High-density extracellular recordings from the interior of intact brain organoids enable automated high-throughput functional assay

Fan Wu[1], Cuthbert Steadman[1], Liam Argent[1], Brian Jamieson[1], Shamsudheen Karuthedath Vellarikkal[2], Alex M. Tamburino[2], Iya Khalil[2], Vanessa M. Peterson[2], Matthew E. Kennedy[3], Rebecca Mathew[2]*, Eran Stark[4,5]*, Bilal Cakir[2]*

[1]Diagnostic Biochips Inc., Glen Burnie, MD, USA

[2]Data, AI & Genome Sciences, Merck & Co., Inc., Cambridge, MA, USA

[3]Department of Neuroscience, Merck & Co., Inc., Boston, MA, USA

[4]Medical School, University of Haifa, Haifa 3498838, Israel

[5]Sagol department of Neurobiology, University of Haifa, Haifa 3498838, Israel

*Correspondence: rebecca.mathew@merck.com(R.M.), eranstark@sci.haifa.ac.il (E.S.), bilal.cakir@merck.com (B.C.)

ABSTRACT

Bridging the gap between preclinical screening and clinical outcomes remains a major challenge in drug development for neurological disorders. Brain organoids, derived from human induced pluripotent stem cells, offer a scalable and physiologically relevant platform to model human neural circuits. We develop a fully automated system to record neural activity from the interior of intact human cortical organoids using a high-density microfabricated probe. The robotic system completes insertion within minutes while preserving organoid integrity and enables immediate recording of spontaneous spikes. We extract physiologically grounded and deterministic spike features, and train a long short-term memory classifier to distinguish between organoids derived from healthy individuals and those harboring familial Alzheimer's disease (AD) mutations in the amyloid precursor protein. Despite intra-class variability, the classifier differentiates between organoid classes with 100% accuracy. The model classifies AD organoids treated with a drug candidate that reduces amyloid-β levels as retaining an AD-like electrophysiological phenotype, demonstrating that functional readout can contradict molecular markers. The findings establish a high-throughput functional framework that may complement and extend existing drug screening assays.

INTRODUCTION

Improving the translation of preclinical screening assays into clinical outcomes remains a central challenge in the development of therapies for neurological diseases[1–3]. Recently, in vivo studies have emphasized the importance of circuit-level phenomena in neurological diseases[4–6]. However, traditional animal electrophysiology studies often require time-intensive surgeries and analyses, are difficult to scale, and are unsuitable for drug screening, where automation and high throughput are critical. Effective tools to bridge this gap are lacking.

Brain organoids derived from human induced pluripotent stem cells (hiPSCs) have emerged as a promising in vitro model, offering scalability, genetic fidelity, and the potential to recapitulate human-specific neurodevelopmental and disease-relevant features[7–9]. Unlike animal models, organoid models are versatile, potentially enabling large-scale drug screening that is not feasible with labor-intensive animal studies. Moreover, physiological data from human-based organoids can be more translatable than those from animal models, which often fail to predict human responses[10,11]. Organoids are three-dimensional (3D) structures that mimic diverse brain region identities, establish synaptic networks, and achieve increasingly mature phenotypes[12–17]. Despite these advantages, the functional interrogation of brain organoids remains underdeveloped, particularly with respect to electrophysiology.

Presently, functional electrophysiology in brain organoids relies primarily on patch clamp and surface microelectrode arrays (MEAs)[18–20]. Single electrode patch clamp techniques offer focal intracellular readout from individual cells, but lack scalability in terms of simultaneously recorded cells, limiting their utility for circuit or network-level analysis. While initially restricted to academic usage due to their high complexity and low throughput, patch clamp techniques saw widespread industrial adoption following the development of automated technologies[21,22]. By contrast, MEAs can include thousands of electrodes[23] but the devices are planar and designed for 2D cultures[24], and therefore provide limited signal fidelity when used with 3D structures[25,26]. The employment of MEAs for recording from the depth of organoids requires slicing, dissociation, or adhesives, procedures that compromise tissue integrity[20].

To fully leverage the 3D cytoarchitecture of brain organoids, functional recordings must access their interior non-destructively. Silicon probes microfabricated by advanced semiconductor processes have long been used in systems neuroscience to record from live animals, offering dense, minimally invasive, multichannel access to circuit-level dynamics in deep brain structures[27,28]. While multi-channel depth arrays are indispensable in neuroscience research, their adoption in industrial and translational settings remains limited due to the manual, low-throughput nature of surgeries and the technical burdens of analysis. Several studies demonstrated the feasibility of using silicon probes in organoids, but these efforts relied on manual probe insertion and techniques adopted directly from animal surgeries[19,29,30].

Here, inspired by the analytical power of silicon probes for depth recordings and the transformative impact of automation for patch clamp techniques, we developed a fully automated platform for high-throughput, depth-resolved electrophysiology in intact brain organoids. Our system automates the entire pipeline: from micrometer-precision probe insertion, through spontaneous multi-site spike recording, and all the way to feature extraction and machine learning-based classification at the system level. Here, we demonstrate the utility of this approach by recording from human cortical organoids (hCO) with and without known genetic perturbations, namely isogenic Alzheimer's disease (AD) model carrying the amyloid precursor protein (APP) Swedish (KM0/671NL) and Indiana (V717F) mutations[31]. Despite inherent biological variability, electrophysiological features vary more between organoid classes than within each class. Machine learning classifiers accurately predict organoid class identity from short novel recordings. These results establish the feasibility of a rapid functional assay, providing a framework for integrating electrophysiology-based phenotyping into high-throughput preclinical pipelines.

RESULTS

An automated instrument allows recording from the inside of brain organoids

To enable high-throughput and consistent recordings from within intact brain organoids, we developed a fully automated instrument capable of precisely inserting a high-density neural probe into the organoid interior and acquiring spontaneous electrophysiological activity. The system is housed in a custom chamber that provides electromagnetic shielding, with internal modules for organoid confinement, probe actuation, and recording (**Fig. 1A**).

The core of the system consists of a custom-designed well and a motorized probe head (**Fig. 1B**). The well is 9 mm deep, featuring conical geometry with a 45° taper and a 0.5 mm-diameter base, engineered to capture organoids with diameters ranging 0.5-2 mm. This geometry ensures that organoids dropped from a pipette anywhere within the well will fall to its center by gravity. The slanted walls provide lateral support that confines and stabilizes the organoid during probe insertion and recording, while minimizing physical contact to preserve media exchange and tissue health. Inside the enclosure, the probe head incorporates a microfabricated silicon neural probe (P64-1, Diagnostic Biochips, Glen Burnie, MD) and an integrated amplifier (RHD2164, Intan Technologies, Los Angeles, CA), mounted on a precision motorized single-axis actuator (KMTS50E, Thorlabs, Newton, NJ). After calibration, the assembled mechanical system maintains ≤100 µm lateral misalignment in the X/Y directions, 0.8 µm Z direction steps, and 15 µm repeatability in the Z direction.

The silicon probe comprises four 9 mm-long shanks, each carrying 16 microelectrodes (**Fig. 1C, left**). Recording sites distribute across a 750 µm lateral span (X axis) over the four shanks, and over 300 µm depth (Z axis) per shank (**Fig. 1C, right**). This coverage is well-suited to the typical organoid cytoarchitecture, which is approximately spherical with a diameter of 0.5 to 2 mm, and with differentiated neurons primarily found within a 150-300 µm thick "shell" near the surface (**Fig. 1D, E**). Standard silicon shanks with a uniform cross-section (roughly 60 µm-wide, XZ plane; and 15 µm-thick, YZ plane) are routinely used to penetrate the rodent pia mater[32]. While this form factor allowed penetrating organoids, we found it essential to sharpen the tips along the YZ plane to minimize tissue compression just before the probe tips penetrate (**Fig. 1C right, inset**). The sharpened tips improve recording stability and allow reliable control of probe descent with predictable signal changes in response to incremental depth adjustments.

V_{RMS}-based algorithm enables closed-loop and precise probe insertion into organoids

To optimize electrode placement within active neuronal regions of intact brain organoids, we developed an automated insertion algorithm guided by real-time measurements of the root mean square (RMS) extracellular voltage (V_{RMS}). During the initial phase of insertion, the Z axis manipulator lowers the probe to a depth above the organoid, until all electrodes are fully immersed in media (BrainPhys, STEMCELL Technologies, BC, Canada). At this stage, we establish a baseline V_{RMS} for each recording channel by computing the mean of 100 consecutive V_{RMS} samples (one sample every 100 ms) from band-pass filtered (300-6,000 Hz) signal from that channel. Then, a spike detection threshold is defined for each channel as five SDs above the mean.

In the second phase, the Z axis manipulator advances the probe in 50 µm steps, pausing for 3 seconds at each step to acquire a new set of 30 V_{RMS} measurements. As the probe enters the organoid, the bottom-most electrodes begin transitioning from the media into the tissue and reliably exhibit a characteristic increase in V_{RMS}. Once this transition is detected at the bottom-most electrodes, the step size is reduced to 20 µm, the vertical spacing between adjacent electrodes. An electrode is marked "active" when its updated V_{RMS} exceeds its own out-of-organoid pre-computed threshold. This incremental, depth-resolved approach enables the algorithm to track the progressive recruitment of active electrodes and to halt insertion once the majority of the electrodes become active, indicating optimal probe placement within the neuronal shell.

The elevated V_{RMS} inside the organoid does not require the presence of local spiking activity; it likely reflects the presence of neurons and neuronal processes (axons and dendrites) and their subthreshold signals (**Fig. 1D-E; Fig. S1A**). The higher impedance of the tissue compared with the fluid medium contributes to the effective impedance of the electrodes, increasing the Johnson noise, further increasing the RMS. To isolate these effects from spike contributions, we excluded spike samples (40 samples centered around every spike extremum) and then computed an RMS value within the first 0.1 s window. Log-transformed RMS values from 6,656 electrodes

across 120 recordings were fit with a two-component 1D Gaussian Mixture Model (**Fig. S1B**). One Gaussian (mean=0.366; SD=0.042; n=4,073; weight=0.591) approximates log-RMS distribution of electrodes in the media (outside organoid), while the other Gaussian (mean=0.677; SD=0.157; n=2,583; weight=0.409) reflects electrodes inside organoids (**Fig. S1B**). The clear separation between the two Gaussians demonstrates that V_{RMS} can reliably indicate electrode penetration into an organoid even in the lack of spontaneous spiking.

To illustrate the RMS changes in depth including contribution by spikes, **Figure 1D** shows the probe shank schematic (left), with each electrode aligned to the corresponding voltage time series (middle) and V_{RMS} measurements (right). In this representative recording from an hCO, the four top channels (13-16) remain in the media, while the other 12 channels penetrate the organoid, as indicated by their V_{RMS} values exceeding the predefined thresholds. Notably, channel 12 marks the uppermost electrode inside the organoid, but is too distant from active neurons to clearly detect spikes. Channels 5-11 display the most robust spiking activity, corresponding to a further increase in V_{RMS}. Channels 1-4 are deepest within the organoid, showing sufficiently high V_{RMS} to be labeled as active, but do not monitor spikes. When electrodes are advanced beyond this depth, V_{RMS} occasionally reduces to near-media levels (not shown). We hypothesize that this reduction reflects low-density progenitor regions or necrotic zones near the organoid core. These findings suggest that the most reliable method for targeting active neurons is to maximize the number of electrodes positioned within the outer neuronal shell of the organoids (150–300 μm), using V_{RMS} feedback to avoid overshooting and passing this layer.

As a second example, **Figure 1E** illustrates a recording from an organoid of the same cell line that has been modified to incorporate APP mutations (hCO^{APP}). In the hCO^{APP} preparation, rapid and spatially dense spiking was observed. The activity was concentrated between channels 2-12, in stark contrast to the relatively less active band observed in the hCO (**Fig. 1D**). However, in both preparations, the thickness of the electro-physiologically identified neuronal shell was similar, corresponding to about 200 μm. In both preparations, the RMS-based algorithm was effective in reliably positioning the probe within the active layer, characterized by clear signal boundaries above and below.

