics (Fig. 5).

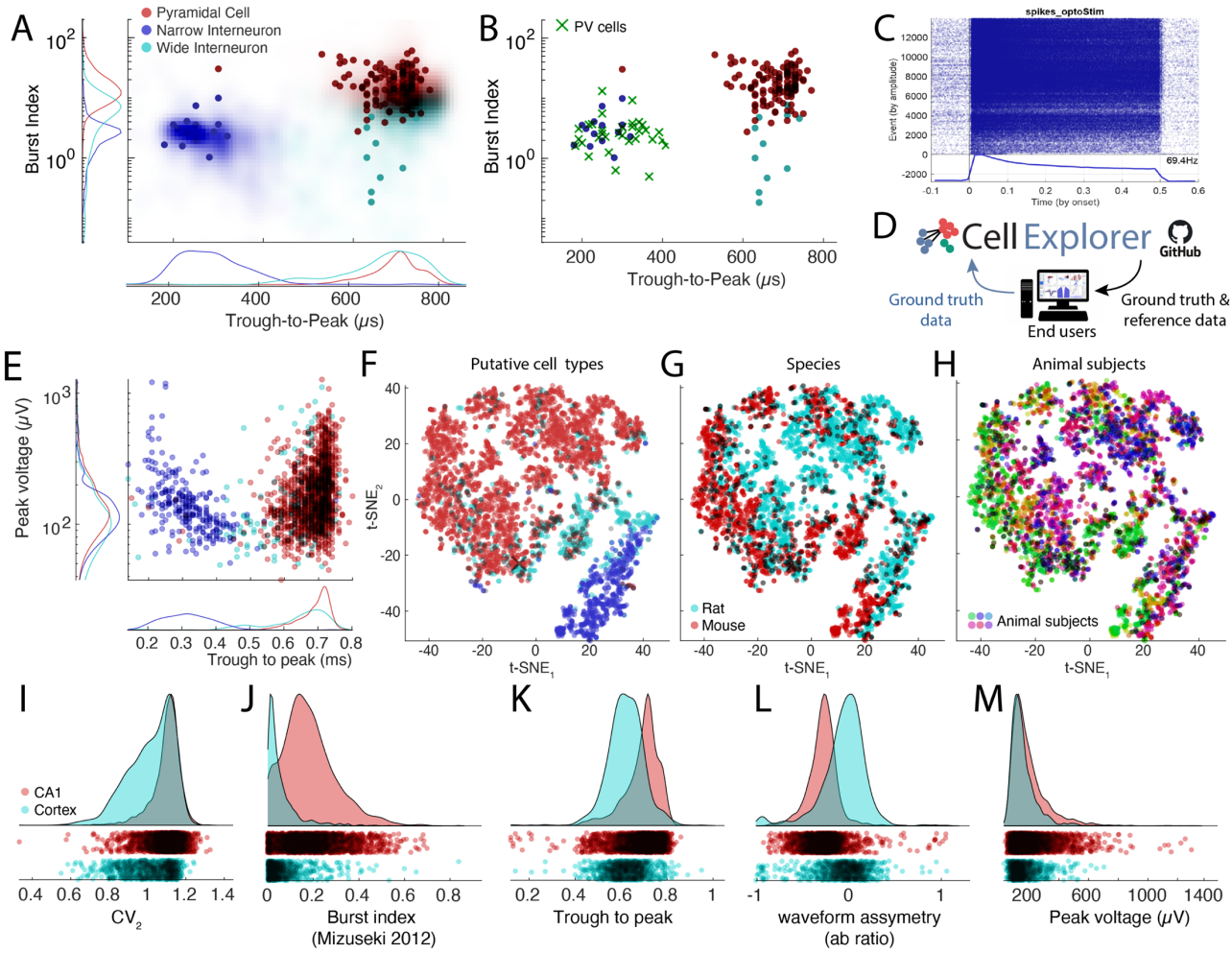


Figure 5. Reference data, ground truth data and comparison of basic metrics. A. Single session (dots) compared with reference data from 30 reference sessions (shaded zones). **B.** Ground truth cells (optogenetically identified parvalbumin [PV]-cells marked with green crosses) projected onto the same sessions as in panel A. **C.** Example of PSTH of a PV-expressing cell. Raster plot and average responses to light pulses visualized in CellExplorer. **D.** The CellExplorer framework allows for sharing ground truth and reference data directly with the end user. End users can upload their ground truth data to the CellExplorer GitHub repository for communal sharing. **E.** Distributions of spike amplitudes and waveform width (quantified by the trough to peak metrics) for the three groups from multiple CA1 recording sessions. Note inverse relationship between spike amplitude and waveform for putative interneurons. **F-H.** t-SNE representations of putative cell types (F), species (G, rat and mouse in magenta and red, respectively) and subjects (H, colors scaled across subjects) for hippocampal neurons. **I-M:** Comparison of spike features of neurons recorded from CA1 pyramidal cells and visual cortex pyramidal cells. Significant differences are observed across several basic metrics, including CV_2 (I), burst index (J), trough-to-peak (K), waveform asymmetry (L) and waveform peak voltage (M).

Benchmarking

Several tests were performed to characterize the speed performance of CellExplorer under various conditions, testing display features and computation of various features. The results are summarized in Fig. 6.

