NEUROeSTIMator: Using Deep Learning to Quantify Neuronal Activation from Single-Cell and Spatial Transcriptomic Data

Ethan Bahl[1,2], Snehajyoti Chatterjee[3,4], Utsav Mukherjee[3,4,8], Muhammad Elsadany[1,2], Yann Vanrobaeys[2,3], Li-Chun Lin[3,4], Miriam McDonough[4,9], Jon Resch[4], K Peter Giese [5], Ted Abel[3,4], and Jacob J. Michaelson[1,3,6,7*]

[1]Department of Psychiatry, University of Iowa.
[2]Interdisciplinary Graduate Program in Genetics, University of Iowa.
[3]Iowa Neuroscience Institute, University of Iowa.
[4]Department of Neuroscience & Pharmacology, University of Iowa.
[5]Department of Basic and Clinical Neuroscience, King's College London, London, UK.
[6]Department of Biomedical Engineering, University of Iowa.
[7]Department of Communication Sciences & Disorders, University of Iowa.
[8]Interdisciplinary Graduate Program in Neuroscience, University of Iowa.
[9]Interdisciplinary Graduate Program in Molecular Medicine, University of Iowa.
* correspondence to jacob-michaelson@uiowa.edu

ABSTRACT

Neuronal activity-dependent transcription directs molecular processes that regulate synaptic plasticity, brain circuit development, behavioral adaptation, and long-term memory. Single cell RNA-sequencing technologies (scRNAseq) are rapidly developing and allow for the interrogation of activity-dependent transcription at cellular resolution. Here, we present NEUROeSTIMator, a deep learning model that integrates transcriptomic signals to estimate neuronal activation in a way that captures electrophysiological features and is robust against differences in species, cell type, and brain region. We demonstrate this method's ability to accurately detect neuronal activity in previously published studies of single cell activity-induced gene expression. Further, we applied our model in a spatial transcriptomic study to identify unique patterns of learning-induced activity across different brain regions. Altogether, our findings establish NEUROeSTIMator as a powerful and broadly applicable tool for measuring neuronal activation, whether as a critical covariate or a primary readout of interest.

INTRODUCTION

Activity-dependent expression of transcription factors is critical in eliciting an intact neurophysiological response, whereas dysregulation of transcriptional activation has been extensively implicated in several disorders of the nervous system[1-6]. Currently, a major focus in neuroscience research is aimed at understanding the cell-type and neuroanatomical specificity of activity-responsive expression of transcription factors and their modulation of downstream targets. Single cell RNA-sequencing (scRNAseq) technologies are rapidly developing and allow for the interrogation of activity-dependent transcription at the resolution of individual neurons[7,8]. However, properties of scRNAseq data pose challenges to characterizing and quantifying gene expression signatures of neuronal activity. Sparsity and variability in gene expression measurements are common caveats of scRNAseq, due to a combination of biological and technical factors such as cellular RNA content, individual gene abundance, and sequencing depth[9-12]. Consequently, genes that are truly expressed can go undetected in single cell data, thus diminishing the confidence in, and utility of, individual marker genes for transcriptionally defining neuronal activity state. Importantly, these challenges pose a significant barrier to analyzing data where neuron activity states are unknown, such as in post-mortem human tissue, as well as for controlled experimental data where manipulations are rarely expected to elicit a uniform response across or within cell types. Further, activity marker genes display basal expression that is detectable even in the absence of stimulation or activity[1]. Because individual genes are unreliable markers in single cell data, robustly estimating the degree of individual neuronal activity requires integrating information from multiple activity marker genes.