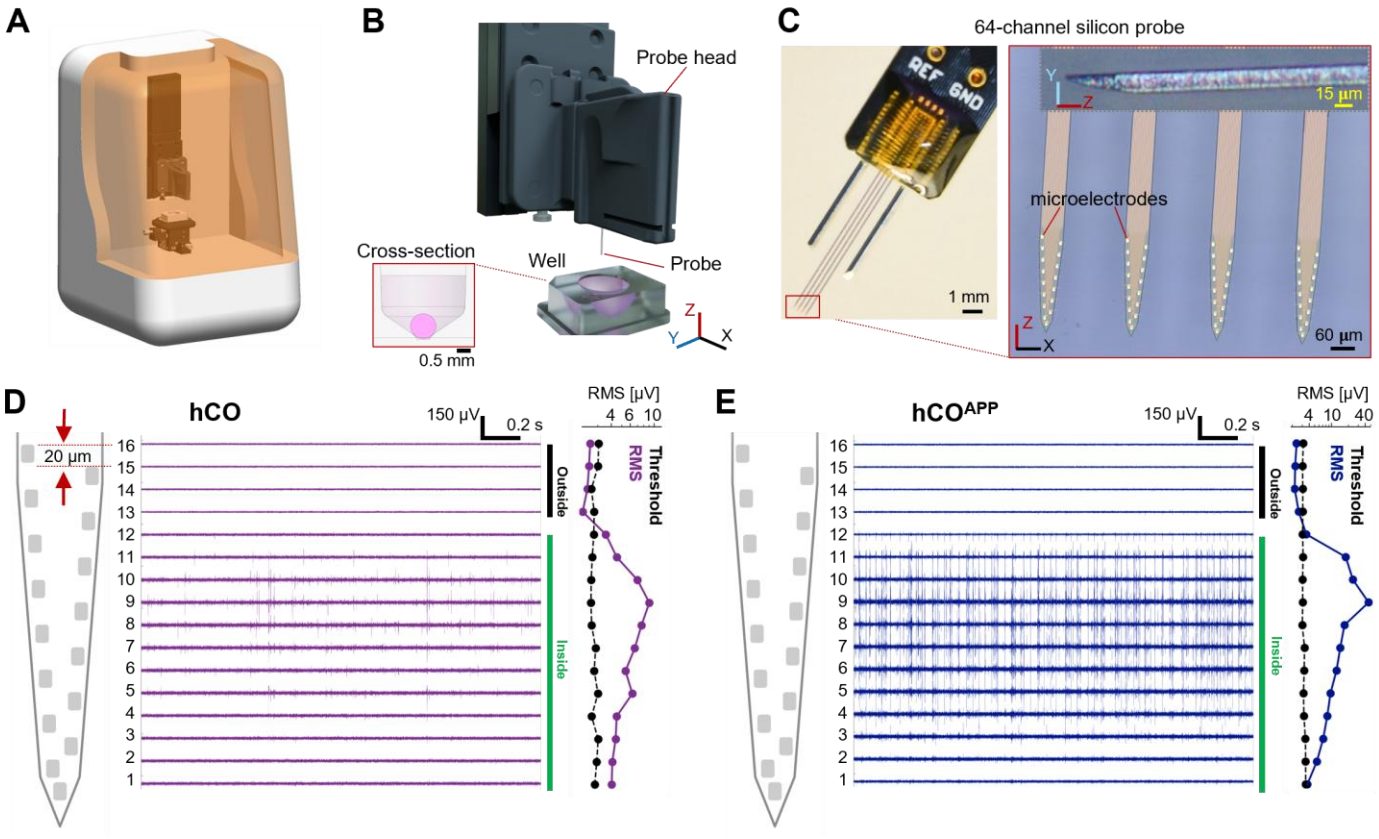


Figure 1. High-throughput instrument automatically inserts a 64-channel microfabricated neural probe and immediately records spiking activity from the inside of intact organoids. (**A**) External design of the instrument chamber. (**B**) Internal design, showing a drivable probe head and a custom designed stationary well. (**C**) Left, Four-shank microfabricated probe assembled on a printed circuit board. Right, Magnified view of the probe tips showing 64 gold microelectrodes integrated on four penetrating probe shanks with sub-micrometer precision. Inset, Sharpened shank tip. (**D**) Recording from a representative human cortical organoid (hCO). Spikes are recorded immediately after insertion and are concentrated on a continuous subset of recording electrodes that are inside the intact organoid. Left, Shank schematic showing 20 µm spacing along the Z-axis between adjacent electrodes. Middle, Bandpass filtered (300-6,000 Hz) traces across one probe shank over a 2 s time window. Right, V_{RMS} (purple) and pre-computed thresholds (black) for every channel distinguish electrodes inside from outside of the organoid. (**E**) Same as **D**, for a representative hCO mutant for the amyloid precursor protein (hCO^{APP}).

hCO variants permit disease modeling and assessment of drug treatments

We generated hCOs using an established protocol[33], and selected distinct variants to test the sensitivity of our electrophysiology platform around day 150 (**Fig. 2A**). Consistent with prior studies[7,34,35], immunofluorescence analysis showed strong expression of the excitatory neuron marker vesicular glutamate transporter 2 (VGLUT2), while the inhibitory neuron marker vesicular GABA transporter (VGAT) remained limited on day 120, consistent with a predominantly excitatory neuronal identity (**Fig. S2A**).

To assess cellular diversity and the maturation process, we performed single-cell RNA sequencing at days 50 and 115. Dimensionality reduction using uniform manifold approximation and projection (UMAP) revealed nine transcriptionally distinct clusters, which we classified as neural progenitor cells (NPCs), cortical neurons (CNs), interneurons (INs), astrocytes (AS), and five other previously described[36,37] populations (**Fig. 2B**; **Fig. S2B**, **S2C**). As a reference, we utilized the datasets from the Human Neural Organoid Cell Atlas (HNOCA) and its computational tools[38], enabling us to automatically define the cell types present in the organoids and further validate their presence. The identity of the annotated cells was highly similar to the automated annotation prediction by the HNOCA model, suggesting that the cell types are indeed relevant to the forebrain **(Fig. S2D**). Thus, the hCO model recapitulates the cellular diversity observed in previously published cortical organoid models, providing an in vitro system for modeling human cortical development and function.

To test disease modeling using our recording platform, we developed hiPSCs carrying familial mutations in the APP (XCL-1_APP Ind (SwHomo)-3G05-E01) associated with early-onset AD[31], and generated a baseline AD hCO (hCO^{APP}). Given that deposition of amyloid beta (Aβ) and neurodegeneration occur sequentially in the human AD brain[39], we examined Aβ accumulation in hCO and hCO^{APP} organoids at different developmental stages (**Fig. S3**). Compared with hCOs, the hCO^{APP} organoids exhibited excessive accumulation of large Aβ oligomers (diameters above 10 μm) at all examined stages (days 50, 90, 120 and 150; $p<0.001$; n=16 hCO and n=16 hCO^{APP} analyzed at each developmental stage; 3 differentiations; two-tailed Mann-Whitney U-test; ***$p<0.001$), with the levels plateauing after day 120 ($p=0.63$, between hCO^{APP} at day 120 and day 150; **Fig. S3**). This profile mirrors early pathogenic features of AD, including the Aβ deposition that precedes neurodegeneration in human brains[40,41]. To generate AD hCOs with oligodendrocytes ($hCOL^{APP}$), we mixed 20% three-transcription factor (3-TF) induced control iPSCs with 80% APP mutation-containing iPSCs (**Fig. 2A**). At day 120, hCO^{APP} demonstrated a higher number of large Aβ oligomers compared with hCO (n=16 organoids; 3 differentiations; two-tailed Mann-Whitney U-test; ****$p<0.0001$). Although $hCOL^{APP}$ also showed significantly elevated Aβ levels compared to hCO (n=16 organoids; 3 differentiations; Mann-Whitney U-test; **$p=0.004$; **Fig. 2C**), these levels were lower than those observed in hCO^{APP} (**Fig. 2C**). This suggests that the presence of oligodendrocytes may alter pathological progression.

To probe therapeutic responsiveness of hCO, we treated a subset of the hCO^{APP} with a beta-site APP cleaving enzyme 1 (BACE1) inhibitor, MK-8277[42]. We conducted electrophysiological recordings after 30 consecutive days of treatment using 10 nM and 100 nM concentrations. Immunofluorescence microscopy revealed a reduction in 6E10 antibody staining, confirming decreased Aβ accumulation in treated samples ($p<0.001$; n=16 organoids in each of the four groups; **Fig. 2D**). These results demonstrate that our disease-modeling platform recapitulates hallmark features of AD pathology a d responds to pharmacological interventions.

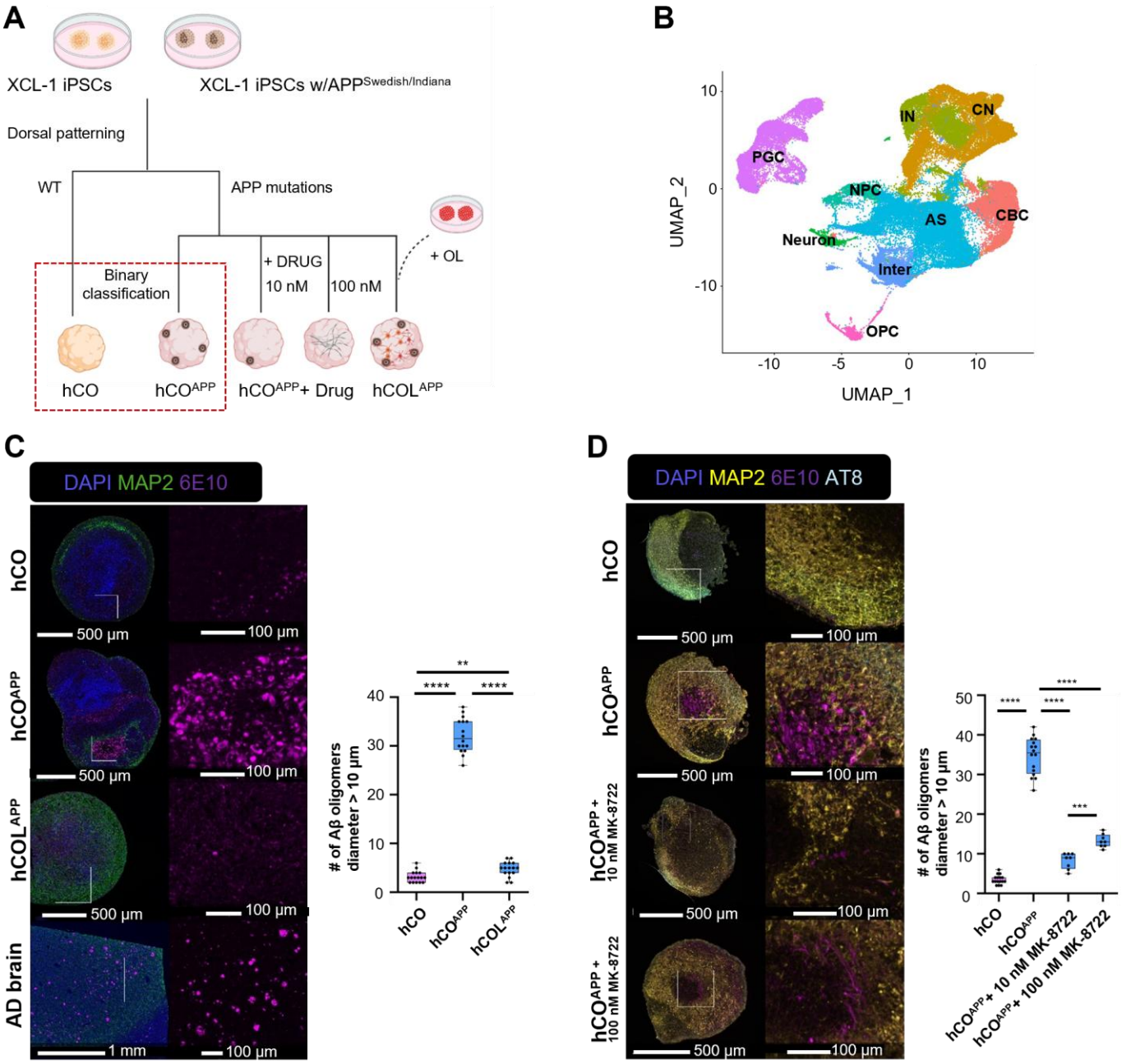


Figure 2. Generation and characterization of human cortical organoid and APP mutants. (**A**) Schematic representation of the generation of isogenic variants of hCOs and the treatments used in this study. hCO and hCO^{APP} are used for the binary classification task. OL, oligodendrocytes. (**B**) UMAP projection of single-cell RNA sampled at day 120 reveals nine distinct clusters, corresponding to distinct cell types, including neural progenitor cells (NPCs), cortical neurons (CNs), interneurons (INs), astrocytes (AS), oligodendrocyte progenitors (OPC), intermediate cells (Inter), proteoglycan expressing cells (PGC), and cilia-bearing cells (CBC). (**C**) Left, Immunostaining of hCO and hCO^{APP} variants at day 120, and AD patient post-mortem brain for neuronal marker MAP and Aβ antibody 6E10. Right, Quantification of Aβ oligomers in organoid variants. (**D**) Left, Co-staining of Aβ (6E10), phospho-tau (AT8) and neuronal marker MAP2 in hCO and hCO^{APP} with and without drug treatments (10 nM and 100 nM MK-8722) at day 150. Right, Quantification of Aβ oligomers in organoid variants. In **C** and **D**, box plots show medians and interquartile range over three independent batches, n=16 organoids in each group. **/***/****, p<0.01/0.001/0.0001, two-tailed Mann-Whitney U-test.