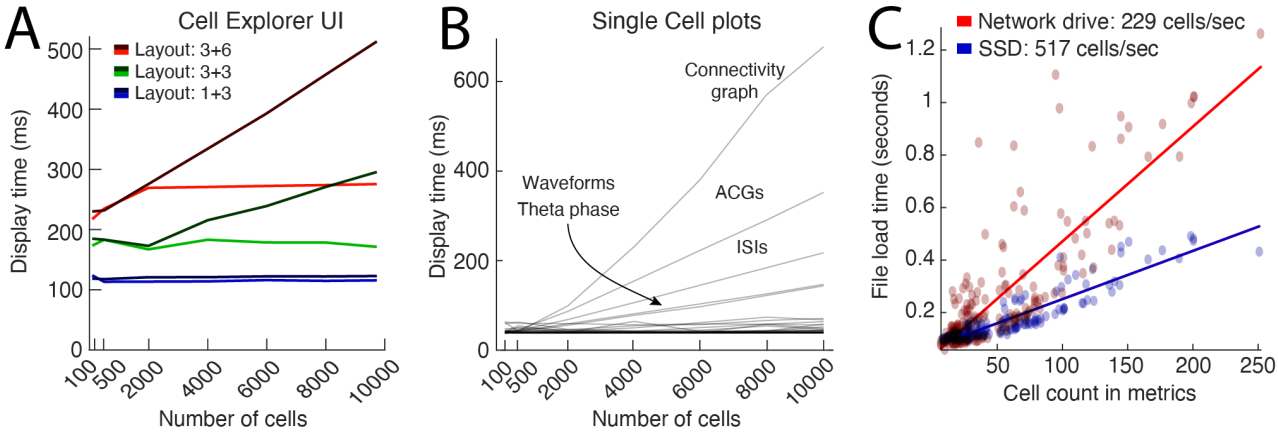


Figure 6. Benchmarks of the CellExplorer user interface (UI). A. UI display times when switching between units for the three layouts shown in figure 3B (approximately 110 ms for layout 1+3; blue lines. 180 ms for layout 3+3; green lines) and 290 ms (layout 3+6; in red), respectively. Dark gradient colored lines (dark red, green and blue) indicate where there were no limits on the number of traces plotted for single cell plots, and the light gradient lines show screen update times with a maximum of 2000 random traces. **B.** Display times for single cell plots, quantified by the number of cells displayed. The plots contributing most to an increased display time are the plots with trace representations for each cell (ACGs, ISIs, waveforms, ISIs, theta phase) and the connectivity graph. By default, a maximum of 2000 traces are drawn capping the processing time below ~80 ms for all plots except the connectivity graph. **C.** Benchmarking of cell metrics file readings. On average, 230 cells can be loaded per second quantified across 180 sessions with various cell count (red dots and linear fit in red). By storing the data on a local SSD, the loading time could be decrease and attain cell loading above 500 cells per second.

Open source database

The CellExplorer takes advantage of web-based resources (Chon et al., 2019; Petersen et al., 2018) https://atlas.brain-map.org/, https://buzsakilab.com/wp/public-data/) for discovering, viewing and comparing physiological features of single neurons. Data sharing allows building large data banks for discovery science, cross-laboratory interactions and reproducibility control.

DISCUSSION

We have developed CellExplorer, a transparent, open source, MATLAB-based resource for characterizing single neurons and neuron types, using physiological features. The CellExplorer platform enables visualization and analysis for users without the need to write code. Its modular format allows for fast and flexible comparisons of a large battery of preprocessed physiological characteristics of single neurons and their interactions with other neurons as well as their correlation with experimental variables. Code is publicly available on GitHub for users to download and use the same standardized processing module on their local personal computer (Windows, OS X and Linux).

CellExplorer offers step-by-step online tutorials for first-time users. It is linked to the Allen Institute reference atlas (Chon et al., 2019), https://atlas.brain-map.org/) and can be expanded to include other online resources that provide annotated data on putative neuron types. The open source design of the CellExplorer permits the optimization of current standardized components tested against an ever-growing community-contributed data base and will accelerate linking genetically identified neuron types with their physiological properties in the intact brain.

Multiple-level characterization and classification of single neurons

To correctly interpret neuron firing-behavior/cognition relationships, numerous controls are needed to rule out or reduce the potential contribution of spurious variables. The Processing Module generates a battery of useful metrics. In addition to first-level description of the biophysical and physiological characteristics of single neurons, it computes brain state-dependent firing rates, insterspike interval variation, and relationships between single neurons and spiking activity of the population and LFP (second level). When spontaneous (control) behaviors are also available, it describes the relationship between single neuron firing patterns and routine behavioral parameters, such as immobility, walking, running speed, respiration and pupil diameter (third level). These third level descriptors may assist in appropriately attributing spiking activity to inferred behavior, such as perception or cognition. Because these 3-level descriptors are independent of particular experimental paradigms, they can be used as benchmarks for assessing consistencies across experiments performed by different investigators in the same laboratory or across laboratories. Concatenating numerous data sets obtained from the same brain regions and layers will create a continuously growing data bank. In turn, these data-rich sets may allow identifying and quantifying reliable boundaries among putative clusters and suggest inclusion and exclusion of parameters for more refined separation of putative neuronal classes. Sets from different brain regions can be readily compared and differences recognized.

Although several statistical tests are available in the CellExplorer, it is not meant to be a substitute for rigorous quantification. Instead, it is designed as a tool for facilitating interpretation and discovery. It is a complementary approach to dimensionality reduction and population analysis methods. Because assemblies of neurons consist of highly unequal partners (Buzsáki and Mizuseki, 2014), knowledge about neuron-specific contribution to population measures is critical in many situations (Nicolelis and Lebedev, 2009). Such inequality may stem from unknowingly lumping neurons of different classes together into a single type and because even members of the same type belong to broad and skewed distribution and may contribute to different aspects of the experiment (Grosmark and Buzsáki, 2016).

Two key features make the CellExplorer platform highly efficient: flexibility and speed. Flexibility is provided by the numerous parameters as outputs of the Processing Module. High speed is achieved by using precomputed metrics and limiting the computation time in the user interface. A caveat is that online alteration of the plots and metrics is not allowed. Yet, such changes can be performed by working with the raw spike data. Multiple features of single neurons, displayed on the same screen, can be compared. Moving from one neuron to the next requires only mouse click. These features can be compared separately or superimposed against another neuron, all neurons in a session or the entire data base. Unexpected features can be noted, obvious artifacts can be deleted. When unusual sets are discovered in any display, all other features of the same set can be rapidly compared and contrasted to other sets. Neuron clouds can be selected by drawing polygons around them and regrouped in any arbitrary configuration. Inspection of data sets containing even several thousand neurons (Steinmetz et al., 2019) is realistic because minimal computing time is required in the graphical interface and because in most conditions only small subsets need individual inspection and quality control.

Various classification schemes have been developed to assign extracellular spikes to putative pyramidal cells, interneurons and their putative subtypes, on the basis of a variety of physiological criteria. These include waveform features, firing rate statistics in different brain states, embeddedness in various population activities, firing patterns characterized by their autocorrelograms, and putative monosynaptic connections to other neurons (Barthó et al., 2004; Csicsvari et al., 1999; Fujisawa et al., 2008; Mizuseki et al., 2009; Okun et al., 2015; Sirota et al., 2008). Increasingly larger data sets will likely improve such physiology-based classification. Yet, the 'ground truth' for these classifying methods is largely missing. Optogenetic tagging (Boyden et al., 2005) offers such grounding by connecting putative subtypes based on physiologically distinct features to their molecular identities. Because in a single animal only one or few neuron types can be tagged optogenetically or by other direct methods (Fosque et al., 2015; Klausberger and Somogyi, 2008), refinement of a library of physiological parameters should be conducted iteratively, so that in subsequent experiments the various neuron types can be recognized reliably by using solely physiological criteria (English et al., 2017; Royer et al., 2012; Senzai and Buzsáki, 2017, 2017; Roux and Buzsáki, 2015). In turn, knowledge about the molecular identity of the different neuronal components of a circuit can considerably improve the interpretation of correlational observations provided by large-scale extracellular recordings.

Data sharing

Above, we described one of the many possible examples that can benefit from large databases. Currently, tens to hundreds of thousands of pre-processed neurons exist across laboratories in different brain regions, which can be streamlined by the Processing Module in an identical fashion, and displayed and compared in the same coordinate system. Our laboratory welcomes shared data sets from other research groups for enhancement and comparison with our publicly-accessible database (buzsakilab.com/wp/public-data/). The single prerequisite for quantitative comparison of data across laboratories is to make wideband data available (≤ 3 Hz to ≥ 8 kHz; ideally ≥ 20 kHz sampling rate) so that all data are processed the same way.

