A multiparametric activity profiling platform for neuron disease phenotyping and drug screening

Bruno Boivin[a,b,†], Kasper C. D. Roet[a,b,†], Xuan Huang[a,b], Kyle W. Karhohs[c], Mohammad H. Rohban[c], Jack Sandoe[d], Ole Wiskow[d], Rie Maeda[a], Alyssa Grantham[a], Mary K. Dornon[a], Jenny Shao[a], Devlin Frost[a], Dylan Baker[d], Kevin Eggan[d], Anne E. Carpenter[c], and Clifford J. Woolf[a,b,*]

[a]F. M. Kirby Neurobiology Center, Boston Children's Hospital, Boston, MA 02115; [b]Department of Neurobiology, Harvard Medical School, Boston, MA 02115; [c]Broad Institute of Harvard and MIT, Imaging Platform, Cambridge, MA 02142; [d]Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Cambridge, MA 02138

ABSTRACT Patient stem cell–derived models enable imaging of complex disease phenotypes and the development of scalable drug discovery platforms. Current preclinical methods for assessing cellular activity do not, however, capture the full intricacies of disease-induced disturbances and instead typically focus on a single parameter, which impairs both the understanding of disease and the discovery of effective therapeutics. Here, we describe a cloud-based image processing and analysis platform that captures the intricate activity profile revealed by GCaMP fluorescence recordings of intracellular calcium changes and enables the discovery of molecules that correct 153 parameters that define the amyotrophic lateral sclerosis motor neuron disease phenotype. In a high-throughput screen, we identified compounds that revert the multiparametric disease profile to that found in healthy cells, a novel and robust measure of therapeutic potential quite distinct from unidimensional screening. This platform can guide the development of therapeutics that counteract the multifaceted pathological features of diseased cellular activity.

INTRODUCTION

Neurodegenerative diseases are among the most difficult to treat. One such disease, amyotrophic lateral sclerosis (ALS), is associated with a progressive loss of upper and lower motor neurons, leading to a gradual loss of control over the muscles vital for walking, talk-ing, swallowing, and breathing, a debilitating and typically rapidly fatal outcome for patients. ALS remains a clinical challenge with only two US Food and Drug Administration (FDA)-approved drugs, both though, with minimal life-prolonging effects. The prognosis for patients diagnosed with ALS is poor, with most patients succumbing to the disease within 3–5 years. The lack of effective treatments stems from an incomplete understanding of the biological basis of the disease as well as the use of preclinical drug screening approaches that are not predictive of efficacy in patients.

In its familial form, ALS is caused by mutations in the *SOD1, C9orf72, FUS,* and *TARBP1* genes, among many others (Renton *et al.*, 2014), and motor neuron electrical activity is severely impacted, as detected in both patients and patient-derived motor neurons (Wainger *et al.*, 2014; Iwai *et al.*, 2016). Mouse models as well as patient iPSC-derived motor neurons reveal alterations in the excitability phenotype across multiple parameters, including membrane potential, sodium and potassium peak currents, action potential firing, and synaptic activity (Kuo *et al.*, 2005; van Zundert *et al.*, 2008; Kiskinis *et al.*, 2014). The development of therapies targeting aberrant electrophysiological properties usually focuses only on a single electrophysiological property, the action potential firing rate (Catterall, 2018). A recent study from our group successfully applied the approach of using a reduction in ALS human motor neuron firing

rate to validate two known ALS activity modulators (Kv7.2/3 ion channels and AMPA receptors) and identify D2 dopamine receptors as a novel target (Huang *et al.*, 2021). However, the excitability phenotype in ALS motor neurons is closely linked to multiple changes in cell health, such as endoplasmic reticulum stress and unfolded protein response (Kiskinis *et al.*, 2014), and to changes in ion channel composition, molecular architecture and posttranslational state, membrane trafficking, intra- and extracellular ion concentrations, and synapse strength, which impact diverse aspects of neuronal activity, especially in neurodegenerative disorders (Kiskinis *et al.*, 2014). We hypothesized that a multifaceted disease etiology requires a multiparameter analysis method and that simultaneous measurements of distinct cellular activity parameters such as peak amplitude, frequency, rise and fall time, and other kinetics might provide a more comprehensive, and therefore superior, basis for defining complex disease phenotypes. In turn, multiparameter compound screening may also better detect disease-modifying actions of drugs than traditional single parameter–based assays. Molecules that counteract the disease pathology across multiple dimensions may provide superior rescue from a disease state and have a greater impact on disease progression than simply focusing on the restoration of a single feature, opening the possibility for novel and more precise therapeutics targeting the complexity of the disease state.

Here, we describe the development and application of a high-throughput cloud-based image processing and unbiased multiparametric activity profiling analytic platform. We call it CELLXPEDITE for its ability to process large volumes of cellular imaging data, thereby accelerating drug screening, and make it open source for use by the neuroscience community. By analyzing 153 parameters simultaneously we can automatically capture the complex activity profiles produced by diseased cells (ALS patient–derived carrying the *SOD1(A4V)* mutation) and the effects of candidate compounds on such cells and from this identify compounds that convert the complex disease profile to one similar to that present in healthy cells. We show that such a platform can identify subtle perturbances in the activity of motor neurons and enables a robust selection of compounds that reverse the complex disease phenotype.

Most drugs for neurodegenerative diseases fail in clinical trials due to poor efficacy or unforeseen side effects, which is costly for the pharmaceutical industry and a health risk for the patients. Here we combine advances in human stem cell–derived neuronal models, fluorescent reporters, high-throughput live-cell imaging systems, and cloud analysis platforms to enable the discovery of molecules that can transform a complex disease phenotype to a healthy phenotype.

RESULTS

Extraction of cellular activity

During action potential firing, voltage-dependent calcium channels are activated and deactivated, leading to changes in intracellular calcium that allow the measurement of neuronal activity through the genetically encoded calcium reporter GCaMP6 (Chen *et al.*, 2013). We developed a robust high-throughput 384-well single-cell GCaMP6-based activity assay to record the spontaneous firing of iPSC-derived human motor neurons with and without the highly penetrant *SOD1(A4V)* mutation (Figure 1A). To capture a multiparametric representation of GCaMP activity in each individual cell, we developed a scalable cloud-based live-cell image processing pipeline that automatically compensates for imaging artifacts, performs photobleaching corrections, identifies individual neurons, and extracts and denoises calcium transients (Figure 1, B and C). In contrast to existing commercial or open-source software packages

(Mukamel *et al.*, 2009; Pnevmatikakis *et al.*, 2016; Moein *et al.*, 2018; Giovannucci *et al.*, 2019), our workflow is seamless, extracts multiple complex parameters, and is suitable for large-scale drug screening. Our aim was to better capture the intricacies of cellular activity for disease modeling and enable high-throughput disease phenotype correction screening.