Single unit analysis supports diverse neuronal subtypes and connectivity within intact organoids

To assess the foundational assumption that brain organoids recapitulate key cellular and circuit features of the brain, we isolated multiple single units (ISI index[43] below 0.2) from the high-density extracellular records. **Figure 3A** shows the spatial distribution of the sorted units from a representative hCO recording. Cell body locations were estimated via triangulation, leveraging the dense electrode layout to detect spikes on multiple nearby channels. Units can be classified using a combination of two metrics (**Fig. 3B**) derived from waveform shape and spike timing[44]. We used the trough-to-peak time (TPT) of the bandpass filtered waveforms (300-6,000 Hz) and a 1D decision separatrix (0.425 ms) to separate narrow-waveform units (putative inhibitory, pInh) from wide-waveform units (putative excitatory, pExt). Classification can be fine-tuned using timing metrics derived from the spike train auto-correlation histogram (ACH; Y-axis, **Fig. 3B**). Together, the features covered by waveform shape and spike timing demonstrate broad coverage of neurons recorded from hCO, consistent with neuronal heterogeneity observed in cortical recordings from live subjects[44], and highlighting the physiological relevance of intact brain organoids.

To illustrate cell type diversity in the hCOs, **Fig. 3Ca** shows a pExt with a broad waveform, while **Fig. 3Cb** shows a fast-spiking pInh with a narrow waveform recorded at the same time. **Figure 3Cc** depicts a bursting pExt, characterized by a bimodal ISI distribution with short intra-burst intervals (~10 ms) and longer inter-burst intervals (~0.2 s). As a fourth example, **Figure 3Cd** shows a unit with a positive-polarity waveform recorded in parallel to the units depicted in panels **a-c**, consistent with non-somatic axonal spikes[45]. To assess functional connectivity, we computed cross-correlation histograms (CCHs) between pairs of spike trains. A representative pair is shown in **Figure 3D**, exhibiting a short-latency (<3 ms) peak indicative of putative monosynaptic coupling. The spike transmission gain for this pair was 0.089, indicating that every presynaptic spike was followed, on average, by 1/11 of a spike in the postsynaptic train, which is consistent with a relatively strong synaptic interaction[46]. These findings suggest that spike-derived features may be useful for building machine learning models for capturing critical information about the composition and function of the underlying circuitry.

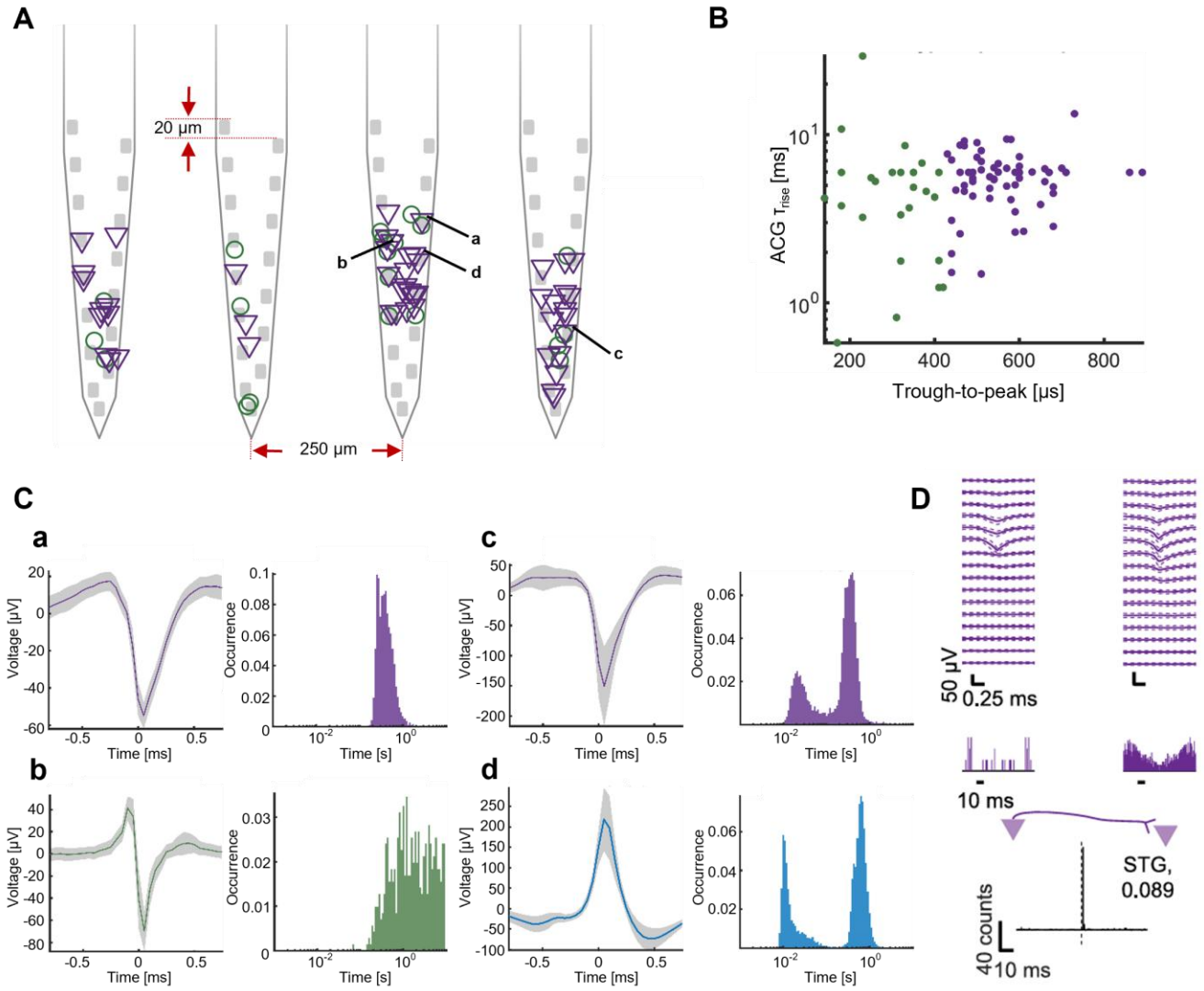


Figure 3. Parallel recording of multiple well-isolated single units indicates neuronal diversity and inter-neuronal connectivity within intact organoids. (**A**) Triangulated putative excitatory (pExt; purple triangles) and putative inhibitory (pInh; green circles) neurons relative to the recording electrodes (gray). The distance between the shanks is not drawn to scale. (**B**) Single units are classified as pExt and pInh based on spike features. (**C**) Single units characterized by their waveform shapes and timing. Individual units, **a-d**, indicated in **A** are shown with spike waveforms (mean±SD, gray band) and inter-spike interval histograms. (**D**) Example of monosynaptic connection between a cell pair shows connectivity at the cellular-network level. Top, Spike waveforms (mean±SD) on all channels from a shank. Middle, Auto correlation histograms of the two cells. Bottom, Schematic showing putative excitatory connection and cross-correlation histogram. STG, spike transmission gain.

Automated processing yields four deterministic spike features from raw recordings

An assay that enables high-throughput screening must extract meaningful patterns from neural recordings that distinguish control from test organoids accurately with minimal human intervention. Single-unit analyses (**Fig. 3**) provide valuable insights into the identity and dynamics of individual units and the neuronal circuitry. However, spike sorting and the associated manual curation are typically time-consuming, subjective, and require large numbers of spikes to reliably separate between spike clusters[47]. To circumvent these inherent limitations, we streamlined a feature extraction pipeline that draws on insights from single-unit analyses without requiring any spike sorting.

We first bandpass-filter the raw data (300-6,000 Hz) on every channel and detect spikes using a 5 SD threshold from the mean. For each detected spike, we define the time of the spike as the maximum of the absolute value. We then extract a 2 ms waveform segment that starts 1 ms before the time of the spike (40 samples at 20 kHz). Then, we use the time of the spike and the extracted waveform to compute four features (**Table 1**).

1. Inter-spike interval (ISI): The time duration between the spike and the most recent spike on the same channel (**Fig. 4A**). Clearly, that preceding spike may originate from a different source neuron. The first spike on every channel is excluded. The ISI captures gross firing patterns at the multi-unit level, including bursting behavior.

2. Amplitude: The voltage difference between the trough of the spike waveform and the first ensuing peak ("Peak B"; **Fig. 4B**). We use the term "trough" in this definition, since extracellular electrophysiology studies typically focus on somatic spikes, which exhibit a negative extremum that occurs just before the positive extremum of the intracellular spike. However, we also extract positive spikes, which have a positive extracellular extremum (e.g., **Fig. 3Cd**). In those cases, the sign of every sample waveform is inverted before deriving the three waveform-based features. Positive spikes have been shown to originate from non-somatic protheses[45]. Therefore, the amplitude feature is signed, corresponding to negative (typically somatic) or positive (typically non-somatic) spikes.

3. Peak A: The (signed) voltage of the nearest peak that precedes the trough. A prominent Peak A defines a biphasic spike waveform, which is associated with axonal potentials[45] (for example, see **Fig. 3Cb**).

4. Trough-to-peak time (TPT): The time duration from the extremum (typically, the trough) to Peak B (**Fig. 4B**). The TPT is an easy to compute proxy for the width of the spike, which is often used as one of the features for cell type classification[44,48].

We extracted and analyzed these four features for each of five organoid classes: hCO, hCO^{APP}, hCO^{APP} + 10 nM MK-8722, hCO^{APP} + 100 nM MK-8722, and $hCOL^{APP}$. Overall, we collected data from 89 sessions (89 organoids), yielding 25,815,311 spike events. Notably, many spike events are detected on multiple channels at the same time, but for the purposes of feature extraction and classification, they are not differentiated from isolated events. We computed median and inter-quartile (IQR) statistics for each of the four features of all organoid classes (**Table 1**). Although this set of features does not explicitly characterize spikes as clusters belonging to individual neurons, it captures meaningful differences in the structure and dynamics of neural activity at the individual spike level. For example, the small median ISI value of hCO^{APP} indicates faster spiking compared to hCO, as observed in **Fig. 1**, suggesting that this AD model is hyperactive[49]. The features serve as the foundation for downstream classification in a scalable, automated manner.

Organoid class	Sessions	Spikes	Amplitude [μV][a]	Peak A [μV]	ISI [ms]	TPT [ms]
hCO	24	2,282,223	-58.78 [36.64]	8.91 [15.99]	174 [304.5]	0.48 [0.3]
hCO^{APP}	33	10,464,522	-76.7 [69.72]	11.94 [25.72]	42.1 [82.55]	0.38 [0.28]
hCO^{APP} + 10 nM MK-8722	12	6,516,509	-63.06 [41.85]	12 [17.01]	58.35 [162.05]	0.43 [0.26]
hCO^{APP} + 100 nM MK87-22	13	6,099,906	-67.6 [53.62]	14 [26.04]	38.55 [56.6]	0.39 [0.36]
$hCOL^{APP}$	7	452,151	-65.1 [32.02]	8.84 [14]	51.1 [204.7]	0.43 [0.3]
Summary	89	25,815,311	-68.5 [53.53]	11.97 [21.97]	46.8 [115.1]	0.39 [0.3]

Table 1. Feature statistics for every organoid class.

[a] Values indicate median [IQR] over all spikes.

