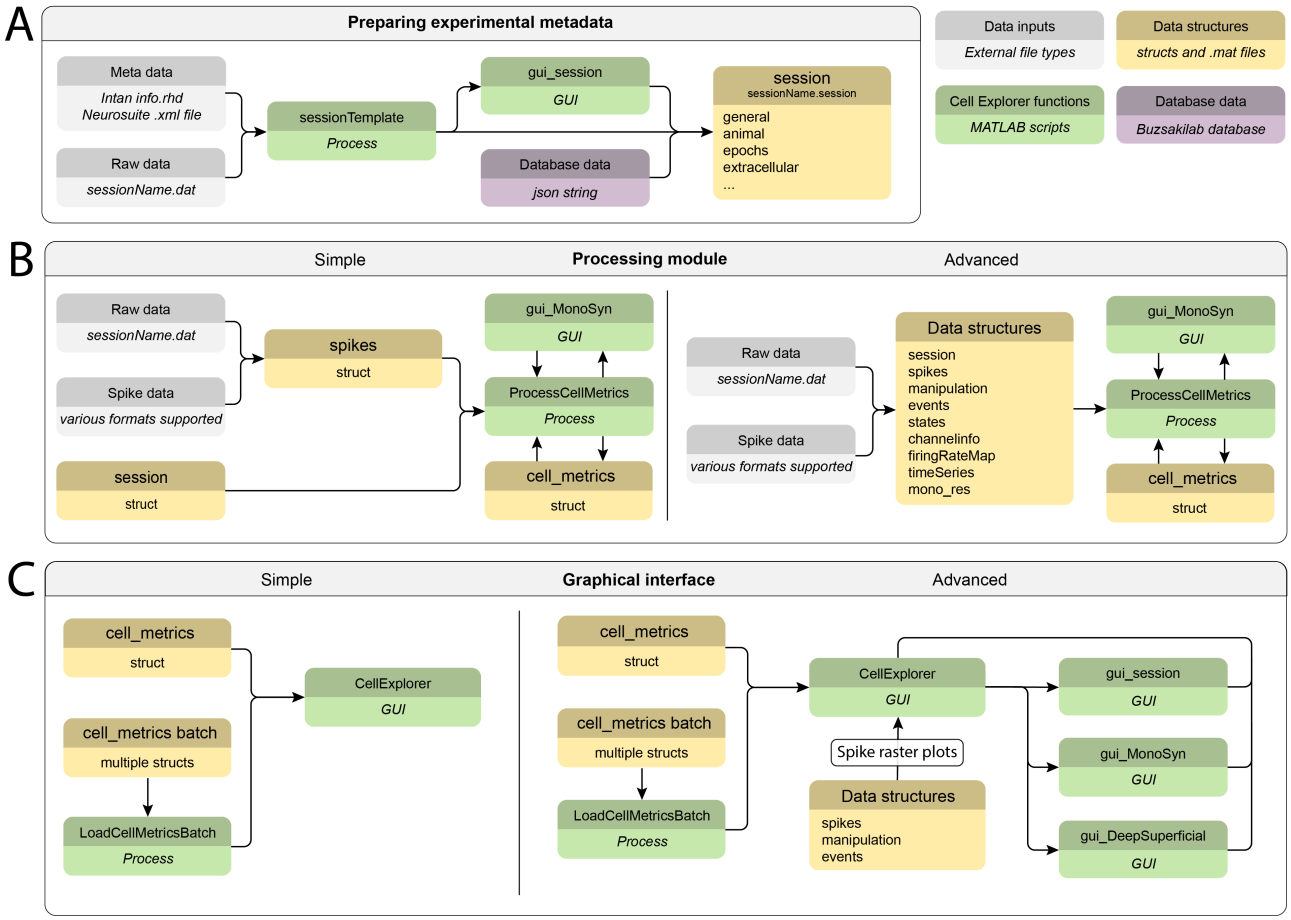
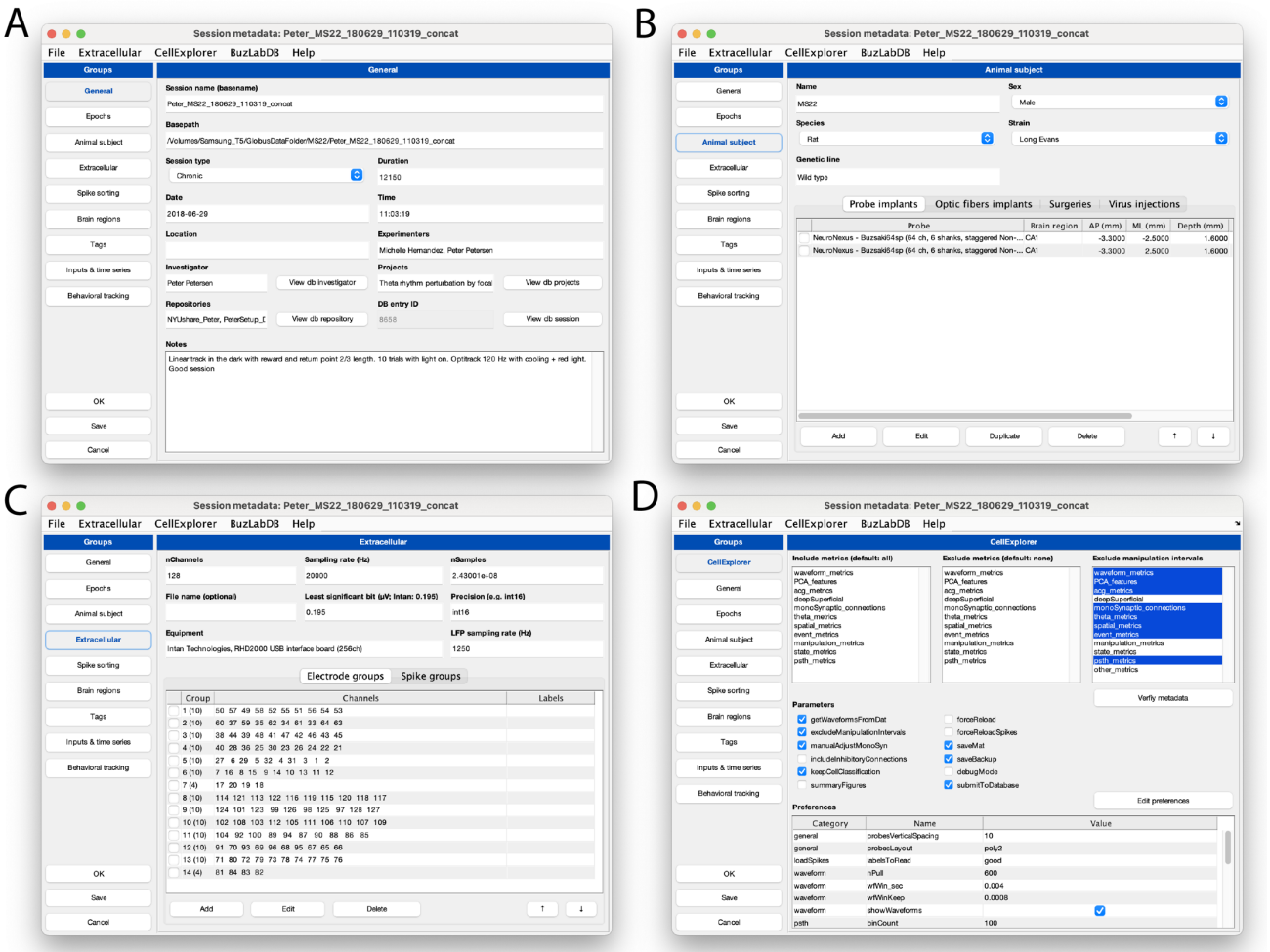
SUPPLEMENTARY MATERIAL

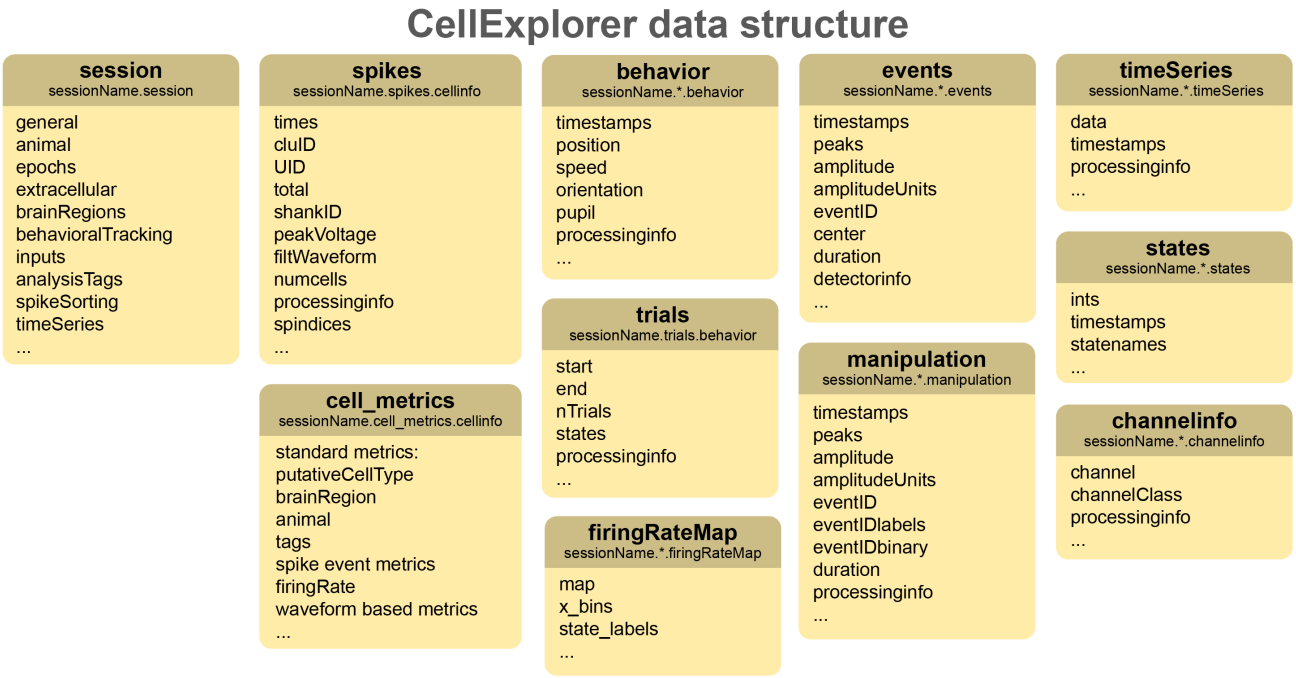
Supplementary Figures and Tables



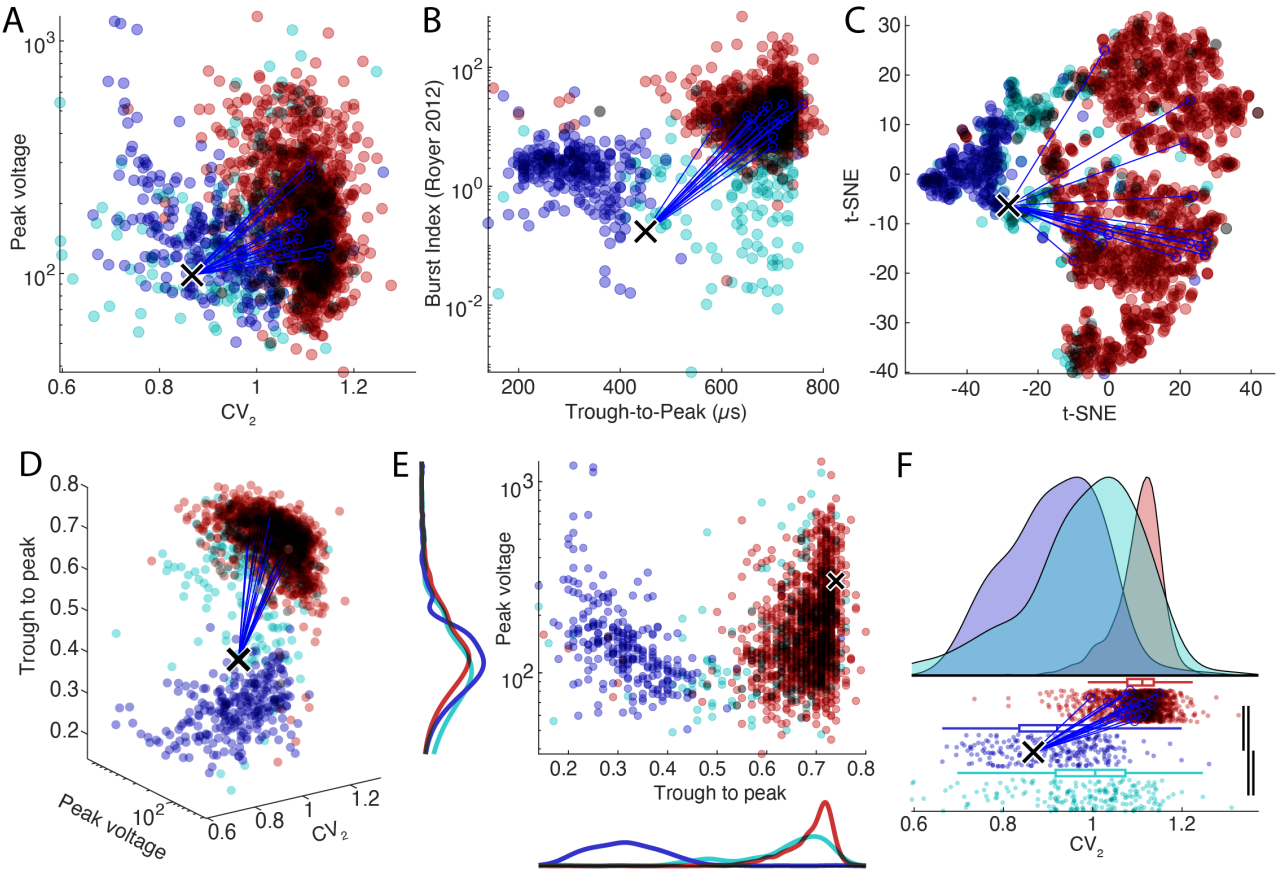
Supplementary Figure 1. Flow charts, related to figure 2. A) Generating the metadata structure for a recording session. **B**) Running the processing pipeline. **C**) Running