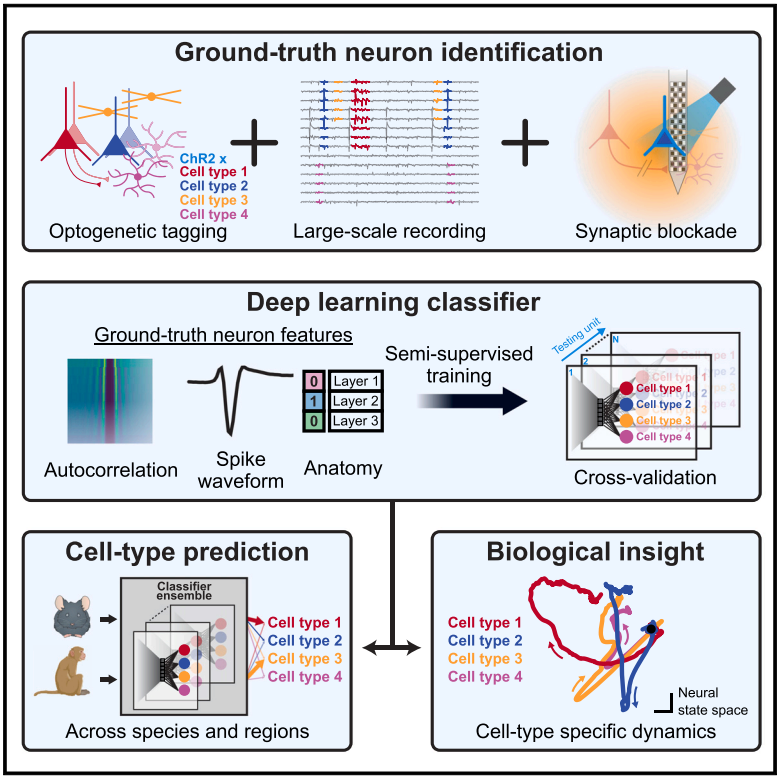
Cell

Article

A deep learning strategy to identify cell types across species from high-density extracellular recordings

Graphical abstract



Highlights

- Ground-truth library of extracellular features of optotagged cerebellar cell types
- Deep learning classifier identifies cell type from extracellular recordings
- Successful cell-type classification across cerebellar regions in mice and monkeys
- Different cerebellar cell types show distinct population dynamics across tasks

Authors

Maxime Beau, David J. Herzfeld, Francisco Naveros, ..., Court Hull, Michael Häusser, Javier F. Medina

Correspondence

jfmedina@bcm.edu

In brief

We constructed a deep learning classifier that solves the problem of neuron-type identification from simultaneous extracellular recordings of many neurons. The classifier was trained on a ground-truth library of extracellular features from optogenetically identified neuron types. It accurately predicts neuron identity across multiple cerebellar regions and two species, allowing the assessment of unique contributions to behavior from different neuron types.

Beau et al., 2025, Cell *188*, 1–17
April 17, 2025
https://doi.org/10.1016/j.cell.2025.01.041

CellPress

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

Cell

CellPress

Article

A deep learning strategy to identify cell types across species from high-density extracellular recordings

Maxime Beau,[1,9] David J. Herzfeld,[2,9] Francisco Naveros,[3,4,9] Marie E. Hemelt,[2,9] Federico D'Agostino,[1,9] Marlies Oostland,[1,5,9] Alvaro Sánchez-López,[3,9] Young Yoon Chung,[1] Michael Maibach,[1] Stephen Kyranakis,[3] Hannah N. Stabb,[1] M. Gabriela Martínez Lopera,[1] Agoston Lajko,[1] Marie Zedler,[1] Shogo Ohmae,[3] Nathan J. Hall,[2] Beverley A. Clark,[1,10,11] Dana Cohen,[6,10,11] Stephen G. Lisberger,[2,10,11] Dimitar Kostadinov,[1,7,10,11] Court Hull,[2,10,11] Michael Häusser,[1,8,10,11] and Javier F. Medina[3,10,11,12,*]

[1]Wolfson Institute for Biomedical Research, University College London, London, UK
[2]Department of Neurobiology, Duke University School of Medicine, Durham, NC, USA
[3]Department of Neuroscience, Baylor College of Medicine, Houston, TX, USA
[4]Department of Computer Engineering, Automation and Robotics, Research Centre for Information and Communication Technologies, University of Granada, Granada, Spain
[5]Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, the Netherlands
[6]The Leslie and Susan Gonda Multidisciplinary Brain Research Center, Bar-Ilan University, Ramat-Gan, Israel
[7]Centre for Developmental Neurobiology, King's College London, London, UK
[8]School of Biomedical Sciences, The University of Hong Kong, Hong Kong, China
[9]These authors contributed equally
[10]These authors contributed equally
[11]Senior author
[12]Lead contact
*Correspondence: jfmedina@bcm.edu
https://doi.org/10.1016/j.cell.2025.01.041

SUMMARY

High-density probes allow electrophysiological recordings from many neurons simultaneously across entire brain circuits but fail to reveal cell type. Here, we develop a strategy to identify cell types from extracellular recordings in awake animals and reveal the computational roles of neurons with distinct functional, molecular, and anatomical properties. We combine optogenetics and pharmacology using the cerebellum as a testbed to generate a curated ground-truth library of electrophysiological properties for Purkinje cells, molecular layer interneurons, Golgi cells, and mossy fibers. We train a semi-supervised deep learning classifier that predicts cell types with greater than 95% accuracy based on the waveform, discharge statistics, and layer of the recorded neuron. The classifier's predictions agree with expert classification on recordings using different probes, in different laboratories, from functionally distinct cerebellar regions, and across species. Our classifier extends the power of modern dynamical systems analyses by revealing the unique contributions of simultaneously recorded cell types during behavior.

INTRODUCTION

The nervous system comprises many cell types defined by their molecular, anatomical, morphological, and physiological properties.[1–6] Powerful modern molecular techniques have revealed multiple subtypes even within known anatomical cell classes.[7–13] Identification of cell type at multiple levels will be crucial to understand how the brain works and to develop selective, targeted therapeutics for brain dysfunction. Therefore, it is crucial to develop strategies to determine cell type and to cross-reference cell type across levels of analysis.[3,5,6,12,14,15]

Extracellular recording of the action potentials of neurons during behavior is an essential approach for understanding how neural circuits generate behavior. Thus, it is necessary to achieve reliable cell-type identification based on the features of extracellular recordings during specific, quantified behaviors. Specifically, simultaneous large-scale electrophysiological recordings using high-density multi-contact recording probes[16,17] coupled with cell-type identification *in vivo* would facilitate characterization of circuit-level processing that drives behavior.

Identification of cell type is a particularly difficult challenge for extracellular recording technologies that deliver only spike waveforms and firing statistics. Previous attempts at extracellular neuron identification by measuring specific features of the recordings have not proven robust across laboratories.[18,19] While optogenetic approaches to cell-type identification[20–23]

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

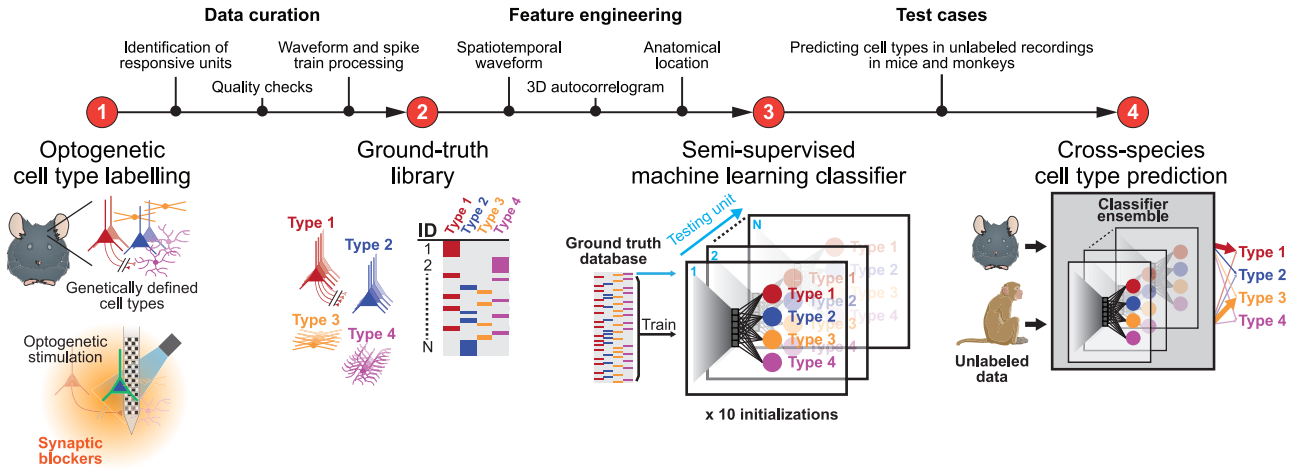


Figure 1. A strategy for cell-type identification from extracellular recordings in neural circuits
See details in main text.

offer a viable solution for the creation of a library of extracellular properties, direct optogenetic identification cannot be deployed at adequate scale: optogenetic identification currently is routine only in mice and can be used to target only one or two cell types at a time in a given preparation.[24] In general, the challenges of cell-type identification from extracellular recordings cannot be resolved by revealing the transcriptional or anatomical properties of neurons[25] because the requisite technologies are not compatible with electrode recordings from multiple cell types simultaneously in awake animals.

Our goal was to enable cell-type identification solely from extracellular recordings in awake animals by developing a strategy that could scale across labs, probes, and species. We chose to pioneer the strategy in the cerebellar cortex. The cerebellum has a crystalline architecture with well-defined neuronal connectivity and a small number of anatomically defined cell types[1,26] that are evolutionarily conserved,[27,28] allowing direct comparison of recordings across species. It has a range of neuron sizes from among the smallest and most densely packed (granule cells) to the largest (Purkinje cells) in the brain, allowing us to test the resolution of our recording approaches. It has many spontaneously firing neurons,[29–31] some with high spontaneous rates, allowing rigorous characterization of their electrophysiological properties. Genetically defined mouse Cre lines are available for all major cell types in the cerebellum,[32–36] allowing us to leverage optogenetic strategies for ground-truth cell-type identification.[20] Finally, the cerebellum has a long history of neurophysiological recording,[37] allowing us to reference our measurements and automated cell-type classifications against hard-won human expertise. Strategies to solve the challenges of cell-type identification in such a testbed should provide a roadmap for application to other structures, including the cerebral cortex, the hippocampus, and the basal ganglia.

We accomplished our goal by creating a ground-truth library of identified cerebellar cell types recorded in awake mice and developing a semi-supervised deep learning classifier that accurately predicts cell type for the ground-truth library based on the waveform, discharge statistics, and anatomical layer of the recording. The classifier identifies cell type with high confidence in a high fraction of expert-labeled cerebellar recordings from two different laboratories in behaving mice and monkeys. Dynamical systems analysis enables biological insights by revealing the distinct temporal dynamics of simultaneously recorded populations of identified cell types during complex behaviors in mice and monkeys.

RESULTS

General approach

We deployed the multi-step strategy outlined in Figure 1. (1) We created a ground-truth library of cell types based on optogenetic activation of genetically defined neurons using synaptic blockade to confirm that neurons were activated directly. (2) We identified features of electrophysiological recordings that could be used to train a semi-supervised deep learning classifier on the ground-truth library. (3) We tested the generality of the classifier by predicting cell types in independent datasets of expert-classified recordings from mice and monkeys.

Multi-contact probe recordings and data curation

In the cerebellar cortex, morphologically distinct cell types reside in different layers (Figure 2A). Purkinje cells comprise a monolayer and extend their planar dendrites through the molecular layer. Molecular layer interneurons (MLIs) reside across the extent of the molecular layer and include basket cells that innervate the Purkinje cell's soma and stellate cells that innervate the Purkinje cell's dendrites. The granule cell layer includes mossy fiber terminals, Golgi cells, and granule cells. Other, less-common cell types exist in the different layers,[26] but we focused on the primary cell types of the cerebellar circuit (Figure 2A) due to the availability of Cre lines for expression of opsins in those cell types.

Purkinje cells allow ground-truth identification from their extracellular electrical signature alone (Figure 2B, left). They have "simple spikes" that fire at high rates and "complex spikes," driven by their climbing fiber input,[38–40] that occur at ~1 Hz

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

CellPress

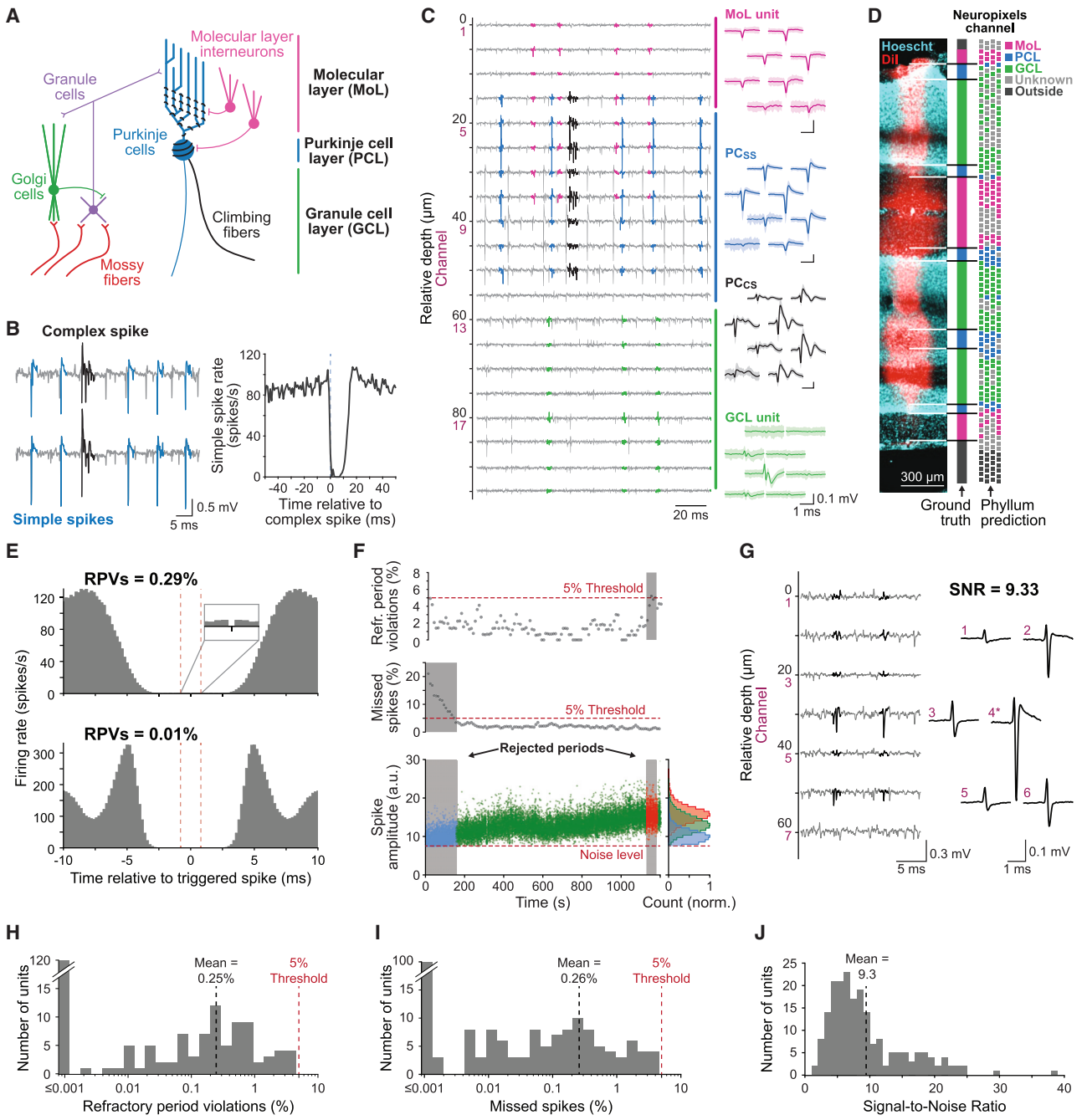


Figure 2. Curation of Neuropixels recordings in the mouse cerebellar cortex

(A) Schematic diagram of the canonical cerebellar circuit.

(B) Example simple spikes (light blue) and complex spikes (black) in a Purkinje cell. Cross-correlogram on the right documents a complex-spike-triggered pause in simple spikes.

(C) Example recordings from 20 channels of a Neuropixels probe. Traces highlight a single unit recorded in the molecular layer (magenta), a Purkinje cell's simple spikes (blue), the same Purkinje cell's complex spikes (black), and a unit recorded in the granule cell layer (green).

(D) Comparison of example histology labeled with DiI and Hoechst with the layers predicted by Phyllum from the electrical recordings. Different colors on the Neuropixels schematic denote Phyllum-predicted layer.

(E) Typical autocorrelograms for two units with very few refractory period violations (RPVs).

(F) Analysis of quality of isolation as a function of time during a recording session. From top to bottom, graphs show the percentage of RPVs, the estimated percentage of missed spikes, and spike amplitude. Horizontal dashed lines show thresholds for acceptance. Gray regions show periods that were rejected from

(legend continued on next page)

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

and trigger a characteristic 10–50 ms pause in simple spikes.[41] Thus, Purkinje cells can be admitted into the ground-truth library if they show a pause in a complex-spike-triggered histogram of simple-spike firing (Figure 2B, right).

Recordings with Neuropixels probes detect neural activity on many of the 384 channels, and spike-sorting yields many units. The magenta waveforms in Figure 2C arose from a neuron in the molecular layer that would be a candidate to be a MLI. The green waveforms came from a neuron recorded in the granule cell layer that could be a mossy fiber, a Golgi cell, or a granule cell. The blue and black waveforms were the simple and complex spikes of an identified Purkinje cell. We highlight the largest-amplitude units in the recordings shown in Figures 2B and 2C, but we also sorted the smaller amplitude units and subjected them to our analysis and curation pipeline, detailed below.

The first step in our analysis pipeline was the objective identification of the layer of each channel. The typical recording trajectory documented with DiI staining in Figure 2D crossed 3 molecular layers, 5 Purkinje cell layers, and 3 granule cell layers. We assigned each channel to a layer using Phyllum, a Phy plugin that analyzes recordings across the channels on a probe to infer the layer recorded by each channel (see STAR Methods). The layer structure inferred by Phyllum agreed well with histological layer identification based on correspondence between the DiI track of the probe and nuclear staining (Figure 2D). We validated Phyllum across 21 histologically confirmed penetrations, and its conclusions agreed with the histology at 99%, 95%, and 98% of 776, 367, and 1,140 recording sites, respectively, in the molecular, Purkinje cell, and granule cell layers.

We next ensured that each unit we admitted for further analysis was a well-isolated single neuron. We manually curated the output from Kilosort2 with Phy and performed automated quality checks to ensure isolation. We strove to ensure that we neither missed many spikes from the neuron under study nor included noise or spikes from neighboring neurons.

- We analyzed the assumed 0.8 ms refractory period from each isolated neuron to assess the level of potential contamination from other neurons or noise.[42] The example autocorrelograms in Figure 2E have few refractory period violations and, respectively, represent the mean (0.25%) and median (0.01%) in our dataset. We rejected any units with greater than 5% refractory period violations (Figure 2F, red symbols and histogram), and almost all accepted neurons had fewer than 1% refractory period violations (Figure 2H). In combination with the mean firing rate, refractory period violations can be used to estimate the fraction of a neuron's spikes that may be contaminated, an analysis we provide later in the paper, once we have identified cell types.
- We estimated the number of missed spikes by fitting the spike amplitude distribution with a Gaussian function and quantifying the fraction of the area under the curve that was clipped at noise threshold[43,44] (Figure 2F). In Figure 2F, we excluded the first ~150s of the recording because we estimated that more than 5% of spikes were missed (blue symbols and histogram). Among the recordings we accepted, the percentage of missed spikes averaged 0.26%, and almost all neurons showed fewer than 1% missed spikes (Figure 2I).

As extracellular signals are important inputs to our classifier, stringent criteria are essential to ensure reliable cell-type identification. In addition to the quality criteria described above, we took steps to ensure uniform and appropriate preprocessing of extracellular data (Figure S1) and optimize temporal alignment of individual action potentials (see STAR Methods). The requirement for a few violations of the refractory period and small numbers of missed spikes ensured that the units we accepted had high signal-to-noise ratios. The mean signal-to-noise ratio on the channel with the largest unit potential was 9.3 in our accepted sample (almost identical to that of the example recording in Figure 2G). Over 90% of the neurons had signal-to-noise ratios larger than 4 (Figure 2J). Our criteria were designed to explicitly exclude any neural units with action potentials that might come from a different neuron, be noise, or represent the superposition of two or more neurons.

Combination of optogenetics and pharmacology for ground-truth cell-type identification

To allow cell-type identification by photostimulation,[20] we introduced Neuropixels probes into the cerebellum of mice expressing opsins (usually Channelrhodopsin-2, ChR2, with some use of inhibitory opsins in GABAergic MLIs; see STAR Methods) in specific cell types. We performed optogenetic stimulation in multiple phases, including before, during, and after pharmacological application of synaptic blockers (Figure 3A): *baseline*, *control*, *infusion*, and *blockade* (details in STAR Methods).

We accepted neurons as activated directly by photostimulation only if we had strong evidence that they were within the region of synaptic blockade and they continued to have reliable and short-latency responses to light (response latency <10 ms, Figure S2A). By contrast, we excluded neurons when ground truth could not be established, either because they had a long latency response to optogenetic stimulation ($\geq$10 ms), they lost their response in the presence of synaptic blockers, or we could

analysis. Symbol color indicates spikes that came from intervals that had too many missed spikes (blue), acceptable isolation (green), and too many RPVs (red). Marginal histograms on the right show the distribution of spike amplitudes to document clipping at the noise level in the blue histogram that would be cause for rejection of a time interval.

(G) Example recording traces and spatial footprint of a representative recording with a signal-to-noise ratio (SNR) of 9.33, with the waveforms numbered according to their channel. Asterisk (*) denotes the channel with the largest peak-to-trough amplitude, used to compute the SNR.

(H) Distribution of the percentage of RPVs across neurons accepted to the ground-truth library.

(I) Distribution of estimates of the percentage of spikes that were missed across neurons accepted to the ground-truth library.

(J) Distribution of SNRs on the channel with the largest-amplitude waveform across neurons accepted into the ground-truth library.

See also Figure S1.

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

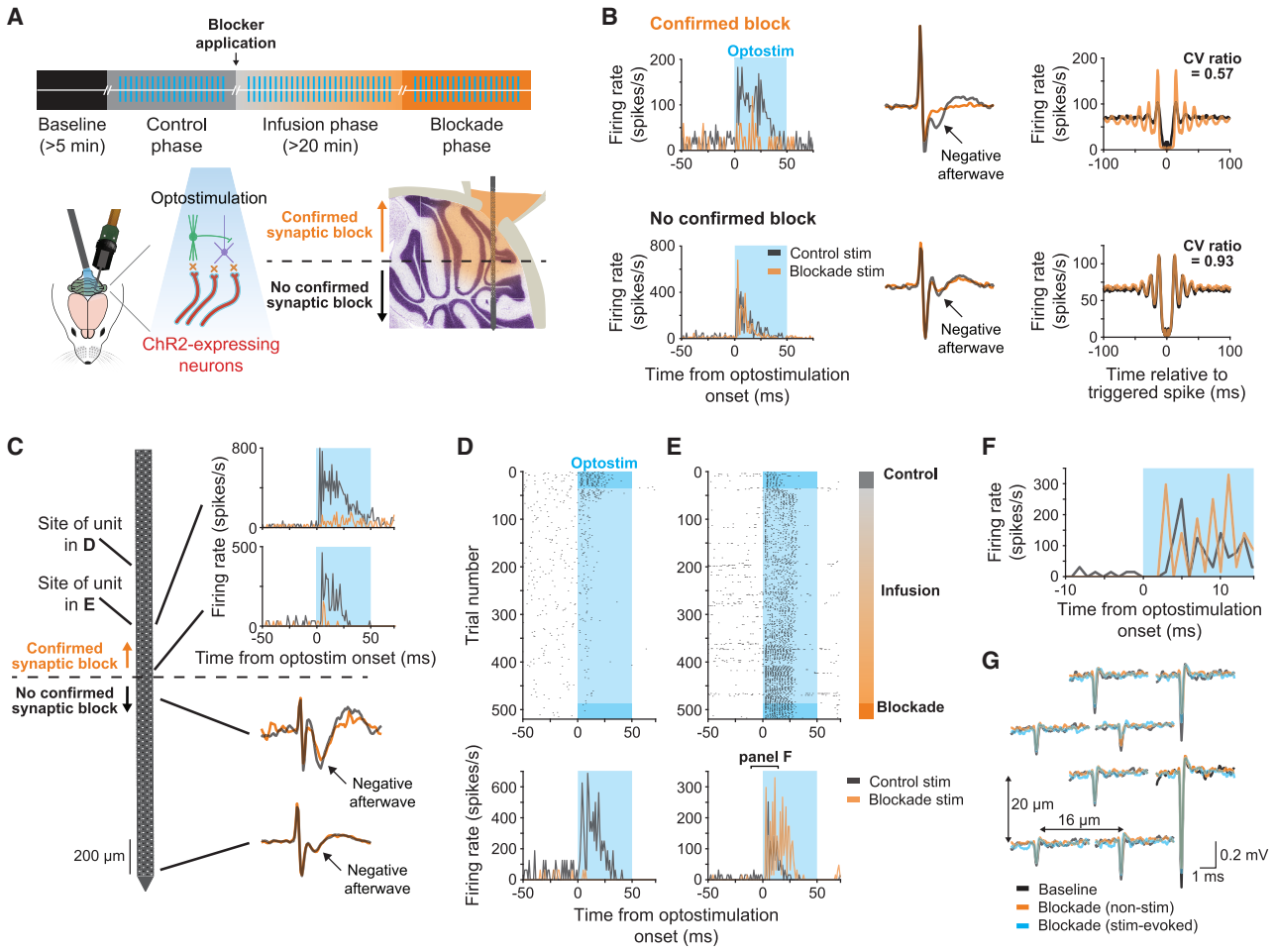


Figure 3. Strategy for ground-truth identification of cell type

(A) Schematic showing the phases in experiments to test for optogenetic activation in the presence of synaptic blockade.