Several methods exist for aggregating transcriptomic data across multiple genes, and they are commonly used in single cell pre-processing steps to categorize cell types and visualize data. Dimensionality reduction methods such as principal component analysis (PCA) and non-negative matrix factorization (NMF) combine information from multiple genes into summarized components. However, such methods are unsupervised, and do not guarantee that the resulting components will be relevant to neuronal activity. Furthermore, these methods do not capture information from non-linear interactions that result from complex layers of biological regulation. Non-linear dimensionality reduction methods like t-stochastic neighbor embedding (tSNE)[13], and uniform manifold projection (UMAP)[14], similarly attempt to explain variability in single cell data by summarizing gene expression patterns into components. In many scRNAseq datasets derived from the brain, these components represent neuron-glia and glutamatergic-GABAergic axes. Despite non-linear capabilities, they are typically applied to linear components from PCA and still do not guarantee identification of a component that indexes neuronal activity state.

Neural networks have been developed for many applications in single cell data such as dimensionality reduction and imputation[15]. DCA[16] is a gene expression-oriented autoencoder for learning a reduced dimensional space, also known as an information bottleneck, which must then reconstruct the input data. Leveraging the inherently destructive nature of dimensionality reduction and statistical noise error models

simultaneously strips noise from input data while retaining informative features in the bottleneck. This approach has the attractive qualities of learning non-linear relationships of input genes and addressing noise from sparsity of gene detection in a supervised manner. Furthermore, neural networks like DCA allow flexibility in the choice of genes targeted for reconstruction and the information capacity of the bottleneck.

Here, we developed a neural network that produces an estimate of neuronal activity based on expression of thousands of genes. The network distills transcriptional signatures into a 1-dimensional information bottleneck before reconstructing expression profiles of 20 well-established, robust markers of neuronal activity. The bottleneck value is bound between 0 and 1 and represents a cell-type-invariant summary of activity-responsive gene expression magnitude. Using Patch-seq data, we demonstrate that this 1-dimensional bottleneck, hereafter referred to as the 'activity score', links transcriptomic signatures of cell activity directly to electrophysiological features of increased excitability. We then demonstrate the broad applicability of the activity score in predicting neuronal activation caused by seizure, cocaine administration, and sensory experience. We use the activity score to accurately classify cells by experimental manipulation, highlight genes involved in these predictions, and identify neuron type-specific expression signatures of activity. Furthermore, we apply our approach to new spatial transcriptomic data from brain slices following learning. To allow the neuroscience community to take advantage of these efforts, we have developed NEUROeSTIMator, an R package with an accompanying tutorial that demonstrates applications of our model to single cell data.

RESULTS

Predicting Activity-Dependent Marker Gene Expression

We used publicly available single cell and nuclei datasets (**Table 1**) generated by the Allen Institute of Brain Science[17] (AIBS), consisting of mouse and human samples, to develop a neural network model trained to predict expression of 20 activity-dependent genes. Activity-dependent genes were identified by intersecting differential expression results of three studies of experimental manipulation of neuron activity. All three studies examined different brain regions, used different methods of neuronal stimulation, and were published from independent groups. Single cell data was sampled with weighting to increase representation of less common neuronal cell types, species, sex, and technical characteristics. Five hundred thousand neurons were selected and partitioned into cell type-balanced training and testing sets, and the training set was further split into 5 folds for cross validation. We trained the neural network to predict expression of the 20 activity-dependent target genes through a 1-dimensional hidden bottleneck layer with sigmoidal activation (see methods for a detailed description of model architecture and training). To evaluate model performance, we applied it to a diverse test set of 56,578 neurons held out from the training process. We found the model performance on the test set was comparable to performance observed through cross validation.

For further analyses, model output at the bottleneck activation layer was extracted to index activity level for each neuron passed through the model. Hereafter, we refer to

this output as the predicted activity or activity score. We first examined the distribution of predicted activity across cell subclasses in the held-out test set (**Fig. 1a, left**). Most cell subclasses exhibited of a distribution concentrated near zero with a tail skewed towards one. A few cell subclasses exhibited higher average predicted activity, including L4/5 IT CTX, L4 RSP-ACA, L5 PPP, and SUB-ProS neurons.