Through our web resource (Petersen et al., 2018), we host > 1,000 publicly shared data sets of long (4 to 24 hrs), large-scale recordings of single units from multiple brain structures, including hippocampus, entorhinal, prefrontal, somatosensory and visual cortices, thalamus, amygdala and septum (buzsakilab.com/wp/public-data/). Long-recordings have the advantages of defining brain state-dependent characteristics of neurons, such as their firing rates and patterns during waking and sleep, unmasking the 'hidden' or relatively silent majority of neurons (Mizuseki and Buzsáki, 2013; Shoham et al., 2006) and discovering their connectivity patterns (English et al., 2017). These data already provide benchmarks assessing the reliability of initial neuron classification into broad pyramidal cell and interneuron groups, many of which are identified physiologically by their monosynaptic connections. They also offer normative data about spikes features, firing rates and spike dynamics. These features can serve as benchmarks for comparison with data collected in any other laboratory.

Development and availability

Development takes place in a public code repository at github.com/petersenpeter/CellExplorer. All examples in this article have been calculated with the pipeline and plotted with the CellExplorer. Extensive documentation, including installation instructions and tutorials, is hosted at petersenpeter.github.io/CellExplorer/. The CellExplorer is available for MATLAB 2017B and forward, and for the operating systems Windows, OS X and Linux. More information can be found at petersenpeter.github.io/CellExplorer/.

ACKNOWLEDGEMENTS

We thank Sam McKenzie, Mihály Vöröslakos, Michelle Hernandez, Thomas Hainmueller, for helping with documentation or implementation of the CellExplorer code and Sam McKenzie, Daniel English,

Sebastian Royer for providing ground truth datasets. Further, thanks go to Thomas Hainmueller, Antonio Hernandez-Ruiz, Viktor Varga and Omid Yaghmazadeh for comments on the manuscript.

REFERENCES

Barlow, H.B. (1972). Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology? Perception *1*, 371–394.

Barthó, P., Hirase, H., Monconduit, L., Zugaro, M., Harris, K.D., and Buzsáki, G. (2004). Characterization of neocortical principal cells and interneurons by network interactions and extracellular features. J. Neurophysiol. *92*, 600–608.

Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. Nature Neuroscience *8*, 1263–1268.

Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. Nature Neuroscience *7*, 446–451.

Buzsáki, G., and Mizuseki, K. (2014). The log-dynamic brain: how skewed distributions affect network operations. Nat Rev Neurosci *15*, 264–278.

Buzsáki, G., Stark, E., Berényi, A., Khodagholy, D., Kipke, D.R., Yoon, E., and Wise, K. (2015). Tools for probing local circuits: high-density silicon probes combined with optogenetics. Neuron *86*, 92–105.

Chon, U., Vanselow, D.J., Cheng, K.C., and Kim, Y. (2019). Enhanced and unified anatomical labeling for a common mouse brain atlas. Nature Communications *10*, 1–12.

Chung, J.E., Magland, J.F., Barnett, A.H., Tolosa, V.M., Tooker, A.C., Lee, K.Y., Shah, K.G., Felix, S.H., Frank, L.M., and Greengard, L.F. (2017). A Fully Automated Approach to Spike Sorting. Neuron *95*, 1381-1394.e6.

Ciocchi, S., Passecker, J., Malagon-Vina, H., Mikus, N., and Klausberger, T. (2015). Selective information routing by ventral hippocampal CA1 projection neurons. Science *348*, 560–563.

Csicsvari, tJozsef, Hirase, H., Czurkó, A., Mamiya, A., and Buzsáki, G. (1999). Fast Network Oscillations in the Hippocampal CA1 Region of the Behaving Rat. J. Neurosci. *19*, RC20–RC20.

Csicsvari, J., Henze, D.A., Jamieson, B., Harris, K.D., Sirota, A., Barthó, P., Wise, K.D., and Buzsáki, G. (2003). Massively Parallel Recording of Unit and Local Field Potentials With Silicon-Based Electrodes. Journal of Neurophysiology *90*, 1314–1323.

English, D.F., McKenzie, S., Evans, T., Kim, K., Yoon, E., and Buzsáki, G. (2017). Pyramidal Cell-Interneuron Circuit Architecture and Dynamics in Hippocampal Networks. Neuron *96*, 505-520.e7.

Fosque, B.F., Sun, Y., Dana, H., Yang, C.-T., Ohyama, T., Tadross, M.R., Patel, R., Zlatic, M., Kim, D.S., Ahrens, M.B., et al. (2015). Labeling of active neural circuits in vivo with designed calcium integrators. Science *347*, 755–760.

Fujisawa, S., Amarasingham, A., Harrison, M.T., and Buzsáki, G. (2008). Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. Nature Neuroscience *11*, 823–833.

Grosmark, A.D., and Buzsáki, G. (2016). Diversity in neural firing dynamics supports both rigid and learned hippocampal sequences. Science *351*, 1440–1443.

Hazan, L., Zugaro, M., and Buzsáki, G. (2006). Klusters, NeuroScope, NDManager: A free software suite for neurophysiological data processing and visualization. Journal of Neuroscience Methods *155*, 207–216.

Kepecs, A., and Fishell, G. (2014). Interneuron cell types are fit to function. Nature *505*, 318–326.

Klausberger, T., and Somogyi, P. (2008). Neuronal Diversity and Temporal Dynamics: The Unity of Hippocampal Circuit Operations. Science *321*, 53–57.

McBain, C.J., and Fisahn, A. (2001). Interneurons unbound. Nat. Rev. Neurosci. *2*, 11–23.

McGinley, M.J., David, S.V., and McCormick, D.A. (2015). Cortical Membrane Potential Signature of Optimal States for Sensory Signal Detection. Neuron *87*, 179–192.

Mizuseki, K., and Buzsáki, G. (2013). Preconfigured, Skewed Distribution of Firing Rates in the Hippocampus and Entorhinal Cortex. Cell Reports *4*, 1010–1021.

Mizuseki, K., Sirota, A., Pastalkova, E., and Buzsáki, G. (2009). Theta Oscillations Provide Temporal Windows for Local Circuit Computation in the Entorhinal-Hippocampal Loop. Neuron *64*, 267–280.

Mizuseki, K., Diba, K., Pastalkova, E., and Buzsáki, G. (2011). Hippocampal CA1 pyramidal cells form functionally distinct sublayers. Nature Neuroscience *14*, 1174–1181.

Nicolelis, M.A.L., and Lebedev, M.A. (2009). Principles of neural ensemble physiology underlying the operation of brain–machine interfaces. Nature Reviews Neuroscience *10*, 530–540.

Okun, M., Steinmetz, N.A., Cossell, L., Iacaruso, M.F., Ko, H., Barthó, P., Moore, T., Hofer, S.B., Mrsic-Flogel, T.D., Carandini, M., et al. (2015). Diverse coupling of neurons to populations in sensory cortex. Nature *521*, 511–515.