High-throughput plate readers generally have an uneven spatial distribution of light across the field of view, which impacts both image segmentation and fluorescence quantification. We therefore incorporated into our pipeline modules from the open-source software CellProfiler (McQuin *et al.*, 2018) to perform illumination correction and enhance the visibility of neurons expressing GCaMP6 (Supplemental Figure 1). Our approach spatially resolved all cells, regardless of temporal activity, allowing us to determine the proportion of active cells and assess well contamination and compound toxicity. Calcium traces were automatically extracted for each neuron and compensated for background luminescence and photobleaching effects (Supplemental Figure 2). To validate that the residual signals were reflective of spiking activity, we extracted fluorescence traces in the presence and absence of tetrodotoxin (TTX), a sodium channel blocker that inhibits the firing of action potentials in neurons (Narahashi *et al.*, 1967). TTX eliminated the spontaneous calcium wave fluctuations observed in control cells (dimethyl sulfoxide [DMSO] vehicle) (Figure 1D). Baseline fluorescence was also lower in the presence of TTX, suggesting that the measurements are sensitive to both absolute and relative differences in calcium levels.

Multiparametric GCaMP phenotypes

The comparison of GCaMP calcium transients between neurons, as well as changes due to drug action, has traditionally relied on counting the total number of peaks in fluorescence as the sole parameter (Figure 2A), an extension of spike counting by electrophysiological recordings. However, because GCaMP6 fluorescence fluctuations are heterogeneous and unevenly dispersed in time (Figure 1, C and D), simple metrics such as peak count or amplitude convey an incomplete picture (Figure 2A). The disparity between the complex nature of the cellular activity and its historical unidimensional characterization raises concerns over the accuracy of such measurements for reflecting the disease state of the neuron and its response to drug treatment.

To provide a more comprehensive readout of excitability in diseased and healthy motor neurons, we set out to parameterize spontaneous fluctuations of the calcium reporter signal. We identified peaks in activity using a differential approach with high sensitivity to rapid fluctuations as well as continuous wavelet transforms (Du *et al.*, 2006) to account for multiscale peaks (Figure 2B). Peaks were then individually parsed to quantify subtle kinetic differences and fluorescence shifts (Figure 2C). Relative and absolute changes were detected by establishing three baselines for the measurements, namely the well background intensity, the cell minimum intensity, and the intensity at peak onset (Figure 2D). To complement the individual peak dissection with signal-wide features, statistical dispersion metrics were computed along with the area under each signal (Figure 2D). Finally, the power spectrum of the calcium traces, that is, the distribution of power into frequency components composing the signals, was obtained by computing their discrete Fourier transform. A total of 153 parameters spanning both time and frequency domains were acquired to describe the activity of each cell (Figure 2E), providing a comprehensive breakdown of activity and widening the scope of the activity phenotype across multiple axes.
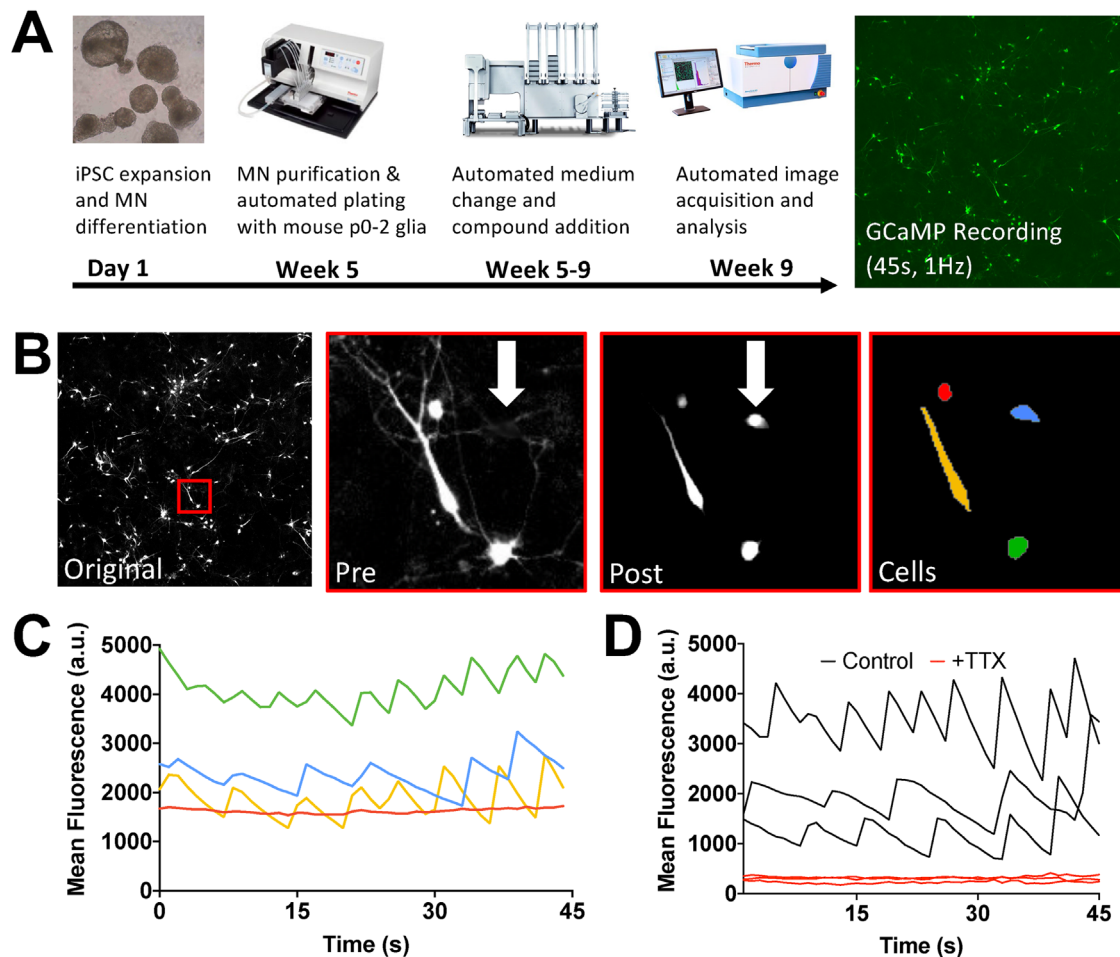
FIGURE 1: Cellular activity extraction pipeline. (A) Schematic overview of the differentiation, plating, maturation, and imaging (whole field, 5×) of GCaMP6-positive human ALS patient–derived motor neurons. (B) Analysis of the fluorescence imaging data at a single time point does not allow for the identification of inactive cells ("original" & "pre"). Temporal projection of the calcium imaging data over 45 s, combined with spatial filtering, enables the identification of all cells regardless of activity ("post" & "cells"). (C) Fluorescence traces of cells identified in B. (D) Spontaneous neuronal activity in DMSO-treated cells is eliminated by TTX treatment.