(B) Examples from 6 different units of the signs used to determine the region of effective synaptic blockade. From left to right, the example units in each row demonstrate the effect of blockade on the response to optogenetic stimulation, the negative afterwave of a mossy fiber waveform, and the discharge statistics defined by autocorrelograms and the value of CV.

(C) Example of how we determined whether recordings were within the region of synaptic blockade. Along the Neuropixels probe, the top histograms on the right show sites that were within the region of blockade because units recorded there lost their responses to optogenetic stimulation, and the lower waveforms show mossy fibers that were below the region of blockade because they retained their negative afterwaves.

(D) Raster and peri-stimulus time histogram for a neuron that lost its response to optogenetic stimulation with synaptic blockade. Black versus orange histograms show responses before versus during synaptic blockade. Blue shading indicates the time of photostimulation.

(E) Same as (D), except for a neuron that retained its response to optogenetic stimulation during synaptic blockade.

(F) Fast timescale records showing the short-latency responses to optogenetic stimulation of the neuron in E in both the control phase and the blockade phase.

(G) Spatial footprint of the neuron in (D). Black, orange, and blue traces show waveforms recorded during the baseline period, synaptic blockade without optogenetic stimulation, and synaptic blockade with optogenetic stimulation. The waveforms are spaced according to the relative locations of their contacts, and the two double-headed arrows indicate the horizontal and vertical spacing of the contacts on the Neuropixels probes.

See also Figures S2 and S3.

not confirm that synaptic blockade was effective at the location the neuron was recorded. Synaptic blockade at a given recording depth was deemed effective only if we could record neurons at or below that depth showing at least one of the 3 signs illustrated in the top row of Figure 3B:

1. Loss during the *blockade phase* of responses to optogenetic stimulation present in the *control phase* (Figure 3B, top row, left).
2. Putative mossy fibers with loss of negative afterwaves in the *blockade phase* (Figure 3B, top row, middle). Prior evidence showed that the negative afterwave is a postsynaptic response of granule cells.[45,46]
3. Substantial changes in a neuron's autocorrelogram or coefficient of variation (CV), usually due to an increase in regularity caused by a shift from synaptically and intrinsically driven spiking to purely intrinsically generated spiking[29] (Figure 3B, top row, right).

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

To illustrate our strategy, we present one example experiment in a transgenic mouse line that expressed ChR2 in mossy fibers[35,47] (Thy1-ChR2 line 18). Figure 3C shows the probe location of six neurons that were good candidates for ground-truth mossy fibers because they were photoactivated with a short latency in the control phase. We excluded three of the six neurons from the ground-truth library because they lost their responses to optogenetic stimulation during the blockade phase (Figures 3C and 3D). Two of the neurons had putative mossy fiber waveforms (Figure 3C), but we also excluded them because synaptic blockade could not be confirmed at their location: they retained the negative afterwave during the blockade phase. Only one of the six neurons satisfied all the requirements necessary to be classified as a ground-truth mossy fiber: it responded with short latency to optogenetic stimulation in both the control and blockade phases (Figures 3E and 3F), and it was located in a region of effective synaptic blockade, as confirmed by the loss of photostimulation responses during the blockade phase in two neurons that were recorded at the same location or just below (Figure 3C). The extracellular waveforms of this neuron were constant across the entire experiment (Figure 3G), indicating a stable recording. We present a quantitative analysis of synaptic blockade across all recordings in Figures S2B–S2F.

Some of the mouse lines we used were less specific than the Thy1-ChR2 line in the example of Figures 3C–3G, a complication that could lead to direct photoactivation of multiple cell types and misidentification of some of our recorded neurons. Figure S3 illustrates how we combined identification of the recording layer based on Phyllum with optogenetic activation to establish cell type unambiguously, even in the GlyT2-Cre line that we chose to label Golgi cells[33] where the problem of off-target expression was most pronounced.

The ground-truth library

Across 188 Neuropixels recordings in two laboratories, we recorded a total of 3,652 neurons that survived the spike-sorting and curation pipeline (Figure 4A). Most failed our criteria for inclusion in the ground-truth library (details in figure legend): only 97 retained responses in the presence of synaptic blockade (e.g., Figure 3E, fast timebase records in Figure 3F; latency distributions in Figure S2A). An additional 28 neurons lost their responses during synaptic blockade and therefore were driven with short latencies (<10 ms) by indirect, synaptic activation (e.g., Figure 3B, top: 5 ms, both neurons in Figures 3C: 3 ms and 3D: 2 ms; latency distributions in Figure S2A). To the 97 directly activated neurons, we added the simple and complex spikes of 62 Purkinje cells identified by a complex-spike-triggered pause in simple spikes and removed 5 units that failed final objective quality checks and 14 with ambiguous layer assignment.

The neurons in the ground-truth library (right side of Figure 4A) all were well isolated. They had more than 95% uncontaminated spikes (Figure 4B), and almost all had close to 100% by a metric detailed in the STAR Methods.[48] For comparison with previous reports,[19,49–53] Table S1 reports the electrophysiological properties of different cell types in our ground-truth library using a range of metrics.

The proportion of units in the ground-truth library is skewed toward Purkinje cells (Figure 4C), and we compare that distribution with better estimates of recording probability in Table S2. We had very limited success in recording from granule cells even though we made 82 recordings in mice with opsins expressed in granule cells (see STAR Methods). We did record multiple possible granule cell units with activity that was responsive to photostimulation in the region of confirmed synaptic blockade, but nearly all units failed one or more of our criteria for good isolation. After curation, we retained 3 units from 82 recordings (0.04 granule cells per recording). Our sample is too small to include them in the classifier we will develop next. A combination of factors may contribute to the inability to record regularly from granule cells: their comparatively small size,[54,55] the spatially restricted closed-field of their extracellular potential, and the low electrode impedance[49] of Neuropixels[16] (150 kOhms).

Armed with a ground-truth dataset, we next developed an accurate classification method based on consistent differences in electrophysiological features across cell types.[56] We used both waveform[19,57,58] and discharge statistics[50–52] as features for cell-type classification.

Waveform

Patch-clamp recordings *in vitro* confirmed our expectation that the different biophysical properties and morphology of different cell types would lead to different waveforms[57–60] (Figure S4). Yet, normalized waveforms have much more variable shapes in extracellular recordings *in vivo* (Figure 4D) than in patch-clamp recordings *in vitro* (Figure S4C). We attribute the higher variation in waveforms in extracellular recordings to the fact that the *in vitro* recordings are made under highly uniform conditions with the electrode in direct contact with the cell membrane. By contrast, extracellular recordings *in vivo* vary in distance from the recording contact to the neuron, the orientation of the electric field relative to that of the probe, and background noise. Further, the shapes of the waveforms were very different *in vitro* versus *in vivo*. Thus, a cell-type classifier trained on the waveforms recorded *in vitro* was unable to classify neurons in the *in vivo* ground-truth library (Figure S4F), meaning that we could not have supplemented our ground-truth library with waveforms recorded *in vitro*.

Discharge statistics

It is common for different cell types to have different discharge statistics throughout the brain.[61,62] In the ground-truth library, discharge statistics estimated by autocorrelograms varied across cell types (Figure 4E). We did not include discharge statistics of neurons recorded in anesthetized animals or *in vitro* because they are not representative of the awake state.[19,63]

Cell-type identification from a semi-supervised deep learning classifier

Figure 4F and Table S1 reveal that it is difficult to guess which specific measures of waveform and firing statistics would be most informative to successfully distinguish cell types in awake animals. Instead, raw data (1) contain richer information, (2) provide unbiased inputs for cell-type identification, and (3) are likely to generalize across regions, tasks, and species. Further, we used "three-dimensional autocorrelograms" (3D-ACGs,

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

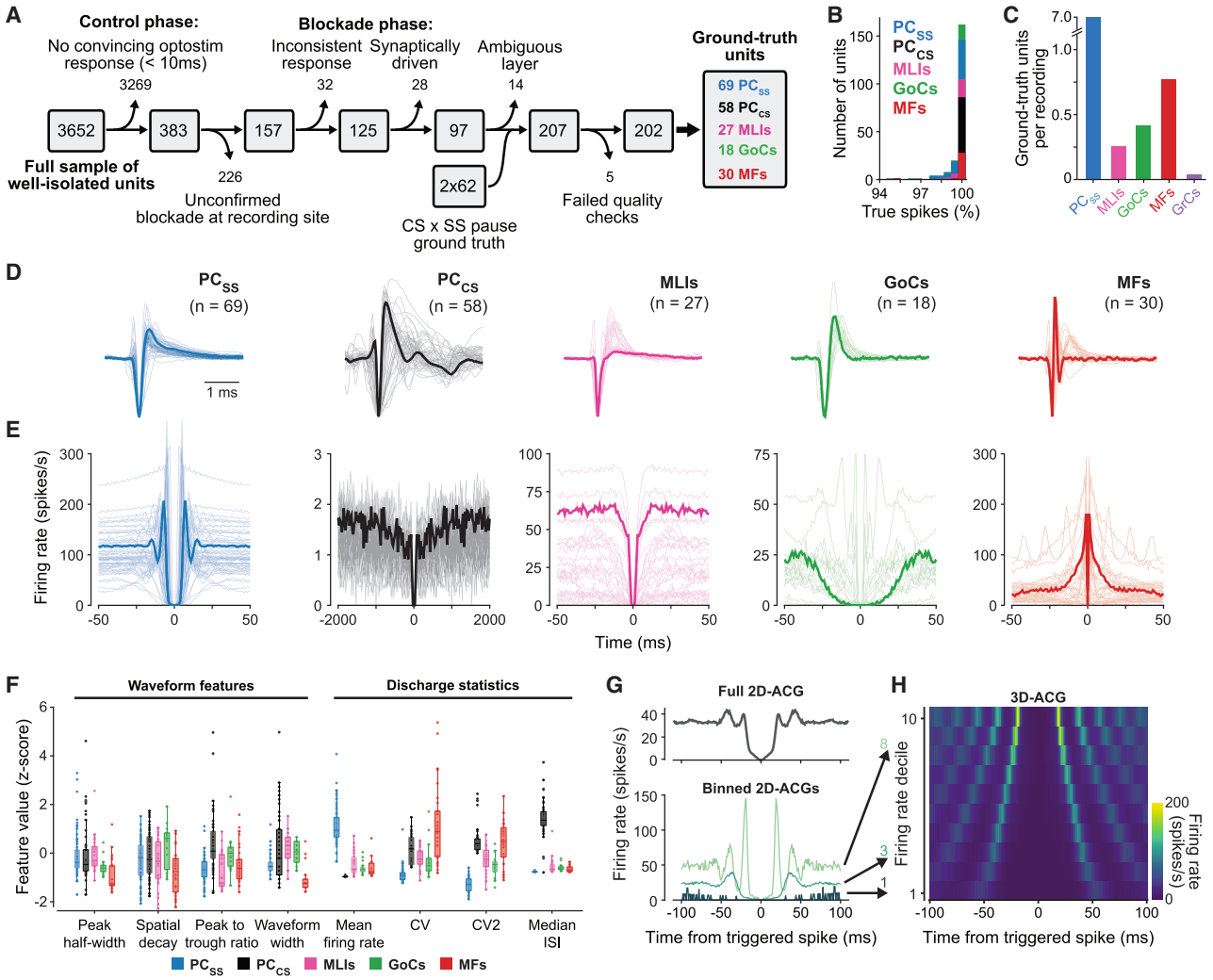


Figure 4. Selection criteria and properties of the ground-truth library of cerebellar cell types

(A) Curation process to create the ground-truth library, including the numbers that were retained or deleted at each stage of the curation.

(B) Distribution of the estimates of the % of spikes that actually belong to a given neuron, segregated according to color by cell type. See STAR Methods for computation of % true spikes. Abbreviations are PC_{SS}, Purkinje cell simple spikes; PC_{CS}, Purkinje cell complex spikes; MLIs, molecular layer interneurons; GoCs, Golgi cells; MFs, mossy fibers; and GrCs, granule cells.

(C) Histogram showing the number of ground-truth units of each cell type normalized for the number of recordings. Abbreviations as in (B).

(D) Superimposed waveforms for each cell type in the ground-truth library. Abbreviations as in (B). The bold traces indicate the neurons in example 3D-ACGs in Figure S7C. Waveform amplitudes are normalized and flipped to ensure the largest peak is negative. Note that flipping causes some mossy fibers (MFs) to possess a negative afterwave that is a positive deflection (as in Figure 3B). Other MFs do not show a negative afterwave at all, probably because of variability in the relative location of the probe contact and the MF. In general, the negative afterwave appeared on only one contact, as might be expected given that the negative afterwave is thought to reflect the postsynaptic granule cell field potential and therefore should occur only on a channel more proximal to the glomerulus. Also note the two distinct waveforms for molecular layer interneurons.

(E) Same as (D), but showing autocorrelograms of ground-truth neurons.

(F) Failure of traditional measurements of waveform or discharge statistics to differentiate cell types. Each symbol shows *Z* scored values of different features from a single neuron, and different colors indicate different cell types, per the key at the bottom of the plot. *Z* scores were computed separately for each feature but across cell types within each feature. Abbreviations as in (B).

(G) 2D-ACGs without regard for firing rate (top) and with spikes segregated according to instantaneous firing rate (bottom). The 2D-ACG (top graph) created without regard for firing rate has multiple shoulders at irregular intervals, whereas the 3 ACGs created for different mean firing rates (bottom graph) are more regular.

(H) 3D-ACG created from the 2D-ACGs in G to normalize for variations in firing rate. Arrows show the destination of each 2D-ACG. The 3D-ACG plots 10 rows that contain 2D-ACGs for 10 different deciles of mean firing rate.

See also Figure S4.

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

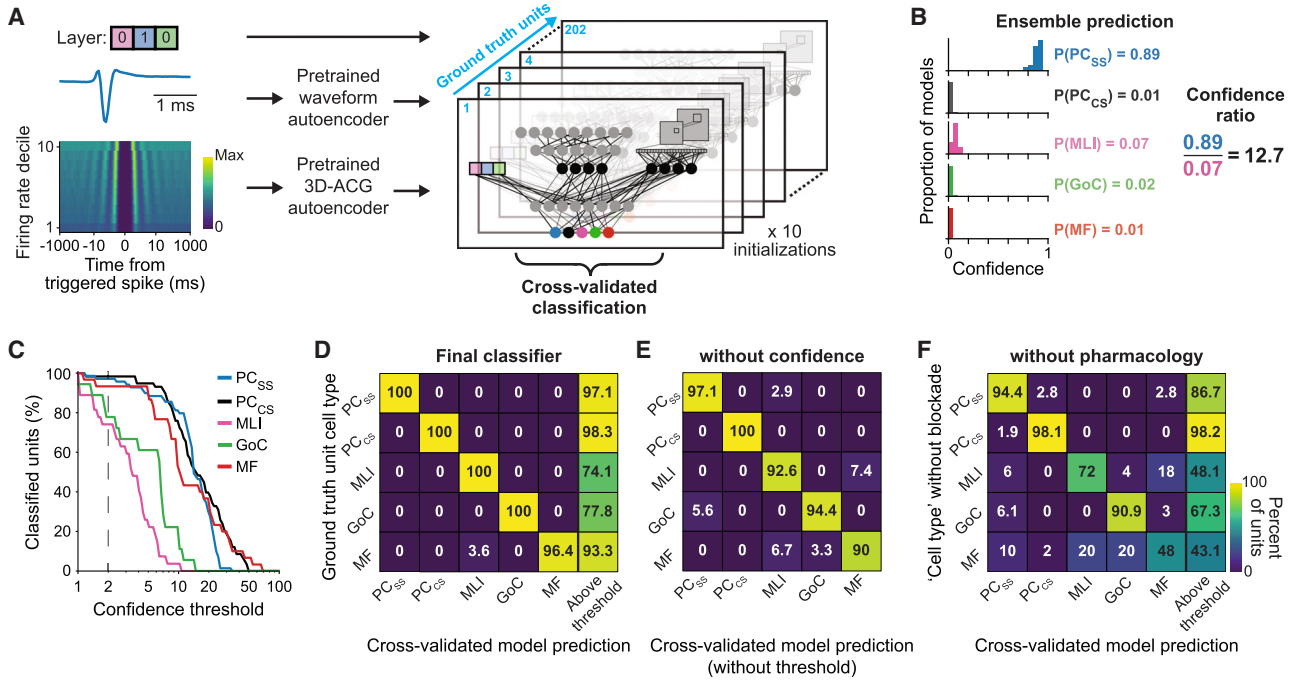


Figure 5. Performance of a deep learning classifier on cell-type identification for the ground-truth library

(A) Classifier architecture. We trained the classifier with 10 different initializations for each of the 202 ground-truth units, symbolized by the 202 pages in the classifier.

(B) Histograms showing the predictions of the classifier for 10 different initial conditions for one left-out neuron.

(C) Percentage of units classified as a function of the ratio we chose as a confidence threshold. Different colors show data for different ground-truth cell types.

(D) Confusion matrix showing the ground-truth cell type of a left-out testing unit as a function of the predictions of the classifier on that unit. The numbers in each cell indicate the percentage of cell-type predictions for each ground-truth cell type. The rightmost column shows the percentage of ground-truth neurons that received a confidence greater than 2, required for inclusion in the confusion matrix.

(E) Same as (D), but with confidence threshold = 0.

(F) Same as (D), but classifier trained on all optogenetically activated units without requiring synaptic blockade.

See also Figure S5 and S6.

Figures 4G and 4H, details in figure legend) to quantify discharge statistics in a way that normalizes for variations in firing rate across a recording session and between species.

Our classifier (Figure 5A) consists of (1) a multi-headed, normalized input layer that accepts the 10-dimensional outputs from 2 pretrained variational autoencoders, one for the 3D-ACGs and one for the full time course of the average optimally aligned, amplitude-normalized waveform (see STAR Methods) on the channel with the largest signal, along with a "one-hot" 3-bit binary code of the unit's cerebellar layer; (2) a hidden layer that processes the 3 inputs; and (3) an output layer with one output unit for each of the 5 cell types. The value of the output units sums to 1 so that the output of the classifier is the probability that a given set of inputs is from each of the 5 cell types. We trained the weights in the classifier on the data in the ground-truth library using gradient descent with leave-one-out cross-validation.

Before optimizing the classifier, we trained the variational autoencoders on large samples of unlabeled data to reduce the high dimensionality of the waveforms and 3D-ACGs to 10 dimensions. This and other steps (see Figure S5 and STAR Methods) helped to mitigate overfitting, given the relatively small number of training examples in the ground-truth library.

We evaluated not only the accuracy but also the "confidence" of the output from the classifier. For each leave-one-out sample (n = 202 ground-truth units), we averaged the cell-type probabilities across an ensemble of 10 models with random initial conditions. In Figure 5B, the classifier repeatedly predicted that the held-out unit was a Purkinje cell simple spike (average probability 0.89). We quantified the classifier's confidence with the "confidence ratio," computed as the ratio of the mean probability of the most likely cell type to the mean probability of the second most likely cell type. The percentage of ground-truth units that could be classified decreased as a function of the confidence ratio we chose as the confidence threshold (Figure 5C). Classifier confidence was higher for Purkinje cell simple and complex spikes and for mossy fibers compared to Golgi cells or MLIs. We chose a confidence threshold of 2 in the remainder of our analysis because it provided a compromise between classifying as many neurons as possible and classifying them accurately.

The classifier showed high accuracy when applied to the units in the ground-truth library. For each held-out neuron that exceeded the confidence threshold, we assigned it the cell type with the highest average probability across the 10 classifier runs. For 74% and 78% of ground-truth MLIs and Golgi cells, confidence exceeded threshold, and the classifier assigned

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

them correctly, i.e., values of 100% along the diagonal of the confusion matrix (Figure 5D). More than 90% of mossy fibers, Purkinje cell simple spikes, and complex spikes exceeded confidence threshold, and nearly all were identified correctly. The accuracy of the classifier degraded without a confidence threshold but still exceeded 90% accuracy on all cell types (Figure 5E). The classifier performed less well without layer information, mainly because layer information allowed it to better distinguish between Golgi cells and certain MLIs (Figure S6). Finally, classifier performance degraded if we trained it on all optogenetically activated units without regard for successful synaptic blockade (Figure 5F). Thus, it was important to include in the ground-truth library only those cells that passed the requirement of persistent optogenetic activation during confirmed synaptic blockade.

Classifier validation of expert-labeled datasets

We next evaluated how well the ground-truth classifier generalized by predicting the cell type for a sample of expert-classified, non-ground-truth neurons recorded from mice and monkeys.

Confidence is a particularly important metric for non-ground-truth data. We took advantage of the 2020 versions of our classifier that emerged from training 202 ground-truth units by 10 instantiations of the classifier per unit. We averaged the probabilities for each unit in the expert-classified datasets across the 2020 classifiers and plotted the probability assigned by the classifier as a function of the 5 possible cell types (Figures 6A and 6B). Units appear in exactly one of 5 different plots, namely the one for cell type assigned by the classifier as the highest probability. For example, the leftmost graph reports probability versus cell type for all units that were classified as most probable to be simple spikes of Purkinje cells. The ground-truth classifier agreed with the human experts about the cell types of almost all units that were above confidence threshold in both mice and monkeys (Figures 6C and 6D). Further, it identified correctly the mouse and monkey Purkinje cell simple spikes and complex spikes from recordings with a complex-spike-triggered pause in simple spikes (Figure 6E).

Similar properties within cell types across species and cerebellar regions

Three additional analyses indicate that the success of the ground-truth classifier on the expert-classified data is based on statistical similarity of the waveforms and firing statistics of each cell type across datasets. First, the percentage of units that were classified with confidence decreased similarly as a function of the confidence threshold for the two samples of expert-classified cells and the ground-truth dataset (Figure 6F, thick gray and black traces versus colored traces). Second, analysis of the output of the classifier's autoencoders revealed excellent agreement between the reduced-dimension representation of expert-classified and ground-truth data (Figure 6G). Third, inspection of the waveforms, 2D-ACGs, and 3D-ACGs reveals impressive similarity across the ground-truth data, the non-ground-truth mouse data, and the monkey recordings (Figure S7).

Functional dissection of cerebellar circuits enabled by cell-type identification

Different neuron types often exhibit distinct temporal response dynamics during behavior. Our four labs recorded in four different areas of the cerebellum known to be engaged during our respective behaviors (Figure 7A): reward conditioning in the lateral cerebellum of mice,[64] eye blink conditioning in the deepest regions of paravermis lobules HV/HVI of mice,[65,66] locomotion in the simplex lobule of mice,[67] and smooth pursuit eye movements in the floccular complex of monkeys.[68] Different temporal responses across cell types appear in traditional trial-averaged peri-stimulus time histograms of each cell type (Figure 7B). Most cell types showed relatively small variation across individual units (error bands in Figure 7B), suggesting functional homogeneity within cell types. At the same time, some cell types in some areas demonstrated multiple functional discharge patterns (e.g., Purkinje cells during eyeblink conditioning or MLIs during smooth pursuit) (Figure 7C).

To reveal whether neural dynamics of different cell type-specific populations are distinguishable from each other, we performed dimensionality reduction (Figure 7D) on the neural state-space for trial-averaged data collected from a single session's recordings in the 3 mouse preparations and a larger pseudo-population recorded across many sessions in monkeys.[69–74] We optimally aligned the principal components computed separately for each cell-type population with those for the full "cell-type-agnostic" population (see STAR Methods).

The trial-averaged neural trajectories of different cell types showed statistically different dynamics from each other and from the cell-type-agnostic population during reward conditioning and smooth pursuit but not eyeblink conditioning or locomotion (Figure 7D). Two statistical analyses verified that any difference in the dynamics between cell types was not simply the result of unequal or limited population sizes (Figure 7E) or gross misidentification of cell types by the classifier (Figure 7F). Failure of the trajectories of different cell types to diverge in two of the behaviors we studied does not necessarily imply that all cell types had the same dynamics. Rather, it could result from the diversity of dynamics within a cell type that matches or exceeds the diversity across cell types, or from the existence of 2 or more subgroups within each cell type. Further clustering within cell types based on functional properties would be required to disentangle functional subclasses. Together, the analyses in Figure 7 show how cell-type identification can provide biological insight about how interactions among different cell types allow neural circuits to control behavior.