Pachitariu, M., Steinmetz, N.A., Kadir, S.N., Carandini, M., and Harris, K.D. (2016). Fast and accurate spike sorting of high-channel count probes with KiloSort. In Advances in Neural Information Processing Systems 29, D.D. Lee, M. Sugiyama, U.V. Luxburg, I. Guyon, and R. Garnett, eds. (Curran Associates, Inc.), pp. 4448–4456.

Petersen, P.C., and Berg, R.W. (2016). Lognormal firing rate distribution reveals prominent fluctuation–driven regime in spinal motor networks. ELife *5*, e18805.

Petersen, P.C., and Buzsáki, G. (2020). Cooling of medial septum reveals theta phase lag coordination of hippocampal cell assemblies. BioRxiv 2019.12.19.883421.

Petersen, P.C., Hernandez, M., and Buzsáki, G. (2018). Public electrophysiological datasets collected in the Buzsaki Lab. https://buzsakilab.com/wp/public-data/ http://doi.org/10.5281/zenodo.3629881 (Zenodo).

Peyrache, A., Lacroix, M.M., Petersen, P.C., and Buzsáki, G. (2015). Internally organized mechanisms of the head direction sense. Nat Neurosci *18*, 569–575.

Quiroga, R.Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised Spike Detection and Sorting with Wavelets and Superparamagnetic Clustering. Neural Computation *16*, 1661–1687.

Radosevic, M., Willumsen, A., Petersen, P.C., Lindén, H., Vestergaard, M., and Berg, R.W. (2019). Decoupling of timescales reveals sparse convergent CPG network in the adult spinal cord. Nat Commun *10*, 2937.

Roux, L., and Buzsáki, G. (2015). Tasks for inhibitory interneurons in intact brain circuits. Neuropharmacology *88*, 10–23.

Royer, S., Zemelman, B.V., Losonczy, A., Kim, J., Chance, F., Magee, J.C., and Buzsáki, G. (2012). Control of timing, rate and bursts of hippocampal place cells by dendritic and somatic inhibition. Nature Neuroscience *15*, 769–775.

Rudy, B., Fishell, G., Lee, S., and Hjerling-Leffler, J. (2011). Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. Developmental Neurobiology *71*, 45–61.

Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K., and Redish, A.D. (2005). Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience *131*, 1–11.

Senzai, Y., and Buzsáki, G. (2017). Physiological Properties and Behavioral Correlates of Hippocampal Granule Cells and Mossy Cells. Neuron *93*, 691-704.e5.

Senzai, Y., Fernandez-Ruiz, A., and Buzsáki, G. (2019). Layer-Specific Physiological Features and Interlaminar Interactions in the Primary Visual Cortex of the Mouse. Neuron *101*, 500-513.e5.

Shoham, S., O'Connor, D.H., and Segev, R. (2006). How silent is the brain: is there a "dark matter" problem in neuroscience? J Comp Physiol A *192*, 777–784.

Sirota, A., Montgomery, S., Fujisawa, S., Isomura, Y., Zugaro, M., and Buzsáki, G. (2008). Entrainment of neocortical neurons and gamma oscillations by the hippocampal theta rhythm. Neuron *60*, 683–697.

Stark, E., Koos, T., and Buzsáki, G. (2012). Diode probes for spatiotemporal optical control of multiple neurons in freely moving animals. J. Neurophysiol. *108*, 349–363.

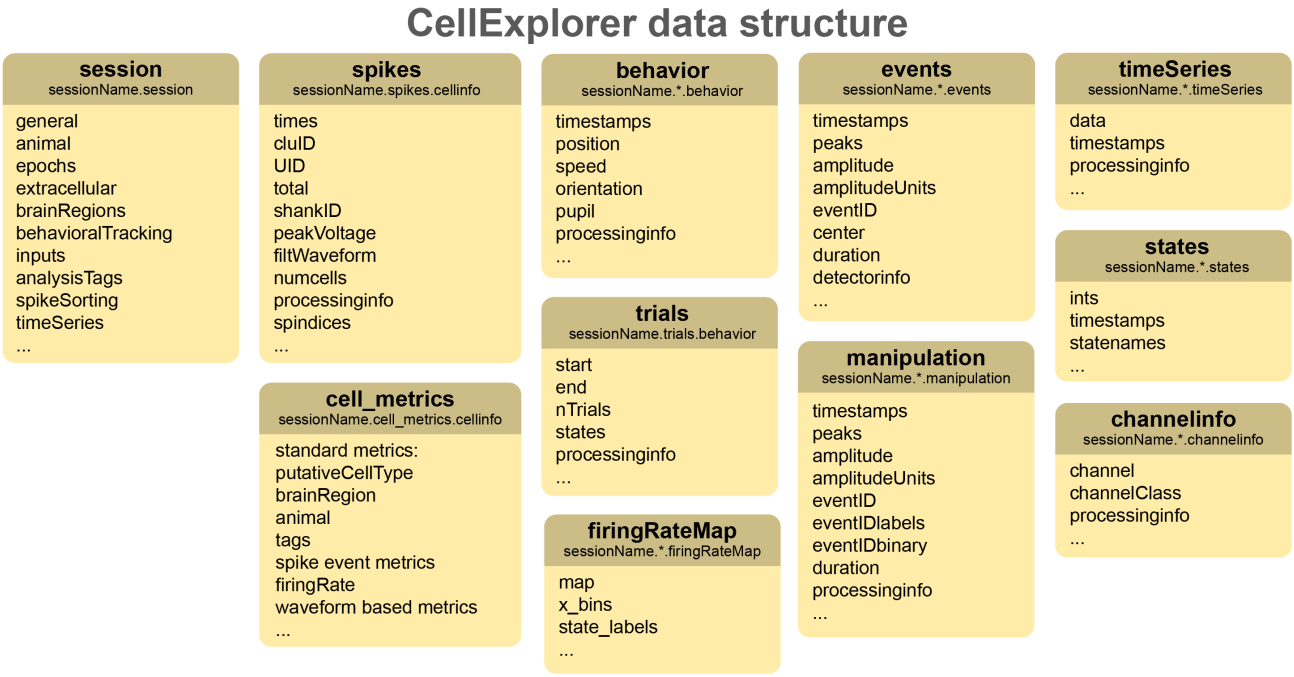
Stark, E., Eichler, R., Roux, L., Fujisawa, S., Rotstein, H.G., and Buzsáki, G. (2013). Inhibition-Induced Theta Resonance in Cortical Circuits. Neuron *80*, 1263–1276.

Steinmetz, N.A., Zatka-Haas, P., Carandini, M., and Harris, K.D. (2019). Distributed coding of choice, action and engagement across the mouse brain. Nature *576*, 266–273.