Healthy and disease activity profiles

To characterize the complex ALS disease phenotype, we utilized the image processing and analysis platform to generate and study all the parameters derived from the spontaneous activity of two cell lines derived from a human ALS patient (39b: ALS disease model and 39b-cor: isogenic corrected healthy control). In addition to the expected increased peak counts in the disease cell line, we discovered changes in activity parameters reflective of slower kinetics and higher peak amplitudes (Figure 2F). With the demonstration that cellular activity is best captured using multiple parameters, we then conducted a full multiparametric analysis of the disease (39b) and control (39b-cor) motor neuron activity phenotypes. These phenotypes were generated by averaging the multiparametric activity profiles across eight replicates for each cell line. From these profiles, we calculated the absolute difference across all parameters between the two cell lines, which resulted in a disease profile consisting of 153 activity parameters (top 25 distinguishing features presented in Figure 3B). Surprisingly, the largest differences were found in features related to peak amplitude as opposed to peak count, though the latter remained in the top five distinguishing features. We found properties from the frequency domain to be of lesser importance. The multiparametric disease profile was then used to identify phe-

notypic rescue of all the disease-distinguishing cellular activity parameters back to those in the healthy control.

ALS multiparametric drug screen

Given the rich multiparametric readout of the ALS hyperexcitability phenotype, we sought to identify disease-correcting drug candidates based on their specific effects on disease activity profiles. We screened 1902 compounds from the Selleck bioactive compound library, which includes FDA-approved compounds, active pharmaceutical agents, chemotherapeutic agents, and natural products. Compounds were mapped to points in the multidimensional space defined by the 153 features in their activity profiles (Figure 2E), with distances between them reflecting the degree of similarity between the phenotypes they produced. We discovered a cluster of compounds that caused the disease-allele–containing cells to change and resemble control healthy cells, providing a pool of promising disease-correcting candidates (Figure 3C). Among these were phosphodiesterase inhibitors and inhibitors of the mTOR pathway, both previously associated with neuroprotection (Nakamizo *et al.*, 2003; Mandrioli *et al.*, 2018). Because our pilot screen suggested that different activity profiles between the disease and control states reflect the modulation of activity, as opposed to a complete blockade of
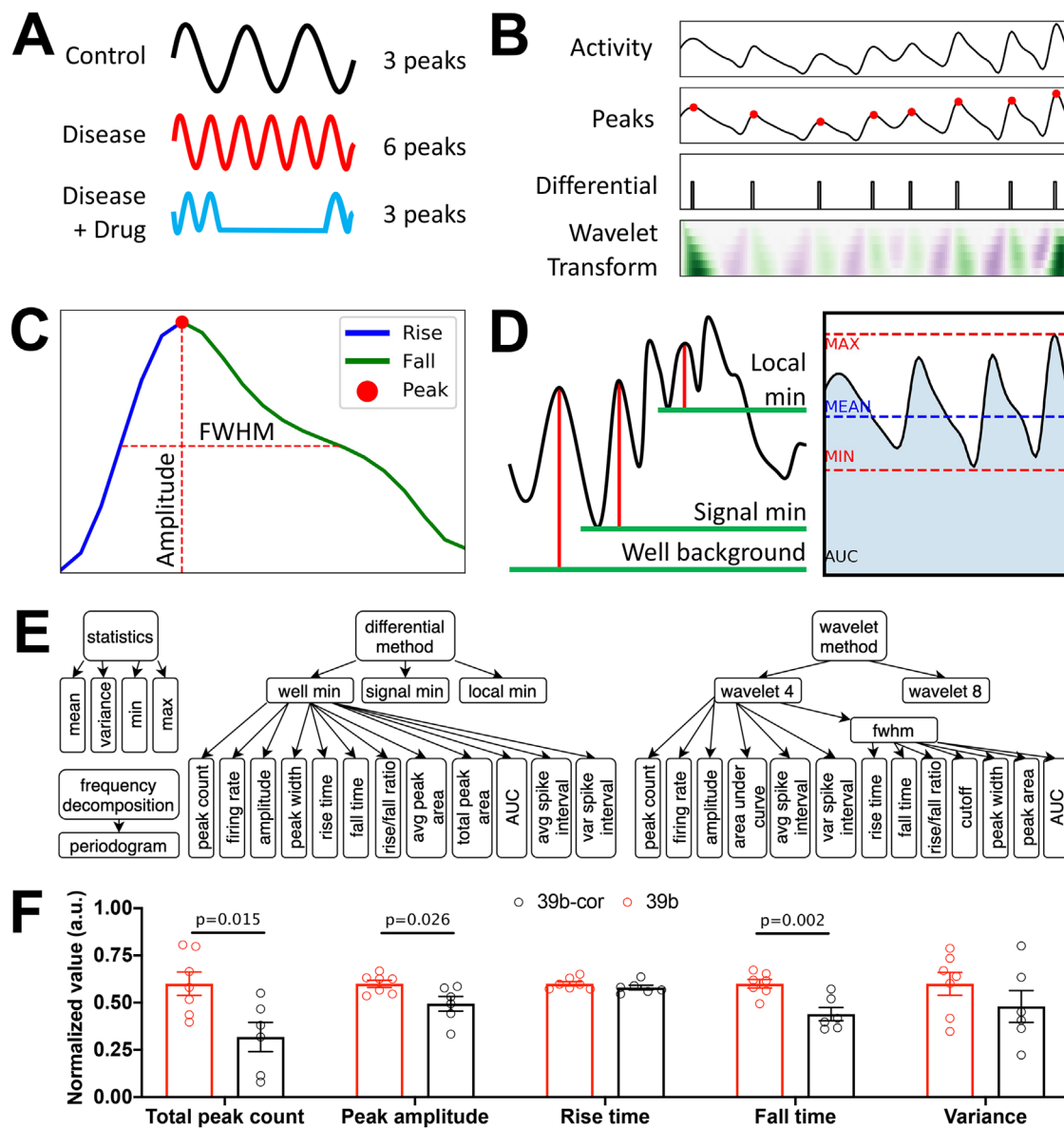
FIGURE 2: Parameterization of cellular activity. (A) Traditional peak counting does not identify temporal (e.g., uneven interspike intervals) or amplitude differences in calcium transients. (B) A combination of differential and continuous wavelet transforms accurately identifies peaks in activity. (C) Automated peak deconvolution quantifies the rise, apex, fall, amplitude, and full width at half maximum of an individual peak. (D) Cellular peak parameters are quantified relative to three baselines: well background intensity, cell minimum intensity, and intensity at the peak onset (left). Signal-wide features of activity, including dispersion metrics (minimum, maximum, mean) and the area under the curve (AUC), are automatically captured (right). (E) Breakdown of the 153 activity parameters computed for each cell. A single node under the differential and wavelet methods is expanded for clarity; the other nodes share the same parameter subtree. (F) Spontaneous GCaMP activity analysis of SOD1^{A4V} human motor neurons compared with isogenic corrected motor neurons (39b-cor) reveals variable differences across five example parameters. Features are normalized to the average of the 39b group and presented as mean ± SEM using Welch's *t* test.