DISCUSSION

Identification of cell type from *in vivo* extracellular recordings is a fundamental requirement in systems neuroscience.[5,6,19,50,51,57,62,75–79] Our approach delivers a highly reliable ground-truth library of the electrophysiological properties of cerebellar cell types in awake mice based on identification through optogenetic stimulation in the presence of synaptic blockers. The ground-truth library consists of the neuron's waveform, statistics of the spike train, and the layer of the cerebellum where we recorded each unit. Our semi-supervised deep learning classifier identifies the cell types

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

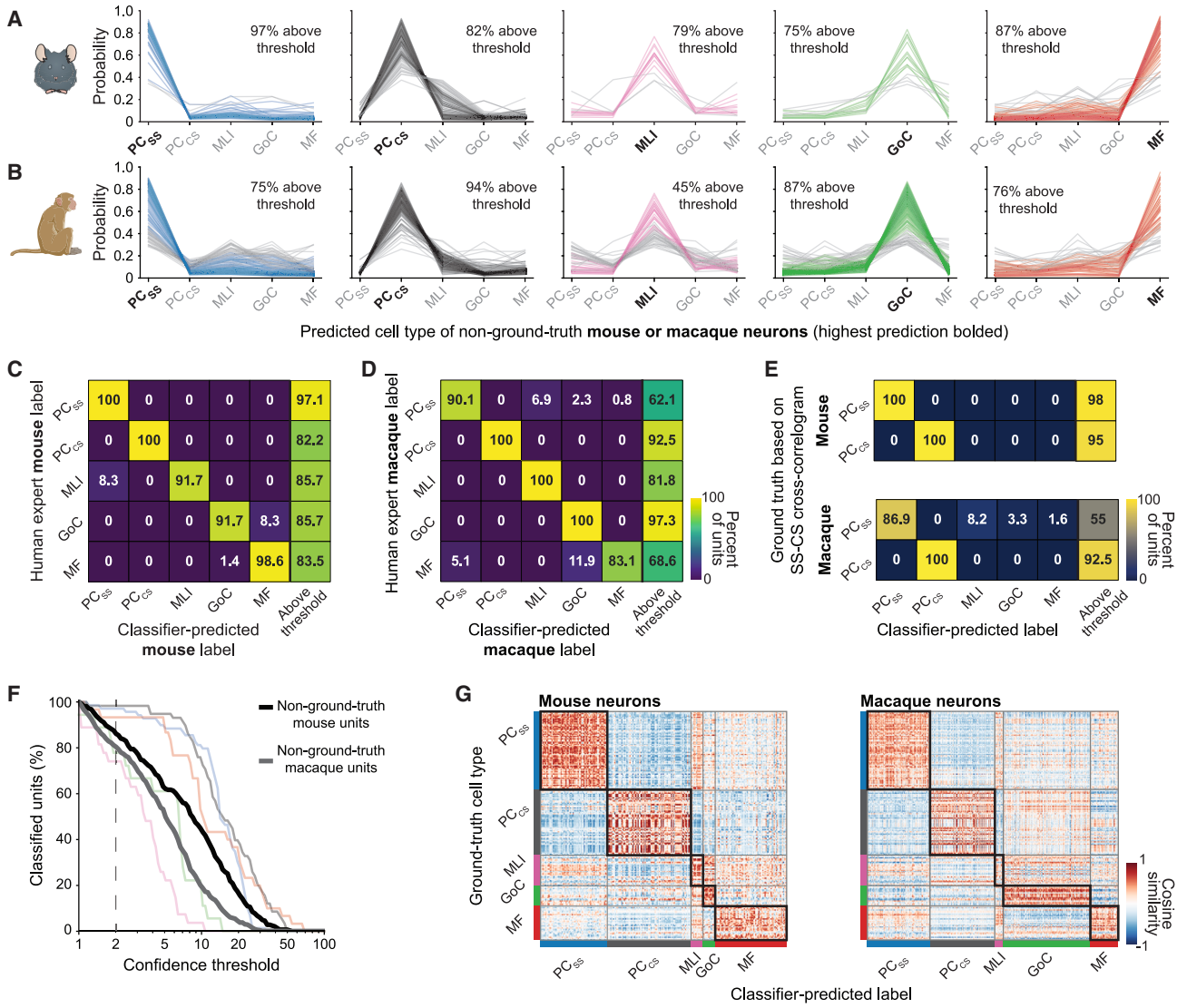


Figure 6. Ground-truth classifier performance on expert-classified datasets from mice and monkeys

(A) Probability assigned by the classifier as a function of cell type for expert-classified neurons from mice, with a separate graph for each cell type assigned the highest probability by the classifier. From left to right, the highest-probability cell type was a Purkinje cell simple spike (PC_{SS}), Purkinje cell complex spike (PC_{CS}), molecular layer interneuron (MLI), Golgi cell (GoC), and mossy fiber (MF). Colored versus gray traces represent neurons that exceeded versus failed the confidence threshold of 2.

(B) Same as (A), but for expert-classified neurons from monkey floccular complex.

(C) Correspondence matrix comparing the predictions of the classifier with expert-labeled cell type from recordings in mice. The numbers in each cell indicate the percentage of expert-classified cell types on the y axis as a function of the cell type predicted by the classifier on the x axis. The rightmost column shows the percentage of expert-classified neurons that received a confidence greater than 2 from the classifier.

(D) Same as (C), for expert-classified neurons from monkey floccular complex.

(E) Confusion matrices showing good agreement between the output from the classifier and the ground-truth identification in mice and monkeys of Purkinje cell simple and complex spikes from the presence of a complex-spike-triggered pause in simple-spike firing.

(F) Comparison of percentage of classified units as a function of confidence threshold for 3 preparations. Faint-colored traces show the same curves for the ground-truth library, copied from Figure 5C. Bold black and gray traces show results for the non-ground-truth units in mouse and monkey, respectively.

(G) Congruence of the output from the autoencoders for ground-truth versus expert-classified neurons across preparations. Each row corresponds to a single ground-truth-identified neuron. Each column corresponds to a single classifier-identified neuron from mouse (left) or monkey (right). Colors at the intersections for each row and column indicate the cosine similarity of the concatenated outputs from the autoencoders for waveform and autocorrelograms, where redder colors indicate greater similarity.

See also Figure S7.

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

CellPress

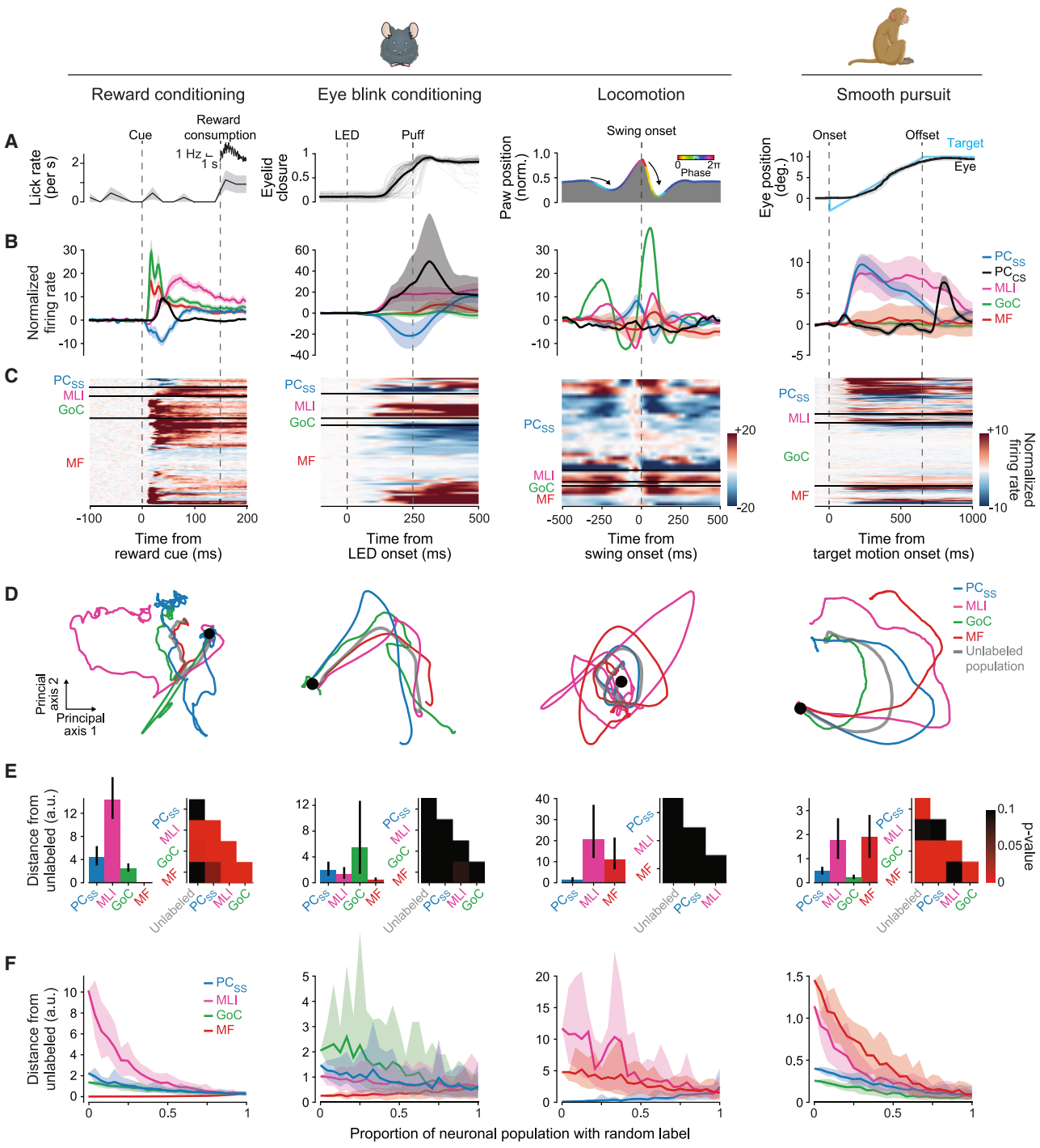


Figure 7. Neural population dynamics of different cell types

(A) Temporal profile of motor responses in four behavioral tasks in 4 different labs and 2 species. From left to right: licks per second when a reward is cued by a solenoid click (mouse), and inset shows lick rate on an extended time scale; eyelid closure after eye blinks have been conditioned by pairing an LED and an air puff (mouse); paw position aligned on the onset of swing phase during locomotion (mouse); and eye position during pursuit of smooth target motion (monkey).

(B) Average firing rate of different cell types during the corresponding behaviors. Different colors indicate normalized responses for the different cell types. Error bands show mean ± SEM across neurons.

(C) Neural firing rates divided according to cell type. Colors in each line show the firing rate of individual neurons as a function of time during the four behaviors, normalized so that the standard deviation in the baseline period was 1.

(legend continued on next page)

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

in the ground-truth library accurately while also reporting its confidence in each identification. The cell types predicted by the classifier for recordings from mouse and monkey agree with experts' assessments. The success of the classifier is unexpected evidence that the properties of different cerebellar cell types are consistent across species and cerebellar regions. We are encouraged by the accuracy and precision of our classifier and expect that it will be possible in the future to align the cell type obtained from extracellular recordings with that obtained from other levels of analysis, including anatomical and molecular fingerprints.

Cell-type identification enables important biological insights about circuit organization, circuit dynamics, and the relationship between output dynamics and behavior. For two of the behaviors we study, different cell types showed very different population dynamics, suggesting distinct roles for different cell types in cerebellar processing and circuit computation. For the other two behaviors, population dynamics were shared among all the different cell types, perhaps as a result of the highly heterogeneous neural response profiles within each cell class that underlie these behaviors. Cell-type identification opens avenues to use multiple different strategies for biological insights that will emerge from future analyses. For example, the combination of cell-type identification and simultaneous recordings should allow direct measurements of connectivity strengths and circuit function using cross-correlogram analysis, as well as localization of multiple sites of learning in a single circuit. Alternatively, analysis of the dynamics[70–72,74,80] in different cell types could reveal population dynamics related to motor behavior in output neurons and to context in specific types of interneurons.

Past attempts to identify discrete populations of cerebellar cortical neurons[19,50–53,81] were less successful than ours. Previous studies identified neurons either (1) by qualitative agreement with spiking signatures found *in vitro* or (2) agreement with recordings in anesthetized preparations where neurons were identified anatomically via juxtacellular labeling.[19,49–51,53,81] Our recordings in awake, behaving mice demonstrate large variance in the discrete metrics used by past attempts for summarizing spiking activity both within and across ground-truth classes (Figure 4F; Table S1). Thus, classification schemes reliant on a finite set of specific features are unlikely to generalize well to other tasks or regions,[57] or from anesthetized to behaving preparations.[19]

Several features of the strategy in our classifier were critical to its success:

- Raw waveforms and 3D-ACGs: Raw features are an unbiased input[57] and allow the classifier to take advantage of extensive information in waveform[57,58] and discharge statistics. 3D-ACGs normalize for variations in firing rate and create a statistic that can be compared across cerebellar areas, experimental tasks, and species. Use of single-channel waveforms (rather than spatial footprints) allows the classifier to generalize across electrode types.
- Mitigation of overfitting: A semi-supervised[82–84] deep learning strategy trained the classifier with a relatively small number of ground-truth neurons while reducing the chances of overfitting[85] through the use of a large unlabeled dataset to pretrain the autoencoders (see STAR Methods and Figure S5). Successful predictions of cell types that agree with two expert-classified datasets support the generalizability of the classifier.
- Confidence: We were particularly cognizant of making our classifier trustworthy. We established confidence by training multiple models on the same data.[86] By requiring confidence above a given threshold,[87,88] we improved the accuracy of the model on the ground-truth data as well as for non-ground-truth recordings. Choice of the confidence threshold allows the user to balance whether to include all neurons even if some cell-type assignments might be incorrect or to include fewer neurons with greater certainty in the cell-type assignment.

When it included layer information as an input, the classifier identified units with greater accuracy. However, the use of a layer as an input does not make classification trivial. Rather, it creates a platform that will become even more useful as we achieve ground-truth identification of other cell types in the cerebellum, for example, of granule cells with improved recording probes. Also, because waveform and firing statistics are necessary to distinguish cells that are in the same layer, the classifier makes a statistical decision about cell type rather than relying solely on layer for cell-type identification.[89] Layer is defined in a specific way for the cerebellum,[1] but we think of layer information more generally as a specific example of "local electrical properties." We imagine that there are other ways to quantify those properties, for example, LFPs and current-source-density analysis,[90,91] that will work in brain areas without a laminar structure.

The strategy we developed may be more useful and important than the exact classifier. Our goal at the outset of our project was to achieve cell-type identification from extracellular recordings in the cerebellar cortex across laboratories and species. We think that the classifier can be used with confidence by any cerebellar recording lab that is curating their electrophysiological data with sufficient rigor. However, a failure of rigorous curation will lead to noisy and unnecessarily variable inputs to the classifier and will contaminate its cell-type identifications.[58] The larger challenge is the deployment of our strategy in other brain areas. The use of layer information to improve classification should be relevant

(D) Two-dimensional trajectories of population dynamics. The trajectories start at the black, filled circle, and the gray trajectory shows the full, cell-type-agnostic population, and the colored trajectories show different cell types.
(E) Statistical analysis of the differences among the dynamic population trajectories of different cell types. For each behavior, the histogram shows the distance from each cell type's trajectory to the cell-type-agnostic trajectory. The half-matrices summarize the p values for the comparison of all trajectories with each other. Black squares are not statistically significant.
(F) Effect of re-labeling different fractions of each cell-type population randomly on the distance between that cell type's dynamic population trajectory and the trajectory for the cell-type-agnostic population.

to other structures—cerebral cortex,[92] hippocampus,[93] and superior colliculus[94]—that have layers with measurable local electrical properties. The strategies used in our classifier enable generalization by showing how to reduce the dimensions of raw input data while mitigating the challenges of small numbers of neurons in training sets. Also, the combination of optogenetic activation with synaptic blockade adds rigor to the creation of a ground-truth library.

Limitations of the study

We identify 4 main limitations. (1) Optogenetic identification is limited to the major cell types in the cerebellar cortex because Cre lines are not currently available to allow us to add some of the recently highlighted cell types[26] to the ground-truth library. (2) It was challenging to record from granule cells, in spite of considerable effort. Advancements in recording probes with higher impedance and/or more dense recording sites will eventually resolve this limitation. (3) We expect that our classifier can be used "out-of-the-box" for cell-type identification in the cerebellum, but our strategy will need to be used to develop ground-truth libraries and successful classifiers for other areas. There cannot be a "general," area-agnostic classifier because the biophysical and firing properties of neurons are fundamentally different in different brain areas.[95] (4) At this time, there is no way to align extracellular electrophysiological cell-type identifications with those provided by single-cell RNA sequencing (RNA-seq),[96–98] juxta-cellular labeling,[50,51] or combinations of *in vivo* recording and single-cell imaging.[99] For example, extracellular electrophysiology cannot "mark" recorded cells for post hoc analysis of molecular identity using approaches such as RNA-seq. Despite its limitations, our study demonstrates a strategy that allows different cell types to be identified robustly and reliably from large-scale electrophysiological recordings in behaving mice and monkeys. The strategy should be of great value to the growing community of cerebellar researchers using high-density silicon probes and provides a template for principled semi-automated detection of cell type that can be applied across other neural circuits in the brain.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources, datasets, and protocols should be directed to and will be fulfilled by the lead contact, Javier F. Medina (jfmedina@bcm.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The data that support the findings of this paper are publicly available as of the date of publication via the DOI listed in the key resources table.
- The original code that supports the findings in this paper has been deposited on GitHub and is publicly available as of the date of publication via the DOI listed in the key resources table and at https://c4-database.com.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank Bonnie Bowell, Soyon Chun, Wenjuan Kong, Margaret Conde Paredes, Caroline Reuter, and Stefanie Tokiyama for technical support; Arnd Roth for helpful discussions; Rob Campbell for supporting the build of the custom two-photon tomography platform; and Julie Fabre for advice on the spike-sorting quality metrics. Funding was provided by NIH grants R01-NS112917 (S.G.L., J.F.M., and C.H.), K99-EY030528 (D.J.H.), R01-NS092623 (S.G.L.), and R01-MH093727 (J.F.M.); ERC AdG 695709 (M.H.); Wellcome Trust PRF 201225/Z/16/Z and 224668/Z/21/Z (M.H.) and Career Development Award 225951/Z/22/Z (D.K.); EMBO ALTF 914-2015 (D.K.); the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement no. 844318 (M.O.) and no. 891774 (F.N.); and the SYNCH project funded by the European Commission under H2020 FET Proactive program-Grant agreement ID 824162 (D.C.).

AUTHOR CONTRIBUTIONS

Conceptualization, M.B., D.J.H., B.A.C., D.C., S.G.L., D.K., C.H., M.H., and J.F.M.; methodology, M.B., D.J.H., F.N., F.D'A., and A.S.-L.; software, M.B., D.J.H., F.N., F.D'A., M.G.M.L., and A.L.; formal analysis, M.B., D.J.H., F.N., M.E.H., F.D'A., M.O., A.S.-L., D.K., M.G.M.L., and N.J.H.; investigation, M.B., D.J.H., M.E.H., M.O., A.S.-L., Y.Y.C., M.M., H.N.S., M.Z., S.O., D.K., and C.H.; writing – original draft, S.G.L.; writing – review & editing, M.B., D.J.H., S.G.L., D.K., C.H., M.H., and J.F.M.; visualization, M.B., D.J.H., M.E.H., D.K., and J.F.M.; supervision. M.B., S.O., B.A.C., D.C., S.G.L., D.K., C.H., M.H., and J.F.M.; funding acquisition, B.A.C., S.G.L., C.H., M.H., and J.F.M. See Table S3 for detailed author contributions.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Ethical statement
  - Animals
- METHOD DETAILS
  - Surgery
  - Extracellular recording procedures
  - Reconstruction of Neuropixels probe tracts
  - Optogenetic stimulation and pharmacology
  - Histological assessment of opsin expression
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Analysis of extracellular recordings
  - Classifier design
  - Evaluation of classifier performance
  - Analysis of neural dynamics during behavior

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.01.041.

Received: May 16, 2024
Revised: November 20, 2024
Accepted: January 28, 2025
Published: February 28, 2025

REFERENCES

1. Ramón y Cajal, S. (1909). Histologie du système nerveux de l'homme & des vertébrés, Ed. française rev. & mise à jour par

l'auteur, tr. de l'espagnol par L. Azoulay. (Maloine) https://doi.org/10.5962/bhl.title.48637.