the CellExplorer module for manual curation and exploration. CellExplorer data structures are shown in yellow, MATLAB functions in green, and the input data in grey. Input from the Buzsáki lab database is shown in purple (Petersen et al., 2020).



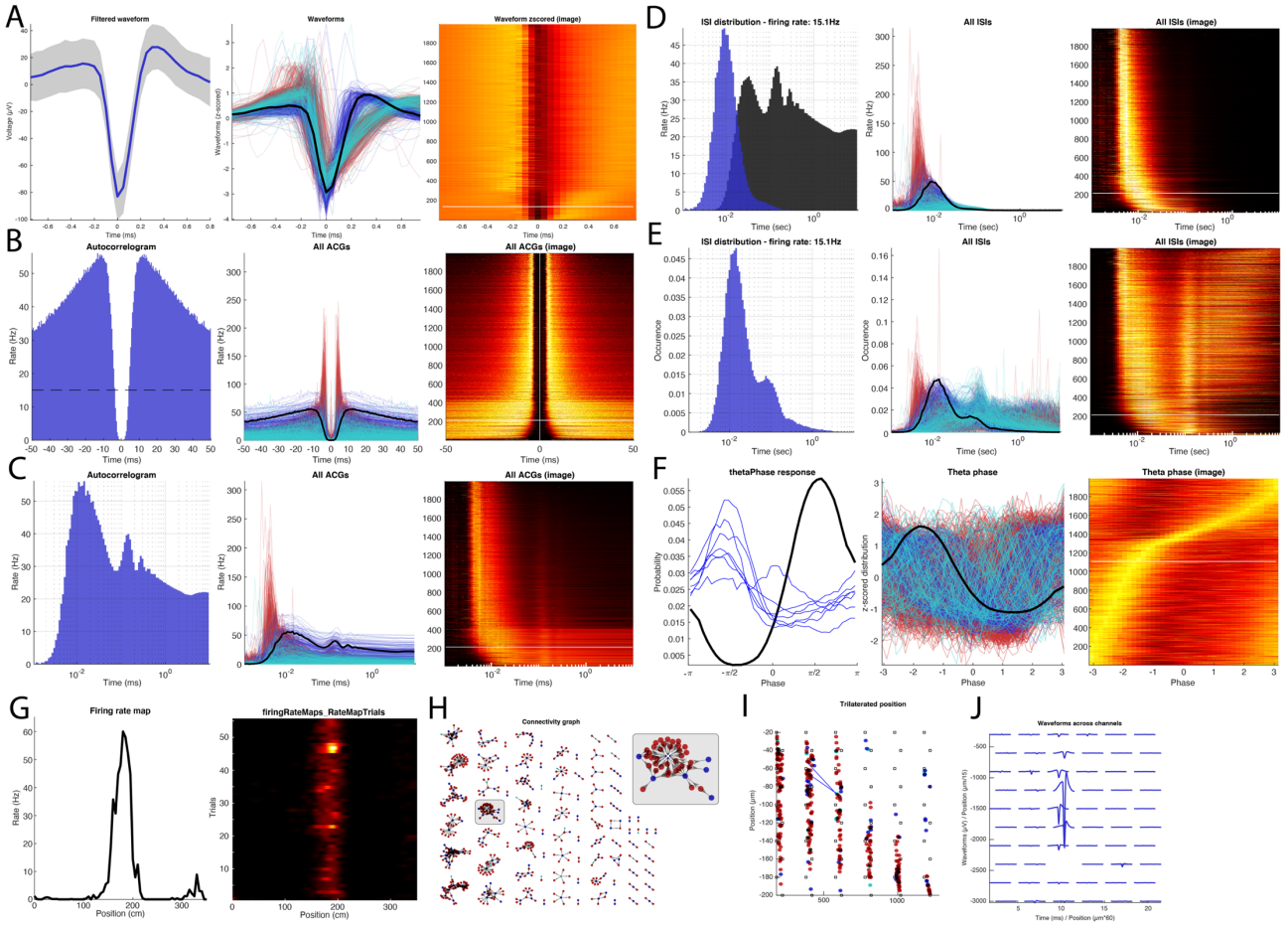
Supplementary Figure 2: Session metadata GUI, related to figure 2. The graphical interface for inspection