activity, we also investigated the effects produced by calcium-channel blockers that abolish the GCaMP signals triggered by action potentials. While these blockers clustered together, they were separated from the control phenotype and were not disease correcting, highlighting the sensitivity of the multiparametric approach to discriminate distinct activity states. Multiple morphological features (46 total) extracted for each cell revealed additional phenotypic differences between the hits identified here and also for calcium-channel blockers and identified differences across groups of compounds with distinct mechanisms of action (Supplemental Figure 4). Mea-

surement of these morphological features can help identify which cell types are most sensitive to a particular compound and will also enable disease-based changes in morphology and its recovery to be determined.

After excluding compounds that resulted in potent inhibition of neuronal activity (including, for example, 5 μM TTX; see Supplemental Table 1), we selected the 65 compounds with the shortest distance (strongest similarity) to the control profile and tested an additional six replicates for each candidate over two independent recording sessions. We advanced the 80% most consistent
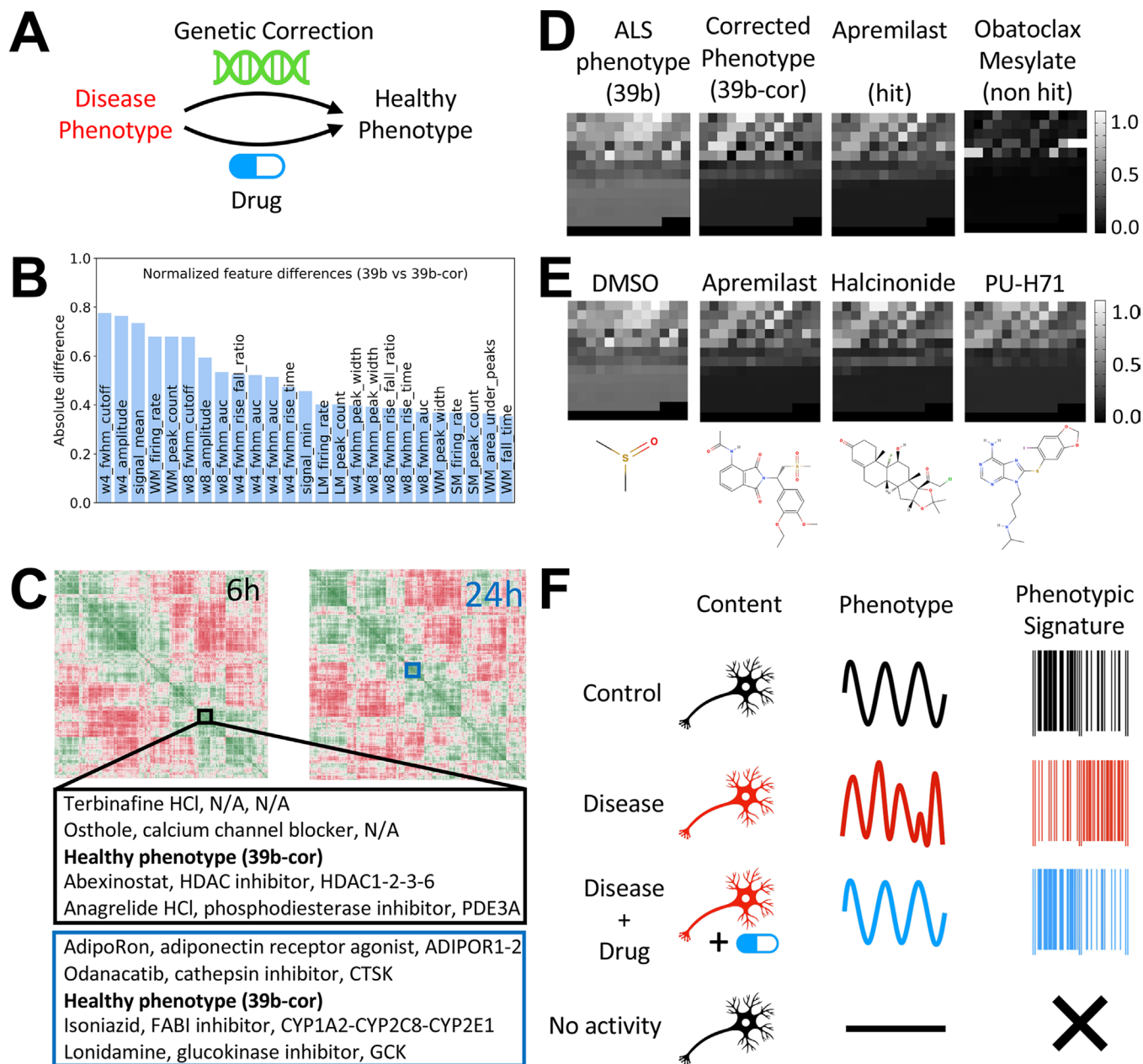
FIGURE 3: Multiparametric drug screening strategy. (A) Schematic illustrating that both genetic mutation correction and drug action can restore a healthy activity phenotype in a diseased cell line. (B) Ranking of the top 25 individual parameter differences between the disease (39b) and corrected cell line (39b-cor) reveals the relative importance of features. (C) Hierarchically clustered heatmap of 1902 compounds from the Selleck bioactive library 6 and 24 h after treatment; each compound is represented using the 153 activity features normalized to the control phenotype. Compound name, mechanism of action, and molecular targets of the four hits that resulted in cellular activity closest to the healthy phenotype at each time point are shown. (D) Comparison of the multiparametric activity signatures for the disease phenotype (39b), the healthy phenotype (39b-corrected), a hit compound (apremilast), and a non–hit compound (obatoclax mesylate) on 39b motor neurons (MN) distinguishes activity patterns. Features are ordered based on preassigned positions in the grid and normalized to fit on the same intensity scale. (E) Comparison of the negative control (DMSO) and three hits (apremilast, halcinonide, PU-H71) that replicated in a validation screen on 39b MN reveals similar activity profiles and distinct molecular structures across the hits. (F) Depiction of phenotypic signatures capturing complex activity phenotypes in control and disease neuronal populations (first two rows). The action of an effective drug on diseased cells restores the control (healthy) phenotype (third row). No phenotypic signatures are produced for inactive cells (fourth row). The comparison of signatures provides a basis for measuring drug efficacy.

compounds (Supplemental Figure 3C) for further analysis. After ordering these compounds by their ability to normalize the disease state back to a healthy phenotype, as measured by the distance between the multiparametric profiles, the three compounds that most strongly and reproducibly normalized activity in the disease motor neurons were apremilast CC-10004 (phosphodiesterase inhibitor), halcinonide (glucocorticoid receptor agonist), and PU-H71 (a heat shock protein inhibitor).