Genes Informing Model Predictions

To identify genes whose expression levels influence model predictions, we calculated integrated gradients[18] (**Supp Fig 1**) for all input genes with respect to predicted activity using data from the held out test set. Integrated gradients attribute model predictions to input features for each cell. We first examined the average impact of target genes on predicted activity for each cell subclass label (**Fig. 1a, right**). The most influential genes were *Egr1* and *Nr4a1*. We further explored gene importance for all non-target input genes and found varying degrees of influence throughout the transcriptome (**Fig. 1b**). Among the most influential were the known activity response genes *Homer1, Egr4,* and *Bdnf*. We observed that many influential genes had high mean expression levels, although lower abundance genes also exerted influence. For example, the gene *Cyr61*, known to regulate dendritic arborization[19], was one of the most influential non-target gene with mean $\log_{10}$ expression less than -1. We also noted relatively few genes whose expression was indicative of reduced predicted activity. To evaluate whether influential genes were enriched for specific annotated biological mechanisms, we performed gene set enrichment analysis on gene ontology annotation sets (**Fig. 1c**). Among the most highly enriched gene sets were regulation of response to external stimulus ($p_{adjusted}$ = 0.041, NES = 1.594), regulation of synaptic plasticity ($p_{adjusted}$ = 0.041, NES = 1.670), circadian rhythm ($p_{adjusted}$ = 0.033, NES = 1.781), regulation of MAPK activity ($p_{adjusted}$ = 0.035, NES = 1.656), telencephalon development ($p_{adjusted}$ = 0.033, NES = 1.806), and cognition ($p_{adjusted}$ = 0.033, NES = 1.729). Many influential genes were members of multiple significant gene sets including *Homer1, Bdnf,* and *Ntrk2*.

Electrophysiological Features of Individual Neurons Predicted by Transcription via NEUROeSTIMator

To demonstrate that NEUROeSTIMator's predictions correspond to actual electrophysiological differences at the individual neuron level, we first integrated RNA assays of a Patch-seq dataset[20] with a SMART-seq dataset[17] of mouse GABAergic neurons, both generated by the Allen Institute for Brain Science (AIBS) (**Supp. Fig. 2a**). Using the standard Seurat workflow[21], we identified seven clusters corresponding to AIBS-defined neuron types with at least 100 cells in each dataset, including Sst, Sst Chodl, Pvalb, Vip, Lamp5, Sncg, and Meis2 neuron clusters. We applied NEUROeSTIMator to both datasets and observed markedly elevated activity score predictions in the Patch-seq dataset relative to the SMART-seq dataset for all seven neuron types (**Fig. 2a**). We obtained empirical p-value estimates of the Kolmogorov-Smirnov (KS) test by permuting expression values in each cell and recalculating activity score for one thousand permutations. All seven neuron types showed significant differences in activity score distributions between datasets (**Fig. 2b**), with Lamp5 neurons

showing the greatest difference ($p < 0.001$, $D = 0.74$, KS test) and Vip neurons showing the smallest difference ($p = 0.044$, $D = 0.39$, KS test). We also noted that several immediate early genes, which serve as predictive targets for NEUROeSTIMator, were among the top genes with higher mean expression in the Patch-seq dataset compared to the SMART-seq dataset (**Supp. Fig. 2b**). Together, these results suggest that the Patch-seq protocol induces significantly greater transcriptional upregulation of activity-dependent genes as compared to the standard SMART-seq protocol, and our model is sensitive to these differences. Notably, this effect is consistent for all neuronal subtypes tested.

We established that our model detects transcriptomic signals that separate the electrically stimulated Patch-seq dataset from the unstimulated SMART-seq dataset. However, it is not clear whether elevated activity was due to electrical stimulation and neuron depolarization, or general features of the Patch-seq protocol. Patch-seq neurons showed high variance in predicted activity, spanning nearly the entire range of possible values from 0.001 to 1. Because all neurons underwent the same protocol, we hypothesized that variability in activity score predictions would be linked to individual neuron differences in electrophysiological responses to stimulation.