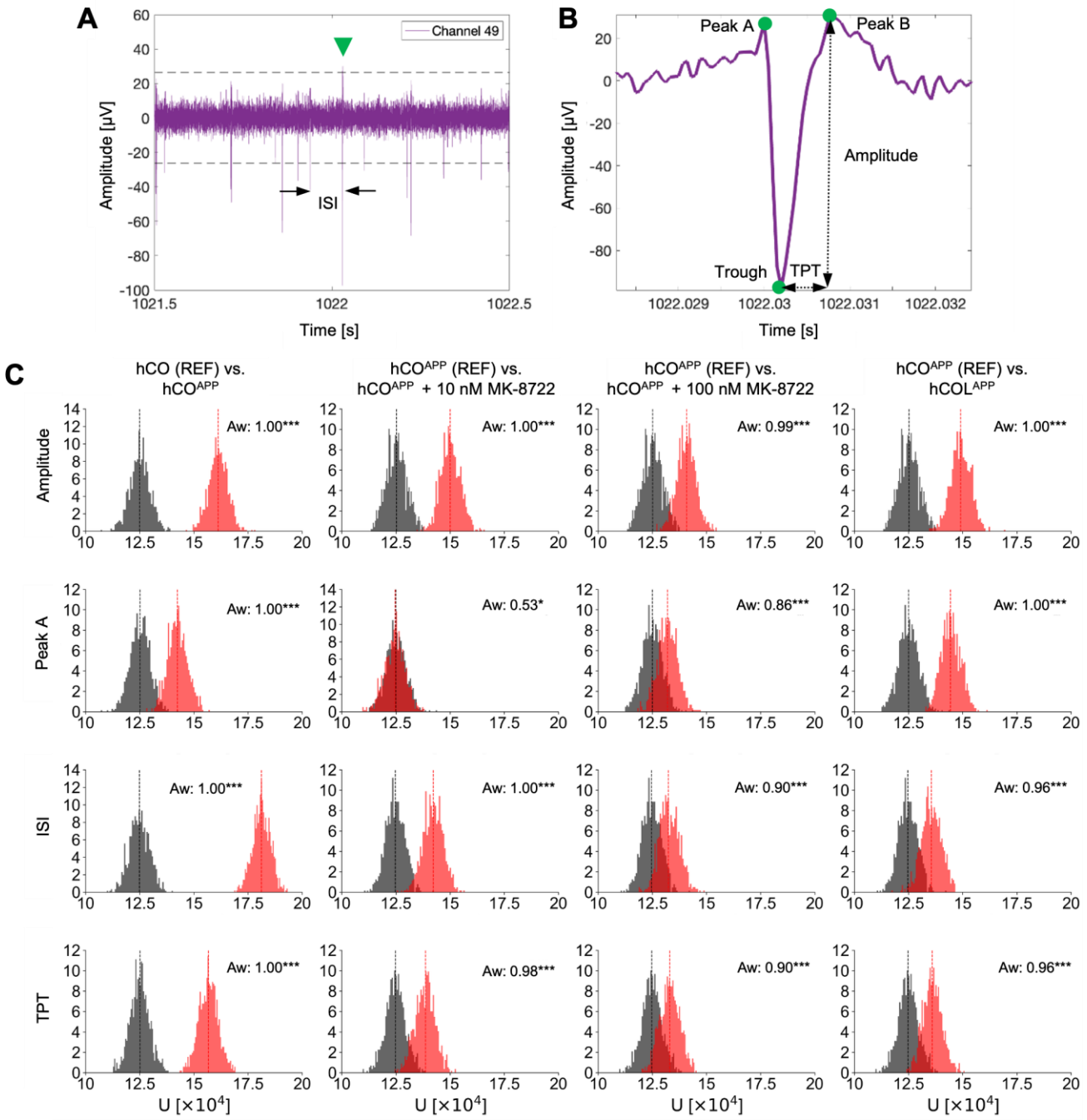


Figure 4. Spike features and their intra- versus inter-class variability distributions. (**A**) For every reference spike (green arrowhead), the inter-spike Interval (ISI) is defined as the time from the most recent spike on the same channel. (**B**) Three waveform-based spike features: Amplitude, Peak A, and trough-to-peak time (TPT). (**C**) Intra- and inter-class feature U-statistic distributions. Each row corresponds to one spike feature, and every column corresponds to a different pair of classes. Gray histograms show the intra-class U-statistic distributions for a reference class (REF), and the red histograms show the inter-class U-statistic distributions, between the reference and the referred classes. Each U-statistic value compares two randomly-sampled groups of 500 spikes from the same class (intra-class distribution) or from two different classes (inter-class distribution). Every histogram is based on 1,000 repetitions, yielding 1,000 U-statistic values. */***: $p<0.05/0.001$, two-tailed Mann-Whitney U-test.

Intra-class feature variability is consistently smaller than inter-class variability

To determine whether the variability of the electrophysiology readout (features) may allow any predictive power of the assay, we first compare spike feature distributions within the same class (intra-class variation) and between different classes (inter-class variation). Intra-class variation reflects differences in structure and circuit dynamics, which may arise from intrinsic (genetic) biological variability, from differences in the environmental conditions during the organoid maturation process, as well as from variability during the data acquisition procedure by an automatically inserted probe. Inter-class variability may arise from the same sources, and from inherent differences between organoid classes. If inter-class variability is consistently larger than intra-class variability, the set of features may provide a useful basis for classification.

To quantify intra-class variability for each organoid class, we pooled and shuffled all spikes from every recording. We then randomly selected two blocks (n=500 spikes each) and computed the Mann-Whitney U-statistic between them. This approach is non-parametric and does not depend on a specific probability distribution. The process was repeated 1,000 times for every feature, generating a distribution of U-statistics (gray histograms in **Fig. 4C**). We chose to compute statistics on small, fixed-size blocks of spikes for two reasons. First, small samples mimic the realistic scenario during high-throughput screening, where only brief stretches of data may be available from each organoid in a large batch. Second, the U-statistic itself depends on the sample size. We used the fixed-size blocks as a building block for training classifiers, ensuring consistency between statistical characterization and predictive modeling. As expected for pairs of similar distributions, the distributions of all intra-class U-statistics were centered around the expected value ($n^2/2$=125,000), consistent with no systematic difference in spike features within each organoid class. The spread of the U-statistics across randomly-selected blocks quantifies intra-class variability, serving as a non-parametric reference for evaluating inter-class comparisons.

To assess inter-class variation, we used a similar procedure. For each pair of organoid classes, we randomly sampled 500 spikes from each class and computed the U-statistic between the two blocks. The U-statistic reflects the difference in ranking order between the two samples. This was repeated 1,000 times, generating a distribution of U-statistics for every pair of classes (red histograms in **Fig. 4C**). To quantify the differences, we compared every intra-class (reference) U-statistic distribution to a corresponding (referred) inter-class U-statistic distribution that included the same reference organoid class (gray and red histograms coplotted in the same subpanel in **Fig. 4C**). Quantification was achieved using the Common Language Effect Size (Aw), which estimates the probability that a randomly selected inter-class U-statistic differs from a randomly selected intra-class U-statistic. Aw ranges from 0.5 (completely overlapping distributions) to 1.0 (no overlap), providing an interpretable measure of effect magnitude. The statistical significance (of rejecting the null hypothesis that Aw is 0.5) was tested using a two-tailed Mann-Whitney U-test. Strikingly, all features comparing hCO and hCO^{APP} showed Aw of approximately 1.00 ($p<0.001$; **Fig. 4C**, first column), suggesting a robust separation between the features of randomly sampled spikes from these two organoid classes.

To determine whether the chosen block size of 500 spikes is critical, we conducted a sensitivity analysis using half- and double-sized blocks and repeated the same Mann-Whitney U-test between inter- and intra- U-statistics. Aw increased slightly for all comparisons as the block size increased to 1,000 spike blocks, indicating a greater magnitude of difference in the feature distributions between the class pairs when sample size is larger. Compared with the 500-spike block test, statistical significance obtained from the 250-spike block test was the same for all comparisons (**Fig. S4**); and statistical significance obtained from the 1,000-spike block test was the same for nearly all comparisons, except for Peak A between hCO^{APP} and hCO^{APP} + 10 nM MK-8722 ($p<0.001$; **Fig. S5**). However, for this specific comparison, Aw increased from 0.53 to only 0.56 when increasing block size from 500 to 1,000 spikes. In summary, at the ranges tested, the feature variability distributions are robust and essentially independent of block size. Statistically significant differences between intra- and inter-class distributions are observed for all four spike features for the organoid class pairs.

A long-short term memory (LSTM) classifier is trained using blocks of spikes from hCO and hCO^{APP}

To allow the same probe to sample a large batch of organoids in rapid succession, the assay must produce accurate phenotypic predictions from short electrophysiological recordings, ideally lasting only a few minutes. We developed a machine learning framework that relies on the deterministic, physiologically relevant spike

features (**Fig. 4**, **Table 1**) to make binary predictions from brief recordings. Specifically, we trained an LSTM classifier[50] to distinguish hCO from hCO^{APP} using the spike features extracted from 500-spike blocks.

We split the dataset at the session level. After discarding sessions with less than 5,000 spikes, we assigned 64% of the sessions to training, 16% to validation, and 20% to held-out testing (**Fig. 5A**). Each recording session represents one unique organoid. Because the number of sessions was limited (**Table 1**), we implemented a block-wise strategy to increase sample size and to increase data variance. Motivated by a previous study[51], we segmented each session into non-overlapping blocks of 500 consecutive spikes, and every block was treated as an independent sample while preserving the temporal ordering of intra-block spikes. We only retained a maximum of 250 blocks (125,000 spikes) per session to avoid overrepresentation of any session. Therefore, the input to the model was therefore an ordered 500-by-4 spike-by-feature matrix, referred to as a "block". This strategy increased the number of training samples while constraining the model to learn local temporal patterns, minimizing overfitting and enabling better generalization to novel sessions.

Block-level predictions were evaluated on spike blocks from the held-out test sessions. At this block level, the model achieved an area under the receiver-operator characteristic (ROC) curve (AUC) of 0.85 (**Fig. 5B**). Using Youden's J statistic[52], we determined that the optimal threshold for binary classification is 0.84. Applying this threshold yielded an accuracy of 82%, with precision of 0.95 and recall of 0.83 for the hCO^{APP} (Class 1), and 0.53 precision and 0.81 recall for the hCO (Class 0; confusion matrix in **Fig. 5B, inset** and full classification report **Fig. 5C**). Despite the block level class imbalance (~4:1, hCO^{APP} to hCO), the model maintains good sensitivity to both classes.

To understand how classification errors and successes are distributed within sessions, we plotted the the weighted block-level prediction scores for every test session (sorted by class: hCO, **Fig. 5D**, purple; and hCO^{APP}, **Fig. 5D**, blue). The block score distributions in every session (histograms shown in **Fig. 5D, right**) revealed that in a few sessions all blocks were classified correctly, appearing at the correct side of the threshold corresponding to the true session label (e.g., **Fig. 5D, top row**). However, other sessions contained blocks which were incorrectly labeled with high confidence. Because the data are inherently heterogenous and because the relevant prediction is at the session (and not the block) level, misclassifications at the block level are both expected and acceptable.