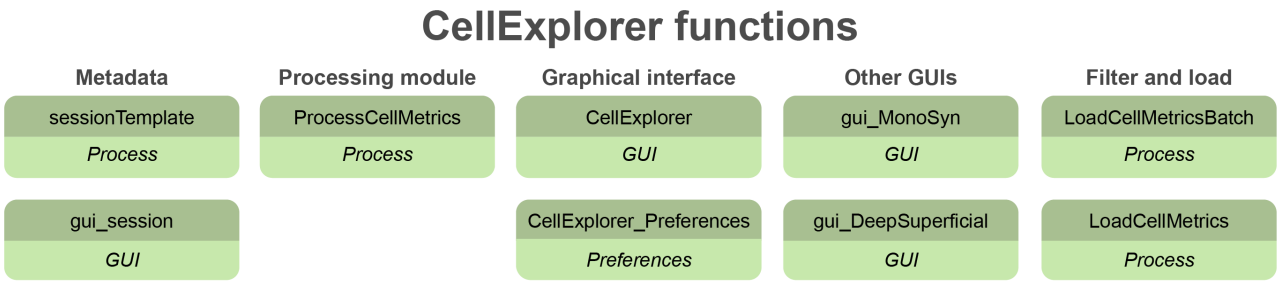
Yger, P., Spampinato, G.L., Esposito, E., Lefebvre, B., Deny, S., Gardella, C., Stimberg, M., Jetter, F., Zeck, G., Picaud, S., et al. (2018). A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo. ELife *7*, e34518.

Zhang, S.-J., Ye, J., Miao, C., Tsao, A., Cerniauskas, I., Ledergerber, D., Moser, M.-B., and Moser, E.I. (2013). Optogenetic Dissection of Entorhinal-Hippocampal Functional Connectivity. Science *340*.