To determine if NEUROeSTIMator predictions of activity could be explained by electrophysiology, we fit a random forest that predicted NEUROeSTIMator output using only electrophysiological features derived from Patch-seq recordings. We extracted 63 electrophysiological features, including features measured at baseline and in response to a range of stimuli. As substantial differences in electrophysiology exist between neuron types, we residualized activity score and derived features for neuron type in order to force the model to learn cell type-agnostic combinations of features that explain variability in activity score. We focused on the four most abundant neuron types: Lamp5, Pvalb, Sst, and Vip neurons. Out-of-bag mean squared error was 0.85, and the random forest predictions were significantly correlated with NEUROeSTIMator output ($R = 0.11$, $p = 7.86 \times 10^{-10}$) (**Fig. 2c**). To interpret the random forest, we then grew a single representative decision tree using the same input features as the random forest and used random forest predictions as an output. The feature used as the first split in this representative decision tree was the difference in membrane potential between resting voltage and action potential threshold voltage for the ramp stimulus. This feature was highly correlated with resting membrane potential as well as the current required to initiate spiking. Hereafter we refer to this feature as the voltage barrier to depolarization. We examined differences in this feature between two groups of cells – high and low activity score neurons that were also accurately predicted as such using the electrophysiological data in the random forest model. We tested for differences in voltage barrier between the high and low activity groups using the KS test and found significant differences for all four major neuron types examined (**Fig. 2d**) (Lamp5: $D = 0.37$, $p = 3.85 \times 10^{-2}$), (Pvalb: $D = 0.67$, $p = 1.21 \times 10^{-9}$), (Sst: $D = 0.6$, $p = 1.22 \times 10^{-15}$), (Vip: $D = 0.7$, $p = 4.68 \times 10^{-9}$). An example of voltage barrier differences in a low and high activity score neuron is depicted in **Fig. 2e**. Taken together, these results demonstrate that NEUROeSTIMator captures transcriptomic signatures of activity that are associated with electrophysiological properties measured in response to electrical stimulation in individual neurons.

Having clearly demonstrated the core premise of NEUROeSTIMator, we further vetted its utility and generalizability in other applications, using five diverse datasets (**Table 1**).

Detecting Pharmacological Activation of Neurons

We applied our model to three datasets containing neurons from rodents and human cell lines treated with powerful stimulating, pharmacological agents. From the first dataset, we computed activity score for neurons of mice treated with either saline or pentylenetetrazol (PTZ), a depolarizing agent used to model status epilepticus and induce seizures[7]. As expected, we observed significant differences in distributions of predicted activity (KS test) for several neuronal subclasses, including excitatory cortical layer 2/3 (L2/3) neurons, Lamp5 interneurons, and Sst interneurons (L2/3 IT CTX: $D = 0.43$, $p = p < 2.22\text{x}10^{-16}$), (L2/3 IT ENT: $D = 0.45$, $p < 2.22\text{x}10^{-16}$), (Lamp5: $D = 0.53$, $p = 2.30\text{x}10^{-9}$), (Sst: $D = 0.39$, $p = 1.61\text{x}10^{-9}$) (**Fig. 3a**).

Next, we compared activity score between neurons from the nucleus accumbens (NAc) of rats treated with either saline or cocaine, a stimulant acting on dopaminergic neurotransmission[22]. We found neuron subtype-specific increases in activity score (**Fig. 3b**). Drd1/3-type medium spiny neurons (MSN) and Sst interneurons were the most profoundly affected neuron subtypes (Drd1-MSN: $D = 0.21$, $p < 2.22\text{x}10^{-16}$), (Drd3-MSN: $D = 0.20$, $p = 2.70\text{x}10^{-3}$), (Sst: $D = 0.21$, $p = 3.04\text{x}10^{-2}$). Because our model was not trained using any rat or dopaminergic neurons, these findings provide further endorsement for robust estimation of neuronal activity induced by potent pharmacological agents of stimulation.