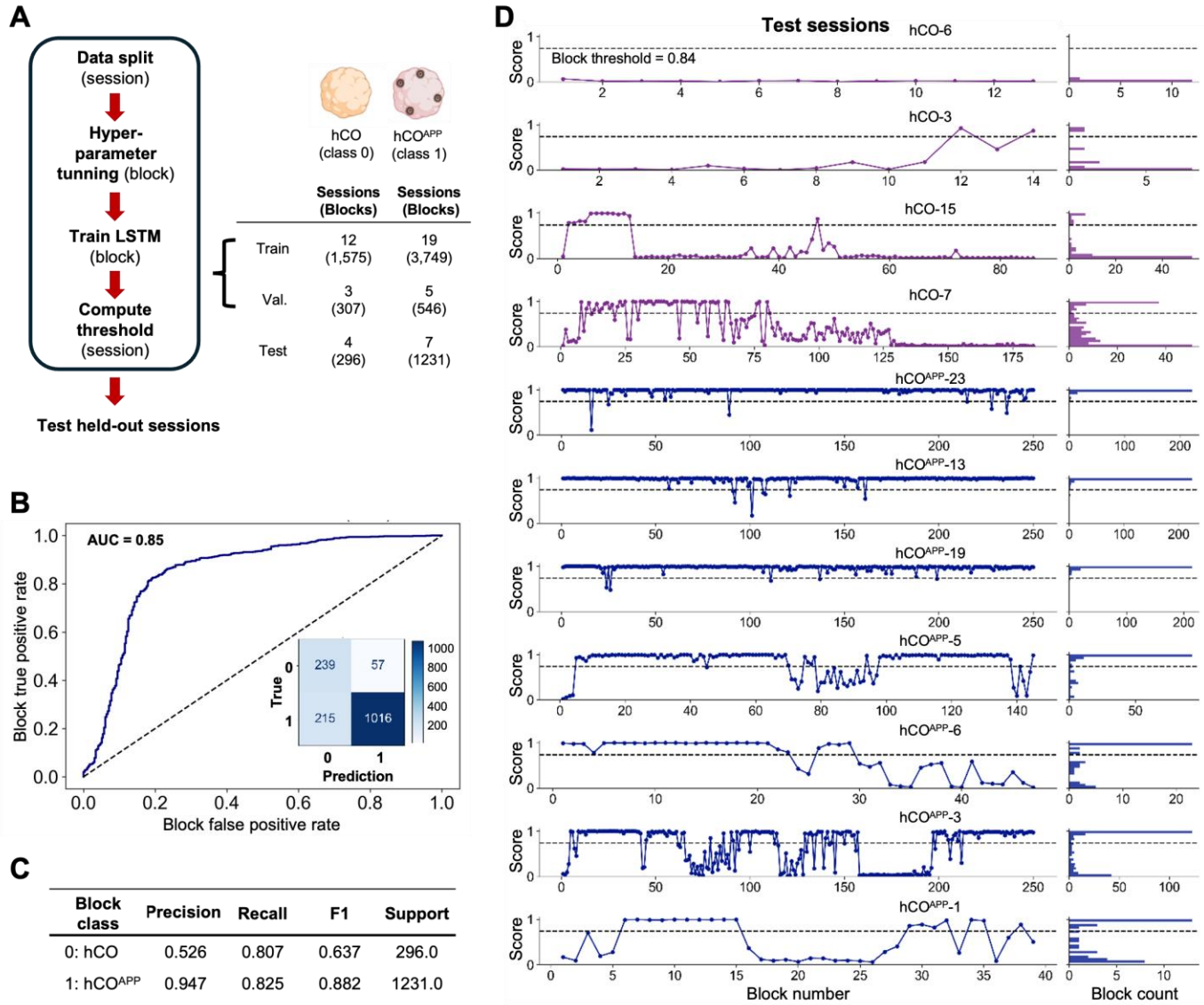


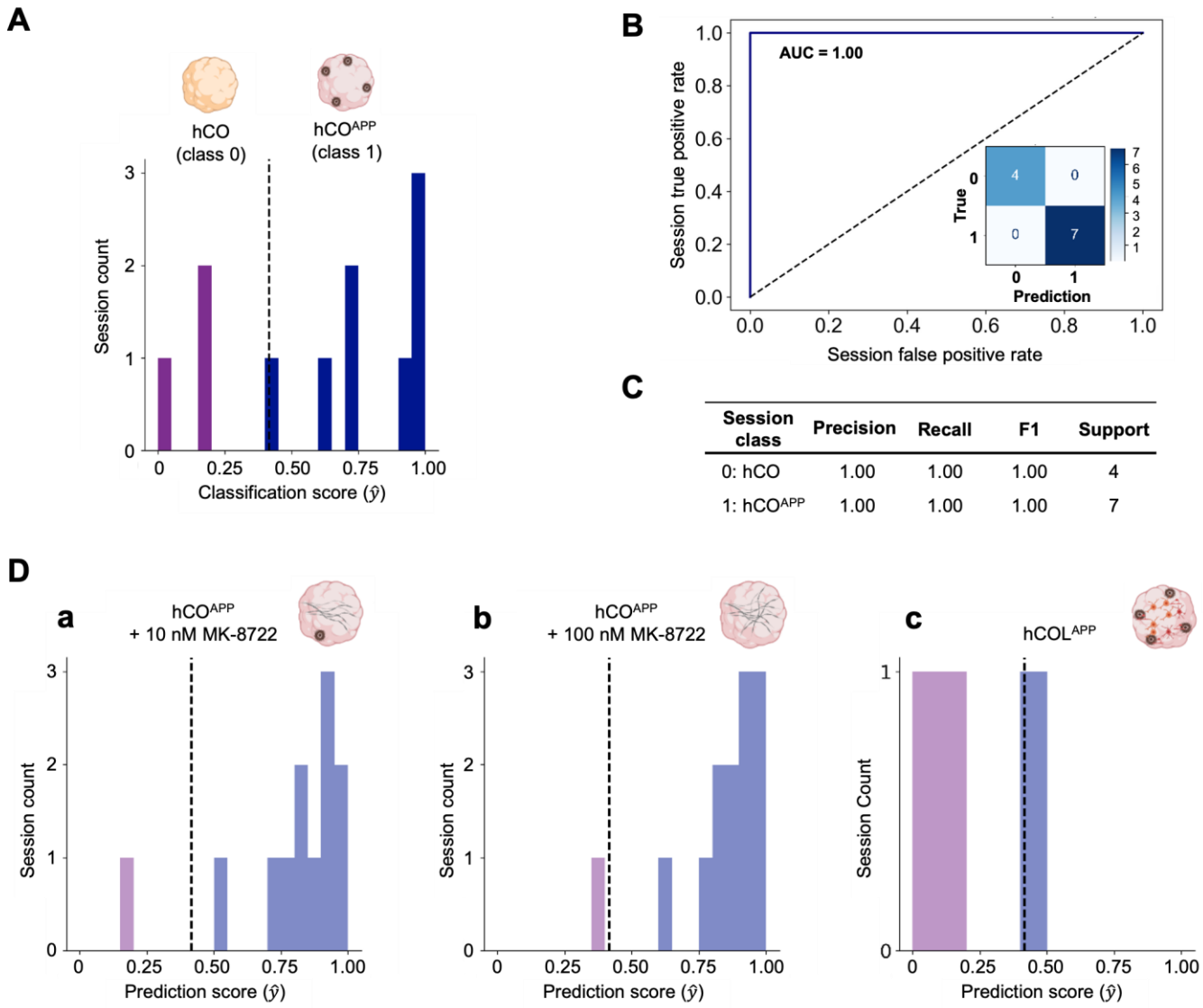
Figure 5. Long short-term memory (LSTM) classifier trained to distinguish hCO and hCO^{APP} spike blocks. (**A**) Overview of model training and testing strategy. The dataset was first split by session, reserving 20% of the sessions for testing. The remaining sessions were split into blocks of 500 consecutive spikes and used to train the LSTM model. A validation split was performed within the training set to guide optimization. During testing, predictions were made at the block level. (**B**) Receiver operating characteristic (ROC) curve for block level predictions on the test sessions. The area under the curve (AUC) reflects the ability of the model to discriminate between classes based on blocks of 500 spikes, independent of a specific threshold. Inset, Confusion matrix for binary classification using the optimal threshold (0.84), highlighting class-specific misclassifications. (**C**) Model classification report for block-level predictions using binary classification showing precision, recall, F1 score, and support for each class. While the model achieved high overall accuracy, precision for class 0 (hCO) was substantially lower (0.53), indicating that many hCO blocks were misclassified as hCO^{APP}. (**D**) Block-level prediction scores across test sessions. Left, Scores for each block are plotted chronologically for every hCO (top, purple) and hCO^{APP} (bottom, blue) session. Right, Session-specific histograms of prediction scores illustrate that despite overlap, block distributions are skewed towards their respective class extremes (0 for hCO, 1 for hCO^{APP}). Thus, accurate classification may be achieved by aggregating information over blocks.

Session-level classification via confidence-weighted voting distinguishes hCO from hCO^{APP}

After training the block-level LSTM model, we derived session-level classification by post hoc aggregation of the block-level model outputs. Specifically, each block classification score $s_i \in [0,1]$ was transformed using a confidence-weighting function, $w(s_i) = e^{c \cdot |s_i - 0.5|}$ where c is a scaling constant (set to 4 in this study) that emphasizes high-confidence predictions near 0 or 1. The final session score $\hat{y}$ was defined as the weighted average of all blocks from that session:

$$\hat{y} = \frac{\sum_i w(s_i) \cdot s_i}{\sum_i w(s_i)}$$

To determine the optimal session-level classification threshold for $\hat{y}$, we applied Youden's J statistic to the training and validation sessions, yielding a cutoff of 0.416. We then evaluated model performance on held-out test sessions. Despite the block-level inconsistencies shown in **Fig. 5D**, all test sessions were correctly classified (**Table S1**). Session classification achieved perfect performance on the test set (**Fig. 6A-C**), yielding an AUC of 1.0. Given the small size of the test set, these results should be interpreted as a proof-of-concept for the hierarchical prediction strategy rather than as final performance metrics. The results illustrate how, despite inherent ambiguity at the block level (**Fig. 5C**), aggregating predictions across multiple noisy blocks enables robust session-level classification.



Session class	Precision	Recall	F1	Support
0: hCO	1.00	1.00	1.00	4
1: hCO^{APP}	1.00	1.00	1.00	7

Figure 6. Trained model accurately classifies held-out test sessions and generalizes to novel organoid perturbations. (**A**) Session-level classification scores ($\hat{y}$) for labeled test sessions (hCO and hCO^{APP}) using a threshold of 0.416 (determined from the training and validation data) yield perfect classification. Histogram bins are colored by their corresponding true class labels (purple for hCO and blue for hCO^{APP}). (**B**) ROC for the session-level classification shows perfect performance, with an AUC of 1. (**C**) Classification report details the session-level classification. (**D**) The trained LSTM model is applied to unseen organoid classes to evaluate generalization. **a** hCO^{APP} treated with 10 nM MK-8277 are classified as hCO^{APP} in 11/12 (91.7%) sessions. **b** hCO^{APP} treated with 100 nM of the same drug are classified as hCO^{APP} in 12/13 (92.3%) sessions. Both treatments were applied to the same hCO^{APP} batch and did not result in a detectable functional rescue (classification as hCO), despite the observed reduction in Aβ levels (see **Fig. 2D**). **c** In contrast, only 2 out of 6 hCOL^{APP} organoids (with enhanced oligodendrocyte differentiation) are classified as hCO^{APP}. Histogram bins are colored by their corresponding predicted class labels (light purple for hCO and light blue for hCO^{APP}).

In the hCO^{APP} model, the electrophysiological and molecular phenotypes are inconsistent

To test the utility of the proposed pipeline for capturing the effect of treatments to a disease model, we applied the model to novel organoid classes that were not used during model training. These classes included: hCO^{APP} organoids treated for 30 days with 10 nM or 100 nM of MK-8277, and $hCOL^{APP}$ (**Fig. 2A**). The rationale is that if a treatment induces a detectable change in the electrophysiological phenotype, the classification yielded by the model may shift accordingly. In contrast, if the treatment has no system level electrophysiological effect, classification results would remain consistent with the baseline hCO^{APP} class.

We found that nearly all MK-8277-treated organoids were classified as hCO^{APP}. Specifically, MK-8277-treated organoids were classified as hCO^{APP} and not as hCO in 11/12 (91.7%) sessions for the 10 nM group, and in 12/13 (92.3%) sessions for the 100 nM group (**Fig. 6Dab**). These results indicate that under these specific conditions and despite the observed reductions in Aβ levels (**Fig. 2D**), the treatments do not produce a detectable functional rescue.

In contrast, most of the oligodendrocyte-enriched organoids were classified as hCO. Specifically, 4/6 (66%) of the $hCOL^{APP}$ sessions were classified as hCO (**Fig. 6Dc**). Furthermore, even the two sessions classified as hCO^{APP} had prediction scores of 0.4162 and 0.4737. These scores are close to the classification threshold of 0.4161, suggesting a shift towards the control (hCO) phenotype. This result implies that differentiation enriched by oligodendrocyte may have a stronger impact on circuit-level electrophysiology than MK-8277. Clearly, these results pertain to the particular hCO^{APP} model and the specific compound and concentrations used in the experiments, and cannot be used to derive overarching conclusions about BACE1 inhibitors as a drug class.

High-throughput assay is feasible using data recorded only during probe insertion

To fully exploit the speed and scalability of our electrophysiology platform for high-throughput drug screening, it is essential that the trained models deliver accurate predictions within a minimal recording time. In principle, data collection and model inference can begin as soon as the probe begins to enter the organoid. As described (**Fig. 1**), our system already implements a closed-loop detection algorithm that estimates the high-pass filtered baseline, identifying when the probe channels enter the tissue based on the V_{RMS} pre-computed thresholds. A parallel strategy can be implemented for classification: rather than only serving to detect the transition into the organoid, the algorithm can detect spikes, extract features and feed those into a pre-trained classification model to generate block-level predictions as the probe descends in real time. Here, we assess this possibility.

To determine the capability of the system to classify organoids on the fly, we used only the spike data acquired during the insertion phase to evaluate how early and confidently the pre-trained model (**Fig. 5B**) could predict the class of an organoid. These brief datasets, collected during probe insertion (before the probe has reached its final recording depth), were not part of any of the sessions used for model training or testing shown in **Fig. 5A**. The routine begins after a short preparation period involving organoid loading and baseline calibration (**Fig. 7A**). Once the insertion routine is initiated, the probe advances stepwise into the organoid while an algorithm continuously performs spike detection, feature extraction, and block-wise predictions. Early during the insertion, only a few channels are engaged, and block accumulation is initially slow. As more electrodes enter the neuronal part of the organoid, the active channel count increases, spike yield increases, and block acquisition accelerates.

Representative wideband traces from one probe shank (of four) are presented in **Fig. 7B**, corresponding to the final two seconds of blocks #1, #4, and #7 from **Fig. 7A** (an hCO^{APP} session). These traces illustrate the progressive increase in neural activity as the probe advances into the neuronal region of the organoid. Low-frequency artifacts are occasionally observed during the descent, likely resulting from the mechanical movement of the probe (**Fig. 7Ba**). However, these artifacts do not have a discernible impact on the spikes, which are detected in the high-pass filtered traces (not shown). Although electrode motion poses a challenge for stable single-unit isolation, it does not impede feature extraction for our classifier which is based on individual spikes. On the contrary, the dynamic sampling process may enrich the diversity of spike waveforms available to the model as input.

For the hCO^{APP} session illustrated, the first 500 spike block yielded a score below threshold. However, after the second block (1,000 cumulative spikes), the cumulative confidence-weighted average score already crossed the session-level threshold, and then continued to increase as additional data were accumulated (**Fig. 7A**, middle

panel). This suggests that reliable classification can be achieved with minimal data under realistic insertion-time conditions.

Finally, we assessed predictions for hCO^{APP} and MK-8277 treated hCO^{APP} organoids that were recorded during the insertion phase (**Fig. 7C**). We found that both (2/2, 100%) hCO^{APP} sessions were correctly classified; that 3/4 (75%) of the 10 nM MK-8277 treated hCO^{APP} sessions were classified as hCO^{APP}; and that all (5/5; 100%) of the 100 nM MK-8277 treated hCO^{APP} sessions were classified as hCO^{APP}. Overall, probe insertion was completed in a median of 4.43 min (IQR: [4.00 4.63] min), during which a median [IQR] of 7 [1 32] blocks were acquired, each comprised of 500 spikes. The results demonstrate that our automated platform yields rapid and accurate phenotype classification using only spike data collected during the probe insertion phase. This underscores the practicality of the approach for real-time, high-throughput screening.