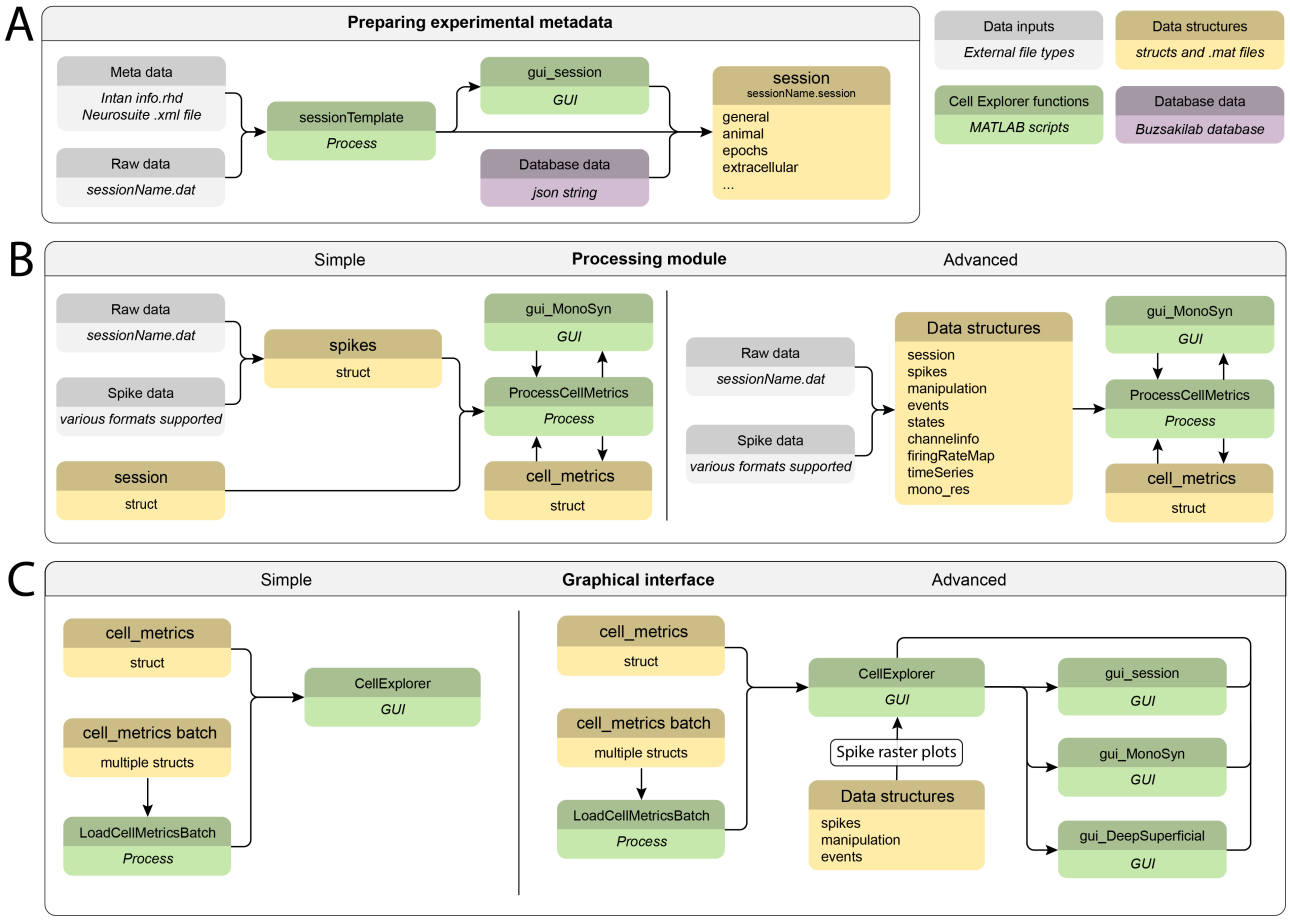
SUPPLEMENTARY SECTION



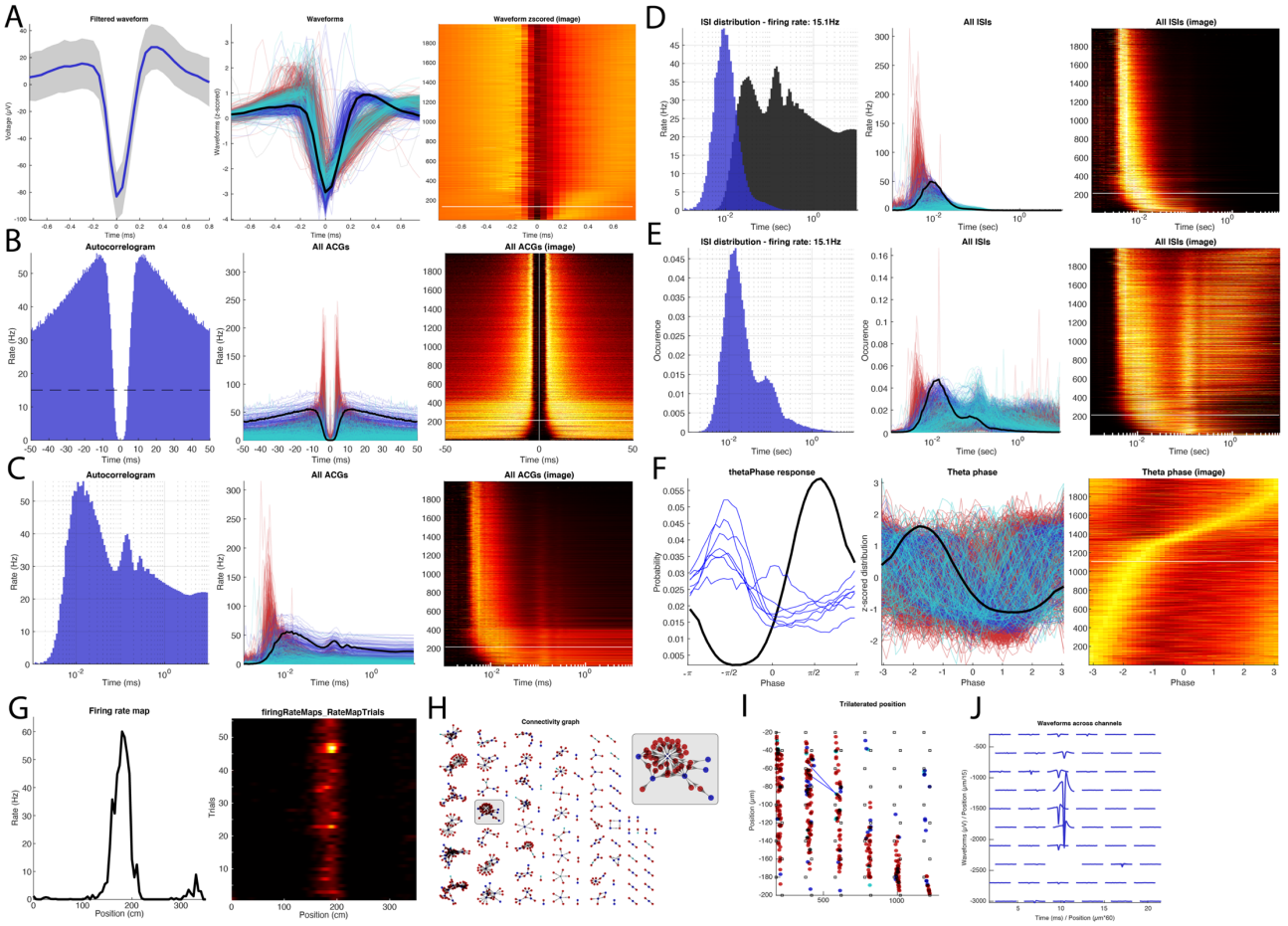
Supplementary figure 1: Data types. The data structure. A detailed description is available online at: petersenpeter.github.io/CellExplorer/datastructure/data-structure-and-format/



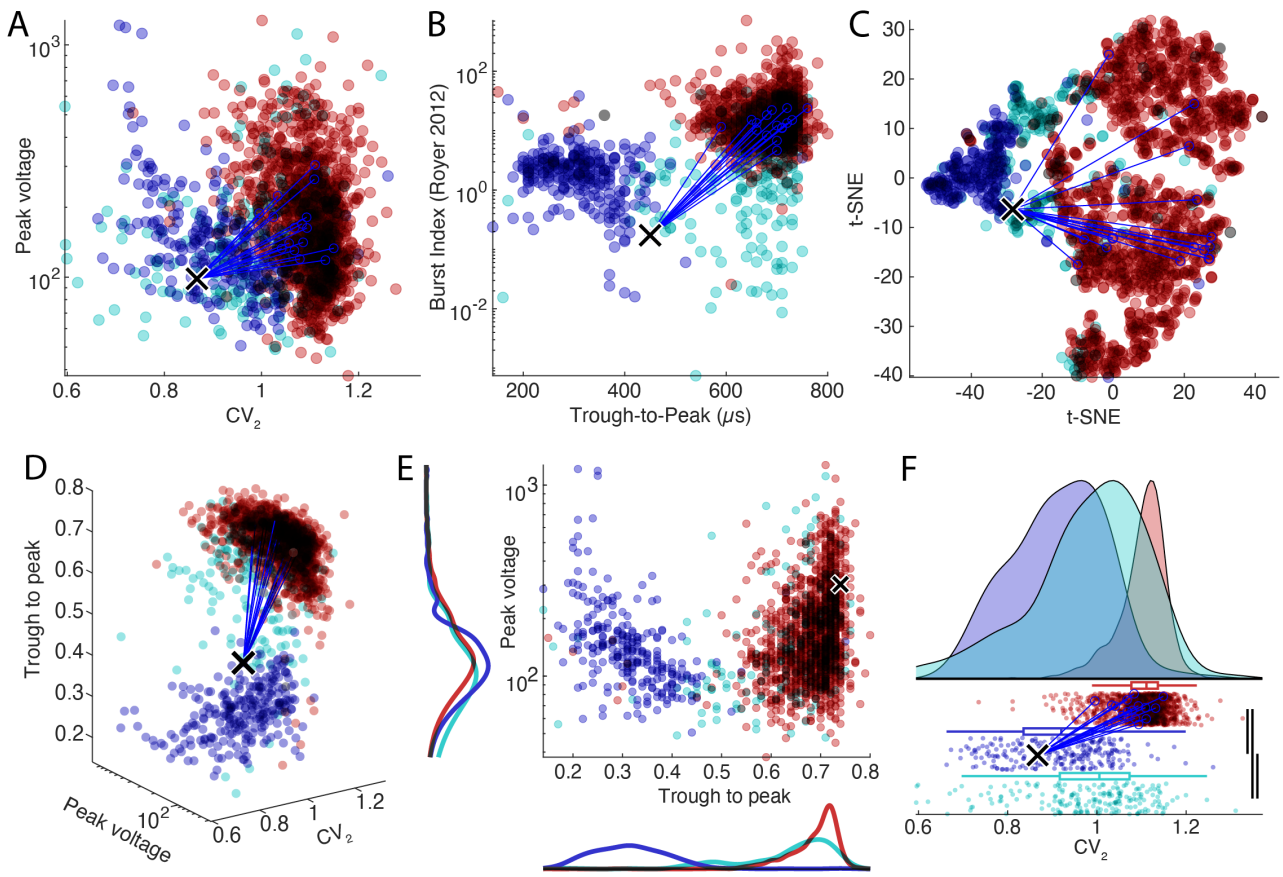
Supplementary figure 2: Primary MATLAB functions of the CellExplorer framework. From left to right: **sessionTemplate:** a template script which automatically can extract and import relevant metadata. **gui_session:** a graphical user interface (GUI) for manual inspection and entry of metadata; **ProcessCellMetrics**: the processing module. **CellExplorer**: the main graphical interface of CellExplorer. **CellExplorer_Preferences**: a preference file for the graphical interface. **gui_MonoSyn**: GUI for manual curation of monosynaptic connections. **gui_DeepSuperficial**: a GUI for manual curation of the depth assignment of neurons based on depth-related changes of sharp-wave-ripples (Mizuseki et al., 2011). **LoadCellMetricsBatch**: Batch loading script for combining cell_metrics structs across sessions. **LoadCellMetrics**: Script for loading cell_metrics with built-in common text filters (putative cell type, brain region, synaptic effect, label, animal, tags, groups, etc).