and entry of session level metadata follows the organization of the Matlab struct, with a tab for each field type. **A.** General information about the session, including name, data, duration, location, and notes. **B.** Animal metadata capturing sex, species, strain and genetic line, but also action performed on the animal including probe implants, optic fiber implants, surgeries, and virus injection. **C.** Basic metadata for the extracellular data, including channel count, sampling rate, equipment and electrode groups. **D.** The session GUI is also used as a graphical interface for the processing pipeline (with a dedicated tab), allowing the user to change parameters, view settings, validate metrics, and see and edit the full session metadata structure, that serves as input to the pipeline.



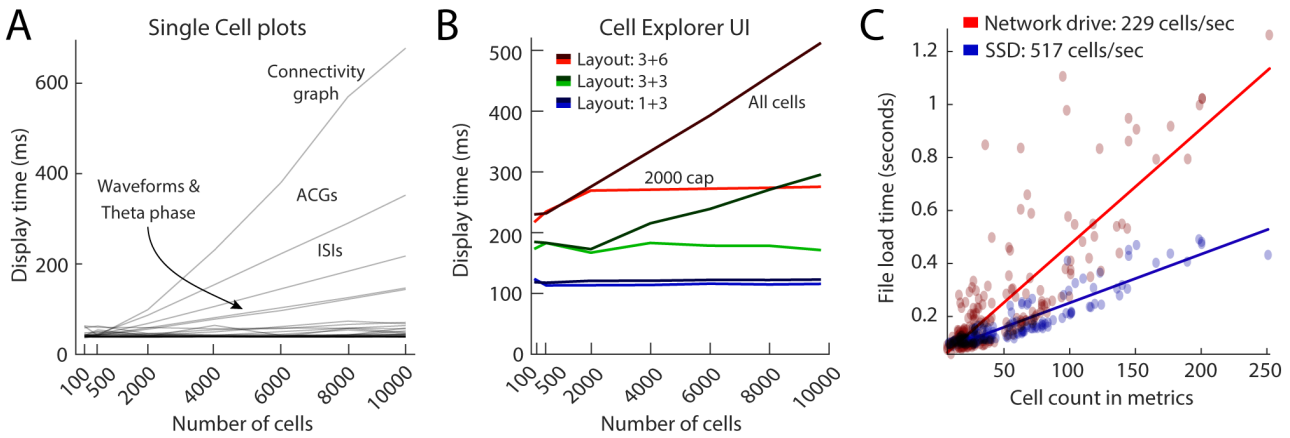
Supplementary Figure 3: Datatypes related to figure 2. The data structure. A detailed description is available online at CellExplorer.org/datastructure/data-structure-and-format. session, spikes, cell_metrics, trials are defined data types, while behavior, firingRateMap, events, manipulation, timeseries, states, channelinfo are data containers.