Next, we wanted to confirm the benefits of using a multiparametric method over the traditional unidimensional approach. In addition to the 48 compounds that abolished neuronal activity (Supplemental Table 1), that is, compounds that resulted in complete activity suppression as opposed to a normalization of the disease activity profile, we also identified hits that reduced only the total number of calcium transient peaks, the top three of which included dexmedetomidine (adrenergic receptor agonist), bimatoprost (prostanoid receptor agonist), and tianeptine (atypical antidepressant). The hit compounds obtained using the multiparametric approach were approximately 20% closer to the targeted healthy control than those obtained from a one-dimensional peak count metric. These results demonstrate that utilization of multiparametric selection criteria offers superior specificity by imposing strict constraints, reducing the number of hits to only those that reverse the disease phenotype across all 153 dimensions. Therefore, screening against multiple disease-related parameters, as opposed to a single one, is likely to lead to the identification of completely different sets of compounds with distinct mechanisms of action and molecular targets and which are therefore likely to have quite different efficacy profiles for the disease in patients.

Activity fingerprints

Next, we compared the hit compounds by converting their activity profiles into quantifiable visual representations (phenotypic signatures), negating the need to perform dimensionality reduction by providing a lens into the multidimensional activity space. We found the activity signatures of the ALS disease phenotype and the healthy phenotype to be quite different, while the signatures from compounds that closely matched the control phenotype were similar to each other and to the healthy profile (Figure 3D). Compounds that introduced changes distinct from the disease or healthy state (potential undesired effects) had very different signatures (Figure 3D). Interestingly, while the activity signatures of the disease-phenotype–correcting hits were comparable, their molecular structures and targets were different (Figure 3E). These results suggest either that multiple different molecular mechanisms can be utilized to normalize pathological excitability profiles in SOD1^{A4V} motor neurons or that the off-target effects of these quite distinct compounds might be similar, although the latter is statistically most unlikely.

DISCUSSION

We have designed a computational approach that combines the temporal and spatial resolution of high-content microscopy imaging and automated high-throughput screening with the scalability of cloud computing to accelerate the search for new molecules capable of rescuing an ALS motor neuron disease phenotype. The pipeline processes imaging data from individual wells in parallel to maximize throughput. Imaging artifacts are removed by a multilayered correction algorithm, and objects are identified by CellProfiler modules combined with a temporal profile analysis. While we collect information about the morphology, eccentricity, and surface area of cells, our analytic approach focuses on the dissection of temporal dynamics of GCaMP activity. The calcium dynamics encoded in each time series are parameterized into 153 features that are normalized and combined to form an overall representation of cellular activity (an activity fingerprint or profile). The generation of a multiparametric disease profile combined with unbiased hierarchical clustering of drug efficacy over 153 parameters simultaneously allowed us to identify drugs that can convert a disease profile into a healthy profile and also identify potential undesired side effects from the perspective of a distortion in the multidimensional representation of healthy activity.

Through our large-scale bioactive compound drug-disease phenotype correcting screen, we identified apremilast, halcinonide, and PU-H71 as compounds that normalize the multifaceted GCaMP-based disease activity phenotype back to a healthy phenotype, providing a proof-of-principle for a new drug screening paradigm for ALS and other diseases, though a future study is needed to identify the molecular targets underlying the action of these three compounds. Although there are similarities in the chemical structures of apremilast, halcinonide, and PU-H71, they are not members of the same class of molecules and have different annotated molecular targets, specifically phosphodiesterase 4, corticosteroid hormone receptor, and heat shock protein 90, respectively. It is, therefore, possible or even likely that the observed drug effects are the result of interactions with different targets; thus any structure–activity relationship (SAR) campaign following this kind of multidimensional phenotypic screen will have to embrace the complexity of the disease and include a multitarget deconvolution process. This strategy may be a necessity for the development of successful therapeutics for complex neurological disorders such as ALS and represents a move from single-target identification.

The ability to generate complex stem cell–derived neuronal models for diseases such as ALS has provided a powerful tool for understanding the known genetic forms of the disease. However, for most ALS patients the cause of their disease is unknown and classified as sporadic. The development and execution of an automated and unbiased high-throughput multiparameter activity profiling platform provides an unprecedented opportunity to interrogate neurons derived from many sporadic patient stem cell lines. Such an interrogation could lead to the identification of patients with a shared abnormal activity profile, and in this way, aid our understanding of common complex changes in neuronal activity in disease conditions, as well as the discovery of compounds that can rescue a comprehensive disease phenotype to a healthy one.

We evaluated the performance of our computational approach in terms of its usability, speed, and accuracy. In comparison to previous approaches, the automated measurement of 153 features of activity for each cell removes the inherent bias associated with manual selection of a subset of features and provides a highly sensitive framework for identifying activity disturbances. From a usability standpoint, this negates the need for user-specified parameters that are often arbitrarily chosen and biased. Additionally, our cloud-based design can theoretically scale to any number of wells provided that sufficient computing resources are available. Because wells are processed in parallel and the results combined in constant time, increasing the number of compounds in a drug screen only marginally increases total processing time. Prior assays that include manual interventions are time-consuming, error-prone, and often not reproducible. Once adjusted for the type of cells being studied and the recording frequency, our approach does not require human intervention and can be applied to any number of recordings, provides comprehensive activity profiles, and is entirely reproducible. This is particularly useful when studying complex genetic diseases such as ALS, because it allows for comparison of the activity profiles of different mutations under the same treatment. Therefore, we argue that this computationally efficient and deterministic approach, one that can be launched by the push of a button, is superior to previous methods in many regards.

In principle, our approach can be used at any imaging frequency with any cell for any cellular functional activity that can be fluorescently tagged and for which dynamics is a biologically relevant property. Because it is readily deployable on cloud infrastructures, it provides a seamless way to interrogate disease activity and the

corrections in this phenotype introduced by drug actions. However, the reliance on dynamic data alone excludes morphological changes such as cell size and dendritic arbor changes, which may be important for disease progression. Previous studies, as well as our own observations, indicate that the soma size of motor neurons carrying the *SOD1^{A4V}* mutation is altered as the disease progresses (Iwai *et al.*, 2016; Dukkipati *et al.*, 2018). Researchers interested in looking at broader phenotypes beyond electrophysiological properties, such as image-based features (Chandrasekaran *et al.*, 2021), could readily incorporate these into the pipeline, as a route to uncovering unexpected and diverse phenotypes in an unbiased way.