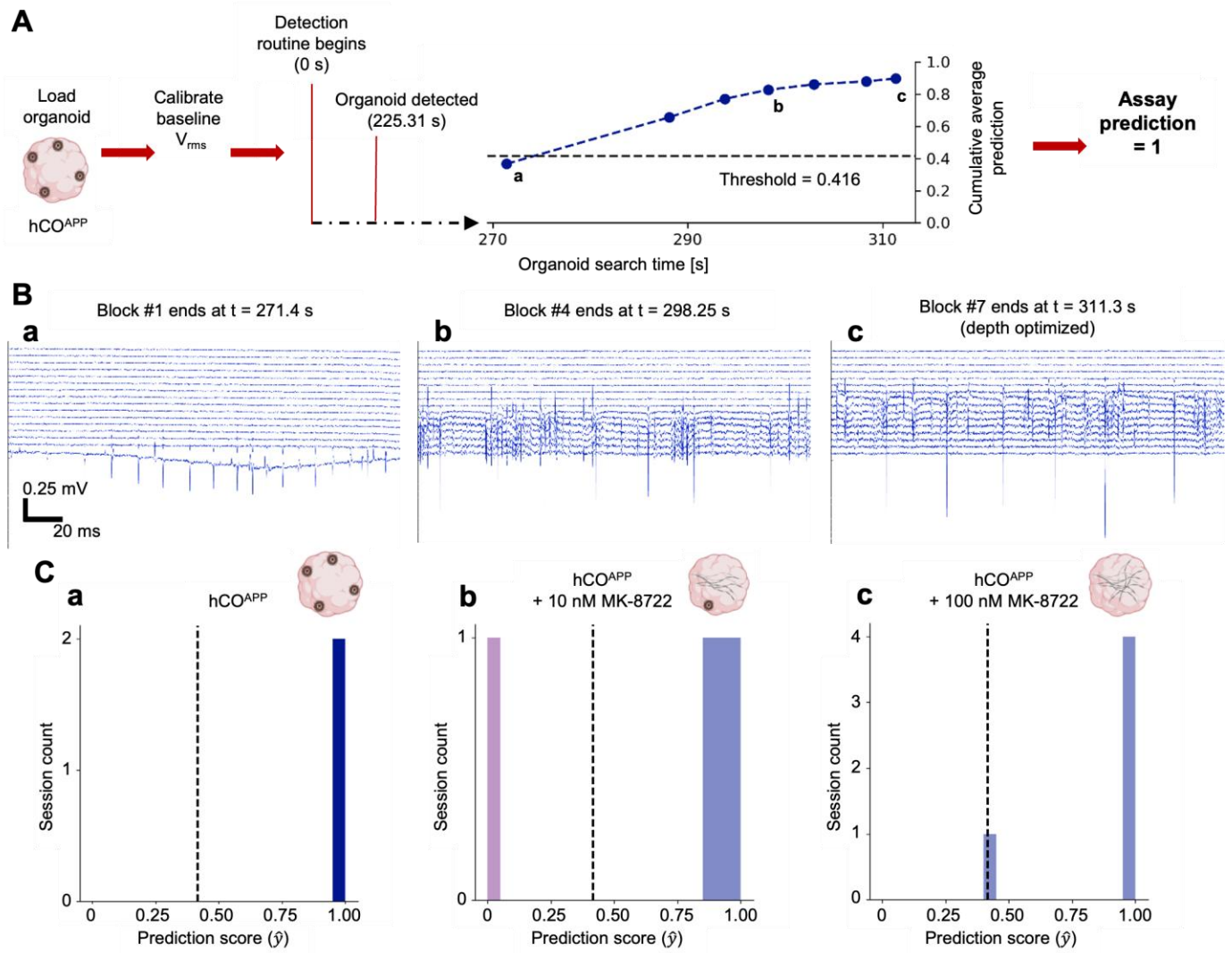


Figure 7. A rapid assay using the pre-trained LSTM model can accurately classify organoid phenotype in real-time during probe insertion. (**A**) Illustration of the automated classification routine. After the organoid is loaded and probe calibration is completed, the detection routine begins (t=0 s). In this example, the organoid is detected at 225.31 s when the first channel exceeds the RMS voltage threshold, and the first spike is detected shortly afterwards. The circles indicate the times at which the acquisition of every 500-spike block has finished. The session level score crosses threshold after the second block (1,000 spikes total) and remains high at the end of the insertion (311.3 s), enabling early and correct organoid classification as hCO^{APP}. (**B**) Example wideband recordings (0.2 s windows) captured at the times marked in **A**, illustrating the diversity of neural activity encountered as the probe moves deeper into the organoid. (**C**) Summary of prediction outcomes using insertion-phase data only. **a** 2/2 (100%) hCO^{APP} sessions are correctly classified. **b** 3/4 (75%) of hCO^{APP} organoids treated with 10 nM MK-8722 are classified as hCO^{APP}. **c** 5/5 (100%) of hCO^{APP} organoids treated with 100 nM MK87-22 are classified as hCO^{APP}.

DISCUSSION

Accessing the functional circuits of 3D brain organoids requires moving beyond the constraints of surface electrodes. While 2D MEAs have served as standard tools, their limited depth access fails to capture the complex, spatially distributed activity of organoids[19,20]. Attempts to extend the reach of MEAs using protruding electrodes have fallen short due to their shallow, fixed-length geometry and the lack of active insertion mechanisms to precisely access deeper layers[53]. These limitations are increasingly untenable as organoids become larger and more structurally complex[8,16,30,37].

By inserting high-density silicon probes into intact organoids, we overcome the depth limitations and unreliable electrode-neuron coupling inherent to planar systems. Originally developed for in vivo recordings, minimally invasive silicon probes provide reliable access to interior layers without disrupting the structure of the organoid[27]. This approach does not merely improve yield; it fundamentally changes what aspects of organoid physiology are accessible, particularly network-level activity that emerges within interior regions.

To make deep recordings practical and scalable, we developed an automated system that eliminates user-dependent variability by driving the probe with a motorized stage under real-time feedback control. As the probe advances into the organoid, an RMS-based signal detection algorithm identifies the onset and progression of neuronal activity, halting insertion at the depth that maximizes active channel count. This closed-loop approach ensures consistent, optimized access to interior regions, enabling standardized recordings across organoids and users without requiring prior experience with silicon probe implantation.

Automated extracellular recording from the organoid interior opens the door to high-throughput functional screening. Importantly, this data acquisition capability must be matched by an analysis framework that supports comparable scalability. Rather than relying on time consuming single-unit analysis, which involves spike sorting and often requires a human expert in the loop, we trained a machine classifier using deterministic spike features. We chose four physiologically meaningful features, derived from spike waveform and timing, known from in vivo studies to correlate with cell type and circuit dynamics[44].

Our initial results established a reusable and robust framework for binary classification of brain organoids. We employed an LSTM neural network to classify organoid phenotypes based on sequences of spike-level features. The LSTM was chosen over other artificial neural networks (ANNs) due to its ability to model temporal dependencies[50]. For example, two classes may have the same number of isolated versus bursting spikes, but in one class, the bursts may consist of two spikes, whereas in the other the bursts may consist of three spikes. Higher order temporal statistics from individual ISIs may hold circuit-level information essential for classification.

While LSTMs offer advantages, they are also more computationally intensive to train and more susceptible to overfitting in small-sample regimes[50]. In our case, training was performed offline, separately from the experimental workflow, which allowed sufficient time and resources for model development. Once trained, the classifier can be used in many new experiments, supporting real-time classification without retraining. As data acquisition is scaled up, the risk of overfitting is expected to decrease. We acknowledge that a simpler ANN might have been sufficient for the current binary task, but this was not tested exhaustively. Looking forward, more challenging classification problems such as subtle drug responses or complex genetic models may also benefit from the temporally aware architecture of the LSTM classifier, which is aligned with the temporally dynamic nature of brain function and circuit-level computations[28].

As a proof of concept, we generated cortical organoids from hiPSCs with and without Swedish/Indiana mutations in the APP gene, mutations widely used in familial AD modeling due to their effect on Aβ overproduction. The resulting hCO and hCO^{APP} lines formed the basis for the binary classification task, serving as an initial benchmark for validating the functional screening framework. For simplicity, binary classifiers are often sufficient for functional assays, where the objective is to determine whether a compound elicits a specific phenotype, such as therapeutic efficacy or toxicity[54]. The same feature extraction and LSTM-based framework is extensible to multiclass settings, enabling more refined classification when warranted. Indeed, our primary goal in this study was not to optimize the classification of a specific disease model, but rather to establish a framework that generalizes to other applications, including assessing genetic mutations, pharmacological perturbations, or environmental toxicants[55]. For example, the U.S. Environmental Protection Agency's ToxCast program has identified thousands of uncharacterized neurotoxicants that remain unscreened[56].

Our model was trained on short blocks of 500 spikes, which were derived from longer continuous sessions. This design allowed for consistent input dimensions and high training sample counts[51]. However, it introduced a challenge: spike blocks were only weakly labeled (inherited from the parent session) and may not uniformly reflect the session phenotype. Variability in organoid maturation, spontaneous circuit states, and random sampling by the probe introduced considerable heterogeneity across blocks. As expected, some blocks yielded confident scores aligned with the session label; others hovered near chance or even misaligned with the session class. This is not a model failure, but rather a reflection of real-world variance. Instead of discarding these fluctuations, we exploited them using a confidence-weighted aggregation strategy, amplifying the influence of high-confidence blocks. This can be especially useful to separate two classes with identical mean feature values, but one would be more variable over blocks than the other.

In one experiment, treatment with a BACE1 inhibitor (MK-8722) reduced Aβ levels in hCO^{APP}, as confirmed by immunostaining, yet the electrophysiology-based model continued to classify the treated organoids as AD-like. While the results pertain to the specific compound and concentrations employed, the discrepancy between immunostaining and neuronal activity underscores the value of a functional readout. This distinction is particularly relevant for diseases like Alzheimer's, where molecular hallmarks may not map directly onto behavioral or physiological outcomes[40,41,57]. Complementary to quantifying the biochemical state, our assay measures functional activity at the circuit level, which underlies all brain functions and dysfunctions.

To push the limits of throughput, we evaluated whether classification could be achieved using only data collected during the probe insertion phase, without requiring extended stationary recording. This test revealed that the model could accurately classify sessions using just a few minutes of insertion-phase activity. As the probe descended and encountered more active neurons, block accumulation accelerated and predictions converged rapidly. This insertion-only mode points toward true real-time screening applications, where the classifier operates concurrently with data acquisition, which is in turn synchronized with the instrument robotics.

While this study demonstrates the feasibility of automated, single-organoid targeting, the platform presented is readily extensible to multi-organoid probing via X-Y actuation, allowing the probe to access multiple wells within the same apparatus. Although a single probe may eventually be designed to target multiple wells simultaneously, full-plate coverage is unlikely due to limited depth optimization, noise accumulation, system complexity, and cost. Thus, it is essential that useful information will be acquired rapidly, minimizing probe dwell time per well. The feasibility of this scalability is already evident from the results reported in the present study: because the full workflow completes within minutes, sequential probing across wells is possible.

Looking ahead, a key feature of this platform lies in its ability to support the cumulative development of a growing portfolio of pre-trained classifiers. As we expand the range of tested compounds, disease models, and experimental conditions, the resulting database can be used to train increasingly accurate and generalizable models tailored to specific toxicants, genetic mutations, or therapeutic targets. We expect that over time, this resource will enable plug-and-play classification across new assays. Continuous addition of new sessions supports iterative refinement, transfer learning, and cohort-level aggregation, paving the way for robust, reproducible, and scalable phenotyping across diverse organoid systems.

ACKNOLWEDGEMENTS

This work was partially funded by Merck Sharp & Dohme LLC, a subsidiary of Merck & Co., Inc. We thank Danny Bitton, Michael Wurst, and Jakub Arnold from MSD Prague for helpful discussion on machine learning model.

AUTHOR CONTRIBUTIONS

Conceptualization: R.M., B.C., B.J., E.S.; Instrument design: F.W., L.A., B.J.; scRNAseq library generation and analysis: S.K.V., V.M.P., A.M.T, B.C.; conceptualization of AD disease modeling: M.E.K., I.K., B.C; Data acquisition: F.W., L.A., B.C.; Electrophysiology analysis: F.W., E.S.; Machine learning framework: F.W., C.S.; Writing: F.W., E.S., B.J., B.C.

COMPETING INTEREST

F.W and B.J are stakeholders in Diagnostic Biochips, Inc.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

METHODS

This study complies with all relevant ethical regulations approved by Merck Research Laboratories (MRL). All experiments involving human iPSCs were approved by the Merck Stem Cell for Drug Discovery (SCDD) Committee. The human brain tissues assessed in this study were obtained from Dx Biosamples LLC (Sample ID # 531888A(3), San Diego, CA) based on MRL Human Investigation Committee. It has been approved for use with a waiver of consent.

Human iPSCs culture

XCL-1, XCL-1_APP Ind (SwHomo)-3G05-E01, and BC1 hiPSCs were cultured on Matrigel (Cat # 356230, Corning, NY) coated cell culture dishes with mTeSR1 plus media (Cat #100-0276, STEMCELL Technologies, BC, Canada). hiPSCs were passaged every week by treatment with Dispase (0.83 U/ml, Cat # 07913, STEMCELL Technologies, BC, Canada).