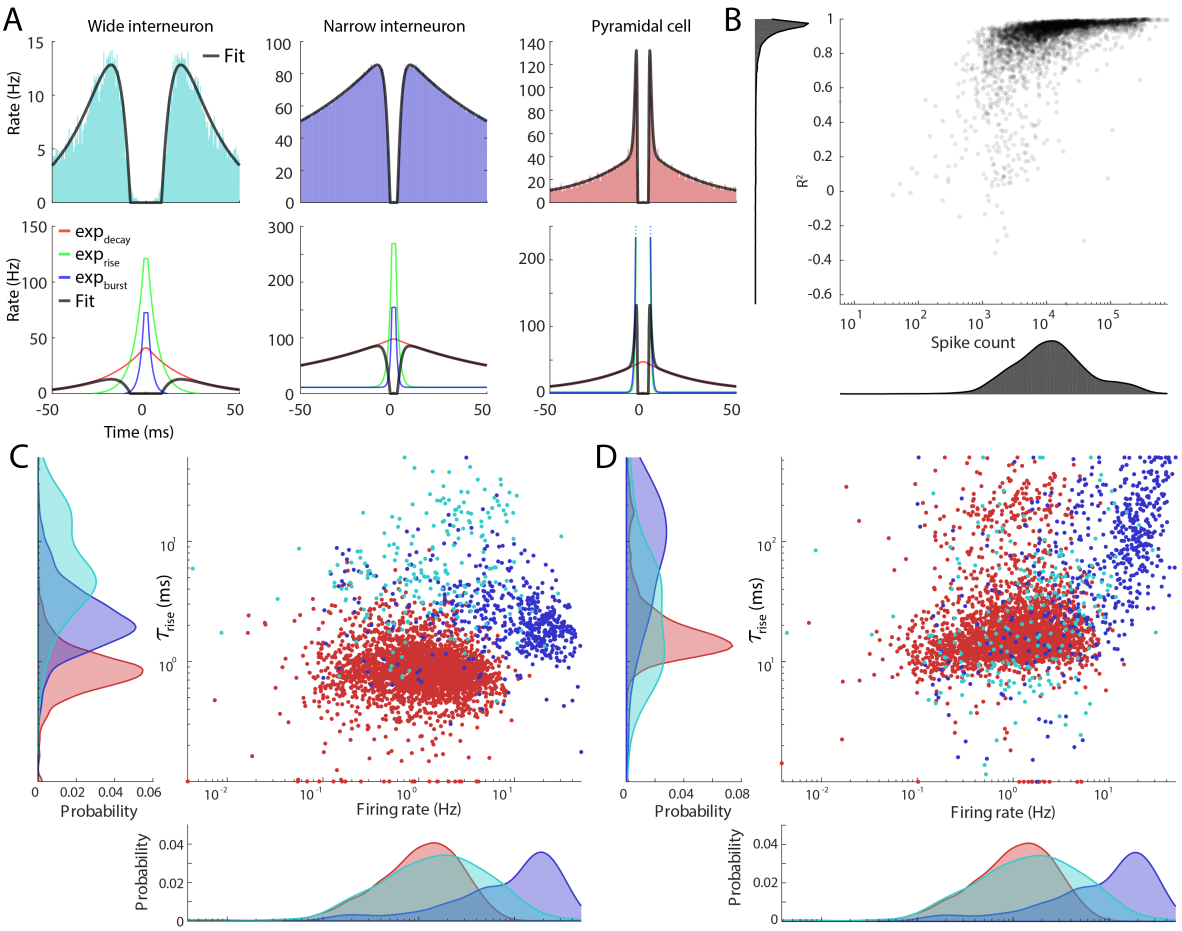
Supplementary Figure 4. Population data plots, related to figure 3. Top row: The three standard representations: custom plot (**A**), classic representation (**B**), and t-SNE plot (**C**). **Bottom row:** The custom plot has 3 further data representations: a 3-dimensional plot with custom marker size (**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).