We anticipate that sophisticated phenotypic screens that embrace the complexity of cellular function and delineate multiple facets of drug action will accelerate the discovery of successful and novel therapeutic interventions. Our findings demonstrate the existence of multiple complex and previously unexploited facets of the ALS excitability phenotype and the discovery of compounds that robustly correct many parameters of this phenotype simultaneously. This discovery, coupled with the sensitive cloud-based high-throughput multiparametric framework for capturing and analyzing subtle changes in cellular activity, represents a substantial leap forward for comprehensive disease profiling and drug screening in the pursuit of novel treatments for ALS and other diseases. Improvements to the earliest phases of the translational process, which are inextricably linked to the success of clinical trials, have the potential to both significantly reduce costs and health risks for patients as well as improve efficacy. Ultimately, the combination of human stem cell models and novel multiparametric screening approaches have, we argue, the promise to lead to safer and more effective therapeutics for ALS.

MATERIALS AND METHODS
Request a protocol through *Bio-protocol*.

Single-cell GCaMP6 screen
iPSC's were expanded and differentiated into motor neurons using an embryonic body–based (EB) protocol as previously described (Kiskinis *et al.*, 2014). At day 24, EBs were dissociated to single cells with accutase, frozen, and stored in liquid nitrogen until fluorescence-activated cell sorting (FACS) purification of motor neurons that were NCAM (BD Biosciences; #561903) positive and EpCAM (BD Biosciences; #347198) negative. Motor neurons were cocultured in Greiner Bio-one 384-well cell culture plates at a density of 7500 cells per well with 10,000 glial cells obtained from P0–P2 C57BL/6 pups (Jackson Laboratory) as described in Di Giorgio *et al.* (2008). Mouse procedures were approved by Boston Children's Hospital Institutional Animal Care and Use Committee (IACUC).

Cells were plated in the presence of 10 µM ROCK inhibitor Y-27632 (Tocris) and 1 µM EDU (Thermo) in motor neuron media (Neurobasal media) (NB; Invitrogen, Carlsbad, CA), supplemented with B27, N2 supplement, glutamax, and nonessential amino acids (GIBCO, Thermo Fisher), 10 ng/ml each of BDNF, GDNF, and CNTF (R&D Systems, Minneapolis, MN), and 0.2 mg/ml ascorbic acid (Sigma). After incubation for 2 d, media was replaced every 3 or 4 d by NB. LV-synapsin-GCaMP6 viral infections were performed in weeks 2 and 3 at a multiplicity of infection (MOI) of 5. After 3–4 wk, synaptic blockers were added to each well at the following concentrations: bicuculline 25 µM, strychnine 10 µM, AP-5 100 µM, CNQX 10 µM. Synaptic blockers are used to prevent cross-talk and network activity between adjacent cells, allowing us to study single-cell excitability. After 2 h, chemogenomic library compounds from the Selleck Bioactive Compound Library (plates 3651–3656; ICCB Long-wood Screening Facility) were added at a final concentration of 3 µM. DMSO was used as a negative control at 0.03%. TTX 5 µM in 0.03% DMSO was used as a positive control. After 6 and 24 h, the plates were recorded for 4–5h in the Arrayscan XTI (Thermo Fisher) with excitation at 485 nm and emission at 521 nm. Recordings were acquired at 1 Hz for 45 s over two independent sessions. We found 1 Hz to be sufficient to capture major calcium transients in motor neurons, but the platform is not specific to this imaging rate and can work with any acquisition frequency.

Illumination correction
Uneven illumination was corrected across the field of view using a CellProfiler (McQuin *et al.*, 2018) pipeline that sequentially applies two built-in modules. The module *CorrectIllumCalculate* was first included to calculate the illumination function from all images across cycles. The smoothing method was set to *Median Filter* with default settings to capture global illumination disparities rather than cell-specific variations in brightness. The other parameters of the module were left unchanged. The *SaveImages* module was then added to save the output upon the last cycle as a 32-bit floating-point tiff image.

Maximum projection
The maximum value of each pixel across all frames in the recording was computed and saved as an 8-bit frame representing the maximum projection of the recording using version 3.7 of the Python programming language (Python Software Foundation, 2018).

Cellular fragment identification
Cells were first identified and morphologically characterized from the maximum projection image by applying a CellProfiler (McQuin *et al.*, 2018) pipeline that combines 13 sequential modules. A sample maximum projection subimage is followed through each module in Supplemental Figure 1B. Starting with the original image (b1), a median filter is applied to blur away small artifacts, with the typical artifact parameter set to 3 (b2). Neurites are then computationally enhanced to become more visible and identifiable using the *EnhanceOrSuppressFeatures* module, with the tubeness enhancement method and smoothing scale set to 2 (b3). The *ImageMath* module is used to average the enhanced features with the blurred image, generating a cleaner delineation of the contours of interest (b4). Clustering-based image thresholding is performed using the Otsu method to separate foreground from background pixels using the *IdentifyPrimaryObjects* module (b5). Simultaneously, objects with diameters ranging from 5 to 35 pixels are identified and their contours outlined (b6). The *IdentifySecondaryObjects* module is used to identify secondary objects such as neurites using the propagation method, which helps find dividing lines between clumped objects and refine cell segregation (b7). The brightness, morphology, and texture of individual fragments are characterized using a combination of modules, specifically *MeasureImageIntensity*, *MeasureObjectIntensity*, *MeasureObjectSizeShape*, and *MeasureTexture*. Consolidation of the location of all pixels belonging to cellular fragments generates a numerical mask saved as a 16-bit tiff image that can be used to identify all or specific subsets of cells (b8). Background pixels are assigned a value of 0 while other pixels are assigned a unique fragment-specific integer identifier between 1 and K, where K is the total number of fragments identified in the well.

Spatial proximity assessment
Pairwise spatial proximity of fragments is assessed by measuring the Euclidean distance between their centroids. Centroids are

calculated using Python's SciPy library. This generates an N×N matrix, where N is the number of fragments and each entry in the matrix the distance in pixels between two fragments.

Activity similarity assessment
Pairwise time series similarity is assessed using the standard implementation of the dynamic time warping algorithm included in version 3.5 of the *mlpy* Python library (Albanese *et al.*, 2012) with default parameters. An unnormalized minimum-distance warp path is computed for each pair of signals. An N×N matrix is thus constructed, where N is the number of fragments and each entry in the matrix the cost associated to the warping path between the signals of two fragments.