Generation of hCO variants

As described earlier[33], we generated hCOs and hCO^{APP} by using XCL-1 and XCL-1_APP Ind (SwHomo)-3G05-E01 human iPSCs, respectively. Briefly, after dissociating cells via Accutase, a total of 9000 cells were plated into a well of U-bottom ultra-low-attachment 96-well plate in neural induction medium (DMEM-F12, 15% (v/v) KSR, 5% (v/v) heat-inactivated FBS (Life Technologies),1% (v/v) Glutamax, 1% (v/v) MEM-NEAA, 100 µM β-Mercaptoethanol) supplemented with 10 µM SB-431542, 100 nM LDN-193189, 2 µM XAV-939 and 50 µM Y27632). FBS and Y27632 were removed from day 2 and 4, respectively. The medium was replenished every other day until day 10, where organoids were transferred to the ultra-low-attachment six-well plate. The organoids were cultured in spinning hCO, dorsal patterning (day 10 to day 18) medium with minus vitamin A (1:1 mixture of DMEM-F12 and Neurobasal media, 0.5% (v/v) N2 supplement, 1% (v/v) B27 supplement without vitamin A, 0.5% (v/v) MEM-NEAA, 1% (v/v) Glutamax, 50 µM β-Mercaptoethanol, 1% (v/v) Penicillin/Streptomycin and 0.025% Insulin). The dorsal patterning medium was replenished every other day until day 18, when media was switched to the maturation media, hCO medium with vitamin A (the same composition as described above except B27 with vitamin A) supplemented with 20 ng/ml BDNF and 200 µM ascorbic acid. The maturation medium was changed every 3 days after day 18.

To generate AD forebrain variants, dorsal and ventral forebrain organoids were developed using XCL-1_APP Ind (SwHomo)-3G05-E01 and 3-TF-expressing BC1-iPSCs. AD control organoid variants utilized 100% XCL-1_APP Ind (SwHomo)-3G05-E01) while AD variants with OL-lineage cells were created by mixing 20% BC1-iPSCs, driving OL-lineage cells, and 80% non-infected XCL-1_APP Ind (SwHomo)-3G05-E01) and followed by dorsal patterning and maturation.

Automated multiplex immunofluorescence staining

Organoids were collected in a 6-well plate and fixed in 4% paraformaldehyde (PFA) for 24 hours at 4 °C as described previously[33]. Following fixation, single organoids were transferred to individual cassettes for paraffin embedding. The tip of a tube was placed on the embedding mold, and paraffin wax was melted; once melted, organoids were carefully moved into the mold[58]. Each paraffin block contained three organoids per condition, along with AD patient brain tissue. The blocks were sectioned at 5 µm thickness onto charged glass slides, allowing for serial sections to be obtained from nine organoids across three different conditions. For immunofluorescence staining, slides were baked at 60 °C for 1 hour, deparaffinized using the BOND Dewax protocol, and subjected to heat-induced epitope retrieval with BOND Epitope Retrieval Solution 2 for 20 minutes. Staining was performed on the Leica Bond RX stainer using BOND Ready-To-Use primary antibodies of monoclonal mouse or rabbit origin. The immunofluorescence protocol (Opal) was executed sequentially, with chromogens mixed on-board the BOND RX as part of the automated staining process, utilizing enzymatic conversion with BOND Polymer Refine Detection (HRP) and BOND Polymer Refine Red Detection (AP). Finally, images were scanned at 20x magnification using the Vectra Polaris and unmixed for analysis. Slides were

incubated in primary antibodies (anti-AT8 1:100, anti-MAP2 1:100, anti-6E10 1:100, anti-SOX2 1:200, anti-Ki67 1:1000, anti-vGAT 1:100, anti-vGLUT2 1:100, anti-OLIG2 1:200, anti-MBP 1:1000).

Library preparation of scRNAseq

Cortical organoids 50 and 115 day old were randomly collected from 3 different culture dishes, with 6-10 organoids pooled together. As previously described[37], the organoids were initially dissociated using the papain dissociation system according to the instructions of the manufacturer (Worthington Biochemical Corporation, NJ). Subsequently, after washing with Hank's balanced salt solution, the organoids were dissected into small pieces in papain solution and oxygenated with 95% O_2:5% CO_2 for 5 minutes, and then incubated at 37 °C for 1 hour. Following the generation of a single-cell suspension via trituration, the single cells were suspended in 1% BSA/PBS supplemented with 10 μM Y27632 and stained for Aqua Dead Cell stain (Cat. No. 50-112-1525). FACS sorted live cells were re-suspended in 0.04% BSA/PBS (128 cell/μl) and used to generate cDNA libraries by utilizing the Single Cell 5' Reagent Kits on a Chromium-X platform. In brief, the cells were partitioned into nanoliter-scale Gel Bead-In-Emulsions (GEMs), and microfluidic cells were streamed at limiting dilution into a stream of Single Cell 5' Gel Beads and then a stream of oil. After cell lysis, primers, an Illumina P7 and R2 sequence, a 14 bp 10xBarcode, a 10 bp randomer, and a poly-dT primer were released and mixed with the cell lysate and a bead-derived Master mix. These cDNAs were used to generate sequencing libraries for Illumina platforms using library generation kit according to manufacturer supplied user guides (10X Genomics, CA). The libraries were sequenced using the Illumina Novaseq 6000 platform.

Data processing of scRNA-seq

scRNA-seq FASTQ files were processed and generated count matrix over human reference version hg38 using CellRanger (v 8.0.1) with default parameters. Before processing scRNA-seq analysis, we confirmed the low doublet frequency of our scRNA-seq libraries (mean±SEM, 0.82±0.28%) by counting cells expressing both TBR1 and GFAP, which are usually exclusively expressed in cortical neurons and astrocytes, respectively.

The scRNA-seq libraries from forebrain organoid variants and human fetal brains[59] were collectively analyzed using Seurat (v5.0) within the R (v4.4.1) environment. Initially, as a part of quality control measures, cells with fewer than 1,000 detected genes and genes expressed in less than 5 cells were excluded. For each organoid study, the feature counts were normalized to the total counts and multiplied by a factor of 10,000. Highly variable features (genes) were then selected using variance stabilizing transformation. Following the prioritization of features based on their variance across scRNA-seq libraries, the top 2,000 features were retained for downstream analysis. Cell pairwise anchor correspondences between different single-cell transcriptome studies were identified using 30-dimensional spaces from canonical correlation analysis. These anchors were then utilized to integrate scRNA-seq datasets and transform them into a shared space. Gene expression values were scaled for each gene across all integrated cells and used for principal component analysis. Subsequently, using the first 30 principal components (PCs), all cells were projected onto a two-dimensional UMAP space. Graph-based cell clustering was implemented with the shared nearest neighbor method from the first 30 PCs and a 0.5 resolution value. Differentially expressed genes in each cluster were identified by employing a threshold of more than 1.25 average fold change and $p<0.05$ through a two-sided t test. Overrepresented gene ontology terms were identified using GOstats (v2.24.0), and the false discovery rate (FDR) was estimated using the Benjamini-Hochberg method with the p.adjust function. An FDR below 0.05 was used as a threshold for statistical significance.

Cell types were assigned to the clusters with slight modifications from a previous method[33]. Initially, the clusters were categorized into two main groups, neuronal and non-neuronal, based on general neuronal markers (STMN2, GAP43, and DCX) and early neurogenesis genes (VIM, HES1, and SOX2). Within the neuronal group, further classification into cortical neurons (CN) and interneurons (IN) was conducted based on the expression of specific markers, such as TBR1 and amino acid synthetic enzymes and transporters (GAD1, GAD2, vGLUT1, and vGAT). Additionally, glutamatergic and GABAergic neurons were identified by the expression of their transporter genes (SLC17A7, SLC17A6, and SLC32A1), while neuronal clusters lacking these specific genes were annotated as immature neurons.

Cell clusters expressing biglycan and decorin were allocated as proteoglycan-expressing cells (PGC). Furthermore, clusters enriched with genes related to “Cilium assembly (GO:0044458)” and “myofibril genes” (GO:00300016) were labeled as cilium-bearing cells (CBC). Clusters exhibiting significant overrepresentation of “Glia cell differentiation (GO:0010001)” were classified as glial cells and further annotated as oligodendrocyte progenitors (OPC), oligodendrocytes (OL), and astrocytes (AS) based on unique markers for each cell type (PDGRFA, PTPRZ1, BCAN, OLIG1/2, PLP1, GFAP, SLC1A3, and S100B) and relevant GO terms. The remaining cluster was designated as intermediate.

To validate the cell type annotation, gene signatures of neuron, astrocyte, and OPC obtained from single-cell transcriptome in human fetal and adult brains[60] were used to pre-rank genes by relative expression and evaluated using GSEAPY software, as described previously[36].

To compare the cell types in our study across different time points with those in established organoid protocols from the literature, we utilized transcriptomic data from the Human Neural Organoid Cell Atlas (HNOCA)[38]. We accessed the publicly available section of HNOCA, which encompasses all integrated protocols except for three that remain unpublished. We employed the CellHint tool[61] v0.1.1 to harmonize cell types across various organoid protocols and to compute a cross-dataset distance matrix. This matrix included 1,665,880 cells and 161 cell types from 26 different protocols, as indicated in the “assay differentiation” section of the HNOCA metadata. During the harmonization process with CellHint, each cell was assigned to the most similar cell type from each dataset, resulting in a comprehensive assignment matrix. The distance score in this matrix quantifies the transcriptomic dissimilarity between cell types across datasets, computed as the cosine distance between their gene expression profiles, with higher values indicating greater differences[38]. For ease of analysis, we focused on specific comparisons: excluding all cell types labeled as ‘unknown’ and retaining only those comparisons that involved our datasets and had a distance score greater than 0.5 for at least one cell type. Finally, we visualized the inferred similarities among cell populations using a UMAP plot, which displayed unsupervised hierarchical clustering

Instrument design and operation

We designed an automated system to precisely drive a microfabricated silicon probe into stationary brain organoids. The core sensor is a 64-channel silicon probe (P64-1, Diagnostic Biochips Inc., Glen Burnie, MD) connected to a custom printed circuit board housing a front-end ASIC (RHD2164, Intan Technologies, Los Angeles, CA). This assembly interfaces with a data acquisition system (Open Ephys, Lisbon, Portugal) via SPI communication. The probe and ASIC are mounted on a Z-axis micromanipulator (KMTS50E, Thorlabs, Newton, NJ), which is computer-controlled. A custom Python script asynchronously controls the Open Ephys system to start/stop recordings and reads recent data to compute RMS voltage for organoid detection. Upon detection, the system halts further probe advancement.

Data acquisition

Recordings were performed with the system set inside an incubator maintained at 37 °C and 5% CO_2 (95% air) (Hercell vios 160i, ThermoFisher, Waltham, MA). Organoids were transferred manually from their culture plate into a recording well using a pipette. The well was preloaded with 0.3 ml of warm BrainPhys medium and topped off with another 0.5 ml post-transfer. After loading the well and the probe into their respective mechanical fixtures, the center of the probe tips (midpoint between shank 2 and shank 3) was automatically aligned to the center of the well by design. A custom Python script then controlled the Z-axis actuator to insert the probe 2 mm above the bottom of the well to provide baseline RMS measurements. The baseline measurement depth was configured based on predetermined maximum size of the organoids. Automated descent proceeded at steps of 50 µm, until one or more channels became active, at which point step size was reduced to 20 µm for a more precise depth control. The insertion stopped if any of the following conditions were met: (1) the probe reached the bottom of the well; (2) no new active channels were detected during three consecutive steps; or (3) the leading channel near the tip of any of the shanks became inactive (fell below the pre-computed threshold). This closed-loop, precision driving helped to avoid overshooting into necrotic or immature zones. Recordings were acquired for at least one hour at a sampling rate of 20 kHz, corresponding to approximately 9 GB/hour.

Single-unit analysis

Spike sorting was performed on a representative hCO session using Kilosort4[62]. Out of 121 clusters identified, we only kept 48 well isolated single-units that were labeled by Kilosort4 as “Good Units” and had ISI index[63] below 0.2. Features were visualized using CellExplorer[44], a MATLAB-based pipeline for single-unit analysis and classification in 2D space defined by spike width and ACH-based metrics.

Spike feature extraction

A custom Python algorithm processed raw recordings by applying a bandpass filter (300-6,000 Hz), followed by spike detection using a 5 SD threshold. Each detected spike waveform consisted of 40 sample points spanning a 2 ms window centered at the trough (maximum absolute voltage deflection). We computed four spike features (**Fig. 4**) for every spike and extracted spike-level metadata, including timestamp, channel, and session ID. These were stored in concatenated Pandas DataFrames[64], grouped by organoid class and encompassing all spikes from corresponding sessions.

Binary classification of spike blocks using an LSTM Model

We split all hCO and hCO^{APP} recording sessions into training (64%), validation (16%) and testing (20%) sets. The LSTM model was trained and optimized using the training/validation sets and all performance metrics were measured using the testing set. Since a limited number of sessions was available, we segmented each session into 500-spike blocks as inspired by a previous work[51], in order to increase data variance. Each block was treated as a separate training instance while maintaining intra-block temporal order. To avoid overrepresentation of long sessions, we retained a maximum of 250 blocks (first 125,000 sequential spikes) per session. The resulting model input was a 500 by 4 spike-by-feature matrix for every block. Separate files tracked metadata associated with every block, including unique session identifier, class label, and split category (train/validation/test).