To further evaluate whether our model could be successfully applied to human data, we examined a dataset of human induced pluripotent stem cell-derived neurons[23] (**Fig. 3c**). Neuron cultures were either unstimulated or treated with KCl depolarization buffer for 1, 2, or 4 hours. Our model predicted low neuronal activity for the unstimulated group, which was consistent among cell types and biological replicates. Cells treated with KCl for 1 hour demonstrated substantial and significant increases in predicted activity. Although significant shifts in predicted activity were observed for all cell types ($p < 0.05$), the activity score of post-mitotic neurons displayed a stronger response to KCl compared to mitotic progenitors. The most responsive neuron types were the mature striatal neurons and the pallial neurons (striatal: $D = 0.65$, $p < 2.22\text{x}10^{-16}$), (pallial: $D = 0.70$, $p < 2.22\text{x}10^{-16}$. Notably, all cell clusters followed a similar temporal pattern of predicted activity modestly declining at 2 hours relative to peak activity at 1 hour, with a further decline at 4 hours. Despite these activity predictions reducing after 1 hour, none of the neuron clusters completely returned to basal activity at 4 hours, the final time point in the experiment.

Together, these analyses suggest our model can robustly assign higher estimates of activity to cells subjected to chemical exposures that are expected to elicit strong and generally ubiquitous transcriptional responses to stimulation.

Activity Score as a Generalizable Classifier of Neuronal Stimulation

Next, we asked if our model could detect neuronal activation by more subtle forms of stimuli such as sensory experience. We applied our model to a dataset containing visual cortex neurons from mice exposed to light for 0, 1, or 4 hours (**Fig. 4a**). We observed a significant increase in predicted activity for neurons from mice exposed to light, relative to controls. To elucidate temporal patterns of activity we tested differences in activity score between pairs of each time point. Activity score was significantly increased at 1 hour for many neuron types. At 4 hours of light exposure, predicted activity began to show diverging trends which were foreshadowed by predicted activity at 1 hour. Cell types weakly activated at 1 hour showed decline in activity towards the baseline at 4 hours, while cell types strongly responsive to light exposure 1 hour declined less. Although trending towards a return to baseline, activity scores of neurons at 4 hours were not significantly different from neurons at 1 hour.

As we observed similar trends in temporal activity predictions between the unstimulated (0h) and 1h group, we investigated the degree to which the activity score derived from our model could be used as a classifier of experimental group. The degree to which activity score is predictive of experimental group in a particular cell type is expected to represent the robustness of the response in that cell type. Using the visual cortex dataset (VIS) mentioned above, we constructed receiver-operator curve (ROC) plots for neuronal and non-neuronal cell types (**Fig. 4b**). For both excitatory neuron and interneuron subtypes, the activity score demonstrated varying degrees of predictive power. For example, the activity score alone was able to almost perfectly separate stimulation groups when considering excitatory cortical layer cell types, though it could only separate hippocampal neurons into stimulation groups with an accuracy slightly better than random chance, though we suspect this reflects a lack of hippocampus responsiveness to simple light exposure. Despite the model being trained on purely neuronal cell type populations, the activity score was able to separate stimulation groups for astrocytes and endothelial cells just as accurately as it could for neurons.