Cellular fragment consolidation
The matrices described in Supplemental Figure 1D are first normalized to their respective means through division. The geometric mean of their normalized forms is then computed from the Python implementation included in the *scipy.stats* package included in SciPy version 1.2.1 (Jones *et al.*, 2001), yielding a matrix with each entry being a unified distance between each pair of fragments. The resulting matrix is fed into an agglomerative hierarchical clustering algorithm (Müllner, 2011) that utilizes complete linkage to merge the subtrees. More specifically, the *scipy.cluster.hierarchy* library is used to compute the linkage matrix and to retrieve flat clusters from it, with a threshold of 0.1. The threshold corresponds to the height at which the dendrogram resulting from the hierarchical clustering is cut. Branches below the cut threshold are pruned, yielding consolidated fragments. Branches above the threshold are deemed part of distinct cells. The cell mask previously generated is updated so that pixels belonging to the same cell are assigned the same unique integer identifier. The identifiers assigned range from 1 to N, where N is the total number of cells after the fragments have been recombined.

Extraction of cellular activity
From the whole-cell mask obtained by combining cellular fragments, a mapping from unique cell identifiers to their respective set of pixels is constructed. The activity of individual cells is extracted by averaging at each time point the brightness of their respective pixels.

Decay correction
The decay in fluorescence is estimated by averaging the decay of randomly selected pixels from each well. Specifically, 100 pixels from each well were randomly selected and grouped. A first-degree exponential decay function was fitted to the aggregated data using the *scipy* curve fitting module, with the model function $y = ae^{-bx} + c$, the bounds constricted to the positive range, and an initial guess of the parameters of $a = 2000$, $b = 0.01$, and $c = 200$. The resulting fit curve is normalized through division by its maximum value. The activity of each cell over time is then divided by this normalized decay curve.

Background subtraction
Using the cell mask to identify background pixels, a sample of 10,000 background pixels are randomly selected and their intensity averaged at each time point. This generates a time series that estimates background intensity over time. This curve is then subtracted from the activity of each cell.

Peak identification
Peaks in cellular activity were identified using code derived from the *find_peak* function of the *scipy* package. Peaks were labeled on their rising edge, with a minimum interpeak distance of 1. Troughs were identified by applying the same approach onto the inverse of the time series. Peaks were also identified using the continuous wavelet transform method provided in the *scipy* package. Wavelets of width up to 4 (W4) and up to 8 (W8) were used separately to identify distinct sets of peaks, each time with a gap threshold of 3 and a maximum distance of half the maximum width. The W4 approach used a minimum ridge line length of 3, a minimum signal-to-noise ratio (SNR) of 1, and a noise percentile of 10. The W8 approach used a minimum ridge line length of 5, a minimum SNR of 2, and a noise percentile of 10.

Baselines
Three baselines were established to compute individual peak characteristics. Because background intensity was subtracted from each signal, measurements relative to the $y = 0$ line were used to obtain features relative to the background intensity. The respective minimum fluorescence value of each cell across the 45-s recording was used a second baseline. The third baseline was determined for each peak by taking the fluorescence value at the onset of the peak.

Peak threshold
For each baseline, a peak threshold was computed to distinguish noise from actual peaks. The noise level was established by taking the amplitude distribution of all peaks in the positive control wells, here those treated with TTX. The spontaneous peak level was established by taking the amplitude distribution of all peaks in the negative control wells, here those treated with DMSO. The peak threshold was then calculated using the Otsu method from the combined distribution of peak heights using the SciPy Python package.

Peak features
Under the differential method, for each peak and each baseline, amplitude was measured relative to the baseline. Peaks whose height fell below the minimum peak threshold were discarded. Peak count was computed as the remaining number of peaks. Rise time was calculated as the time between the peak apex and the previous trough. Fall time was calculated as the time between the peak apex and the following trough. Peak width was measured as the time between the onset of the rise and the end of the fall. The area under a given peak was computed using the trapezoidal rule along the width of the peak. The average and variance of the time interval between peaks was also calculated.

Under the wavelet methods, the full width at half maximum was used as a reference point for computing the peak features. Rise time was calculated as the time between the half-maximum point along the rise and the peak apex. Fall time was calculated as the time between the peak apex and the half-maximum point along the fall. The other peak features were obtained using the same logic used under the differential method.

Signal features
Statistical metrics including the minimum, maximum, mean, and variance of each trace were computed using the Python *numpy* package version 1.18 (van der Walt *et al.*, 2011). The area under the entire signal was approximated using the trapezoidal rule provided in the same package. The power spectral density of each trace was obtained using the *periodogram* module of the *scipy.signal* package.

Multidimensional activity clustering

Individual cell activity is described by the set of 153 features. Individual well activity is computed as the mean of each feature across all cells in the well. The control phenotype is obtained by averaging the features from the control wells, here the healthy control line (39b-cor). The disease activity phenotype is obtained by averaging the features from the disease wells, here ALS (39b). For each well, the features were normalized to the control phenotype features using Cytominer version 0.1.0 (Singh *et al.*, 2020), which normally distributes the data around the control phenotype. Pairwise comparison of well activity phenotypes is evaluated by measuring the Euclidean distance between the wells in the multidimensional feature space using the SciPy spatial module. The resulting matrix is plotted as a hierarchically clustered heatmap using *seaborn* version 0.9.0 (Waskom, 2021). Compounds that result in similar activity phenotypes are observed as clusters in the heatmap. From the pairwise distance matrix, wells with activity phenotypes most similar to the control phenotype are extracted and ranked.

Activity signatures

For each well, each feature is normalized to the [0;1] range based on all observed values. The normalized feature vector representing each well is converted into a matrix of size 13×13 using the Python package *numpy*. The first 153 entries in the matrix correspond to the normalized features; the remaining entries are set to 0. The resulting matrix is used as the well activity signature and visualized in the form of a color-coded grid generated using the Python package *matplotlib* (Hunter, 2007). Grid squares are colored based on a grayscale gradient where 0 is black and 1 is white. The position of each feature in the grid is preassigned to allow direct comparison of signatures.

Extraction of morphological features

A total of 46 features characterizing individual cell shape were extracted using the CellProfiler module *MeasureObjectSizeShape*. The Euler number and Zernike polynomial coefficients were not included for further analysis. The 15 morphological features depicted in Supplemental Figure 4 were standardized and averaged over two runs for each well. Compounds were grouped by mechanism of action, providing a distribution of values for each morphological feature and each group. The probability density function of each variable was estimated using kernel density estimation and plotted so that each distribution density is normalized independently using the *seaborn* package. The description of each measurement collected is available in the CellProfiler manual at http://cellprofiler-manual. s3.amazonaws.com/CellProfiler-3.0.0/modules/measurement. html#measureobjectsizeshape

Incorporating additional parameters into the analysis pipeline

The computational pipeline presented in this article supports the addition of new features beyond the provided electrophysiological properties. Image-based features could be extracted for each cellular footprint by modifying the provided CellProfiler cell segmentation pipeline. By including modules that characterize the morphology of cells, researchers interested in morphological aberrations could incorporate measurements such as cell size and shape-related metrics. The script responsible for extracting cell and well features is located in the analysis package of the CELLXPEDITE codebase. A mapping from features to values is defined for each cell, providing a complete dictionary of features. By adding key-value pairs to the dictionary, new features could be combined with activity features to define broader phenotypes.