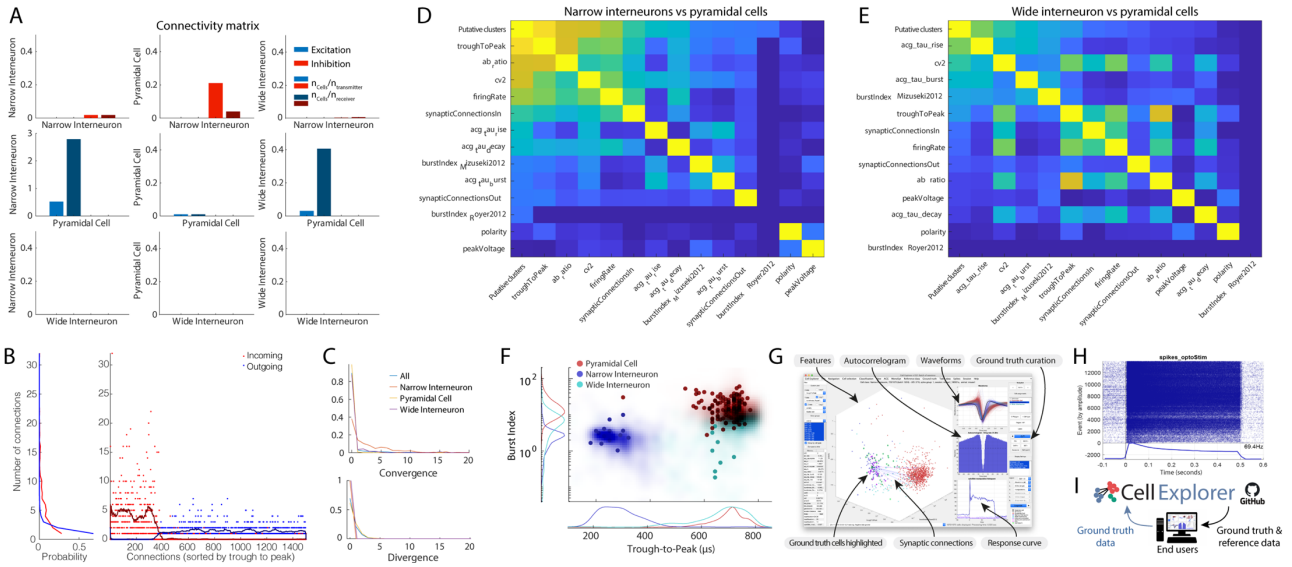
Supplementary Figure 5. The various single-cell plots, related to figure 3. Most single cell data-visualizer have three representations: single neurons (with neuronal connections highlighted for a subset of the plots), all neurons (absolute or normalized representations), and an image representation (normalized data, with selected cell highlighted by a white line). **A.** Waveform representations: waveform of a selected single neuron, waveforms of all neurons (z-scored), and their image representation. The white line in the image representation corresponds to the selected neuron. **B.** Autocorrelograms (ACGs) for the single neuron, ACGs for all neurons and their image representation. **C.** ACGs on a 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 are also shown (blue lines) **J.** Average waveform across channels of the single interneurons shown in most panels. A-F, H-J: a narrow interneuron, G: Spatial firing rate of a pyramidal cell on a linear track.



Supplementary figure 6. Benchmarks of the CellExplorer user interface (UI) related to figure 3. A. Display times for single-cell plots, quantified by the number of cells displayed. The slowest plots are the ones with a trace 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 for which all connections are shown. **B.** UI display times when switching between units for the three layouts shown in figure 3B (approximately 110 ms for layout 1+3 with 4 subplots; blue lines. 180 ms for layout 3+3 with 6 subplots; green lines) and 290 ms (layout 3+6 with 9 subplots; 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 display times with a maximum of 2000 random traces. **C.** Benchmarks of cell metrics file loading time. 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 can be decreased and attain cell loading above 500 cells per second. Graphical benchmarks were performed on an iMac from 2017 with a 4.2GHz Quad-Core Intel i7 with 32GB of ram. File load time tests were performed on a custom PC running Window 10, with a 512GB Samsung 870 EVO SSD (SSD) and the NYU Langone Health network storage solution "Research Isilon" (Network drive).



Supplementary Figure 7. ACG fits related to figure 5. A. Three examples of typical autocorrelograms for a wide interneuron (left column) narrow interneuron (middle column) and a pyramidal cell (right column). The exponential components are plotted in the lower row. **B.** The R^2 values for each fit across the 4000 cells plotted against the number of spikes. **C-D** τ_{rise} (C) and τ_{decay} (D) values plotted against the firing rate. Color coded by putative cell type.