2. Fishell, G., and Heintz, N. (2013). The Neuron Identity Problem: Form Meets Function. Neuron *80*, 602–612. https://doi.org/10.1016/j.neuron.2013.10.035.
3. Masland, R.H. (2004). Neuronal cell types. Curr. Biol. *14*, R497–R500. https://doi.org/10.1016/j.cub.2004.06.035.
4. Migliore, M., and Shepherd, G.M. (2005). Opinion: An integrated approach to classifying neuronal phenotypes. Nat. Rev. Neurosci. *6*, 810–818. https://doi.org/10.1038/nrn1769.
5. Zeng, H., and Sanes, J.R. (2017). Neuronal cell-type classification: challenges, opportunities and the path forward. Nat. Rev. Neurosci. *18*, 530–546. https://doi.org/10.1038/nrn.2017.85.
6. Luo, L., Callaway, E.M., and Svoboda, K. (2008). Genetic Dissection of Neural Circuits. Neuron *57*, 634–660. https://doi.org/10.1016/j.neuron.2008.01.002.
7. Siletti, K., Hodge, R., Mossi Albiach, A., Lee, K.W., Ding, S.-L., Hu, L., Lönnerberg, P., Bakken, T., Casper, T., Clark, M., et al. (2023). Transcriptomic diversity of cell types across the adult human brain. Science *382*, eadd7046. https://doi.org/10.1126/science.add7046.
8. Hodge, R.D., Bakken, T.E., Miller, J.A., Smith, K.A., Barkan, E.R., Graybuck, L.T., Close, J.L., Long, B., Johansen, N., Penn, O., et al. (2019). Conserved cell types with divergent features in human versus mouse cortex. Nature *573*, 61–68. https://doi.org/10.1038/s41586-019-1506-7.
9. Poulin, J.-F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J.M., and Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. Nat. Neurosci. *19*, 1131–1141. https://doi.org/10.1038/nn.4366.
10. Sugino, K., Hempel, C.M., Miller, M.N., Hattox, A.M., Shapiro, P., Wu, C., Huang, Z.J., and Nelson, S.B. (2006). Molecular taxonomy of major neuronal classes in the adult mouse forebrain. Nat. Neurosci. *9*, 99–107. https://doi.org/10.1038/nn1618.
11. Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L., et al. (2008). Application of a Translational Profiling Approach for the Comparative Analysis of CNS Cell Types. Cell *135*, 749–762. https://doi.org/10.1016/j.cell.2008.10.029.
12. Huang, Z.J., and Zeng, H. (2013). Genetic Approaches to Neural Circuits in the Mouse. Annu. Rev. Neurosci. *36*, 183–215. https://doi.org/10.1146/annurev-neuro-062012-170307.
13. Fujita, H., Kodama, T., and du Lac, S. (2020). Modular output circuits of the fastigial nucleus for diverse motor and nonmotor functions of the cerebellar vermis. eLife *9*, e58613. https://doi.org/10.7554/eLife.58613.
14. Ecker, J.R., Geschwind, D.H., Kriegstein, A.R., Ngai, J., Osten, P., Polioudakis, D., Regev, A., Sestan, N., Wickersham, I.R., and Zeng, H. (2017). The BRAIN Initiative Cell Census Consortium: Lessons Learned toward Generating a Comprehensive Brain Cell Atlas. Neuron *96*, 542–557. https://doi.org/10.1016/j.neuron.2017.10.007.
15. Bota, M., and Swanson, L.W. (2007). The neuron classification problem. Brain Res. Rev. *56*, 79–88. https://doi.org/10.1016/j.brainresrev.2007.05.005.
16. Jun, J.J., Steinmetz, N.A., Siegle, J.H., Denman, D.J., Bauza, M., Barbarits, B., Lee, A.K., Anastassiou, C.A., Andrei, A., Aydın, Ç., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. Nature *551*, 232–236. https://doi.org/10.1038/nature24636.
17. Steinmetz, N.A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., et al. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. Science *372*, eabf4588. https://doi.org/10.1126/science.abf4588.
18. Haar, S., Givon-Mayo, R., Barmack, N.H., Yakhnitsa, V., and Donchin, O. (2015). Spontaneous Activity Does Not Predict Morphological Type in Cerebellar Interneurons. J. Neurosci. *35*, 1432–1442. https://doi.org/10.1523/JNEUROSCI.5019-13.2015.
19. Van Dijck, G., Van Hulle, M.M.V., Heiney, S.A., Blazquez, P.M., Meng, H., Angelaki, D.E., Arenz, A., Margrie, T.W., Mostofi, A., Edgley, S., et al. (2013). Probabilistic Identification of Cerebellar Cortical Neurones across Species. PLoS One *8*, e57669. https://doi.org/10.1371/journal.pone.0057669.
20. Lima, S.Q., Hromádka, T., Znamenskiy, P., and Zador, A.M. (2009). PINP: a new method of tagging neuronal populations for identification during in vivo electrophysiological recording. PLoS One *4*, e6099. https://doi.org/10.1371/journal.pone.0006099.
21. Tye, K.M., and Deisseroth, K. (2012). Optogenetic investigation of neural circuits underlying brain disease in animal models. Nat. Rev. Neurosci. *13*, 251–266. https://doi.org/10.1038/nrn3171.
22. Petreanu, L., Huber, D., Sobczyk, A., and Svoboda, K. (2007). Channelrhodopsin-2-assisted circuit mapping of long-range callosal projections. Nat. Neurosci. *10*, 663–668. https://doi.org/10.1038/nn1891.
23. Petreanu, L., Mao, T., Sternson, S.M., and Svoboda, K. (2009). The subcellular organization of neocortical excitatory connections. Nature *457*, 1142–1145. https://doi.org/10.1038/nature07709.
24. Liu, X., Chen, H., Wang, Y., Si, Y., Zhang, H., Li, X., Zhang, Z., Yan, B., Jiang, S., Wang, F., et al. (2021). Near-infrared manipulation of multiple neuronal populations via trichromatic upconversion. Nat. Commun. *12*, 5662. https://doi.org/10.1038/s41467-021-25993-7.
25. Gouwens, N.W., Sorensen, S.A., Berg, J., Lee, C., Jarsky, T., Ting, J., Sunkin, S.M., Feng, D., Anastassiou, C.A., Barkan, E., et al. (2019). Classification of electrophysiological and morphological neuron types in the mouse visual cortex. Nat. Neurosci. *22*, 1182–1195. https://doi.org/10.1038/s41593-019-0417-0.
26. Hull, C., and Regehr, W.G. (2022). The Cerebellar Cortex. Annu. Rev. Neurosci. *45*, 151–175. https://doi.org/10.1146/annurev-neuro-091421-125115.
27. Striedter, G.F., Bullock, T.H., Preuss, T.M., Rubenstein, J., and Krubitzer, L.A. (2016). Evolution of Nervous Systems (Academic Press).
28. Jacobs, B., Johnson, N.L., Wahl, D., Schall, M., Maseko, B.C., Lewandowski, A., Raghanti, M.A., Wicinski, B., Butti, C., Hopkins, W.D., et al. (2014). Comparative neuronal morphology of the cerebellar cortex in afrotherians, carnivores, cetartiodactyls, and primates. Front. Neuroanat. *8*, 24. https://doi.org/10.3389/fnana.2014.00024.
29. Häusser, M., and Clark, B.A. (1997). Tonic Synaptic Inhibition Modulates Neuronal Output Pattern and Spatiotemporal Synaptic Integration. Neuron *19*, 665–678. https://doi.org/10.1016/S0896-6273(00)80379-7.
30. Khaliq, Z.M., Gouwens, N.W., and Raman, I.M. (2003). The contribution of resurgent sodium current to high-frequency firing in Purkinje neurons: an experimental and modeling study. J. Neurosci. *23*, 4899–4912. https://doi.org/10.1523/JNEUROSCI.23-12-04899.2003.
31. Raman, I.M., and Bean, B.P. (1999). Ionic currents underlying spontaneous action potentials in isolated cerebellar Purkinje neurons. J. Neurosci. *19*, 1663–1674. https://doi.org/10.1523/JNEUROSCI.19-05-01663.1999.
32. Amat, S.B., Rowan, M.J.M., Gaffield, M.A., Bonnan, A., Kikuchi, C., Taniguchi, H., and Christie, J.M. (2017). Using c-kit to genetically target cerebellar molecular layer interneurons in adult mice. PLoS One *12*, e0179347. https://doi.org/10.1371/journal.pone.0179347.
33. Gurnani, H., and Silver, R.A. (2021). Multidimensional population activity in an electrically coupled inhibitory circuit in the cerebellar cortex. Neuron *109*, 1739–1753.e8. https://doi.org/10.1016/j.neuron.2021.03.027.
34. Zhang, X.-M., Ng, A.H.-L., Tanner, J.A., Wu, W.-T., Copeland, N.G., Jenkins, N.A., and Huang, J.-D. (2004). Highly restricted expression of Cre recombinase in cerebellar Purkinje cells. Genesis *40*, 45–51. https://doi.org/10.1002/gene.20062.

35. Hull, C., and Regehr, W.G. (2012). Identification of an Inhibitory Circuit that Regulates Cerebellar Golgi Cell Activity. Neuron *73*, 149–158. https://doi.org/10.1016/j.neuron.2011.10.030.

36. Chow, L.M.L., Tian, Y., Weber, T., Corbett, M., Zuo, J., and Baker, S.J. (2006). Inducible Cre recombinase activity in mouse cerebellar granule cell precursors and inner ear hair cells. Dev. Dyn. *235*, 2991–2998. https://doi.org/10.1002/dvdy.20948.

37. Eccles, J.C., Ito, M., and Szentágothai, J. (1967). The Cerebellum as a Neuronal Machine (Springer) https://doi.org/10.1007/978-3-662-13147-3.

38. Eccles, J.C., Llinás, R., and Sasaki, K. (1966). The excitatory synaptic action of climbing fibres on the Purkinje cells of the cerebellum. J. Physiol. *182*, 268–296. https://doi.org/10.1113/jphysiol.1966.sp007824.

39. Thach, W.T. (1970). Discharge of cerebellar neurons related to two maintained postures and two prompt movements. II. Purkinje cell output and input. J. Neurophysiol. *33*, 537–547. https://doi.org/10.1152/jn.1970.33.4.537.

40. Kostadinov, D., Beau, M., Blanco-Pozo, M., and Häusser, M. (2019). Predictive and reactive reward signals conveyed by climbing fiber inputs to cerebellar Purkinje cells. Nat. Neurosci. *22*, 950–962. https://doi.org/10.1038/s41593-019-0381-8.

41. Mathy, A., Ho, S.S.N., Davie, J.T., Duguid, I.C., Clark, B.A., and Häusser, M. (2009). Encoding of Oscillations by Axonal Bursts in Inferior Olive Neurons. Neuron *62*, 388–399. https://doi.org/10.1016/j.neuron.2009.03.023.

42. Hill, D.N., Mehta, S.B., and Kleinfeld, D. (2011). Quality Metrics to Accompany Spike Sorting of Extracellular Signals. J. Neurosci. *31*, 8699–8705. https://doi.org/10.1523/JNEUROSCI.0971-11.2011.

43. Laboy-Juárez, K.J., Ahn, S., and Feldman, D.E. (2019). A normalized template matching method for improving spike detection in extracellular voltage recordings. Sci. Rep. *9*, 12087. https://doi.org/10.1038/s41598-019-48456-y.

44. Fabre, J.M.J., van Beest, E.H., Peters, A.J., Carandini, M., and Harris, K.D. (2023). Bombcell: Automated Curation and Cell Classification of Spike-Sorted Electrophysiology Data, Version 1.0.0 (Zenodo) https://doi.org/10.5281/zenodo.8172822.

45. Walsh, J.V., Houk, J.C., Atluri, R.L., and Mugnaini, E. (1972). Synaptic Transmission at Single Glomeruli in the Turtle Cerebellum. Science *178*, 881–883. https://doi.org/10.1126/science.178.4063.881.

46. Taylor, A., Elias, S.A., and Somjen, G. (1987). Focal synaptic potentials due to discrete mossy-fibre arrival volleys in the cerebellar cortex. Proc. R. Soc. Lond. B Biol. Sci. *231*, 217–230. https://doi.org/10.1098/rspb.1987.0042.

47. Arenkiel, B.R., Peca, J., Davison, I.G., Feliciano, C., Deisseroth, K., Augustine, G.J., Ehlers, M.D., and Feng, G. (2007). In Vivo Light-Induced Activation of Neural Circuitry in Transgenic Mice Expressing Channelrhodopsin-2. Neuron *54*, 205–218. https://doi.org/10.1016/j.neuron.2007.03.005.

48. Llobet, V., Wyngaard, A., and Barbour, B. (2022). Automatic post-processing and merging of multiple spike-sorting analyses with Lussac. Preprint at bioRxiv. https://doi.org/10.1101/2022.02.08.479192.

49. Barmack, N.H., and Yakhnitsa, V. (2008). Functions of Interneurons in Mouse Cerebellum. J. Neurosci. *28*, 1140–1152. https://doi.org/10.1523/JNEUROSCI.3942-07.2008.

50. Ruigrok, T.J.H., Hensbroek, R.A., and Simpson, J.I. (2011). Spontaneous Activity Signatures of Morphologically Identified Interneurons in the Vestibulocerebellum. J. Neurosci. *31*, 712–724. https://doi.org/10.1523/JNEUROSCI.1959-10.2011.

51. Hensbroek, R.A., Belton, T., van Beugen, B.J., Maruta, J., Ruigrok, T.J.H., and Simpson, J.I. (2014). Identifying Purkinje cells using only their spontaneous simple spike activity. J. Neurosci. Methods *232*, 173–180. https://doi.org/10.1016/j.jneumeth.2014.04.031.

52. Prsa, M., Dash, S., Catz, N., Dicke, P.W., and Thier, P. (2009). Characteristics of Responses of Golgi Cells and Mossy Fibers to Eye Saccades and Saccadic Adaptation Recorded from the Posterior Vermis of the Cerebellum. J. Neurosci. *29*, 250–262. https://doi.org/10.1523/JNEUROSCI.4791-08.2009.

53. Holtzman, T., Rajapaksa, T., Mostofi, A., and Edgley, S.A. (2006). Different responses of rat cerebellar Purkinje cells and Golgi cells evoked by widespread convergent sensory inputs. J. Physiol. *574*, 491–507. https://doi.org/10.1113/jphysiol.2006.108282.

54. Badura, A., and De Zeeuw, C.I. (2017). Cerebellar Granule Cells: Dense, Rich and Evolving Representations. Curr. Biol. *27*, R415–R418. https://doi.org/10.1016/j.cub.2017.04.009.

55. D'Angelo, E., Solinas, S., Mapelli, J., Gandolfi, D., Mapelli, L., and Prestori, F. (2013). The cerebellar Golgi cell and spatiotemporal organization of granular layer activity. Front. Neural Circuits *7*, 93. https://doi.org/10.3389/fncir.2013.00093.

56. Ophir, O., Shefi, O., and Lindenbaum, O. (2023). Neuronal Cell Type Classification using Deep Learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.2306.00528.

57. Lee, E.K., Balasubramanian, H., Tsolias, A., Anakwe, S.U., Medalla, M., Shenoy, K.V., and Chandrasekaran, C. (2021). Non-linear dimensionality reduction on extracellular waveforms reveals cell type diversity in premotor cortex. eLife *10*, e67490. https://doi.org/10.7554/eLife.67490.

58. Lee, K., Carr, N., Perliss, A., and Chandrasekaran, C. (2023). WaveMAP for identifying putative cell types from in vivo electrophysiology. Star Protoc. *4*, 102320. https://doi.org/10.1016/j.xpro.2023.102320.

59. Mountcastle, V.B. (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. J. Neurophysiol. *20*, 408–434. https://doi.org/10.1152/jn.1957.20.4.408.

60. Gold, C., Henze, D.A., Koch, C., and Buzsáki, G. (2006). On the Origin of the Extracellular Action Potential Waveform: A Modeling Study. J. Neurophysiol. *95*, 3113–3128. https://doi.org/10.1152/jn.00979.2005.

61. Katai, S., Kato, K., Unno, S., Kang, Y., Saruwatari, M., Ishikawa, N., Inoue, M., and Mikami, A. (2010). Classification of extracellularly recorded neurons by their discharge patterns and their correlates with intracellularly identified neuronal types in the frontal cortex of behaving monkeys. Eur. J. Neurosci. *31*, 1322–1338. https://doi.org/10.1111/j.1460-9568.2010.07150.x.

62. Trainito, C., von Nicolai, C., Miller, E.K., and Siegel, M. (2019). Extracellular Spike Waveform Dissociates Four Functionally Distinct Cell Classes in Primate Cortex. Curr. Biol. *29*, 2973–2982.e5. https://doi.org/10.1016/j.cub.2019.07.051.

63. Holt, G.R., Softky, W.R., Koch, C., and Douglas, R.J. (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. J. Neurophysiol. *75*, 1806–1814. https://doi.org/10.1152/jn.1996.75.5.1806.

64. Heffley, W., and Hull, C. (2019). Classical conditioning drives learned reward prediction signals in climbing fibers across the lateral cerebellum. eLife *8*, e46764. https://doi.org/10.7554/eLife.46764.

65. Heiney, S.A., Wohl, M.P., Chettih, S.N., Ruffolo, L.I., and Medina, J.F. (2014). Cerebellar-Dependent Expression of Motor Learning during Eyeblink Conditioning in Head-Fixed Mice. J. Neurosci. *34*, 14845–14853. https://doi.org/10.1523/JNEUROSCI.2820-14.2014.

66. Ohmae, S., and Medina, J.F. (2015). Climbing fibers encode a temporal-difference prediction error during cerebellar learning in mice. Nat. Neurosci. *18*, 1798–1803. https://doi.org/10.1038/nn.4167.

67. Lee, K.H., Mathews, P.J., Reeves, A.M.B., Choe, K.Y., Jami, S.A., Serrano, R.E., and Otis, T.S. (2015). Circuit Mechanisms Underlying Motor Memory Formation in the Cerebellum. Neuron *86*, 529–540. https://doi.org/10.1016/j.neuron.2015.03.010.

68. Lisberger, S.G., and Fuchs, A.F. (1978). Role of primate flocculus during rapid behavioral modification of vestibuloocular reflex. I. Purkinje cell activity during visually guided horizontal smooth-pursuit eye movements

and passive head rotation. J. Neurophysiol. *41*, 733–763. https://doi.org/10.1152/jn.1978.41.3.733.

69. Vyas, S., Golub, M.D., Sussillo, D., and Shenoy, K.V. (2020). Computation Through Neural Population Dynamics. Annu. Rev. Neurosci. *43*, 249–275. https://doi.org/10.1146/annurev-neuro-092619-094115.
70. Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Foster, J.D., Nuyujukian, P., Ryu, S.I., and Shenoy, K.V. (2012). Neural population dynamics during reaching. Nature *487*, 51–56. https://doi.org/10.1038/nature11129.
71. Perich, M.G., Gallego, J.A., and Miller, L.E. (2018). A Neural Population Mechanism for Rapid Learning. Neuron *100*, 964–976.e7. https://doi.org/10.1016/j.neuron.2018.09.030.
72. Gallego, J.A., Perich, M.G., Miller, L.E., and Solla, S.A. (2017). Neural Manifolds for the Control of Movement. Neuron *94*, 978–984. https://doi.org/10.1016/j.neuron.2017.05.025.
73. Cunningham, J.P., and Yu, B.M. (2014). Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. *17*, 1500–1509. https://doi.org/10.1038/nn.3776.
74. Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature *503*, 78–84. https://doi.org/10.1038/nature12742.
75. Petersen, P.C., Siegle, J.H., Steinmetz, N.A., Mahallati, S., and Buzsáki, G. (2021). CellExplorer: A framework for visualizing and characterizing single neurons. Neuron *109*, 3594–3608.e2. https://doi.org/10.1016/j.neuron.2021.09.002.
76. Senzai, Y., and Buzsáki, G. (2017). Physiological Properties and Behavioral Correlates of Hippocampal Granule Cells and Mossy Cells. Neuron *93*, 691–704.e5. https://doi.org/10.1016/j.neuron.2016.12.011.
77. Lemon, R.N., Baker, S.N., and Kraskov, A. (2021). Classification of Cortical Neurons by Spike Shape and the Identification of Pyramidal Neurons. Cereb. Cortex *31*, 5131–5138. https://doi.org/10.1093/cercor/bhab147.
78. Ardid, S., Vinck, M., Kaping, D., Marquez, S., Everling, S., and Womelsdorf, T. (2015). Mapping of Functionally Characterized Cell Classes onto Canonical Circuit Operations in Primate Prefrontal Cortex. J. Neurosci. *35*, 2975–2991. https://doi.org/10.1523/JNEUROSCI.2700-14.2015.
79. Mitchell, J.F., Sundberg, K.A., and Reynolds, J.H. (2007). Differential attention-dependent response modulation across cell classes in macaque visual area V4. Neuron *55*, 131–141. https://doi.org/10.1016/j.neuron.2007.06.018.
80. Remington, E.D., Egger, S.W., Narain, D., Wang, J., and Jazayeri, M. (2018). A Dynamical Systems Perspective on Flexible Motor Timing. Trends Cogn. Sci. *22*, 938–952. https://doi.org/10.1016/j.tics.2018.07.010.
81. Simpson, J.I., Hulscher, H.C., Sabel-Goedknegt, E., and Ruigrok, T.J.H. (2005). Between in and out: linking morphology and physiology of cerebellar cortical interneurons. In Progress in Brain Research Creating coordination in the cerebellum (Elsevier), pp. 329–340. https://doi.org/10.1016/S0079-6123(04)48026-1.
82. Hady, M.F.A., and Schwenker, F. (2013). Semi-supervised Learning. In Handbook on Neural Information Processing Intelligent Systems Reference Library, M. Bianchini, M. Maggini, and L.C. Jain, eds. (Springer), pp. 215–239. https://doi.org/10.1007/978-3-642-36657-4_7.
83. Peikari, M., Salama, S., Nofech-Mozes, S., and Martel, A.L. (2018). A Cluster-then-label Semi-supervised Learning Approach for Pathology Image Classification. Sci. Rep. *8*, 7193. https://doi.org/10.1038/s41598-018-24876-0.
84. van Engelen, J.E., and Hoos, H.H. (2020). A survey on semi-supervised learning. Mach. Learn. *109*, 373–440. https://doi.org/10.1007/s10994-019-05855-6.
85. Mahmud, M.S., and Fu, X. (2019). Unsupervised classification of high-dimension and low-sample data with variational autoencoder based dimensionality reduction. In 2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM), pp. 498–503. https://doi.org/10.1109/ICARM.2019.8834333.
86. Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., and Suganthan, P.N. (2022). Ensemble deep learning: a review. Eng. Appl. Artif. Intell. *115*, 105151. https://doi.org/10.1016/j.engappai.2022.105151.
87. Hendrycks, D., and Gimpel, K. (2016). Bridging Nonlinearities and Stochastic Regularizers with Gaussian Error Linear Units. Preprint at arXiv.
88. Taha, A.A., Hennig, L., and Knoth, P. (2022). Confidence estimation of classification based on the distribution of the neural network output layer. Preprint at arXiv.org. https://arxiv.org/abs/2210.07745v2.
89. Ioffe, S., and Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37 ICML'15 (JMLR.org), pp. 448–456.
90. Buzsáki, G., Anastassiou, C.A., and Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. *13*, 407–420. https://doi.org/10.1038/nrn3241.
91. 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. https://doi.org/10.1016/j.neuron.2018.12.009.
92. Strominger, N.L., Demarest, R.J., and Laemle, L.B. (2012). Cerebral Cortex. In Noback's Human Nervous System, [Seventh Edition]: Structure and Function, N.L. Strominger, R.J. Demarest, and L.B. Laemle, eds. (Humana Press), pp. 429–451. https://doi.org/10.1007/978-1-61779-779-8_25.
93. Andersen, P., Morris, R., Amaral, D., Bliss, T., and O'Keefe, J. (2007). The Hippocampus Book (Oxford University Press).
94. Basso, M.A., and May, P.J. (2017). Circuits for Action and Cognition: A View from the Superior Colliculus. Annu. Rev. Vis. Sci. *3*, 197–226. https://doi.org/10.1146/annurev-vision-102016-061234.
95. Llinás, R.R. (1988). The Intrinsic Electrophysiological Properties of Mammalian Neurons: Insights into Central Nervous System Function. Science *242*, 1654–1664. https://doi.org/10.1126/science.3059497.
96. Lipovsek, M., Bardy, C., Cadwell, C.R., Hadley, K., Kobak, D., and Tripathy, S.J. (2021). Patch-seq: Past, Present, and Future. J. Neurosci. *41*, 937–946. https://doi.org/10.1523/JNEUROSCI.1653-20.2020.
97. Cadwell, C.R., Palasantza, A., Jiang, X., Berens, P., Deng, Q., Yilmaz, M., Reimer, J., Shen, S., Bethge, M., Tolias, K.F., et al. (2016). Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. Nat. Biotechnol. *34*, 199–203. https://doi.org/10.1038/nbt.3445.
98. Liu, J., Wang, M., Sun, L., Pan, N.C., Zhang, C., Zhang, J., Zuo, Z., He, S., Wu, Q., and Wang, X. (2020). Integrative analysis of in vivo recording with single-cell RNA-seq data reveals molecular properties of light-sensitive neurons in mouse V1. Protein Cell *11*, 417–432. https://doi.org/10.1007/s13238-020-00720-y.
99. Turner, N.L., Macrina, T., Bae, J.A., Yang, R., Wilson, A.M., Schneider-Mizell, C., Lee, K., Lu, R., Wu, J., Bodor, A.L., et al. (2022). Reconstruction of neocortex: organelles, compartments, cells, circuits, and activity. Cell *185*, 1082–1100.e24. https://doi.org/10.1016/j.cell.2022.01.023.
100. Aller, M.I., Jones, A., Merlo, D., Paterlini, M., Meyer, A.H., Amtmann, U., Brickley, S., Jolin, H.E., McKenzie, A.N., Monyer, H., et al. (2003). Cerebellar granule cell Cre recombinase expression. Genesis *36*, 97–103. https://doi.org/10.1002/gene.10204.
101. Foster, E., Wildner, H., Tudeau, L., Haueter, S., Ralvenius, W.T., Jegen, M., Johannssen, H., Hösli, L., Haenraets, K., Ghanem, A., et al. (2015). Targeted Ablation, Silencing, and Activation Establish Glycinergic Dorsal Horn Neurons as Key Components of a Spinal Gate for Pain and Itch. Neuron *85*, 1289–1304. https://doi.org/10.1016/j.neuron.2015.02.028.
102. Gaffield, M.A., Rowan, M.J.M., Amat, S.B., Hirai, H., and Christie, J.M. (2018). Inhibition gates supralinear Ca2+ signaling in Purkinje cell

dendrites during practiced movements. eLife *7*, e36246. https://doi.org/10.7554/eLife.36246.

103. Madisen, L., Mao, T., Koch, H., Zhuo, J.M., Berenyi, A., Fujisawa, S., Hsu, Y.-W.A., Garcia, A.J., Gu, X., Zanella, S., et al. (2012). A toolbox of Cre-dependent optogenetic transgenic mice for light-induced activation and silencing. Nat. Neurosci. *15*, 793–802. https://doi.org/10.1038/nn.3078.

104. Jelitai, M., Puggioni, P., Ishikawa, T., Rinaldi, A., and Duguid, I. (2016). Dendritic excitation–inhibition balance shapes cerebellar output during motor behaviour. Nat. Commun. *7*, 13722. https://doi.org/10.1038/ncomms13722.

105. Wagner, M.J., Kim, T.H., Savall, J., Schnitzer, M.J., and Luo, L. (2017). Cerebellar granule cells encode the expectation of reward. Nature *544*, 96–100. https://doi.org/10.1038/nature21726.

106. Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat. Neurosci. *13*, 133–140. https://doi.org/10.1038/nn.2467.

107. Daigle, T.L., Madisen, L., Hage, T.A., Valley, M.T., Knoblich, U., Larsen, R.S., Takeno, M.M., Huang, L., Gu, H., Larsen, R., et al. (2018). A Suite of Transgenic Driver and Reporter Mouse Lines with Enhanced Brain-Cell-Type Targeting and Functionality. Cell *174*, 465–480.e22. https://doi.org/10.1016/j.cell.2018.06.035.

108. Herzfeld, D.J., Joshua, M., and Lisberger, S.G. (2023). Rate versus synchrony codes for cerebellar control of motor behavior. Neuron *111*, 2448–2460.e6. https://doi.org/10.1016/j.neuron.2023.07.002.

109. Robinson, D.A. (1963). A Method of Measuring Eye Movement Using a Scleral Search Coil in a Magnetic Field. IEEE Trans. Biomed. Eng. *10*, 137–145. https://doi.org/10.1109/TBMEL.1963.4322822.

110. Ragan, T., Kadiri, L.R., Venkataraju, K.U., Bahlmann, K., Sutin, J., Taranda, J., Arganda-Carreras, I., Kim, Y., Seung, H.S., and Osten, P. (2012). Serial two-photon tomography for automated ex vivo mouse brain imaging. Nat. Methods *9*, 255–258. https://doi.org/10.1038/nmeth.1854.

111. Bastianelli, E. (2003). Distribution of calcium-binding proteins in the cerebellum. Cerebellum *2*, 242–262. https://doi.org/10.1080/14734220310022289.