CELLXPEDITE platform

We have released our implementation of the cloud-based processing and analysis platform as an open-source software (https://github.com/brunoboivin/cellxpedite). It is currently designed for deployment on Amazon Web Services, but it is also compatible with local clusters and other cloud computing service providers with minimal changes to the deployment scripts.

REFERENCES

Albanese D, Visintainer R, Merler S, Riccadonna S, Jurman G, Furlanello C (2012). mlpy: machine learning python. arXiv:1202.6548. http://mlpy.sourceforge.net/

Catterall WA (2018). Dravet syndrome: a sodium channel interneuronopathy. Curr Opin Physiol 2, 42–50.

Chandrasekaran SN, Ceulemans H, Boyd JD, Carpenter AE (2021). Image-based profiling for drug discovery: due for a machine-learning upgrade? Nat Rev Drug Discov 20, 145–159.

Chen TW, Wardill TJ, Sun Y, Pulver SR, Renninger SL, Baohan A, Schreiter ER, Kerr RA, Orger MB, Jayaraman V, et al. (2013). Ultrasensitive fluorescent proteins for imaging neuronal activity. Nature 499, 295–300.

Di Giorgio FP, Boulting GL, Bobrowicz S, Eggan KC (2008). Human embryonic stem cell-derived motor neurons are sensitive to the toxic effect of glial cells carrying an ALS-causing mutation. Cell Stem Cell 3, 637–648.

Du P, Kibbe WA, Lin SM (2006). Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. Bioinformatics 22, 2059–2065.

Dukkipati SS, Garrett TL, Elbasiouny SM (2018). The vulnerability of spinal motoneurons and soma size plasticity in a mouse model of amyotrophic lateral sclerosis. J Physiol 596, 1723–1745.

Giovannucci A, Friedrich J, Gunn P, Kalfon J, Brown BL, Koay SA, Taxidis J, Najafi F, Gauthier JL, Zhou P, et al. (2019). CaImAn an open source tool for scalable calcium imaging data analysis. eLife 8, e38173.

Huang X, Roet KCD, Zhang L, Brault A, Berg AP, Jefferson AB, Klug-McLeod J, Leach KL, Vincent F, Yang H, et al. (2021). Human amyotrophic lateral sclerosis excitability phenotype screen: target discovery and validation. Cell Rep 35, 109224.

Hunter J (2007). Matplotlib: a 2D graphics environment. Comput Sci Eng 9, 90–95.

Iwai Y, Shibuya K, Misawa S, Sekiguchi Y, Watanabe K, Amino H, Kuwabara S (2016). Axonal dysfunction precedes motor neuronal death in amyotrophic lateral sclerosis. PLoS One 11, e0158596.

Jones E, Oliphant T, Peterson P (2001). SciPy: open source scientific tools for Python. http://www.scipy.org/.

Kiskinis E, Sandoe J, Williams LA, Boulting GL, Moccia R, Wainger BJ, Han S, Peng T, Thams S, Mikkilineni S, et al. (2014). Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. Cell Stem Cell 14, 781–795.

Kuo JJ, Siddique T, Fu R, Heckman CJ (2005). Increased persistent Na(+) current and its effect on excitability in motoneurones cultured from mutant SOD1 mice. J Physiol 563, 843–854.

Mandrioli J, D'Amico R, Zucchi E, Gessani A, Fini N, Fasano A, Caponnetto C, Chio A, Dalla Bella E, Lunetta C, et al. (2018). Rapamycin treatment for amyotrophic lateral sclerosis: protocol for a phase II randomized, double-blind, placebo-controlled, multicenter, clinical trial (RAP-ALS trial). Medicine (Baltimore) 97, e11119.

McQuin C, Goodman A, Chernyshev V, Kamentsky L, Cimini BA, Karhohs KW, Doan M, Ding L, Rafelski SM, Thirstrup D, et al. (2018). CellProfiler 3.0: next-generation image processing for biology. PLoS Biol 16, e2005970.

Moein M, Grzyb K, Goncalves Martins T, Komoto S, Peri F, Crawford AD, Fouquier d'Herouel A, Skupin A (2018). CaSiAn: a calcium signaling analyzer tool. Bioinformatics 34, 3052–3054.

Mukamel EA, Nimmerjahn A, Schnitzer MJ (2009). Automated analysis of cellular signals from large-scale calcium imaging data. Neuron 63, 747–760.

Müllner D (2011). Modern hierarchical, agglomerative clustering algorithms. arXiv:1109.2378.

Nakamizo T, Kawamata J, Yoshida K, Kawai Y, Kanki R, Sawada H, Kihara T, Yamashita H, Shibasaki H, Akaike A, Shimohama S (2003). Phosphodiesterase inhibitors are neuroprotective to cultured spinal motor neurons. J Neurosci Res 71, 485–495.

Narahashi T, Moore JW, Poston RN (1967). Tetrodotoxin derivatives: chemical structure and blockage of nerve membrane conductance. Science 156, 976–979.

Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, Reardon T, Mu Y, Lacefield C, Yang W, *et al.* (2016). Simultaneous denoising, deconvolution, and demixing of calcium imaging data. Neuron 89, 285–299.

Python Software Foundation (2018). Python Language Reference Version 3.7. http://www.python.org.

Renton AE, Chio A, Traynor BJ (2014). State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci 17, 17–23.

Singh S, Goodman A, Becker T, McQuin C, Rohban MH, Way GP, Cimini BA, Carpenter AE, *et al.* (2020). Cytominer: Methods for Image-Based Cell Profiling [Computer software]. https://cran.r-project.org/package=cytominer

Van Der Walt S, Colbert SC, Varoquaux G (2011). The NumPy array: a structure for efficient numerical computation. Comput Sci Eng 13, 22–30.

van Zundert B, Peuscher MH, Hynynen M, Chen A, Neve RL, Brown RH Jr, Constantine-Paton M, Bellingham MC (2008). Neonatal neuronal circuitry shows hyperexcitable disturbance in a mouse model of the adult-onset neurodegenerative disease amyotrophic lateral sclerosis. J Neurosci 28, 10864–10874.

Wainger BJ, Kiskinis E, Mellin C, Wiskow O, Han SS, Sandoe J, Perez NP, Williams LA, Lee S, Boulting G, *et al.* (2014). Intrinsic membrane hyperexcitability of amyotrophic lateral sclerosis patient-derived motor neurons. Cell Rep 7, 1–11.

Waskom ML (2021). Seaborn: statistical data visualization. J Open Source Softw 6, 3021.