Supplementary Figure 8. Cell type separation, ground truth- and reference data related to figure 5.
A. Connectivity by cell types. Each panel is a projection pattern showing connection from one cell type to another, both excitatory (blue) and inhibitory (red). The two bars in the same color are normalized by the transmitter and receiver population count. **B.** Every synaptic connection is sorted by the spike waveform trough-to-peak, showing a clear separation between which cells transmit and receive on the basis of spike waveform features. **C.** Convergence and divergence by cell types. **D-E.** Correlation between putative clusters and various metrics. **D**: Narrow interneuron vs pyramidal cells, **E**: Wide interneurons vs pyramidal cells. **F.** Single session (dots) data compared with data from 30 reference sessions (shaded zones). **G.** Opto-tagged data can be processed and curated directly in CellExplorer. **H.** Example of a PSTH of a PV-expressing neuron to 500 ms square light pulses. Raster plot and average responses to the light pulses are visualized in CellExplorer. **I.** 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 (see the opto-tagging tutorial at CellExplorer.org).

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
.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)	Name of animal subject	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	Name of the investigator	char
general.processinginfo	Struct containing processing info: date of the processing, version of the script, function name…	
.params	Struct containing the input parameters used by ProcessCellMetrics	
UID	The ID for each cell unique within a session (1:nCells)	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
shankID	Shank number / electrode 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		
spikes.times	struct containing spike times	
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. Hippocampus 2012	1xN double
Waveform metrics		
waveform	struct containing waveform information	struct
.filt	Average filtered waveform from peak chanel (µV)	1xN cell array of numeric vectors
.filt_std	Std of average filtered waveform (µV)	1xN cell array of numeric vectors
.raw	Average raw waveform from peak chanel (µV)	1xN cell array of numeric vectors
.raw_std	Std of average raw waveform (µV)	1xN cell array of numeric vectors
.time	Time vector (ms)	1xN cell array of 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
Autocorellogram (ACG) metrics		
acg	struct containing autocorrelogram information	struct
.wide	[-1000ms:1ms:1000ms]	1xN cell array of numeric vectors
.narrow	[-50:0.5:50]	1xN cell array of numeric vectors
.log10	[log-intervals spanning 1ms:10s]	1xN cell array of 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) over their sum.	1xN double
ACG fit metrics	Fit to the autocorrelogram with a triple-exponential equation (fit = cexp(-x/τ_decay)-dexp(-x/τ_rise))	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
Interspike Intervals (ISI) metrics		
isi	struct with interspike interval information	struct
.log10	[log-intervals spanning 1ms:10s]	1xN cell array of 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 numeric vectors
_modulationIndex	modulation index for each event types	1xN double
_modulationSignificanceLevel	modulation significance level for each event types	
_modulationPeakResponseTime	modulation peak response time for each event types	1xN double
Firing rate map metrics		
firingRateMaps	struct with (spatial) linearized firing rate maps	struct
.ratemap	Primary firing rate map	1xN cell array of numeric vectors
.'ratemapName'	Other firing rate maps	1xN cell array of 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 et al ?).	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
firingRateCV	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. abs(diff(firingRateAcrossTime))	1xN double

Supplementary Table 1: Cell metrics, related to figure 1. An incomplete list of the standard cell metrics. The full list is available online at CellExplorer.org/datastructure/standard-cell-metrics

Functions	Description
sessionTemplate	A template script which automatically extracts and imports relevant metadata
gui_session	A graphical user interface (GUI) for inspection and entry of metadata
ProcessCellMetrics	The processing module
CellExplorer	The main graphical interface of CellExplorer
preferences_CellExplorer	Preferences for the graphical interface
preferences_ProcessCellMetrics	Preferences for the processing module
NeuroScope2	An Ephys visualizer built upon the data structure of CellExplorer
gui_MonoSyn	GUI for manual curation of monosynaptic connections
gui_DeepSuperficial	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.)
loadSpikes	Script for importing and loading spike data
saveStruct	Saving to various file containers (e.g. cellinfo, behavior, session, events, states)
saveCellMetrics	Saving cell metrics

Supplementary Table 2. Primary MATLAB functions of the CellExplorer framework, related to Star method, related to figure 2. All code is available at GitHub: https://github.com/petersenpeter/CellExplorer.