112. Pachitariu, M., Rossant, C., Steinmetz, N., Colonell, J., Winter, O., Bondy, A.G., kushbanga, Bhagat, J., Sosa, M., O'Shea, D., et al. (2020). MouseLand/Kilosort2: 2.0 final, Version v2.0 (Zenodo) https://doi.org/10.5281/zenodo.4147288.

113. Hall, N.J., Herzfeld, D.J., and Lisberger, S.G. (2021). Evaluation and resolution of many challenges of neural spike sorting: a new sorter. J. Neurophysiol. *126*, 2065–2090. https://doi.org/10.1152/jn.00047.2021.

114. Beau, M., D'Agostino, F., Lajko, A., Martínez, G., Häusser, M., and Kostadinov, D. (2021). NeuroPyxels: Loading, Processing and Plotting Neuropixels Data in Python, Version V2.0.4 (Zenodo) https://doi.org/10.5281/zenodo.5509776.

115. Gold, C., Girardin, C.C., Martin, K.A.C., and Koch, C. (2009). High-Amplitude Positive Spikes Recorded Extracellularly in Cat Visual Cortex. J. Neurophysiol. *102*, 3340–3351. https://doi.org/10.1152/jn.91365.2008.

116. blinklab (2024). blinklab/Phyllum_public: Documentation Update, Version 0.0.2 (Zenodo) https://doi.org/10.5281/zenodo.10659384.

117. Lisberger, S.G., and Pavelko, T.A. (1986). Vestibular signals carried by pathways subserving plasticity of the vestibulo-ocular reflex in monkeys. J. Neurosci. *6*, 346–354. https://doi.org/10.1523/JNEUROSCI.06-02-00346.1986.

118. Verstynen, T., and Kording, K.P. (2023). Overfitting to "predict" suicidal ideation. Nat. Hum. Behav. *7*, 680–681. https://doi.org/10.1038/s41562-023-01560-6.

119. Kingma, D.P., Rezende, D.J., Mohamed, S., and Welling, M. (2014). Semi-supervised learning with deep generative models. Preprint at arXiv. https://doi.org/10.48550/arXiv.1406.5298.

120. Kingma, D.P., and Welling, M. (2022). Auto-Encoding Variational Bayes. Preprint at arXiv. https://doi.org/10.48550/arXiv.1312.6114.

121. Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., and Lerchner, A. (2017). beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. In International Conference on Learning Representations. https://openreview.net/forum?id=Sy2fzU9gl.

122. Bowman, S.R., Vilnis, L., Vinyals, O., Dai, A., Jozefowicz, R., and Bengio, S. (2016). Generating Sentences from a Continuous Space. In Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, S. Riezler and Y. Goldberg, eds. (Association for Computational Linguistics), pp. 10–21. https://doi.org/10.18653/v1/K16-1002.

123. Loshchilov, I., and Hutter, F. (2017). SGDR: Stochastic Gradient Descent with Warm Restarts. Preprint at arXiv. https://doi.org/10.48550/arXiv.1608.03983.

124. Hendrycks, D., and Gimpel, K. (2023). Gaussian Error Linear Units (GELUs). Preprint at arXiv. https://doi.org/10.48550/arXiv.1606.08415.

125. Lemaître, G., Nogueira, F., and Aridas, C.K. (2017). Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. J. Mach. Learn. Res. *18*, 1–5. https://dl.acm.org/doi/10.5555/3122009.3122026.

126. Wang, B., and Zou, H. (2021). Honest leave-one-out cross-validation for estimating post-tuning generalization error. Stat *10*, e413. https://doi.org/10.1002/sta4.413.

127. Zhang, Y., and Yang, Y. (2015). Cross-validation for selecting a model selection procedure. Journal of Econometrics *187*, 95–112. https://doi.org/10.1016/j.jeconom.2015.02.006.

128. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Inf. Fusion *76*, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008.

129. Kristiadi, A., Hein, M., and Hennig, P. (2020). Being Bayesian, even just a bit, fixes overconfidence in ReLU networks. Preprint at arXiv. https://doi.org/10.48550/arXiv.2002.10118.

130. Daxberger, E., Kristiadi, A., Immer, A., Eschenhagen, R., Bauer, M., and Hennig, P. (2021). Laplace Redux - Effortless Bayesian Deep Learning. In Advances in Neural Information Processing Systems (Curran Associates, Inc.), pp. 20089–20103.

131. Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., Murthy, V.N., Mathis, M.W., and Bethge, M. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci. *21*, 1281–1289. https://doi.org/10.1038/s41593-018-0209-y.

132. Cawley, G.C., and Talbot, N.L.C. (2010). On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation. J. Mach. Learn. Res. *11*, 2079–2107.

133. Lackey, E.P., Moreira, L., Norton, A., Hemelt, M.E., Osorno, T., Nguyen, T.M., Macosko, E.Z., Lee, W.-C.A., Hull, C.A., and Regehr, W.G. (2024). Specialized connectivity of molecular layer interneuron subtypes leads to disinhibition and synchronous inhibition of cerebellar Purkinje cells. Neuron *112*, 2333–2348.e6. https://doi.org/10.1016/j.neuron.2024.04.010.

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat anti-mCherry	Thermo Fisher Scientific	Cat#M11217; RRID:AB_2536611
Mouse anti-Parvalbumin	Millipore	Cat#MAB1572; RRID:AB_2174013
Donkey anti-Rat-Alexa 594	Invitrogen	Cat#A-21209; RRID:AB_2535795
Goat anti-Mouse-Alexa 633	Invitrogen	Cat#A-21050; RRID:AB_2535718
Bacterial and virus strains		
AAV1.CAGGS.Flex.ChR2.tdTomato	U. Penn Vector Core	N/A
AAV1.Ef1a.Flex.GtACR2.eYFP	Duke Vector Core	N/A
Chemicals, peptides, and recombinant proteins		
Neurotrace 435/455	Thermo Fisher Scientific	Cat#N21479
DAPI, Dihydrochloride	EMD Millipore	Cat#268298
Hoechst 33342	Thermo Fisher Scientific	Cat#H1399
Vybrant DiD, DiI, and DiO	Thermo Fisher Scientific	Cat#V22889
ProLong Diamond Antifade	Thermo Fisher Scientific	Cat#P36965
Gabazine	Tocris	Cat#1262
NBQX disodium	Tocris	Cat#1044
D-AP5	Abcam	Cat#ab120003
MCPG	Tocris	Cat#3696
Deposited data		
Ground-truth database (spiking and waveform data)	This paper	https://doi.org/10.5522/04/23702850
Experimental models: Organisms/strains		
Macca Mulatta	Various	N/A
C57BL/6J	Jackson Labs	RRID:IMSR_JAX:000664
B6.Cg-Kit^{m}1$^{(cre)Htng}$/J ("c-kitIRES-Cre")	Amat et al.[32]	RRID:IMSR_JAX:032923
B6.Cg-Tg(Thy1-COP4/EYFP)18Gfng/J	Jackson Labs	RRID:IMSR_JAX:007612
BACα6Cre-C	Aller et al.[100]	N/A
B6.Cg-Gt(ROSA)26Sortm40.1(CAG-aop3/EGFP)Hze/J	Jackson Labs	RRID:IMSR_JAX:021188
B6.Cg-Tg(Pcp2-cre)3555Jdhu/J	Jackson Labs	RRID:IMSR_JAX:010536
B6.129-Nos1tm1(cre)Mgmj/J	Jackson Labs	RRID:IMSR_JAX:017526
Tg(Slc6a5-cre)1Uze	Foster et al.[101]	RRID:IMSR_JAX:038515
B6.Cg-Tg(Atoh1-cre)1Bfri/J	Jackson Labs	RRID:IMSR_JAX:011104
B6.Cg-Gt(ROSA)26Sortm32(CAG-COP4*H134R/EYFP)Hze/J	Jackson Labs	RRID:IMSR_JAX:024109; IMSR_JAX:012569
Recombinant DNA		
pAAV-EF1a-Flex(loxP) rev-GtACR2.eYFP	Gaffield et al.[102]	N/A
Software and algorithms		
NeuroPyxels (including automated cerebellar classifier)	Github	https://github.com/m-beau/NeuroPyxels
Database visualization/exploration tool	This paper	https://www.c4-database.com/
Maestro	Lisberger Laboratory	https://sites.google.com/a/srscicomp.com/maestro/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SpikeGLX	Github	https://billkarsh.github.io/SpikeGLX
Phyllum	Github	https://github.com/blinklab/Phyllum_public
Phy	Github	https://github.com/cortex-lab/phy
Kilosort	Github	https://github.com/cortex-lab/Kilosort
Full Binary Pursuit	Github	https://github.com/njh27/spikesorting_fullpursuit
Python 3	Python software foundation	https://www.python.org
PyTorch	The Linux Foundation	https://pytorch.org/
ZEN	ZEISS	https://www.zeiss.com/microscopy/en/products/software/zeiss-zen.html?vaURL=www.zeiss.com/zen
BakingTray	Github	https://github.com/SWC-Advanced-Microscopy/BakingTray
Matlab	MathWorks	https://mathworks.com

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Ethical statement

We conducted experiments in four laboratories and on two species, mice and macaque monkeys. All mouse procedures in the Häusser lab were approved by the local Animal Welfare and Ethical Review Board at University College London and performed under license from the UK Home Office in accordance with the Animals (Scientific Procedures) Act 1986 and in line with the European Directive 2010/63/EU on the protection of animals used for experimental purposes. Mouse procedures in the Hull and Medina labs were approved in advance by the *Institutional Animal Care and Use Committees* at Duke University and Baylor College of Medicine, respectively, based on the guidelines of the United States' *National Institutes of Health*. Monkey procedures in the Lisberger lab were approved in advance by the *Institutional Animal Care and Use Committee* at Duke University. Every effort was made to minimize both the number of animals required and any possible distress they might experience.

Animals

Mouse

We used adult (>P60) male and female C57BL/6J wildtype mice and transgenic mice maintained on the C57BL/6J background. Experiments were conducted during the light cycle in the *Häusser* and *Hull* labs and the dark cycle in the *Medina* lab. All mice were housed in an animal facility with standardized temperature and humidity, with 12 h light/dark cycles and food and water ad libitum, except reward conditioning experiments and some optotagging experiments where animals were water restricted to 85% of initial body weight. The following mouse lines were used:

- *Häusser*: Channelrhodopsin-2 (ChR2) was expressed in various cerebellar cell types primarily by crossing Cre lines to a Cre-dependent ChR2-eYFP reporter line[103] (Ai32, B6.Cg-Gt(ROSA)26Sortm32(CAG-COP4*H134R/EYFP)Hze/J), or, in a subset of experiments, by injecting Cre-dependent ChR2 virus (AAV1.CAGGS.Flex.ChR2-tdTomato [UPenn]). Cre lines were: BAC-Pcp2-IRES-Cre (B6.Cg-Tg(Pcp2-cre)3555Jdhu/J), intended to label Purkinje cells[34]; Nos1-Cre (B6.129-Nos1tm1(cre)Mgmj/J), intended to label molecular layer interneurons[104]; Glyt2-Cre (Tg(Slc6a5-cre)1Uze), intended to label Golgi cells[33]; and Math1-Cre (B6.Cg-Tg(Atoh1-cre)1Bfri/J), intended to label granule cells.[105] In addition to the transgenic crosses and viral ChR2 expression, we used the Thy1-ChR2 line 18 (B6.Cg-Tg(Thy1-COP4/EYFP)18Gfng/J) to express ChR2 in mossy fibers.[35] Recordings using each strategy were performed as follows: L7-Cre x Ai32 – 1 recording (1 mouse), Nos1-Cre x Ai32 – 40 recordings (34 mice), Nos1-Cre + AAV1.CAGGS.Flex.ChR2-tdTomato – 3 recordings (3 mice), GlyT2-Cre x Ai32 – 32 recording (31 mice), GlyT2-Cre + AAV1.CAGGS.Flex.ChR2-tdTomato – 3 recordings (3 mice), Math1-Cre x Ai32 – 47 recording (38 mice), Math1-Cre + AAV1.CAGGS.Flex.ChR2-tdTomato – 3 recordings (3 mice), and Thy1-ChR2 line 18 – 26 recordings (22 mice). The specificity of opsin expression in the cerebellum of our Cre transgenic crosses was further investigated by crossing the listed Cre lines to a Cre-dependent tdTomato reporter line, Ai9 (B6.Cg-Gt(ROSA)26Sortm9(CAG-tdTomato)Hze/J),[106] so that we could evaluate expression specificity through cytosolic, rather than membrane-bound, fluorescence.

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

- *Hull*: Mice expressing ChR2 or the inhibitory GtACR2 or ArchT opsins were generated by either: 1) crossing the *c-kit*$^{IRES\text{-}Cre}$ line, intended to label molecular layer interneurons,[32] with a Cre-dependent ArchT-GFP reporter line (Ai40 (B6.Cg-Gt(ROSA) 26Sortm40.1(CAG-aop3/EGFP)Hze/J),[107] or 2) crossing the BACα6Cre-C line, intended to label granule cells,[100] with the Ai32[103] Cre-dependent ChR2 reporter line. Alternatively, we injected the same Cre lines with Cre-dependent viruses: (AAV1.CAGGS.Flex.ChR2-tdTomato into *c-kit*$^{IRES\text{-}Cre}$ and BACα6Cre-C, and AAV1.Ef1a.Flex.GtACR2.eYFP into *c-kit*$^{IRES\text{-}Cre}$). In addition, we used the Thy1-ChR2 line 18 to express ChR2 in mossy fibers. Recordings using each strategy were performed as follows: *c-kit*$^{IRES\text{-}Cre}$ + AAV1.CAGGS.Flex.ChR2-tdTomato – 8 recordings (2 mice), *c*-kit$^{IRES\text{-}Cre}$ x Ai40 – 3 recordings (1 mouse), *c-kit*$^{IRES\text{-}Cre}$ + AAV1.Ef1a.Flex.GtACR2.eYFP – 11 recordings (6 mice), BACα6Cre-C x Ai32 – 22 recordings (12 mice), BACα6Cre-C + AAV1.CAGGS.Flex.ChR2-tdTomato – 10 recordings (4 mice), and Thy1-Chr2 line 18 – 13 recordings (4 mice).
- *Medina*: All experiments were performed in wildtype C57BL/6J mice obtained from Jackson Laboratories.

Monkey

Recordings in non-human primates were conducted in the *Lisberger* lab on three adult male rhesus monkeys (*Macaca mulatta*) weighing 10-15 kg.

METHOD DETAILS

Surgery

Mouse

All labs followed the same procedure to prepare mice for awake in vivo recordings by implanting a headplate/headpost under isoflurane anesthesia in sterile conditions. Pre-operative and post-operative analgesia were administered, and mice were allowed to recover from surgery for at least one week before being habituated to head-fixation and prepared for recordings. Lab-specific details are as follows:

- *Häusser*: We installed a custom-made aluminum headplate with a 5 mm long and 9 mm wide oval inner opening over the cerebellum. Mice received a steroid anti-inflammatory drug at least 1 hour before surgery (dexamethasone, 0.5 mg/kg), followed by an analgesic NSAID (meloxicam, 5mg/kg) immediately before surgery. Anesthesia was induced and maintained with 5% and 1-2% isoflurane, respectively. The headplate was positioned over lobule simplex of the left cerebellar hemisphere, angled at approximately 26° with respect to the transverse plane, and attached to the skull with dental cement (Super-Bond C&B, Sun-Medical). Post-operative analgesia (carprieve, 5 mg/kg) was given for 3 days. After several days of habituation on the recording apparatus, a 1 mm-diameter craniotomy and durotomy were performed to allow access for Neuropixels probes into the lobule simplex (3 mm lateral to the midline, anterior to the interparietal-occipital fissure). Before the craniotomy, a conical nitrile rubber seal (Stock no. 749-581, RS components) was attached to the headplate with dental cement to serve as a bath chamber. The exposed brain was then covered with a humid gelatinous hemostatic sponge (Surgispon) and silicone sealant (Kwik-Cast, WPI) until the experiment was performed (1-2 h after recovery). At the beginning of the experiment, mice were head-fixed, the silicone sealant was removed, and physiological HEPES-buffered saline solution was immediately applied to keep the craniotomy hydrated.
- *Hull*: We installed a titanium headpost (HE Palmer, 32.6x19.4 mm) to the skull and a stainless-steel ground screw (F.S. Tools) over the left cerebellum, both secured with Metabond (Parkell). Mice received dexamethasone (3 mg/kg) 3-4 hours before surgery and an initial dose of ketamine/xylazine (50 mg/kg and 5 mg/kg, IP) and carprofen (5 mg/kg) 20 min before induction with isoflurane anesthesia. Isoflurane was administered at 1-2% throughout surgery to maintain appropriate breathing rates and prevent toe pinch response, which were monitored throughout the duration of the surgery. Body temperature was maintained with a heating pad (TC-111 CWE). Mice received buprenex and cefazolin (0.05 mg/kg and 50 mg/kg respectively, subcutaneously) twice daily for 48 hours after surgery and were monitored daily for 4 days. After 2+ weeks of recovery, mice received dexamethasone (3 mg/kg) 4-24 hours before recordings. Craniotomies (approx. 0.5-1.5 mm) were opened over vermis or lateral cerebellum (relative to bregma: between -6.0 and -7.0 mm AP, and between 1.0 and 2.8 mm ML) on the first day of recording, under 1-2% isoflurane anesthesia, and were sealed between recordings using Kwik-Cast (WPI) covered by Metabond. Craniotomies could be re-opened for subsequent recordings under brief (<30 min) 1-2% isoflurane anesthesia.
- *Medina*: Preoperative analgesia was provided (5 mg/kg meloxicam, 0.02mL 0.5% bupivacaine and 2% lidocaine) and surgery was carried out under sterile conditions. Mice were anesthetized with isoflurane (5% by volume in O_2 for induction and 1-2% by volume for maintenance; SurgiVet) and kept on a heating pad to maintain body temperature. The skull was exposed and leveled to match the stereotaxic plane before two stainless steel screws were implanted (relative to bregma: AP -0.3 mm, ML ±1.4 mm) to anchor the whole preparation. A custom-made stainless steel headplate was placed over the screws and the whole preparation was secured to the skull with Metabond cement (Parkell). Additionally, a craniotomy

was performed (relative to bregma: AP -5.5 mm) consisting of a 5x2 mm section of bone removed to expose the cerebellar vermis and the right anterior and posterior lobes. A chamber was then built with Metabond to cover the exposed bone around the craniotomy, the dura was protected with a thin layer of biocompatible silicone (Kwik-Cast, WPI), and the whole chamber was sealed with silicone adhesive (Kwik-Sil, WPI). Mice were monitored until fully recovered from anesthesia and analgesia was provided during the three days following the surgical procedure.

Monkey

A portion of the primate dataset reported here has been published previously along with corresponding detailed methods.[108] Briefly, monkeys underwent several surgical procedures under isoflurane in preparation for neurophysiological recordings. In succession, we (i) affixed a head-holder to the calvarium, (ii) sutured a small coil of wire to the sclera of one eye to monitor eye position and velocity using the search coil technique[109] and (iii) implanted a recording cylinder aimed at the floccular complex. Analgesics were provided to the monkeys after each surgery until they had recovered.

Extracellular recording procedures

Mouse

All the mouse labs followed the same general procedures for cerebellar recordings. Mice were progressively habituated to head fixation prior to Neuropixels recordings. Neuropixels 1.0[16] probes were inserted into the cerebellar cortex at a speed of 1-4 μm/s while monitoring electrophysiological signals. Probes were typically inserted to a recording depth of 2-3 mm below the cerebellar surface and allowed to settle for at least 20 minutes before commencing data acquisition. The recording chamber surrounding the craniotomy was bathed in ACSF, with or without synaptic blockers. In all three laboratories, Neuropixels data were acquired using SpikeGLX and signals were digitized at 30 kHz.

Monkey

Each day, we acutely inserted either tungsten single electrodes (FHC) or, for the majority of our data, custom-designed Plexon s-Probes into the cerebellar floccular complex. Plexon s-Probes included 16 recording contacts (tungsten, 7.5 μm diameter) spaced in two columns separated by 50 μm. Adjacent rows of contacts were also separated by 50 μm. Once the electrode reached the ventral paraflocculus, we allowed the electrode to settle for a minimum of 30 minutes. We recorded continuous wideband data from all contacts at a sampling rate of 40 kHz using the Plexon Omniplex system. We used a 4th order Butterworth low-pass hardware filter with a cutoff frequency of 6 kHz prior to digitization to eliminate distortion of the recorded signal by the electrical field produced by the magnetic field that enabled measurement of eye movement.

Reconstruction of Neuropixels probe tracts

Neuropixels probes used in the mouse experiments were coated with DiI, DiO, or DiD (Cat.Nos.V22885, V22886, and V22887; Thermo Fisher Scientific) by repeatedly moving a drop of dye along the probe shank using a pipette until a dye residue was visible along its entire length (~20 passes). After each recording, the probe was removed and soaked in Tergazyme, then soaked in distilled water, and finally washed with isopropyl alcohol. After the last recording session, the brains were fixed and processed for histology to verify recording locations.

- *Häusser*: Mice were deeply anesthetized with ketamine/xylazine and perfused transcardially with PBS followed by 4% PFA in PBS. The brains were dissected and post-fixed overnight in 4% PFA, then embedded in 5% agarose. To reconstruct electrode tracts, we imaged full 3D stacks of the brains in a custom-made serial two-photon tomography microscope coupled to a microtome,[110] controlled using ScanImage (2017b, Vidrio Technologies) and BakingTray. Brains were imaged at 20 μm intervals and sectioned at 40 μm (2 optical sections/slice). Images were acquired in two channels (green channel: 500–550 nm, ET525/50; red channel: 580–630 nm, ET605/70; Chroma) through a piezo-mounted (PIFOC P-725, Physik Instrumente) Nikon 16x/0.8NA objective. Each section was imaged in 1025 x 1025 μm tiles at 512x512-pixel resolution with 7% overlap.
- *Hull*: After the last day of recording, mice were deeply anesthetized with ketamine/xylazine (350 mg/kg and 35 mg/kg) and perfused with PBS followed by 4% PFA in PBS. Brains were extracted and post-fixed in 4% PFA in PBS overnight, then sectioned at 100 μm using a vibratome (Pelco 102). Before sectioning, some brains were encased in a 2% agar block for stability. Slices were either stained with DAPI (DAPI, Dihydrochloride, 268298, EMD Millipore) and then mounted with mounting medium (Fluoromount-G, Southern Biotech) or were mounted with a DAPI-containing mounting medium (DAPI Fluoromount-G, Southern Biotech). Electrode tracts were visualized using a confocal microscope (Leica SP8).
- *Medina*: After perfusion with 4% PFA in PBS, brains were extracted, post-fixed in the same solution for at least 12h and then cryoprotected in 30% sucrose solution in PBS for 48h. The brains were aligned so the coronal sections would match the track angle and sectioned at 50 μm on a cryostat (Leica CM1950). Free floating sections were recovered in PBS and incubated in Hoechst solution for 3 minutes (Hoechst 33342, 2μg/mL in PBS-TritonX 0.25%, Thermo Fisher Scientific).

Sections were then washed in PBS three times and coverslipped using fluorescence protectant medium (ProLong Diamond Antifade, Thermo Fisher Scientific). Epifluorescence was acquired at 10x magnification on an Axio Imager Z1 microscope (Zeiss), and track reconstruction and measurements were made using microscopy analysis software (ZEN software, Zeiss).

Optogenetic stimulation and pharmacology

Ground-truth neurons were recorded in mouse experiments, using the same general procedures for optogenetic stimulation and pharmacology in both the *Häusser* and *Hull* labs. This procedure consisted of four phases: (1) A *baseline phase* recorded spontaneous activity. (2) A *control phase* applied light externally to the cerebellum or, in some recordings, through a tapered optic fiber that ran alongside the recording probe to activate opsins in the cell types that expressed them. (3) An *infusion phase* delivered light to the cerebellum while we added synaptic blockers[20] to the surface of the cerebellum. Application of synaptic blockers on the surface of the cerebellum preserves the integrity of the tissue. (4) A *blockade phase*, after the synaptic blockers had permeated well into the tissue, assayed neurons for direct responses to optogenetic stimulation. The details of the procedures for ground-truth identification of cell-type varied slightly between the two labs.