We implemented the LSTM classifier using TensorFlow/Keras. We used Optuna[65] to optimize hyperparameters including the number of recurrent units, dropout rate, dense layer width, and learning rate. The best-performing architecture comprised an LSTM layer with 64 input units, dropout rate of 0.1, a 32-unit ReLU dense layer, and a single sigmoid output unit. Training was performed using binary cross-entropy loss and the Adam optimizer[66] (learning rate = 0.001).

Statistical analyses

All statistical tests used a significance threshold of $\alpha=0.05$. Descriptive statistics (n, mean, median, IQR, SD, SEM) are reported in the text, figure legends, and tables. Group medians were compared using two-tailed Mann-Whitney U-tests. Statistical significance is annotated in figures as follows: ns: $p>0.05$; */**/***: $p<0.05/0.01/0.001$.

REFERENCES

1. Burns, T. C., Li, M. D., Mehta, S., Awad, A. J. & Morgan, A. A. Mouse models rarely mimic the transcriptome of human neurodegenerative diseases: A systematic bioinformatics-based critique of preclinical models. *Eur J Pharmacol* **759**, 101–117 (2015).
2. Zeiss, C. J. From Reproducibility to Translation in Neurodegenerative Disease. *ILAR J* **58**, 106–114 (2017).
3. Polis, B. & Samson, A. O. Addressing the Discrepancies Between Animal Models and Human Alzheimer's Disease Pathology: Implications for Translational Research. *Journal of Alzheimer's Disease* **98**, 1199–1218 (2024).
4. Babiloni, C. *et al.* What electrophysiology tells us about Alzheimer's disease: a window into the synchronization and connectivity of brain neurons. *Neurobiol Aging* **85**, 58–73 (2020).
5. Koch, C. *et al.* Next-generation brain observatories. *Neuron* **110**, 3661–3666 (2022).
6. McGregor, M. M. & Nelson, A. B. Circuit Mechanisms of Parkinson's Disease. *Neuron* vol. 101 1042–1056 Preprint at https://doi.org/10.1016/j.neuron.2019.03.004 (2019).
7. Lancaster, M. A. *et al.* Cerebral organoids model human brain development and microcephaly. *Nature* **501**, 373–379 (2013).
8. Pașca, S. P. The rise of three-dimensional human brain cultures. *Nature* **553**, 437–445 (2018).
9. Uzquiano, A. & Arlotta, P. Brain organoids: the quest to decipher human-specific features of brain development. *Curr Opin Genet Dev* **75**, 101955 (2022).
10. Marshall, L. J., Bailey, J., Cassotta, M., Herrmann, K. & Pistollato, F. Poor Translatability of Biomedical Research Using Animals — A Narrative Review. *Alternatives to Laboratory Animals* **51**, 102–135 (2023).
11. Garner, J. P. The Significance of Meaning: Why Do Over 90% of Behavioral Neuroscience Results Fail to Translate to Humans, and What Can We Do to Fix It? *ILAR J* **55**, 438–456 (2014).
12. Birtele, M., Lancaster, M. & Quadrato, G. Modelling human brain development and disease with organoids. *Nat Rev Mol Cell Biol* **26**, 389–412 (2025).
13. Cakir, B., Kiral, F. R. & Park, I.-H. Advanced in vitro models: Microglia in action. *Neuron* **110**, 3444–3457 (2022).
14. Popova, G. *et al.* Human microglia states are conserved across experimental models and regulate neural stem cell responses in chimeric organoids. *Cell Stem Cell* **28**, 2153-2166.e6 (2021).
15. Ciceri, G. *et al.* An epigenetic barrier sets the timing of human neuronal maturation. *Nature* **626**, 881–890 (2024).
16. Di Bella, D. J., Domínguez-Iturza, N., Brown, J. R. & Arlotta, P. Making Ramón y Cajal proud: Development of cell identity and diversity in the cerebral cortex. *Neuron* **112**, 2091–2111 (2024).
17. Scuderi, S. *et al.* Specification of human brain regions with orthogonal gradients of WNT and SHH in organoids reveals patterning variations across cell lines. *Cell Stem Cell* **32**, 970-989.e11 (2025).
18. Landry, C. R. *et al.* Electrophysiological and morphological characterization of single neurons in intact human brain organoids. *J Neurosci Methods* **394**, 109898 (2023).
19. Sharf, T. *et al.* Functional neuronal circuitry and oscillatory dynamics in human brain organoids. *Nat Commun* **13**, 4403 (2022).
20. Oliva, M. *et al.* Standardizing a method for functional assessment of neural networks in brain organoids. *J Neurosci Methods* **409**, 110178 (2024).
21. Farre, C., George, M., Brüggemann, A. & Fertig, N. Ion channel screening – automated patch clamp on the rise. *Drug Discov Today Technol* **5**, e23–e28 (2008).
22. Farinelli, F. *et al.* Automated high-throughput patch clamp electrophysiology of hiPSC-derived neuronal models. Preprint at https://doi.org/10.1101/2025.05.12.653142 (2025).
23. Cartiglia, M. *et al.* A 4096 channel event-based multielectrode array with asynchronous outputs compatible with neuromorphic processors. *Nat Commun* **15**, 7163 (2024).

24. Thomas Jr, C., Springer, P., Loeb, G., Berward-Netter, Y. & Okun, L. A miniature microelectrode array to monitor the bioelectric activity of cultured cells. *Exp Cell Res* **74**, 61–66 (1972).
25. Passaro, A. P. & Stice, S. L. Electrophysiological Analysis of Brain Organoids: Current Approaches and Advancements. *Front Neurosci* **14**, 1–13 (2021).
26. McCready, F. P., Gordillo-Sampedro, S., Pradeepan, K., Martinez-Trujillo, J. & Ellis, J. Multielectrode Arrays for Functional Phenotyping of Neurons from Induced Pluripotent Stem Cell Models of Neurodevelopmental Disorders. *Biology (Basel)* **11**, 316 (2022).
27. Seymour, J. P., Wu, F., Wise, K. D. & Yoon, E. State-of-the-art mems and microsystem tools for brain research. *Microsyst Nanoeng* **3**, 1–16 (2017).
28. Csicsvari, J. *et al.* Massively parallel recording of unit and local field potentials with silicon-based electrodes. *J Neurophysiol* **90**, 1314–1323 (2003).
29. Quadrato, G. *et al.* Cell diversity and network dynamics in photosensitive human brain organoids. *Nature* **545**, 48–53 (2017).
30. Revah, O. *et al.* Maturation and circuit integration of transplanted human cortical organoids. *Nature* **610**, 319–326 (2022).
31. Chishti, M. A. *et al.* Early-onset Amyloid Deposition and Cognitive Deficits in Transgenic Mice Expressing a Double Mutant Form of Amyloid Precursor Protein 695. *Journal of Biological Chemistry* **276**, 21562–21570 (2001).
32. HajjHassan, M., Chodavarapu, V. & Musallam, S. NeuroMEMS: Neural Probe Microtechnologies. *Sensors* **8**, 6704–6726 (2008).
33. Xiang, Y. *et al.* Fusion of Regionally Specified hPSC-Derived Organoids Models Human Brain Development and Interneuron Migration. *Cell Stem Cell* **21**, 383-398.e7 (2017).
34. Bagley, J. A., Reumann, D., Bian, S., Lévi-Strauss, J. & Knoblich, J. A. Fused cerebral organoids model interactions between brain regions. *Nat Methods* **14**, 743–751 (2017).
35. Birey, F. *et al.* Assembly of functionally integrated human forebrain spheroids. *Nature* **545**, 54–59 (2017).
36. Tanaka, Y., Cakir, B., Xiang, Y., Sullivan, G. J. & Park, I.-H. Synthetic Analyses of Single-Cell Transcriptomes from Multiple Brain Organoids and Fetal Brain. *Cell Rep* **30**, 1682-1689.e3 (2020).
37. Cakir, B. *et al.* Engineering of human brain organoids with a functional vascular-like system. *Nat Methods* **16**, 1169–1175 (2019).
38. He, Z. *et al.* An integrated transcriptomic cell atlas of human neural organoids. *Nature* **635**, 690–698 (2024).
39. Hardy, J. & Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science (1979)* **297**, 353–356 (2002).
40. Gulisano, W. *et al.* Role of Amyloid-β and Tau Proteins in Alzheimer's Disease: Confuting the Amyloid Cascade. *Journal of Alzheimer's Disease* **64**, S611–S631 (2018).
41. Kent, S. A., Spires-Jones, T. L. & Durrant, C. S. The physiological roles of tau and Aβ: implications for Alzheimer's disease pathology and therapeutics. *Acta Neuropathol* **140**, 417–447 (2020).
42. Rasmussen, C. *et al.* P2-002: An Integrated Mechanistic Amyloid Pathway Model Informed by Clinical Biomarker Data from Verubecestat and other BACE Inhibitors. *Alzheimer's & Dementia* **15**, 571 (2019).
43. Fee, M. S., Mitra, P. P. & Kleinfeld, D. Automatic sorting of multiple unit neuronal signals in the presence of anisotropic and non-Gaussian variability. *J Neurosci Methods* **69**, 175–188 (1996).
44. Petersen, P. C., Siegle, J. H., Steinmetz, N. A., Mahallati, S. & Buzsáki, G. CellExplorer: A framework for visualizing and characterizing single neurons. *Neuron* **109**, 3594-3608.e2 (2021).
45. Someck, S. *et al.* Positive and biphasic extracellular waveforms correspond to return currents and axonal spikes. *Commun Biol* **6**, 950 (2023).
46. Spivak, L., Someck, S., Levi, A., Sivroni, S. & Stark, E. Wired together, change together: Spike timing modifies transmission in converging assemblies. *Sci Adv* **10**, (2024).
47. Rossant, C. *et al.* Spike sorting for large, dense electrode arrays. *Nat Neurosci* **19**, 634–641 (2016).

48. Csicsvari, J., Hirase, H., Czurko, A. & Buzsáki, G. Reliability and State Dependence of Pyramidal Cell–Interneuron Synapses in the Hippocampus. *Neuron* **21**, 179–189 (1998).
49. Ghatak, S. *et al.* Mechanisms of hyperexcitability in Alzheimer's disease hiPSC-derived neurons and cerebral organoids vs isogenic controls. *Elife* **8**, (2019).
50. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput* **9**, 1735–1780 (1997).
51. Sukman, L. J. & Stark, E. Cortical Pyramidal and Parvalbumin Cells Exhibit Distinct Spatiotemporal Extracellular Electric Potentials. *eNeuro* **9**, ENEURO.0265-22.2022 (2022).
52. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
53. Choi, J. S., Lee, H. J., Rajaraman, S. & Kim, D.-H. Recent advances in three-dimensional microelectrode array technologies for in vitro and in vivo cardiac and neuronal interfaces. *Biosens Bioelectron* **171**, 112687 (2021).
54. Wu, L. & Ma, X. Binary Classification is Enough: A Lightweight Strategy for Drug Screening with Small Datasets. in *2023 2nd International Conference on Health Big Data and Intelligent Healthcare (ICHIH)* 77–81 (IEEE, 2023). doi:10.1109/ICHIH60370.2023.10396144.
55. Crofton, K. M. *et al.* Developmental neurotoxicity testing: recommendations for developing alternative methods for the screening and prioritization of chemicals. *ALTEX* **28**, 9–15 (2011).
56. Richard, A. M. *et al.* ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology. *Chem Res Toxicol* **29**, 1225–1251 (2016).
57. Moreth, J., Mavoungou, C. & Schindowski, K. Is abeta a sufficient biomarker for monitoring anti-abeta clinical studies? A critical review. *Front Aging Neurosci* **5**, (2013).
58. Yoshimoto, S., Taguchi, M., Sumi, S., Oka, K. & Okamura, K. Establishment of a novel protocol for formalin-fixed paraffin-embedded organoids and spheroids. *Biol Open* **12**, (2023).
59. Braun, E. *et al.* Comprehensive cell atlas of the first-trimester developing human brain. *Science (1979)* **382**, (2023).
60. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences* **112**, 7285–7290 (2015).
61. Xu, C. *et al.* Automatic cell-type harmonization and integration across Human Cell Atlas datasets. *Cell* **186**, 5876-5891.e20 (2023).
62. Pachitariu, M., Sridhar, S., Pennington, J. & Stringer, C. Spike sorting with Kilosort4. *Nat Methods* **21**, 914–921 (2024).
63. Hill, D. N., Mehta, S. B. & Kleinfeld, D. Quality Metrics to Accompany Spike Sorting of Extracellular Signals. *Journal of Neuroscience* **31**, 8699–8705 (2011).
64. McKinney, W. Data Structures for Statistical Computing in Python. in *SciPy* 56–61 (2010). doi:10.25080/Majora-92bf1922-00a.
65. Akiba, T., Sano, S., Yanase, T., Ohta, T. & Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. *ArXiv* (2019).
66. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. (2014) doi:https://arxiv.org/abs/1412.6980.