Supplementary Figure 3. Flow charts. A) Creating meta data structure for each recording session. **B**) Running the pipeline. **C**) Running the CellExplorer module for manual curation and exploration. CellExplorer data structures are highlighted in yellow, MATLAB functions in green and the input data in grey. Buzsáki lab database input is in purple (https://buzsakilab.com/wp/public-data/).



Supplementary Figure 4. Single cell plots. Most data have three representations: single neurons (with neuronal connections highlighted for a subset of the plots), all neurons (absolute or normalized representations) and image representation (normalized data, with selected cell highlighted by a white line). A-F, H-J: a single narrow interneuron, G: Place field of a pyramidal cell on a linear track. **A.** Waveform representations: waveform of a chosen single neuron, waveforms of all neurons (z-scored) and their image representation. White line corresponds to the single neuron. **B.** Autocorrelogram (ACGs) for the single neuron, ACG for all neurons and their image representation. **C.** ACGs on log scale (single, all, image). **D, E.** Interspike interval distributions (ISIs) on a log scale (single, all image) for two different normalizations (**D,** rate (Hz); **E,** occurrence). **F.** Theta phase spike histogram for the single interneuron (black line) and those of pyramidal neurons monosynaptically connected to the interneurons (blue lines; left) and all neurons in the same session (middle and right panels). **G.** Firing rate map for a pyramidal cell. Session average (left) and trial-wise heatmap. **H.** Connectivity graph showing all monosynaptic modules in the dataset. A module is highlighted and enhanced (top right). **I.** physical location of neurons recorded in the same animal using trilateration. Eight-shank silicon probe recording (8 sites on each shank). Red, pyramidal cells. Blue, interneurons. Monosynaptic connections between two pyramidal cells and a target interneuron is also shown (blue lines) **J.** Average waveform across channels of the single interneurons shown in most panels.



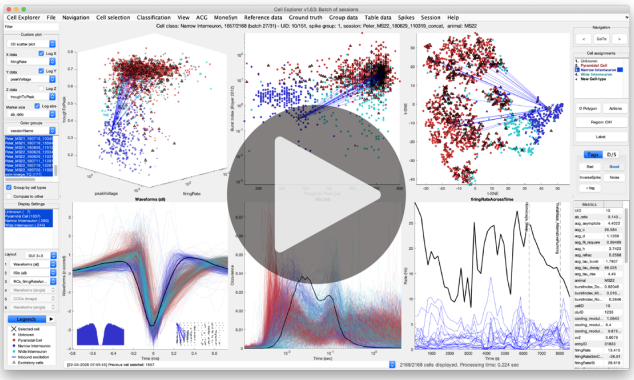
Supplementary Figure 5. Population data plots. Top row: The three standard representations: custom plot (**A**), classic representation (**B**), and t-SNE plot (**C**). **Bottom row:** The custom plot has 3 further data representations: 3-dimensional plot with custom marker size (**D**), 2D plot with marginal histograms (**E**) and one-dimensional raincloud plots (**F**), combining 1D scattered neurons with error bars histogram and KS significance test (line thickness represent significance levels). Color-coded according to cell types: pyramidal cell (red), narrow interneuron (blue), wide interneuron (cyan).

Metrics	Description/Calculation	Type
General metrics		
general	struct containing general information about the session	struct
.basename	the name of the session	char
.basepath	the path to the raw data	char
.clusteringpath	the relative clustering path	char
.cellCount	number of cells in the current session	double
.ccg	cross correlogram matrix between cell pairs within a session	201xNxN double
.ccg_time	time vector describing the time bins in the ccg (standard: -100ms:1ms:100ms)	201x1 double
animal (name)	Unique name of animal	1xN cell array of charactor vectors
general.animal	struct containing animal specific information	struct
.sex	Sex of the animal [Male, Female, Unknown]	char
.species	Animal species [Rat, Mouse,...]	char
.strain	Animal strain [Long Evans, C57B1/6,...]	char
.geneticLine	Genetic line of the animal	char
sessionName	Name of session	1xN cell array of charactor vectors
general.session	struct containing session specific information	struct
.sessionType	[Acute, Chronic]	1xN cell array of charactor vectors
.spikeSortingMethod		char
.investigator		char
UID	The ID for each cell unique within a session	1xN double
cellID		1xN double
cluID	clustering ID from spike sorting pipeline	1xN double
batchIDs	only present in batch sessions. The batch ids the cells	1xN double
putativeCellType	Putative cell type	1xN cell array of charactor vectors
brainRegion	Brain region acronyms from Allan institute Brain atlas.	1xN cell array of charactor vectors
spikeGroup	Spike group: Shank number / spike group	1xN double
labels	Custom labels	1xN cell array of charactor vectors
groups	struct containing groups	struct
tags	struct containing tags	struct
Spike event-based metrics		
spikeCount	Spike count of the cell from the entire session	1xN double
firingRate	Firing rate in Hz: Spike count normalized by the interval between the first and the last spike.	1xN double
cv2	Coefficient of variation	1xN double
refractoryPeriodViolation	Refractory period violation (‰): Fraction of ISIs less than 2ms.	1xN double
burstIndex_Mizuseki2012	Burst index: Fraction of spikes with a neighboring ISI < 6ms as defined in Mizuseki et al.	1xN double
Waveform metrics		
waveform	struct containing waveform information	struct
.filt	Average filtered waveform from peak chanel (µV)	1xN cell array of 1xM numeric vectors
.filt_std	Std of average filtered waveform (µV)	1xN cell array of 1xM numeric vectors
.raw	Average raw waveform from peak chanel (µV)	1xN cell array of 1xM numeric vectors
.raw_std	Std of average raw waveform (µV)	1xN cell array of 1xM numeric vectors
maxWaveformCh	peak channel (0-indexed)	1xN double
maxWaveformCh1	peak channel (1-indexed)	1xN double
maxWaveformChannelOrder	linearized channel position	
polarity	waveform polarity	
troughToPeak	waveform trough to peak interval (µs)	1xN double
ab_ratio	waveform peak to peak ratio	1xN double
peakVoltage	amplitude of the filtered waveform (µV). max(waveform)-min(waveform).	1xN double
troughtoPeakDerivative	derivative of waveform trough to peak interval (µs)	1xN double
ACG metrics		
acg	struct containing autocorrelogram information	struct
.wide	[-1000ms:1ms:1000ms]	1xN cell array of 1xM numeric vectors
.narrow	[-50:0.5:50]	1xN cell array of 1xM numeric vectors
.log10	[log-intervals spanning 1ms:10s]	1xN cell array of 1xM numeric vectors
thetaModulationIndex	defined by the difference between the theta modulation trough (mean of autocorrelogram bins 50-70 ms) and the theta modulation peak (mean of autocorrelogram bins 100-140ms)	1xN double
ACG fit metrics	Fit to the autocorrelogram with a triple-exponential equation (fit = cexp(-x/τ_decay)-dexp(-	1xN double
acg_asymptote	the asymptote of the ACG fit	1xN double
acg_c	ACG fit: amplitude	1xN double
acg_d	ACG fit: amplitude	1xN double
acg_fit_rsquare	ACG fit R-square (the goodness of the fit)	1xN double
acg_h	ACG fit: amplitude	1xN double
acg_refrac	ACG fit: refractory period (ms)	1xN double
acg_tau_burst	ACG fit: tau bursts (ms)	1xN double
acg_tau_decay	ACG fit: tau decay (ms)	1xN double
acg_tau_rise	ACG fit tau rise (ms)	1xN double
burstIndex_Royer2012	Burst index (Royer 2012)	1xN double
burstIndex_Doublets	Burst index doublets	1xN double
ISI metrics		
isi	struct with interspike interval information	struct
.log10	[log-intervals spanning 1ms:10s]	1xN cell array of 1xM numeric vectors
Putative connections		
putativeConnections	putative connections determined from cross correlograms	struct
putativeConnections.Excitatory	excitatory connection pairs	2xP double
putativeConnections.Inhibitory	inhibitory connection pairs	2xP double
synapticEffect	Excitatory' or 'Inhibitory'	1xN cell array of charactor vectors
synapticConnectionsIn	Synatic connections count	1xN double
synapticConnectionsOut	Synatic connections count	1xN double
Event metrics		
events	event time series	struct
.'name'	the event curve	1xN cell array of 1xM numeric vectors
name'_modulationIndex	modulation index for each event types	1xN double
name'_modulationSignificanceLevel	modulation significance level for each event types	
name'_modulationPeakResponseTime	modulation peak response time for each event types	1xN double