- *Häusser*: Optogenetic stimulation was performed using 1 or 2 blue LEDs (470 nm, Thorlabs M470F3) and, in some experiments, a blue laser for surface illumination (Stradus 472, Voltran). Surface illumination was performed by coupling the laser or the LED via a patch cable (M95L01, Thorlabs) to a cannula (CFMXB05, Thorlabs) positioned in contact with the brain surface near the probe. In some experiments a second illumination source - a tapered fiber (Optogenix 0.39NA/200μm) glued directly to the head of the Neuropixels probe - was inserted into the brain. Total power at the fiber tip (surface fiber) and coupling cannula (tapered fiber) was 1-6.9 mW. Each recording session consisted of: (1) a 20 minute baseline period of spontaneous activity, (2) a set of 50 optogenetic stimuli (at 0.1 Hz: 1 stimulation of 250 ms or a train of 5 stimulations of 50 ms at 5 Hz, depending on the experiment), (3) an application of a synaptic blocker cocktail (Gabazine 0.2-0.8 mM, NBQX 0.8 mM, APV 1.6 mM, MCPG 0-1.3 mM) to the surface of the cerebellum followed by a 20 minute incubation, and (4) a second set of 50 optogenetic stimuli in the presence of synaptic blockers. We did not record any neurons in the ground-truth library with the blue laser as a source of photostimulation.
- *Hull*: Neurons expressing ChR2 or GtACR2 were respectively activated or inhibited with a 450 nm laser (MDL-III, OptoEngine) and neurons expressing ArchT were inhibited using a 532 nm laser (MGL-III, OptoEngine). Lasers were coupled with a 400-micron optic patch fiber (FT400 EMT, Thorlabs) that was positioned 4-10 mm from the brain surface. Power at the brain surface was approximately 2-30 mW and was calibrated for each experiment to produce neuronal responses with minimal artifact. Laser stimulations lasted 50 or 100 ms and were delivered at 0.1 Hz throughout the recording after the 20-minute baseline period, with brief pauses to replenish ACSF or apply blockers (Gabazine 0.2-0.8 mM, NBQX 0.6-1.2 mM, AP-5 0.15-0.6 mM, MCPG 1-2.5 mM). Synaptic block was achieved with gabazine alone in recordings targeting MLIs except one recording where blocker was not applied and 3 MLIs were accepted into the ground-truth library based on direct inhibitory GtACR2 responses with short latency (<3 ms, a latency that seems too short to occur because of synaptic inputs rather than direct inhibition).

Histological assessment of opsin expression

To assess the specificity of opsin expression, PFA-fixed brains of the different transgenic mouse lines (Cre-lines crossed to the tdTomato reporter line) were sectioned at 100 μm and prepared for immunohistochemistry. Sections were blocked with 2.5% normal donkey serum / 2.5% normal goat serum / 0.5% Triton X-100/PBS for 4-6 hours at room temperature, primary antibodies for 4-6 days at 4°C, and secondary antibodies overnight at 4°C. Antibodies were diluted in blocking solution. The following antibodies were used: rat anti-mCherry (1:250, ThermoFisher M11217), Mouse anti-Parvalbumin (1:1000, Millipore MAB1572), Donkey anti-Rat-Alexa 594 (1:1000, Invitrogen), and Goat anti-Mouse-Alexa 633 (1:1000, Invitrogen). Neurotrace 435/455 (1:250, ThermoFisher N21479) was added to the secondary antibody solution. Sections were mounted and imaged on a Zeiss LSM 880 using a 20x objective in 425x425 μm tiles at 1024x1024-pixel resolution.

To identify the classes of cerebellar neurons that expressed optogenetic actuators, we determined the layer in which fluorescent neurons were present and whether they expressed parvalbumin (PV), which should be present in all molecular layer interneurons and Purkinje cells.[111] The location of cerebellar layers in each image was identified in the Neurotrace (fluorescent Nissl) channel. The soma locations of neurons expressing tdTomato (as a proxy for Cre expression) and PV were marked manually in grayscale images using Fiji (NIH). Neurons were deemed to express both tdTomato and PV if their somatic locations were less than 5 μm apart, and the layer of each neuron was determined by overlaying the Neurotrace laminar mask to cell locations.

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of extracellular recordings

Spike sorting and curation

- **Mouse**: Following data acquisition, automated spike sorting was performed using Kilosort 2.0[17,112] and a preliminary manual curation was performed using Phy. Then, we applied multiple quality checks to ensure that the resulting clusters selected for further analysis corresponded to single units with physiological waveforms, good isolation properties, and few or no refractory period violations. Rigorous curation was especially important for our long recordings, which could have periods of good isolation intermixed with periods of drift or poor unit isolation. We divided our recordings into overlapping segments (30 second segments computed every 10 seconds) and computed the 'false-positive' and 'false-negative' rates in each segment. False positives were defined as spikes that fell within the refractory period of a unit (± 0.8 ms from a given spike) and termed refractory period violations (RPVs). The proportion of false-positives was estimated as the quotient between the RPV rate and the mean firing rate.[42] False negatives were defined as spikes that were not detected because they fell below the noise threshold of the recording. They were estimated by fitting each unit's spike amplitude distribution with a Gaussian function[43,44] and quantifying the fraction of area under the curve clipped at the noise threshold. A 30-second segment was deemed acceptable if it had less than 5% false positive and false negative rates. Acceptable intervals were concatenated and used for subsequent classifier training. A unit was required to have 3 minutes of acceptable isolation during the baseline period to be included in the sample.

  For each neuron in the ground-truth library, we performed an additional analysis to assess the quality of unit isolation. We used an equation derived by others[48] to estimate the "fraction of uncontaminated spikes" (i.e. not noise or neighboring neurons) from measurements of the number of refractory period violations (n_v), the total number of spikes (N), the refractory period interval (T_r), and the mean firing rate (MFR): $\sqrt{1 - \frac{n_v}{N*MFR*T_r}}$. All the neurons in the ground-truth library, except for only a small subset of mossy fibers, had a fraction of uncontaminated spikes more than 0.97 when we measured refractory period violations $\pm$1 ms from the trigger spike (T_r=0.001). However, mossy fibers have shorter refractory periods and can burst at up to 1,000 spikes/s. Therefore, we re-measured refractory period violations $\pm$0.5 ms from the trigger spike (T_r=0.0005) and estimated the fraction of true spikes from those measurements.
- **Monkey**: Following each recording session, individual action potentials were assigned to putative neural units using the semi-automated "Full Binary Pursuit" sorter,[113] designed to distinguish temporally and spatially overlapping spikes from different neurons. Following automated sorting, we manually curated our dataset, removing neurons with significant inter-spike interval violations or low signal-to-noise ratios. The majority of units in our primate dataset significantly exceeded the metrics used for automated curation of the mouse data, which potentially biases our sample of primate units towards those that are easier to record.

Data harmonization

To achieve consistency of the extracellular recording data acquired across labs and setups, we implemented several procedures (Figure S1):

1. We mitigated the filtering introduced by the hardware filter on Neuropixels probes. It is possible to disable the filter and we did so for some of our mouse recordings. It also was not a factor for the monkey recordings. To harmonize all the recordings, we applied a comparable causal first-order Butterworth high-pass filter (300 Hz cutoff) to the wideband voltage recordings made in monkeys and those obtained without the hardware filter in mice.
2. We used a two-step procedure to construct a high-quality waveform template for each unit: (a) We sub-sampled the spikes of each neuron by grouping waveforms with a similar amplitude on the principal channel, and therefore the same drift-state (i.e. position of probe relative to the recorded neuron): "drift-matching", and (b) We re-aligned the spikes in time by maximizing the cross-correlation of each spike to a high amplitude template: "shift-matching". After alignment, the individual spikes were averaged, resulting in the final mean waveform for the neuron under study. Neuropixels data processing (non-manual curation, filtering, drift-shift-matching) was performed using the NeuroPyxels library.[114]
3. We preprocessed all waveform templates by selecting the mean waveform from the highest amplitude channel, resampling it to 30 kHz (if necessary), aligning it to the peak, and flipping it if necessary to ensure the largest deflection in the waveform was always negative. We did so with the knowledge that the polarity of the action potential waveform depends on a number of factors including the proximity of the recording electrode to the dendrites, soma, and axon[90,115] and relative orientation of the recording contact and the reference. We used the harmonized waveforms to compute summary statistics (Figure 4; Table S1) that have been used previously to classify cerebellar neurons.[18,19,50,51]

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

Assignment of layers with Phyllum

For recordings in the mouse, we assigned the units recorded in each channel of the Neuropixels probe to a layer using *Phyllum*,[116] a custom-designed plugin for the curation software Phy. The algorithm for layer identification in *Phyllum* starts by automatically setting 'anchor' channels whose recorded layer can be unambiguously identified by the presence of Purkinje cell units with simple and complex spikes (Purkinje layer anchor), mossy fiber units with triphasic waveforms (granule layer anchor), or low 1-2 Hz frequency units with wide waveforms indicative of dendritic complex spikes (molecular layer anchor). Then, *Phyllum* fills in the layer of the remaining channels via an iterative procedure based on (1) proximity to the nearest Purkinje cell anchor and (2) allowed layer transitions. Every channel assigned to the Purkinje cell layer must contain at least a Purkinje cell recording within 100 μm, but the channel may also contain additional units located in the neighboring granule or molecular layers. If none of the channels between two consecutive Purkinje cell anchors contain another anchor unit, their layer is set to 'Unknown'. On average, *Phyllum* assigns 82% of all the channels on the Neuropixels probe to a specific layer. Histological reconstruction of 21 recording tracks confirmed that for channels that are assigned a specific layer, the assignment is highly accurate: >99% for molecular layer channels, >98% for granule layer channels, and >95% for Purkinje layer channels.

Identification of units directly responsive to optogenetic stimulation

Units recorded during optogenetic activation experiments were deemed to be directly responsive to photostimulation if they met the following conditions: (1) their firing rate increased (ChR2) or decreased (GtACR2, ArchT) more than 3.3 standard deviations from the pre-stimulus baseline within 10 ms of stimulation onset in the synaptic *blockade* phase (computed using 0.1 ms bins smoothed with a causal Gaussian filter with a standard deviation of 0.5 ms), (2) they were recorded at a depth at which pharmacological blockade was confirmed, and (3) the spike waveforms evoked in the synaptic *blockade* phase matched those recorded during the pre-stimulation *baseline* phase.

Construction of 3D autocorrelograms

All recordings were performed in awake animals that were either head-fixed but otherwise free to move on a wheel (mice) or performing discrete trials of smooth pursuit (primates), resulting in behaviorally-driven modulation of firing rates across the experimental session. To normalize for the impact of changes in firing rate on measures of firing statistics, we constructed "three-dimensional autocorrelograms" (3D-ACGs). At each point in time, we estimated the instantaneous firing rate of the neuron as the inverse interspike interval.[117] We smoothed the firing rates using a boxcar filter (250 ms width) and measured the value of the smoothed instantaneous firing rate timeseries at the time of each spike. Next, we determined the distribution of firing rates, as assayed at the time of each spike in a recording, stratified the distribution of firing rates into 10 deciles, and computed separate 2D-ACGs for the spikes in each decile. We visualized the resulting 3D-ACGs as a surface where the color axis corresponds to the probability of firing, the y-axis stratifies the firing rate deciles so that each 3D-ACG contains 10 rows, and the x-axis represents time from the trigger spike. Note that the spike counts in the autocorrelograms have been divided by the width of the bin so that the y-axis or color map is calibrated in spikes/s. As input to the classifier, we used log-distributed bins relative to *t=0* in contrast to the linearly-spaced bins shown in the Figures and supplemental information.

Human expert labeling of cerebellar units

- **Mouse**. We performed human expert cell-type identification of an unlabeled mouse dataset collected in the *Medina* lab. We used Phyllum to identify the layer of each recording. Most Purkinje cells were identified by the presence of both simple spikes and complex spikes and complex-spike-triggered histograms that showed a characteristic pause in the simple-spike firing rate following the complex spike. We identified putative Purkinje cells by the presence of simple spikes without a complex spike, location in a Purkinje cell layer, and regular firing rate resulting in characteristic "shoulders" present in the autocorrelogram. Putative molecular layer interneurons were identified by their presence in a molecular layer with firing rates above 5 spikes/s, incompatible with the firing properties of the dendritic complex spikes. Putative mossy fibers were in the granular cell layer and some displayed a characteristic triphasic shape due to the negative afterwave recorded near the glomerulus.[45,46] Putative Golgi cells were in the granular cell layer and had broad waveforms and relatively regular firing rates.
- **Monkey**. We performed human expert cell-type identification of an unlabeled monkey dataset collected in the *Lisberger* lab. We classified recordings as ground-truth Purkinje cells if they demonstrated the characteristic post-complex-spike pause in simple-spike firing. Units that exhibited known characteristics of Purkinje cell simple spikes but lacked a complex spike were treated as "putative" Purkinje cells and used in the comparison of classifier-predicted and expert-predicted labels. We included molecular layer interneurons only if they showed spike-triggered inhibition of an identified Purkinje cell's simple spikes at short latency, leaving some potential molecular layer interneurons out of our sample. We included units as putative mossy fibers only if the waveform showed a negative after-wave, characteristic of recording near a single glomerulus.[45,46] We note that our classification of mossy fibers is highly conservative and likely leaves a large subset of mossy fiber recordings not near a glomerulus as unlabeled. Putative Golgi cells were identified by their presence in the granule cell layer, broad waveforms, and highly regular firing, consistent with previous recordings.[50,51] Expert labeling of units in the monkey were performed before collection and analysis of the ground-truth units in the mouse.

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

Classifier design

We began the design of our cell-type classifier by selecting its "hyperparameters": the feature space passed to the model, the model class, and model characteristics such as number of units and learning rate. Our decision to select hyperparameters independently from the ground-truth dataset was critical to ensure generalizability by minimizing overfitting.[118]

General classifier architecture

For generating the inputs to the classifier, we used two pretrained autoencoders to reduce the dimensionality of the waveforms and 3D-ACGs of the neurons in the ground-truth library. The output of the two autoencoders, along with the layer of each neuron, served as inputs used to train the final classifier on the ground-truth dataset. As detailed below, no aspect of the model's feature space or its architecture was chosen based on the model's performance on the ground-truth dataset. Our classifier is a "semi-supervised" model because it uses two variational autoencoders that were tuned and pretrained with *unsupervised learning* on a set of unlabeled neurons while the complete classifier was trained with *supervised learning* on a separate set of ground-truth identified neurons. We derived our strategy from the "M1" model.[119]

Feature engineering

To construct an unbiased feature space to train the model, we decided *a priori* that the model's inputs would be anatomical location, extracellular waveform, and firing statistics present in the 3D-ACG. We elected not to use summary firing statistics because they provide an impoverished set of information compared to the inputs we selected.

Unsupervised pre-training with variational autoencoders

We used two variational autoencoders to reduce the dimensionality of the feature space and optimize the model's architecture fully independently from our ground-truth dataset by leveraging n=3,090 curated but unlabeled units that were recorded in the experiments used to create the ground-truth library but were not activated optogenetically.

We pretrained the two autoencoders to reconstruct the waveforms and the log-scaled 3D-ACGs of our unlabeled units. Ultimately, the autoencoders compressed the input data into two 10-dimensional 'latent spaces' for the 2 input features: 3D-ACG and waveform. The latent spaces had Gaussian priors that encouraged each network unit to have activation values with zero mean and unit variance across the unlabeled dataset. The training objective of the variational autoencoders was the 'Evidence Lower Bound' loss[120] modified to include a β term to encourage disentanglement of the latent space.[121] During training, we employed a Kullback–Leibler divergence annealing procedure to enhance model stability and convergence.[122] Both variational autoencoders were trained through gradient descent with the Adam optimizer, complemented by a cosine-annealing learning rate strategy with periodic warm restarts.[123]

To both facilitate model convergence and yield high-quality reconstructions, we manually tuned the variational autoencoder parameters to adapt the model to our specific data characteristics and improve its performance in subsequent tasks. It is important to note that the parameters of the autoencoders were optimized based on the quality of the reconstruction of features from neurons in the unlabeled dataset, completely independent from the ground-truth dataset.

- The *waveform variational autoencoder* consisted of a 2-layer perceptron (2LP) encoder with Gaussian Error Linear Units (GeLU) non-linearities[124] and a 2LP decoder also with GeLU non-linearities. It was trained for 60 epochs with η=1e-4, β=5 and a mini-batch size of 128.
- The *3D-ACG variational autoencoder* consisted of a 2-layer convolutional neural network (CNN) encoder with average pooling after convolutions, batch normalization, and rectified linear unit (ReLU) non-linearities, and a 2LP decoder with ReLU non-linearities. It was trained for 60 epochs with η=5e-4, β=5 and a mini-batch size of 32.

Semi-supervised classifier

The complete classifier model consisted serially of: (1) the waveform and 3D-ACG variational autoencoders pretrained on unlabeled data to reduce the dimensionality of the input features; (2) a multi-headed input layer that accepted the latent spaces of the waveform and 3D-ACG variational autoencoders, along with a "one-hot-encoded" 3-bit binary code of the neuron's cerebellar layer; (3) a single fully-connected hidden layer with 100 units that processed the 3 inputs; (4) an output layer with one output unit for each of the 5 cell types. The values of the 5 output units sum to 1 via a softmax function so that the output of the classifier is the probability that a given set of inputs is from each of the 5 cell types. Between the input layer and the fully-connected hidden layer, we applied batch normalization[89] to equate the contributions of waveform, discharge statistics, and layer.

Supervised training procedure

We trained the weights of the complete classifier on the data in the ground-truth library using gradient descent with a leave-one-out cross-validation strategy. We trained the models until convergence or for 20 epochs, whichever came first, with η=1e-3, a mini-batch size of 128 and the AdamW optimizer.[123] We allowed the weights in the pre-trained variational autoencoders to change during optimization to allow fine-tuning that caused a small improvement in performance on the downstream classification task.

Strategies to mitigate overfitting

We adopted 6 regularization procedures to minimize the risk of overfitting.

1. We reduced input dimensionality to minimize the number of parameters in the ultimate classifier that needed to be trained *de novo* by pre-training two variational autoencoders using an unsupervised procedure (see above).[120]

2. We fixed the hyperparameters of the classifier before beginning the training procedure.
3. We used data augmentation strategies that are designed to regularize highly expressive models such as deep networks. We built a total of 7 custom data augmentations, 2 specific to the waveforms and 5 specific to the spike trains.
4. We used dropout layers within the classifier to ensure that the model does not rely too heavily on any single feature or neuron.
5. We employed early-stopping by interrupting training of our models at the earliest signs of convergence on the training set.
6. Our final model is an ensemble that combines the predictions of multiple models to produce its output and helps to reduce the variance associated with individual models. The ensemble approach mitigates overfitting by averaging out the errors of the individual models.

The fact that the classifier agrees well with the classification of experts on independent datasets from both monkey and mouse provides additional evidence that the classifier is not significantly overfit.

Assessment of autoencoder operation

To generate the graphs in Figure S5D and evaluate whether the variational autoencoders in the classifier could represent a broad range of input statistics in their latent space and capture the variance in our data, we provided novel inputs to the decoding half of each autoencoder and visualized the resulting reconstructions.[120] Because the values supplied to the decoder could change in 10 dimensions (corresponding to the 10-dimensional latent space), to visualize the reconstructions we identified the two dimensions in the latent space that accounted for the majority of the variance across our training dataset. We utilized the encoder network to encode the waveforms and 3D-ACGs of 3090 unlabeled units onto a latent space. Within this space, we then performed principal component analysis to identify the two components that accounted for the most variance. We generated novel inputs to the decoding half of the autoencoder via the weighted sum of these two principal components.

The location of each reconstructed waveform and 3D-ACG in the 8x8 matrices in Figure S5D corresponds to the relative position of the weights applied to the first (horizontal axis) and second (vertical axis) principal components. The weights were chosen as the octiles of a mean-zero Gaussian distribution whose standard deviation was chosen to yield a representative distribution of the waveforms and 3D-ACGs observed across these principal components. We chose the weights from a Gaussian distribution because the distribution of activations across each unit in the latent space was encouraged to be a zero-mean Gaussian through the assigned standard normal prior, but other choices of weights across a similar range of values would yield qualitatively similar results. We reasoned that if the autoencoders were overfit or if the set of waveforms and 3D-ACGs in the unlabeled dataset were insufficiently diverse, we would observe discontinuities in the reconstructed outputs rather than smooth transitions that could accommodate a wide range of input statistics.

Evaluation of classifier performance

Cell-type classification of ground-truth neurons

We took multiple steps to evaluate the performance of the classifier on the ground-truth library and ensure that it generalizes well across datasets:

1. To account for "class imbalance" created by the different number of neurons in each cell type, we performed random oversampling of the under-represented cell types for every model after splitting into testing and validation data.[125]
2. We assessed the performance of all models through leave-one-out cross-validation, which has a lower bias and comparable variance to other cross-validation methods[126,127] and has been used in the past to assess performance on small datasets.[19]
3. We adopted a strategy to prevent *confidence miscalibration*, the tendency of deep neural networks to exhibit over-confidence in their predictions.[128] We corrected the overconfidence of each model instance by applying a last-layer Laplace approximation to the output layer.[129,130]
4. We quantified classifier confidence by averaging the predicted probability for each cell type across the 10 instantiations of the model. We computed the *confidence ratio* as the ratio of the highest- to second-highest predicted cell-type for the input features from each cell in our samples. We chose a confidence ratio of 2 as the *confidence threshold*, but higher thresholds could be applied to increase confidence in each prediction of cell type. Many ground-truth neurons were assigned high probabilities for the same cell type across 10 instantiations of the model. However, that need not have been the case: if the data for a given unit were compatible with more than one cell type, then the classifier might classify the unit as highly-probable to be cell type #1 in one model instance and highly-probable to be cell type #2 in another instance: the average probabilities across 10 runs of the classifier might be similar and therefore closer to 0.5 for these two cell types, indicative of low classifier confidence.

Cell-type classification of unlabeled mouse and macaque neurons

We predicted the cell type of unlabeled mouse (*Medina*) and macaque (*Lisberger*) cerebellar neurons that were not involved in the classifier training procedures using an ensemble classifier that utilized all ground-truth neurons and initial conditions (202 x 10 = 2020 models in total). Each of the 2020 models made a unique prediction because it was slightly different from the other models due to the use of (i) 10 different initial conditions for training the models and (ii) training on different subsets of neurons because of the 'leave-one-out' procedure. The predicted cell-type of each neuron in the unlabeled sample was chosen as that with the

maximum average prediction across the 2020 models. We applied the *confidence ratio* and *confidence threshold* as we had for the analysis of cell-type classification in the ground-truth library.

Analysis of neural dynamics during behavior

Behavioral paradigms

We leveraged our classifier to investigate the temporal profiles of distinct populations of classifier-labeled neurons during discrete trials in four behavioral paradigms:

- *Häusser*: Data were recorded in mice performing self-initiated locomotion. Five days after surgery, mice began water restriction and habituation to head fixation on the running wheel. Mice were head fixed on the wheel once daily for about 30 min and were rewarded with drops of water for moving forward on the wheel. On the day of the recording, mice typically ran 150 meters in about 1 hour. Locomotion was recorded with a Vision Mako U-130B camera at 100 frames/s and the ipsilateral forepaw of the animals was tracked with DeepLabCut[131]. Forepaw swing onsets were identified by thresholding the paw trajectory in polar coordinates (phase obtained from Hilbert transform) and used to align neural data.
- *Hull*: Mice were water restricted for at least five days prior to experiments and then habituated to head fixation and reward delivery for 6 days on a freely moving wheel. A tube for reward delivery was placed in front of the mouse with an IR LED and photodiode positioned to detect licks. Reward consisted of water sweetened with saccharin (10 mM), delivered every 23-85 seconds via an audible solenoid, for a total of 218 trials. We aligned the simultaneously-recorded responses of neurons in a single session to the time of the solenoid click that cued reward delivery.
- *Medina*: Mice were trained using a classical conditioning protocol in which the presence of an LED (conditioned stimulus) predicted the occurrence of a puff to the cornea (unconditioned stimulus) 220 ms later. The exact experimental protocol has been described previously.[65] Animals had been extensively conditioned and were generating reliable conditioned responses before neural recordings took place. We aligned the simultaneously-recorded responses of many neurons from a single session on the onset of the conditioned stimulus.
- *Lisberger*: We recorded neural data during discrete trials of smooth-pursuit target motion.[108] Animals were seated in front of the CRT monitor and trained to pursue the smooth movement of a black dot as it moved in one of eight directions at a constant velocity. Here, we included only trials in which the target moved in the horizontal direction towards the side of the cerebellum where we were performing neural recordings and where the monkey successfully tracked the target and maintained fixation after the termination of target motion. Data were aligned to the onset of a 650 ms duration target motion.

Neural data trajectory analysis

We included in our analysis each recorded neuron that our classifier was able to label with a confidence ratio > 2, analyzing only a single session for each of the 3 behaviors in mice and a pseudo-population recorded across n=163 sessions for the behavior in monkeys, similar to previous analyses of population dynamics.[70,74] We converted spike trains into firing rates and temporally smoothed them using causal kernels that were appropriate for each behavioral paradigm (10–50 ms smoothing time constants). The differences in time-scales of the behavioral tasks (e.g., approximately 200 ms during reward conditioning but longer than 1,000 ms during smooth pursuit) necessitated the use of task-specific smoothing methods. We note, however, that our general conclusions are robust to a large range of smoothing time constants. Following smoothing, we normalized the firing of each neuron by the standard deviation of its activity during a 100 ms pre-trial baseline period, averaged the responses across discrete trials to form peri-stimulus time histograms for each neuron and subtracted the mean firing rate during the pre-trial baseline period. Changes in the normalized firing rate are therefore expressed relative to their baseline activity in units of baseline firing rate standard deviation. The PSTHs in Figure 7B are averaged across all instances of each cell type and the heat maps in Figure 7C show the individual trial-averaged PSTHs for each cell type in the population.