Firing rate map metrics		
firingRateMaps	struct with (spatial) linearized firing rate maps	struct
.firingRateMaps	The mean firing rate map	1xN cell array of 1xM numeric vectors
spatialCoverageIndex	Spatial coverage index. Defined from the inverse cumulative distribution, where bins are sorted by decreasing rate. The 75 percentile point defines the spatial coverage by the fraction of bins below and above the point (defined by Royer et al., NN 2012)	1xN double
spatialGiniCoeff	Spatial Gini coefficient. Defined as the Gini coefficient of the firing rate map	1xN double
spatialCoherence	Spatial Coherence. Defined by the degree of correlation between the firing rate map and a hollow convolution with the same map	1xN double
spatialPeakRate	Spatial peak firing rate (Hz). Defined as the peak rate from the firing rate map	1xN double
placeFieldsCount	Place field count: Number of intervals along the firing rate map that fulfills a set of spatial criteria: minimum rate of 2Hz and above 10% of the maximum firing rate bin and minimum of 4 connecting bins. The cell further has to have a spatial coherence greater than 0.6 (Mizuseki	1xN double
spatialSplitterDegree		1xN double
placeCell	Place cell (determined from the Mizuseki spatial metrics)	1xN binary
Manipulation metrics		
manipulations	manipulations time series	struct
.'manipulationName'		1xN cell array of charactor vectors
Response curves metrics		
responseCurves	response curves	struct
.'responseCurveName'		1xN cell array of charactor vectors
Quality metrics		
refractoryPeriodViolation	Refractory period violation (‰): Fraction of ISIs less than 2ms	1xN double
isolationDistance	Isolation distance as defined by Schmitzer-Torbert et al. Neuroscience. 2005.	1xN double
lRatio	L-ratio as defined by Schmitzer-Torbert et al. Neuroscience. 2005.	1xN double
Hippocampal sharp wave ripple metrics		
deepSuperficial	Deep-Superficial region assignment [Unknown, Cortical, Superficial, Deep]	
deepSuperficialDistance	Deep Superficial depth relative to the reversal of the sharp wave (µm)	1xN double
Hippocampal theta oscillation metrics		
thetaPhasePeak	Theta phase peak	1xN double
thetaPhaseTrough	Theta phase trough	1xN double
thetaEntrainment	Theta entrainment	1xN double
thetaModulationIndex	Theta modulation index. determined from the ACG	1xN double
Firing rate stability metrics		
firingRateGiniCoeff	The Gini coefficient of the firing rate across time	1xN double
firingRateStd	Standard deviation of the "firing rate across time" divided by the mean'	1xN double
firingRateInstability	Mean of the absolute differential "firing rate across time" divided by the mean.	1xN double
Database metrics		
entryID	database entry id	1xN double
sessionID	database session id	1xN double

Table 1: Cell metrics. An incomplete list of the standard cell metrics. The full list is available online at petersenpeter.github.io/CellExplorer/datastructure/standard-cell-metrics/

Supplementary Movie 1



The supplementary movie is available at: buzsakilab.com/CellExplorer/CellExplorerMovie.mp4

TUTORIALS

A general tutorial on the full pipeline is described below. There are many more detailed tutorials online, covering: how to generate the metadata struct, the manual curation process, generating spike raster plots, the manual curation process of monosynaptic connections, performing opto-tagging, using ground truth data, export figure.

Tutorials are available online at: petersenpeter.github.io/CellExplorer/tutorials/tutorials/

General tutorial

This tutorial shows you the full processing pipeline, from generating the necessary session metadata using the template, running the processing pipeline, opening multiple sessions for manual curation in the CellExplorer, and finally using the cell_metrics for filtering cells, by two different criteria. The tutorial is also available as a Matlab script: (`tutorials/CellExplorer_Tutorial.m`).

1. Define the basepath of the dataset to process. The dataset should consist of a `basename.dat` (a binary raw data file), a `basename.xml` (recommended; not required) and spike sorted data.

   ```
   basepath = '/your/data/path/basename/';
   cd(basepath)
   ```

2. Generate session metadata struct using the template function and display the metadata in a GUI

   ```
   session = sessionTemplate(basepath, 'showGUI',true);
   ```

   In the GUI you can put in relevant metadata. Please pay attention to the general, extracellular and spikesorting tabs and verify all metadata.

3. Run the cell metrics pipeline `ProcessCellMetrics` using the session struct as input

   ```
   cell_metrics = ProcessCellMetrics('session', session);
   ```

4. Visualize the cell metrics in the CellExplorer

   ```
   cell_metrics = CellExplorer('metrics', cell_metrics);
   ```

5. Open several sessions from paths

   ```
   basenames = {'session1', 'session2'};
   clusteringpaths = {'path/to/session1','path/to/session2'};
   cell_metrics = LoadCellMetricsBatch('clusteringpaths', clusteringpaths, 'basenames', basenames);
   cell_metrics = CellExplorer('metrics', cell_metrics);
   ```

6. Curate cells and save the metrics
7. Now to incorporate the cell metrics into your analysis you can use the load function that has filters built-in:
   1. Get cells that are assigned as Interneuron

      ```
      cell_metrics_idxs1 = loadCellMetrics('cell_metrics', cell_metrics, 'putativeCellType', {'Interneuron'});
      ```

   2. Get cells that are has groundTruthClassification as Axoaxonic

      ```
      cell_metrics_idxs2 = loadCellMetrics('cell_metrics', cell_metrics, 'groundTruthClassification', {'Axoaxonic'});
      ```