A more sophisticated approach for studying the temporal structure of large-scale population responses relies on dimensionality reduction techniques to understand consistent temporal signatures that exist across neurons in a population.[70–72,74,80] To answer whether low-dimensional representations of population activity from specific cell types were different from those of a cell-type agnostic population, we derived an analysis pipeline where we could compare the neural trajectories from different sized populations in a common space. In each case, we began by constructing a matrix X with dimensions NxT where each row contained the baseline normalized and smoothed peri-stimulus time histogram with T time points for one of N neurons. To apply principal component analysis, we centered the firing rates in each row of X. As principal component analysis identifies the dimensions in order of variance, neurons that show dramatic differences relative to their baseline firing rate will likely drive a majority of the variance. To mitigate both the over-representation of highly responsive neurons in the absence of normalization as well as the over-representation of non-responsive neurons in the case of z-scoring, we took advantage of a previously described approach for firing rate preprocessing using a "soft" normalization procedure.[70] We reduced the influence of highly active neurons, those with modulation that exceeded their baseline standard deviation by more than 2x, by ensuring that their range was approximately unity. The range of firing rate changes for non-responsive neurons, in contrast, was reduced to values near zero and, therefore, did not contribute meaningfully to the largest principal components.

Please cite this article in press as: Beau et al., A deep learning strategy to identify cell types across species from high-density extracellular recordings, Cell (2025), https://doi.org/10.1016/j.cell.2025.01.041

We decomposed our data matrix $X^T = U\Sigma V^T$ using singular value decomposition. Here, U represents the left singular vectors of X^T, consisting of an orthonormal set of temporal representations of length T (principal axes across time) ordered by their variance. Typically, principal component analysis projects the data matrix X into a lower dimensional subspace by selecting the first C axes of U, weighted by their associated singular values contained in the diagonal elements of Σ, namely $U^{(1:C)}\Sigma^{(1:C)}$. While each directional axis defined in U has unit magnitude, the resultant projection is scaled by the singular values whose values are related to the standard deviation, σ, of the discovered principal axes and thus have a dependence on the population size, N. To account for the dependence of the projection into principal component space on the population size, we transformed the singular value matrix as $\widehat{\Sigma} = k\Sigma / tr(\Sigma)$. In this equation, k is an arbitrary constant (chosen to be $10\sqrt{T}$) and tr denotes the trace operator. We can now project X into principal component space in a manner that retains the relative scaling between the projected axes but remains agnostic to the underlying population size using $U\widehat{\Sigma}$. Given this formulation, we could compute the projection of each population into their respective principal component spaces even though the population sizes differed. Once we obtained our projection $U\widehat{\Sigma}$, we subtracted off the mean of the activity in each principal component during the pre-trial baseline period. Therefore, all trajectories are guaranteed to start near the origin.

We next asked whether the low-dimensional geometry of each cell-type population demonstrated comparable trajectories to the label-agnostic population. To account for the fact that the sign of each principal component projection is arbitrary, we optimally rotated and reflected the low-dimensional trajectory representations of each cell-type population to align its trajectory to the low-dimensional projection of the full population, agnostic to cell-type. We did so by solving the orthogonal Procrustes problem, resulting in an orthogonal matrix R that reflected and rotated cell-type population trajectories about the origin to maximally agree with the cell-type agnostic trajectory, thereby finding the R that minimized $\|U_{Full}\widehat{\Sigma_{Full}} - RU_{Partial}\widehat{\Sigma_{Partial}}\|$. After computing R for each population, we analyzed the Euclidean distance between the optimally rotated cell-type trajectories and the cell-type agnostic population trajectories in the same low-dimensional space.

Because the population sizes were unequal across cell-types, we used permutation testing to assess statistical significance of the distance between optimally rotated and reflected trajectories of each cell type and to the cell-type agnostic trajectories derived from the overall population. We tested the null hypothesis that random selection of equal numbers of cells for each of the compared trajectories would show a similar distribution of Euclidean distances. We performed 1,000 permutations where we sampled neurons from both compared populations, optimally rotated and reflected these permuted trajectories into the same space, and then derived the null distribution of distances. From the empirically derived null distribution we could directly assay the significance of distances between two population trajectories.

We also tested the performance of the trajectory analysis when we introduced random errors in the cell-type labels provided by the classifier. For each given fraction of cells relabeled, we performed 1,000 replicates of randomly selected sets of cells to receive a new cell-type label from labels corresponding to either a Purkinje cell simple spike, Golgi cell, mossy fiber, or a molecular layer interneuron, with equal probability. We then selected all neurons with a given cell-type label, performed dimensionality reduction as described above, and computed the distance between the optimally rotated and reflected low-dimensional trajectory and the trajectory agnostic to cell type labels. We used bootstrapped statistics to derive the 95% confidence intervals as a function of the fraction of cells that were randomly assigned a new cell-type label.

CellPress

Supplemental figures

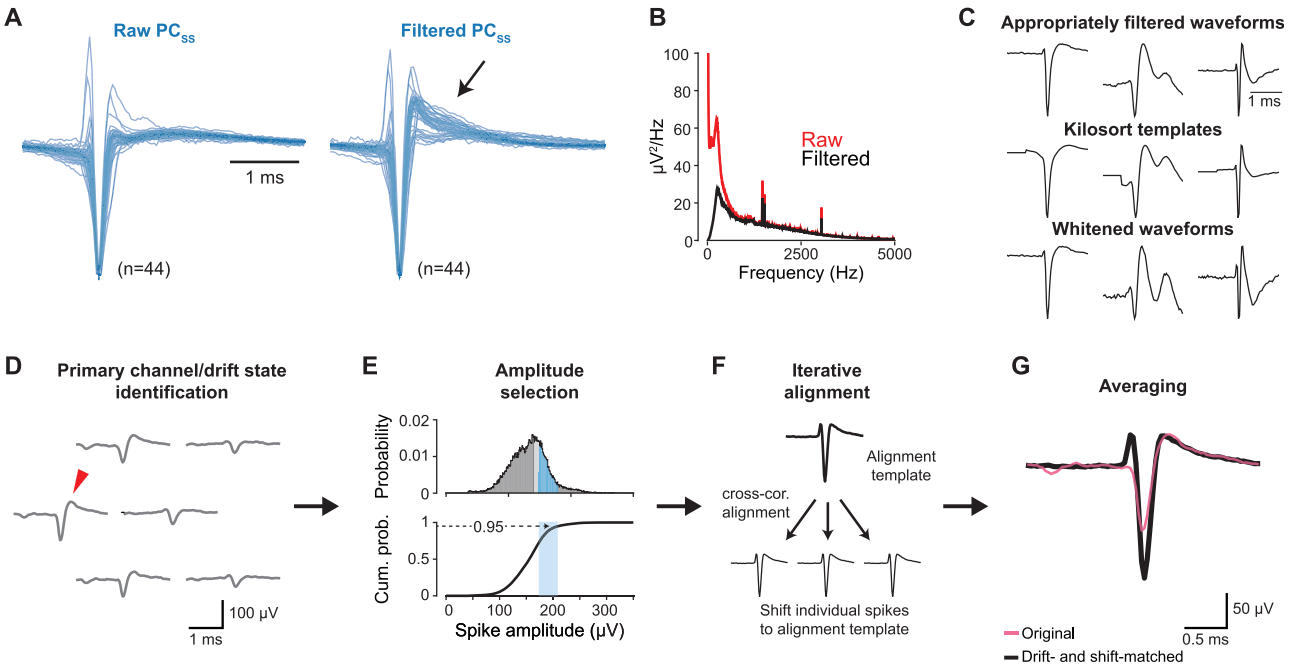


Figure S1. Analysis pipeline for preprocessing of neural waveforms, related to Figure 2

As we compared recordings across labs and understood how the idiosyncrasies of different analysis pipelines work, we learned that we needed to harmonize data across preparations and to ensure that the output of our pipeline provided the best possible estimate of the actual waveform. Careful preprocessing of the waveform reduces the variation in waveform within each cell type and likely improves the performance of the classifier at distinguishing cell types. We detail those procedures—consistent filtering and drift-shift matching—here.

(A) Simple-spike waveforms from 44 ground-truth Purkinje cells, identified through their complex-spike-induced pause in simple-spike firing. Traces on the left display normalized voltage-versus-time traces from the primary channel of Neuropixels recordings with the hardware filter disabled. Traces on the right depict the same waveforms after application of a software filter that is equivalent to the onboard hardware filter on Neuropixels probes (single-pole 300 Hz high-pass Butterworth). The black arrow points out the main difference in waveform shapes before versus after filtering. The procedure to harmonize waveforms across datasets by applying a software version of the hardware filter was critical.

(B) Representation of power in the low-frequency band averaged across the primary channels for raw (red) and filtered (black) waveforms from (A), verifying the impact of the onboard hardware filter.

(C) Comparison of waveforms for three cerebellar neurons, showing how the standard tools used with Neuropixels probes can cause aberrations in the waveforms: top row shows the best reconstruction of waveforms of 3 units, using the analysis pipeline developed in our study; middle row shows the templates created by Kilosort, which can be quite distorted relative to the best waveform identified by our process; and bottom row shows the results of zero phase component analysis (ZCA) whitening by Kilosort, a process that is relevant to performing spike-sorting but that can badly distort the waveform because of the role of activity on neighboring channels in ZCA whitening. The distortions of some waveforms by the standard analysis pipeline underscore the improvements we have made to provide classifier inputs of the highest quality.

(D) Step 1 of the preprocessing pipeline for high-quality neural waveform identification using the drift-shift-matching algorithm, Z-drift matching: identification of the primary channel through the largest peak-to-peak amplitude (red arrow).

(E) Step 2, X-Y drift-matching: sub-selection of neural waveforms on the identified primary channel, with the top plot showing the distribution of peak-to-peak amplitudes and the bottom plot depicting the cumulative probability distribution. A subset of N (user-configurable) action potential waveforms with peak-to-peak amplitudes below the 95th percentile is selected (blue-shaded region), eliminating spikes in the 95th to 100th percentiles to mitigate potential large amplitude artifacts.

(F) Step 3, shift-matching: consecutive/iterative alignment of small batches to waveforms via the peak in cross-correlation to an alignment template, computed as the mean of the largest peak-to-peak waveforms following amplitude selection.

(G) Final averages of the aligned waveforms on the primary channel. The black curve depicts the waveform template following the complete drift-shift pipeline, while the red waveform shows the original mean waveform reproduced from (D). The difference between the red and black waveforms shows the impact of our analysis pipeline.

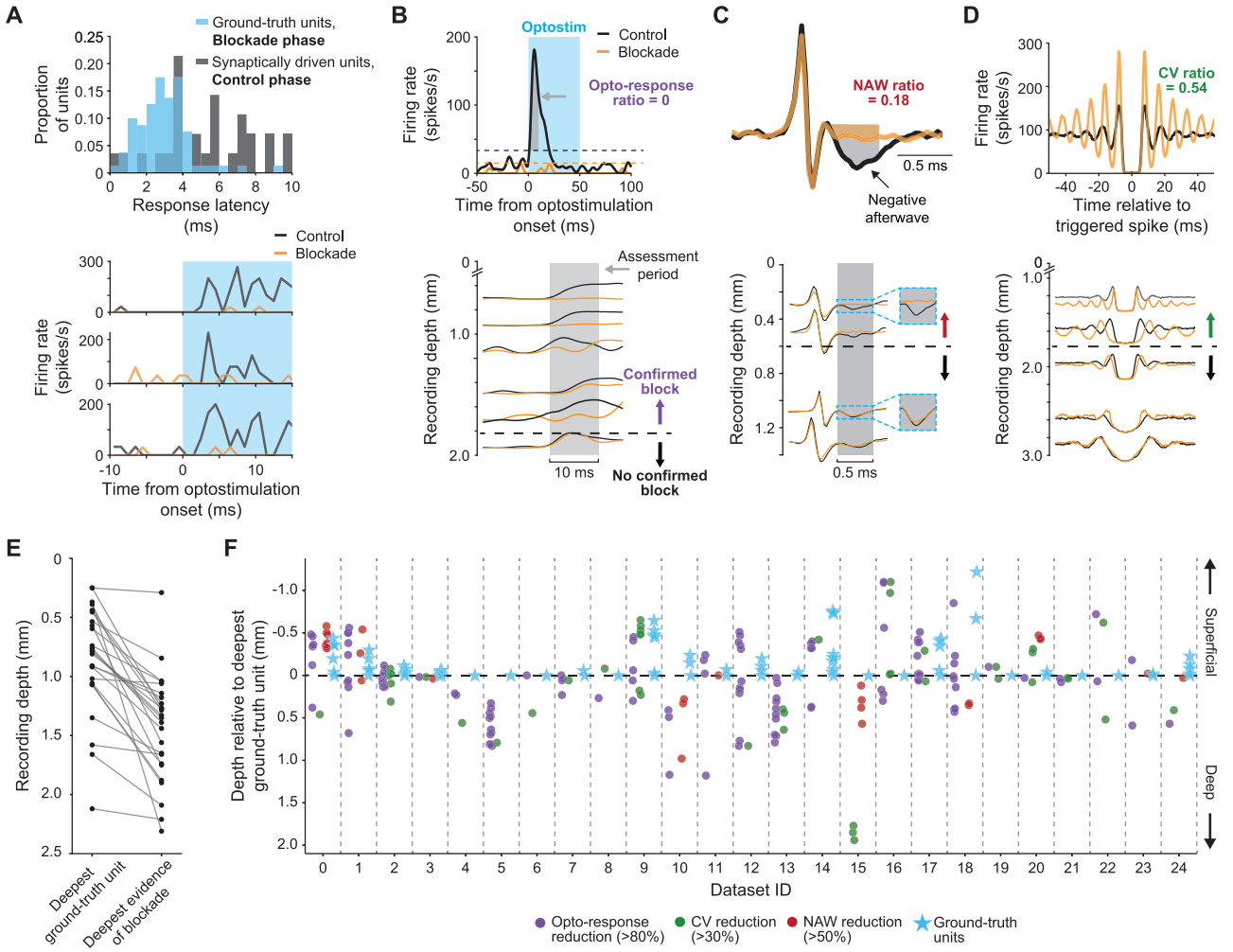


Figure S2. Quantitative assessment of synaptic blockade, related to Figure 3

For each recording that yielded at least one ground-truth neuron, we analyzed all recorded neurons with three quantitative criteria used to establish the region of confirmed synaptic blockade. Criteria were the effect of synaptic blockade on the (i) size of the response to optogenetic stimulation (B), (ii) the size of the negative afterwave on mossy fiber waveforms (C), and (iii) the coefficient of variation of firing rate (D). A recording was considered to potentially contain ground-truth neurons if at least two out of three signs of effective synaptic blockade were found anywhere along the Neuropixels probe. The analysis confirmed that every neuron in the ground-truth library was at or above the deepest recording site with confirmed synaptic blockade (E and F). We conclude that all neurons in our ground-truth library were likely activated directly by optogenetic stimulation and not indirectly through synaptic inputs.

(A) Distribution of optogenetic response latencies. Blue and gray bars indicate latencies for neurons in the ground-truth library versus neurons that were judged to be activated synaptically because they lost their responses to optogenetic stimulation during synaptic blockade.

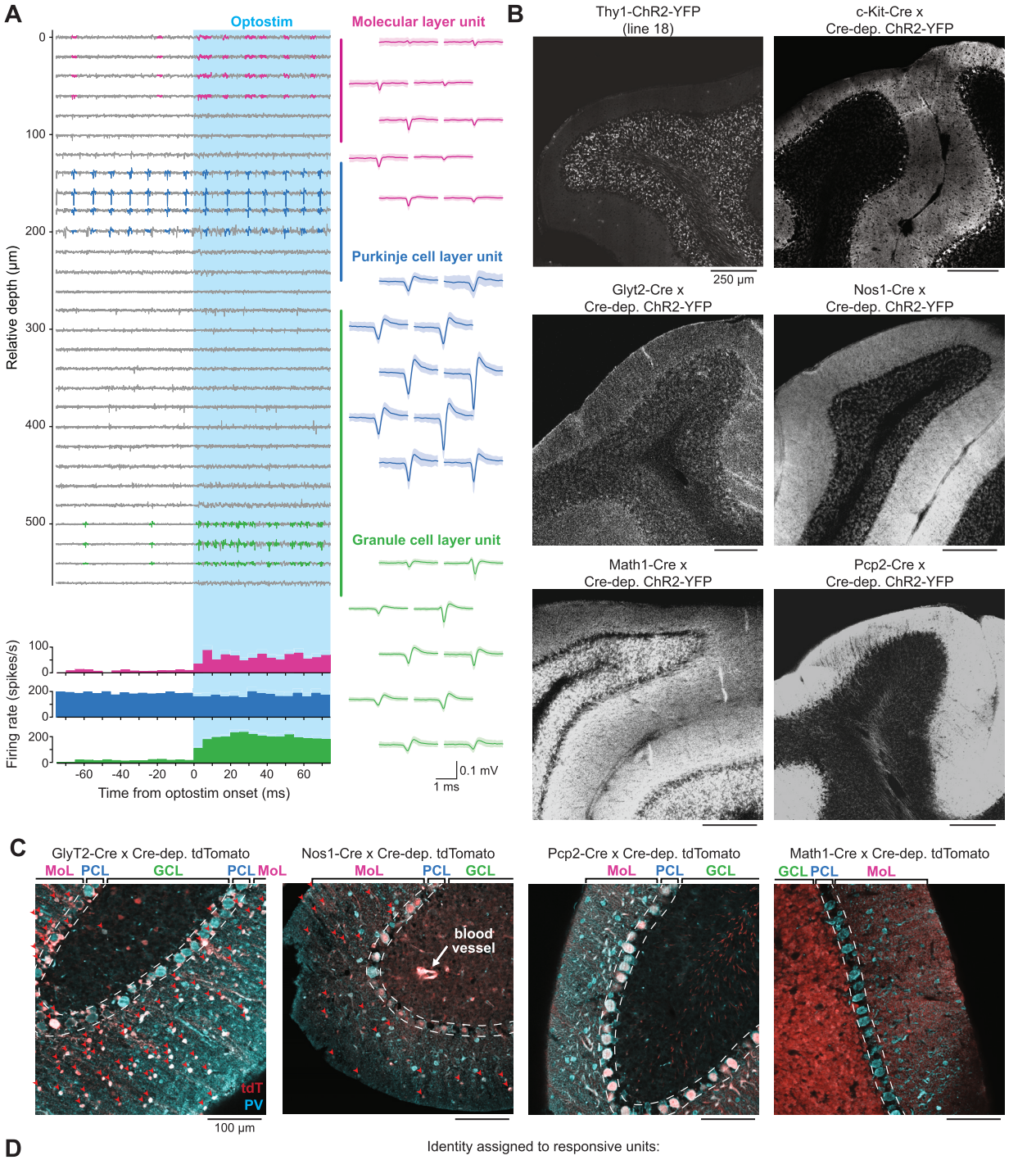
(B) Top panel shows response of an example neuron to illustrate how we quantified the effect of synaptic blockade on responses to optogenetic stimulation. Black and orange traces show data in the control phase versus during blockade, blue shading shows the time of optogenetic stimulation, and gray shading shows the interval used for measurement of the area under the curve. "Opto-response ratio" is defined as the area under the curve in the measurement interval after blockade divided by that before blockade. Horizontal dashed line shows the statistically determined baseline for optogenetic responses. Bottom panel shows responses of multiple neurons along a single recording with a Neuropixels probe and illustrates many neurons that lost their responses to optogenetic stimulation in the region of confirmed synaptic blockade. The horizontal dashed line demarks the line of the deepest evidence of synaptic blockade, here placed at the deepest level where we recorded a neuron that lost its response to optogenetic stimulation during the blockade phase. We defined an opto-response ratio of 0.2 or smaller as evidence for synaptic blockade.

(C) Top panel shows an example mossy fiber waveform to illustrate how we quantified the effect of synaptic blockade on the negative afterwave. Black and orange traces show data in the control phase versus during blockade, and gray shading shows the interval used for quantification of the negative afterwave. "NAW ratio" is defined as the area under the negative afterwave in the measurement interval after blockade divided by that before blockade, in the interval from 1.7 to 2.2 ms after waveform onset. Bottom panel shows effect of synaptic blockade on the negative afterwaves of multiple mossy fiber waveforms above or below the horizontal dashed line that indicates the deepest site of confirmed synaptic blockade. We defined an NAW ratio of 0.5 or smaller as evidence for synaptic blockade.

(D) Top panel shows how synaptic blockade altered the autocorrelogram and the coefficient of variation of an example neuron. Black and orange autocorrelograms show results in the control phase versus during blockade. "CV ratio" is defined as the CV after blockade divided by that before blockade. Bottom panel

(legend continued on next page)

shows effect of synaptic blockade on the autocorrelograms of multiple neurons recorded above or below the horizontal dashed line that indicates the deepest site of confirmed synaptic blockade. The neurons recorded above the dashed line showed regularization of firing rate after synaptic blockade, while the autocorrelograms of those recorded below the dashed line were almost unchanged.[29] We define a CV ratio of 0.7 or smaller as evidence for synaptic blockade.
(E) The graph plots the depth of the deepest ground-truth neuron and the deepest evidence of synaptic blockade for all 24 recordings that yielded at least one ground-truth neuron. Each line corresponds to an individual recording. The negative or zero slopes of all the lines indicate that all ground-truth neurons were within the region of quantitatively verified synaptic blockade.
(F) The scatterplot shows the depth of each neuron that satisfied one of the criteria for effective synaptic blockade for each of the 24 recordings that yielded at least one ground-truth neuron. Cyan stars indicate ground-truth neurons, red symbols indicate mossy fibers where synaptic blockade caused at least a 50% reduction in the amplitude of the negative afterwave, green symbols indicate neurons where synaptic blockade caused at least a 30% reduction in the coefficient of variation, and purple symbols indicate neurons where synaptic blockade caused at least an 80% reduction in the size of the response to optogenetic stimulation. All depths have been plotted relative to the deepest ground-truth neuron recorded in the experiment. In all cases, at least one neuron satisfied the criteria for synaptic blockade at the depth of, or deeper than, the deepest ground-truth neuron.



D

Identity assigned to responsive units:

Layer of responsive unit	Thy1-ChR2-YFP	c-Kit-Cre	α6/Math1-Cre	GlytT2-Cre	Nos1-Cre	L7-Cre
MoL	No labeling	MLI	No labeling	MLI	MLI	Purkinje cell
PCL	Mossy fiber	Excluded (ambiguous)	Excluded (ambiguous)	Excluded (ambiguous)	MLI	Purkinje cell
GCL	Mossy fiber	Golgi cell	Granule cell	Golgi cell	No labeling	Purkinje cell

(legend on next page)

Figure S3. Quantification of potential off-target opsin expression in the mouse lines used in our study and strategy to mitigate it when present, related to Figure 3

Off-target labeling is an important attribute of some of the mouse lines we used. Therefore, we could not simply assume that an optogenetically activated neuron was of the cell type that a line has previously been described to label. This figure describes a strategy to manage off-target labeling based on (i) careful assessment of histology for all the mouse lines we used and (ii) identification of the layer of the recordings by Phyllum.

The problem of off-target expression was most pronounced in the GlyT2-Cre line used previously to image activity in Golgi cells.[33] (C) Shows that the GlyT2-Cre line has substantial off-target expression in molecular layer interneurons and, very occasionally, Purkinje cells; the relative density of molecular layer interneurons was higher than that of Golgi cells (79 versus 20%). Accordingly, we recorded neurons directly responsive to optogenetic stimulation in both the granule cell layer and the molecular layer (PSTH's at the bottom of A). We used Phyllum to identify the recording layers and labeled units in the granule cell layer that were directly activated by optogenetic stimulation as Golgi cells and labeled activated units in the molecular layer as molecular layer interneurons.

The remaining histology in this figure shows that other mouse lines also showed some, but less pronounced, off-target expression. For example, the Math1-Cre line used to label granule cells was generally specific but exhibited rare labeling of Purkinje cells. The c-kit-Cre line also labeled a small number of Golgi cells,[32] and the Nos1-Cre line exhibited occasional labeling of non-neuronal cells in addition to molecular layer interneurons. By contrast, other lines we used were cleaner, such as the Thy1-ChR2-YFP line 18 and Pcp2-Cre lines used to label mossy fibers and Purkinje cells, respectively. Crucially, we did not observe multiple labeled cell types within a single cerebellar layer in any of our lines. Thus, the combination of an identified layer with direct optogenetic activation allowed us to disambiguate cell type for all experiments.

(A) Example optotagging experiment in the GlyT2-Cre line showing raw voltage traces across channels of a Neuropixels probe. Magenta, blue, and green waveforms on the right show the spatial footprint of neurons in the molecular layer (MoL), Purkinje cell layer (PCL), and granule cell layer (GCL). Histograms below the voltage traces show that both the MoL and GCL layer neurons were activated by optogenetic stimulation (cyan shading).

(B) Histology showing the localization of the ChR2-YFP fusion protein in our mouse lines. The top row shows that off-target expression was minimal in the Thy1-ChR2 line used to identify mossy fibers and the c-kit-Cre line used in one laboratory to identify molecular layer interneurons. The second row shows substantial off-target expression in molecular layer interneurons in the GlyT2-Cre line used in one laboratory to identify Golgi cells and non-neuronal expression along with dense labeling in the MoL of the Nos1 line that we used to identify molecular layer interneurons. The third row shows very little off-target expression in the Math1-Cre line used to attempt to identify granule cells and the L7-Cre line that we used to identify some ground-truth Purkinje cells.

(C) Cerebellar sections from Cre lines crossed to a tdTomato reporter mouse immunostained to label Cre-positive cells (red) and parvalbumin (PV; cyan), a marker for Purkinje cells and molecular layer interneurons. Left panel demonstrates that the GlyT2-Cre line drives expression in both Golgi cells in the GCL and molecular layer interneurons in the MoL. Red arrows denote double-labeled cells. Remaining panels show representative sections for the Nos1-Cre, Pcp2-Cre, and Math1-Cre lines.

(D) The table outlines how we used layer information to disambiguate cell types despite off-target expression in certain Cre lines. We used layer information as a sanity check even for Cre lines that had minimal off-target expression, such as the c-kit-Cre line.

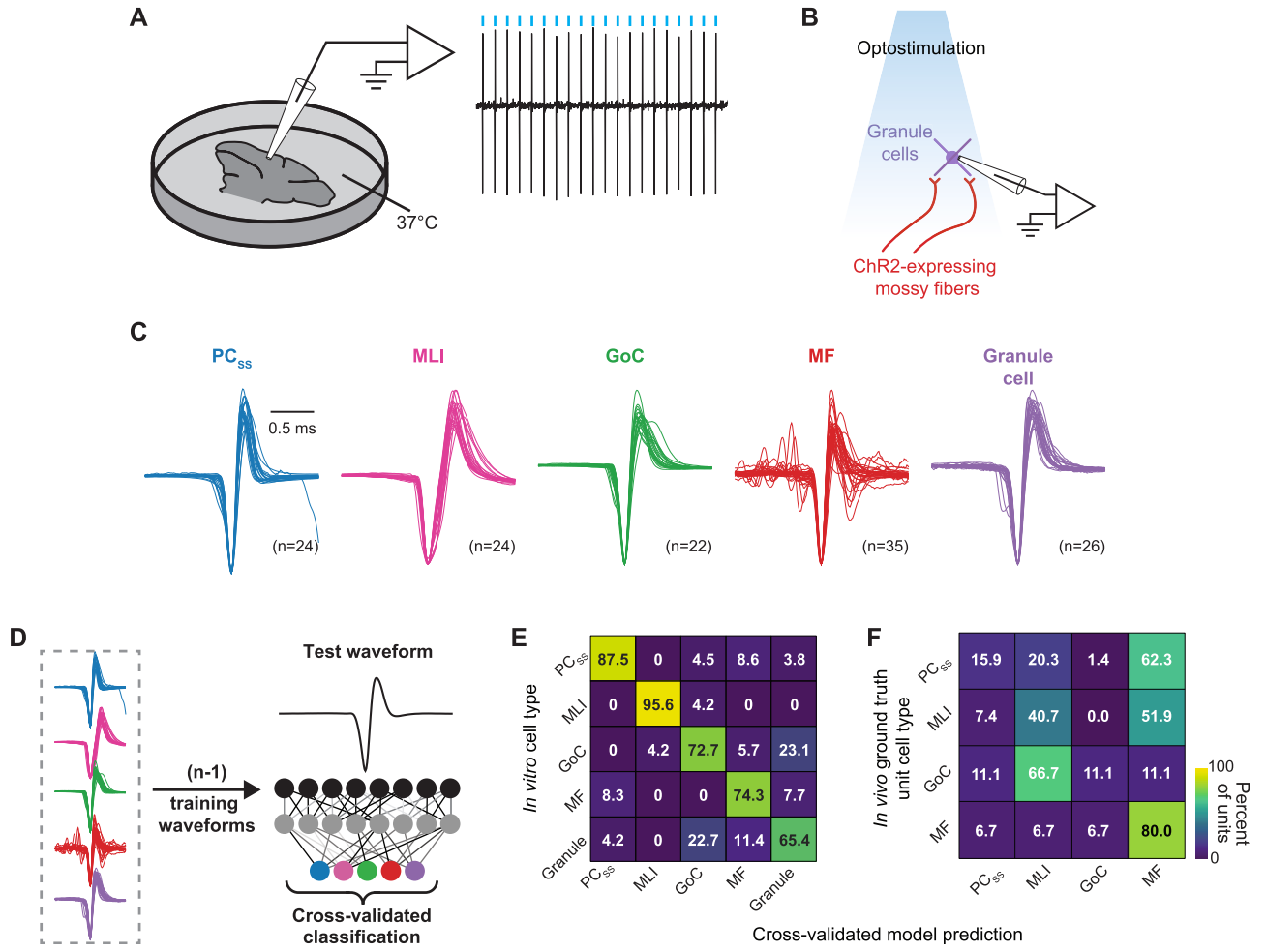


Figure S4. Demonstration from *in vitro* recordings that waveform is informative about cell type, related to Figure 4

We were provoked, by the impression from Figure 4D that waveform from extracellular recordings is a useful indicator of cell type, to perform one more set of experiments to test the informativeness of waveform in a different kind of ground-truth data. We performed cell-attached patch recordings in slices of the mouse cerebellum at 37°C. We identified mossy fibers using optogenetics with the Thy1-ChR2 line 18, and we identified other cell types by visualizing them in the microscope during recording. The currents associated with action potentials were both very uniform within cell types and clearly different between cell types. It is not surprising that the waveforms are more uniform in the *in vitro* recordings compared to the extracellular fields recorded *in vivo* with multi-contact probes. Many uncontrolled factors affect the exact waveform recorded extracellularly. Also, the difference in recording technique and preparation precludes comparison of the waveforms recorded *in vitro* with those in Figure 4D, but the principle that different cell types have distinguishable waveforms remains. We verified the informativeness of waveform from the *in vitro* data by creating a deep learning classifier and validating its performance with a "leave-one-out" strategy. The overall accuracy of the classifier was 78% compared with the 20% expected from random performance. We conclude that there are mechanistic physiological reasons why we can use extracellular waveform as one major feature to classify cell types. At the same time, (F) shows the poor performance of the classifier trained on the *in vitro* waveforms at the identification of cell types from waveforms in the *in vivo* ground-truth neurons. We conclude that we must use the waveforms from extracellular recordings *in vivo* to create a classifier for cell type from extracellular recordings *in vivo*.

(A) Slice recording schematic.

(B) Schematic of optogenetic stimulation of mossy fibers during cell-attached patch-clamp recording. (C) Superimposed waveforms from identified cell types, with different colors showing different cell types: PC_{SS}, Purkinje cell simple spikes; MLI, molecular layer interneuron; GoC, Golgi cell; MF, mossy fiber.

(D) Schematic of a machine learning classifier that we trained to predict cell type based on waveform.

(E) Confusion matrix showing the performance of the classifier on left-out test cell types. The numbers in the entries of the matrix indicate the percentage of cells of a given ground-truth type (y axis) as a function of the prediction of the classifier on the x axis. The diagonal has the highest percentages, meaning that the classifier was accurate: the overall accuracy of the classifier was 78% compared with the 20% expected from random performance.

(F) Confusion matrix showing the poor performance of the classifier trained on the *in vitro* waveforms in identifying the cell types for the waveforms from our ground-truth library, confirming that the waveforms from *in vitro* data cannot be used for cell-type identification from *in vivo* extracellular recordings with multi-contact probes.

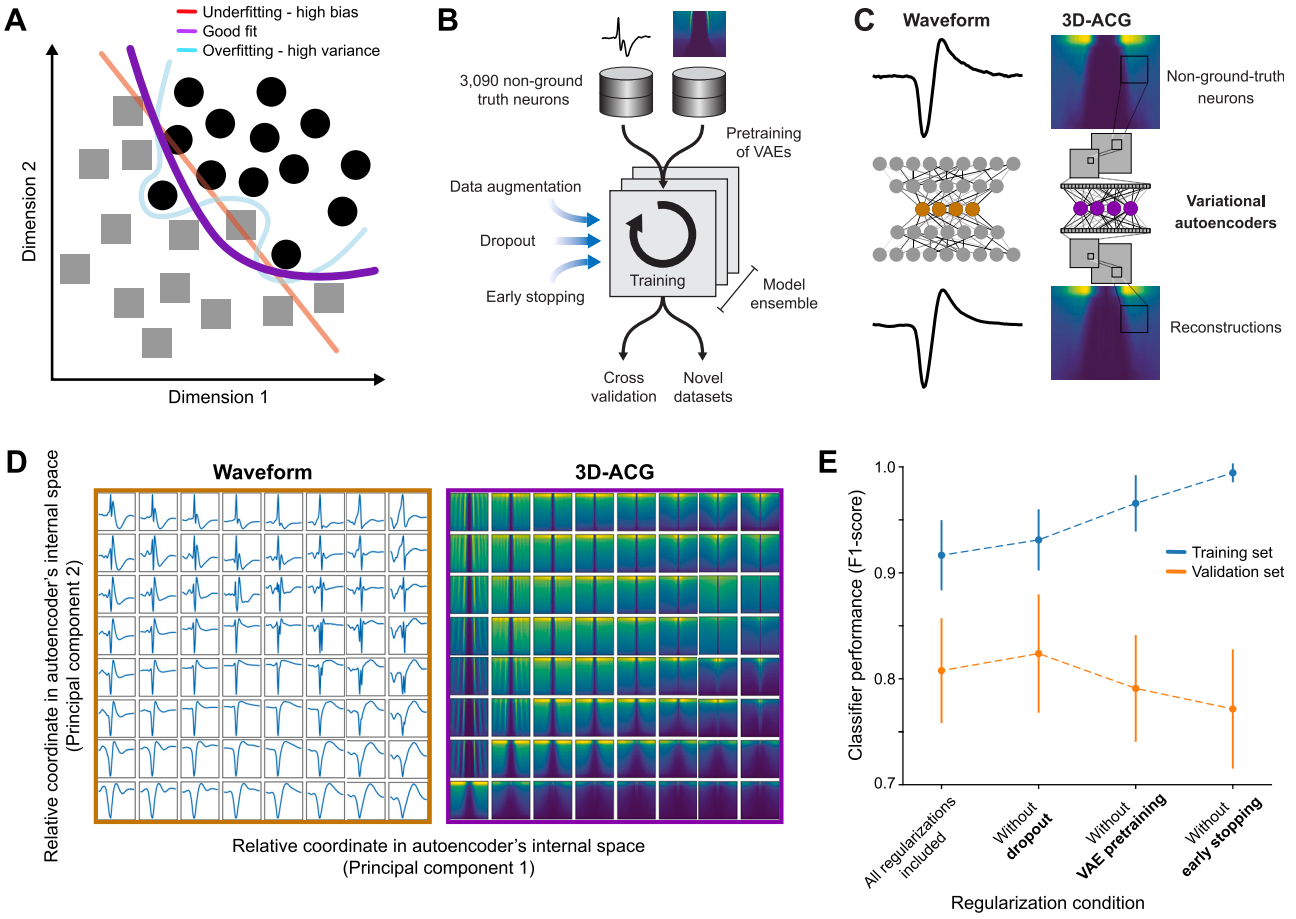


Figure S5. Regularization approaches included in the classifier strategy to mitigate overfitting, related to Figure 5

We were aware that deep learning classifiers are prone to overfitting, particularly in cases where the size of the training dataset is small.[132] Overfitting would result in excellent performance for classification of the ground-truth dataset but poor generalization to novel datasets, including the expert-labeled datasets tested later in the paper. Here, we outline several steps that were designed to avoid overfitting and provide some evidence that those steps were successful.

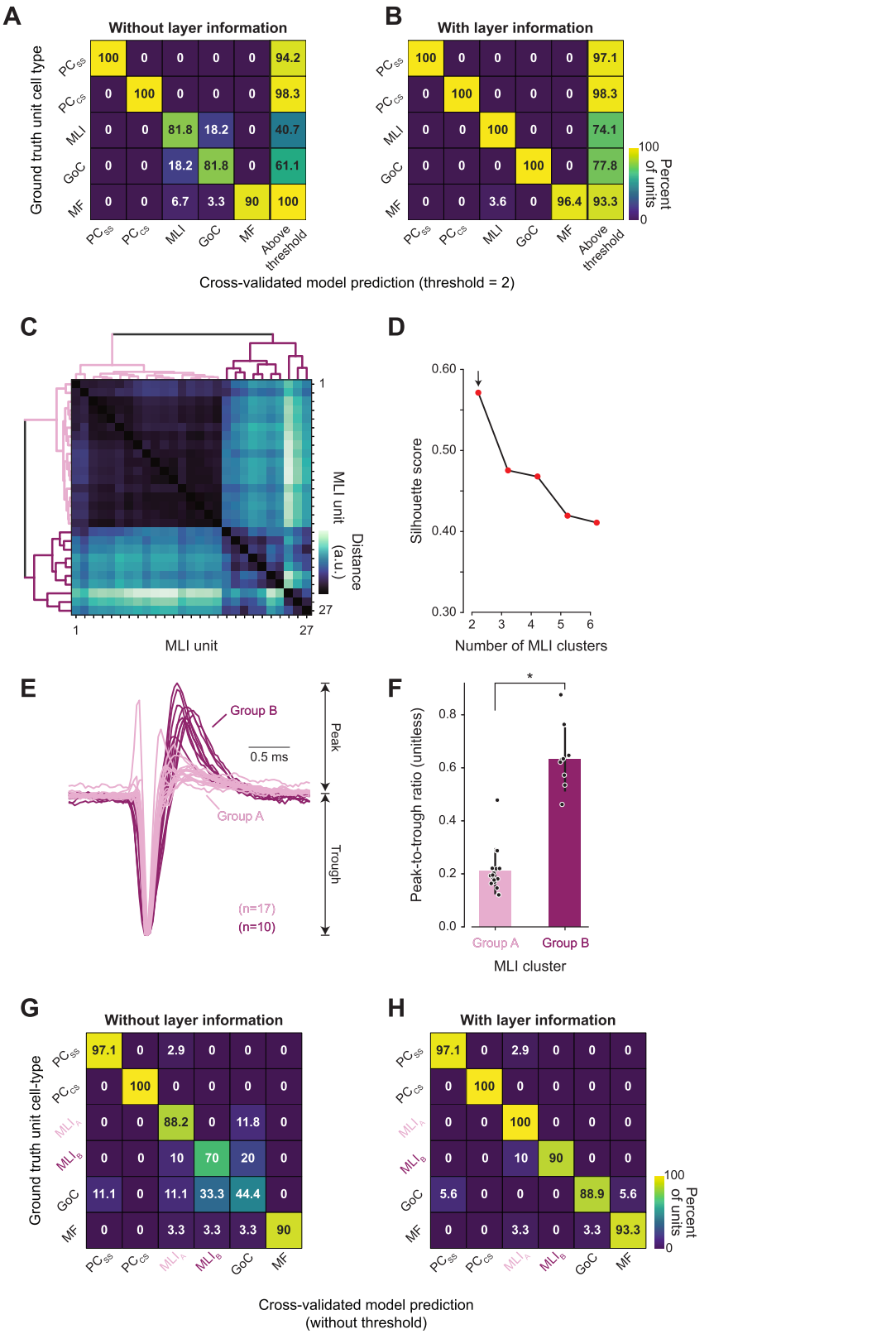
(A) Illustration of the overfitting problem. The two sets of symbols show the distribution of features that might classify two different cell types. The light blue, orange, and magenta curves show the classification that might result from overfitting, underfitting, and proper classification.

(B) Schematic showing the regularization strategies we used to mitigate overfitting, including dimension reduction by input variational autoencoders pretrained in an unlabeled dataset, data augmentation, drop out, early stopping, an ensemble of models, and leave-one-out cross-validation (details in STAR Methods). We also tested classifier performance on independent ("novel") expert-identified datasets, as documented in Figure 6.

(C) Two variational autoencoders, one for the waveform and a separate convolutional variational autoencoder for 3D-ACG, were designed to reduce the dimensionality of their respective inputs to a 10-element vector (colored circles) that could be subsequently supplied as inputs to the final classifier (see Figure 5A). Each variational autoencoder is an artificial neural network that reduces the input dimensionality by placing an information bottleneck ("latent space") between an encoding and decoding network. We trained the weights in the autoencoders with gradient descent (see STAR Methods) using either the waveform (left) or 3D-ACG (right) derived from a set of unlabeled cerebellar neurons. Training minimized the difference between the supplied input and the encoded-decoded output.

(D) The dimensionality reduction performed by the two variational autoencoders captures the diversity of waveform and 3D-ACG statistics present in our ground-truth and expert-labeled datasets (see STAR Methods). Here, we illustrate the diversity of waveforms and 3D-ACGs captured by the autoencoders trained on mouse neurons independent from the ground-truth dataset. The matrices of heatmaps show how waveforms and 3D-ACGs are represented in the latent space of the autoencoders. Most of the low-dimensional representations inside the waveform variational autoencoder were occupied by variations of somatic waveforms, which are indeed the most common in the dataset. However, all other typical spike shapes were also represented, including dendritic waveforms, both bi-phasic and tri-phasic axonal waveforms, and waveforms featuring postsynaptic depolarizations. The same was true for the reconstructed 3D-ACGs, which captured activity profiles corresponding to bursting, oscillations, and both high and low firing rates. Note that some non-biological-looking traces in the figure should not be interpreted as a failure of the model but rather as a by-product of the interpolation process used to visualize the reconstructions.

(E) Post hoc effect of systematically peeling off regularization precautions one by one. As evidence that the regularization procedures mitigated overfitting, the graph shows that peeling them off one by one increases overfitting. It shows a progressive increase in performance on the training set and a decrease on the validation units, quantified using 10 runs of 5-fold cross-validation. Here, classifier performance is quantified by the "F1-score," defined as the harmonic mean of the precision and recall of the classifier.



(legend on next page)

Figure S6. Classifier confusion without layer information as an input and an explanation for the confusion based on the existence of two groups of molecular layer interneurons with different waveform shapes, related to Figure 5

The classifier performs better when we include layer information as an input, especially for molecular layer interneurons and Golgi cells (A and B). When we investigated further, we realized that the waveforms of the full sample of molecular layer interneurons in the ground-truth library suggested a bimodal distribution. Here, we characterize the two groups quantitatively (C–F) and show that the classifier can distinguish between the two groups of molecular layer interneurons. Because one group of molecular layer interneurons (MLI_B) has a waveform very similar to Golgi cells, which are located in the granule cell layer, the classifier performs better with versus without layer information as an input (G and H). The two distinct waveshapes in molecular layer interneurons may align with differences in anatomy or function, but the work of others shows that waveform alone does not discriminate between the two known types of molecular layer interneurons.[133]

(A) Confusion matrix showing the cross-validated performance of the classifier when layer was not provided as input to the classifier. Here, the value in each entry of the matrix shows the percentage of ground-truth cell types on the y axis that were predicted by the classifier to be the cell type on the x axis. Note the confusion specifically between molecular layer interneurons and Golgi cells without layer information.

(B) Same as (A), but showing the improvement when we included layer as an input to the classifier.

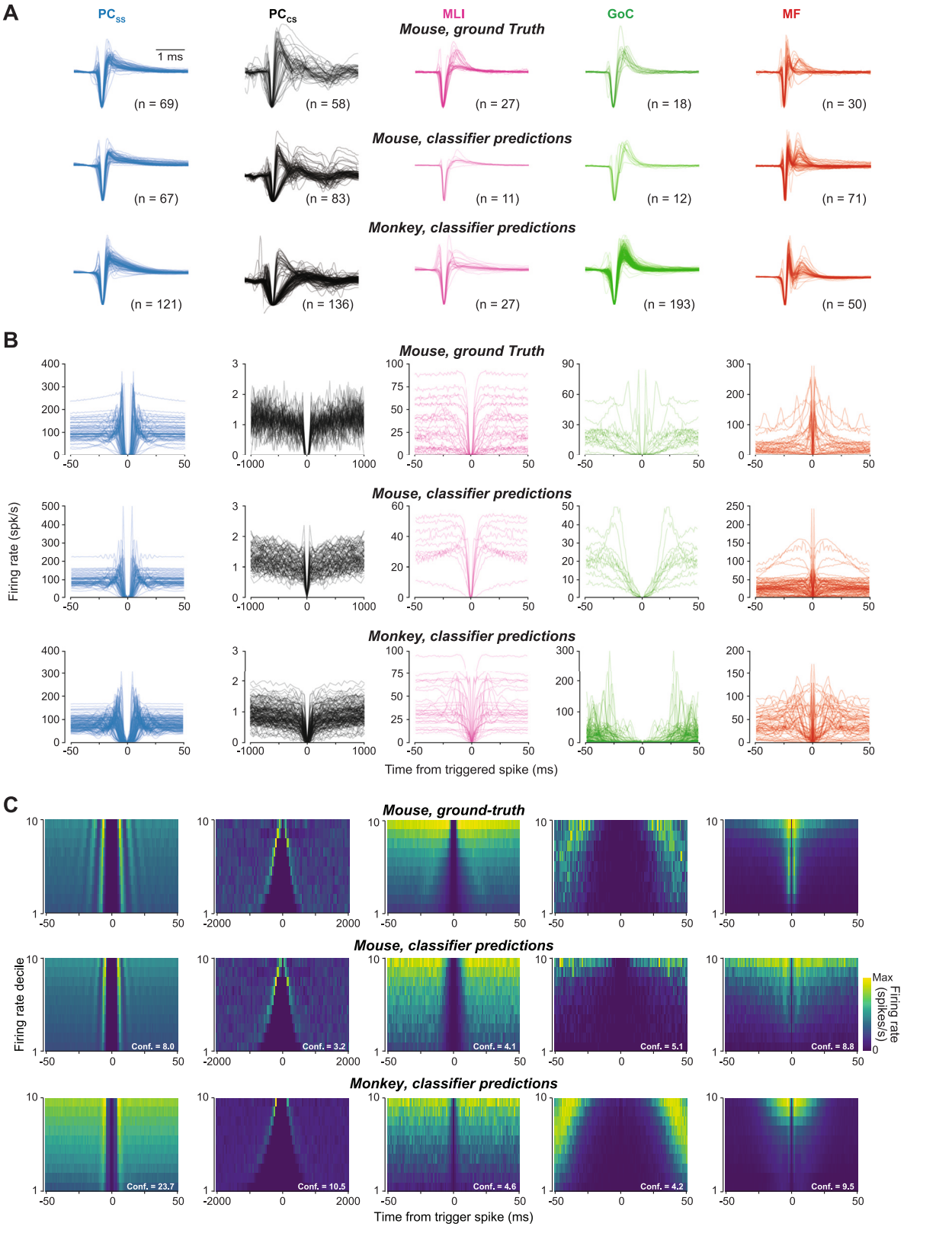
(C) Pairwise distance matrix between normalized waveforms for the entire sample of 27 molecular layer interneurons in the ground-truth library. Lines on the left and top show dendrograms obtained via hierarchical clustering.

(D) Silhouette score as a function of molecular layer interneuron clusters. A high score indicates that a sample matches appropriately to its own cluster and is separated from neighboring clusters. Maximizing the silhouette score provides an unbiased estimate of the number of underlying clusters. Here, the assumption of two clusters maximizes the silhouette score. The maximum silhouette score is +1, and the minimum silhouette score is −1.

(E) Separation of the waveforms of molecular layer interneurons into group A (n = 17) and group B (n = 10), shown in different colors, based on the pairwise distance matrix and dendrograms in (C).

(F) Comparison of peak-to-trough ratios for the two clusters of molecular layer interneurons.

(G) Confusion matrix showing the cross-validated performance of the classifier with separate labels for group A and group B molecular layer interneurons when the layer was not provided as input to the classifier. Here, the value in each entry of the matrix shows the percentage of ground-truth cell types on the y axis that were predicted by the classifier to be the cell type on the x axis. Note that without layer information, the classifier specifically confused molecular layer interneuron group B neurons and Golgi cells. (H) Same as (G), but with layer information.



(legend on next page)

Figure S7. Similarity of waveforms and resting discharge statistics of different cell types across the ground-truth library and the expert-labeled data from mouse and monkey, related to Figure 6

Here, we included only the neurons that were classified with confidence greater than 2.

(A) Waveforms of different cell types across laboratories and species. In the first row, waveforms are divided according to ground-truth cell type in mice. In the second and third rows, cell types are divided according to classifier predictions of cell type for non-ground-truth neurons recorded in mice and monkeys. Recall that all waveforms were flipped so that the largest deflection was negative. Cell-type abbreviations are PC_{SS}, Purkinje cell simple spikes; PC_{CS}, Purkinje cell complex spikes; MLIs, molecular layer interneurons; GoCs, Golgi cells; MFs, mossy fibers.

(B) Same as (A), except showing 2D autocorrelograms. Note that the spike counts in the autocorrelograms have been normalized by the width of the bin so that the y axis is in spikes/s.

(C) Example three-dimensional autocorrelograms (3D-ACGs) for 5 cell types in the ground-truth library from mice and from non-ground-truth recordings in mice and monkeys. For the non-ground-truth recordings, we selected examples where the prediction of the classifier agreed with the experts' identification of cell type.