Utility in Data Modalities Beyond scRNA-seq

Next, we asked whether our model could identify spatial signatures of learning in brain slices of mice following spatial object recognition (SOR) training - a widely used behavioral paradigm to investigate spatial memory mechanisms[24]. Using spatial transcriptomic data from brain slices of SOR-trained and homecage control (HC) mice, we applied our model to predict activity for each spot (**Supp Figure 3a**) and clustered all spots into anatomical regions. The 23 resulting clusters were annotated with brain region naming conventions from the Allen Mouse Brain Atlas (**Fig. 5a**). In baseline homecage samples, we noted weak activation signatures, primarily covering cortical layers of the isocortex and subregions of the hippocampus (**Fig. 5b, left**). Following SOR training, we noted region-specific increases in predicted activity (**Fig. 5b, right**). To identify a spatial activation signature of SOR for the entire brain slice, we tested for differences in predicted activity for each brain region cluster (**Fig. 5c**). We observed significant increases in

predicted activity for several cortical and subcortical regions. Multiple layers of the isocortex and the retrosplenial area showed increases in activity following SOR (**Fig. 5d**), with the greatest increase observed in layers 2/3. We also found a significant activation of the caudoputamen area of the dorsal striatum. The amygdala, hippocampus, and olfactory/piriform areas also showed significant increases in activity, of comparable magnitude. Subregions of the hippocampus were variably activated following SOR, with the strongest induction observed in the CA1 region. Regions predicted to be least activated by SOR include the thalamus, hypothalamus, dentate gyrus of the hippocampus, the lateral ventricle, and fiber tracts. To validate the finding that the CA1 region was preferentially activated by spatial learning over the dentate gyrus, we performed RNAscope[25] on coronal brain slices from SOR-trained mice (1 hour post-training) and a homecage control mouse (**Supp Fig 3b**). We selected three NEUROeSTIMator target genes, *Egr1*, *Egr3*, and *Nr4a1* as probes. For each brain slice, we quantified mean fluorescent intensity (MFI) of each probe in CA1 and dentate gyrus. In line with the spatial transcriptomic data (**Supp Fig. 2c**)., we found that CA1, but not dentate gyrus, showed a significant increase in average probe intensity following spatial learning (**Supplemental Fig. 2d**).

NEUROeSTIMator Outperforms a Scaled Additive Approach in Discerning Stimulation Groups

The NEUROeSTIMator approach to quantifying activity uses an encoding neural network with thousands of input features and nonlinear transformations. We asked if the complexities of our model are necessary to produce a metric that successfully discerns stimulated samples from unstimulated. We compared NEUROeSTIMator to a scaled additive model that sums scaled expression of NEUROeSTIMator's 20 target genes. We evaluated each model by scoring activity in stimulated and unstimulated samples and determined separability of these experimental groups using area under the curve (AUC). This approach was applied to datasets featured in Figures 3 and 4, as well as a novel single nuclei dataset from mouse dorsal hippocampus following SOR. We observed that NEUROeSTIMator substantially improved AUC compared to the scaled additive approach (**Supp Fig. 4a**). This observation was consistent across multiple datasets, stimulation paradigms, cell types, and brain regions. We examined AUC differences between NEUROeSTIMator and the scaled additive model across a range of time points and total counts (**Supp Fig. 4b**). At early time points, each model performs similarly. However, NEUROeSTIMator performance improvement drastically increased as a function of time after stimulation, when differences between control and stimulated are more subtle. Together, these results suggest that NEUROeSTIMator's use of broader transcriptomic signals underlies its increase in performance compared to a less complex approach.

DISCUSSION

We present NEUROeSTIMator, a generalizable tool for *in silico* estimation of neuronal activity from transcriptome-wide single-cell gene expression. The neuronal activity score is an easily interpretable value that quantifies the transcriptional response

359 to stimulation. NEUROeSTIMator can be used to rapidly identify and prioritize subsets of
360 neurons showing transcriptional evidence of a stimulus response. In tests of predictive
361 performance and generalizability, we demonstrate that the neuronal activity score
362 captures electrophysiological features and robustly detect signatures of activation from
363 multiple types of stimulation, neuron subtypes, species, and sequencing technologies,
364 including spatial transcriptomics. We have made this tool freely available to the research
365 community with a tutorial and R package.
366
367 Using permutation analyses, we found that predictive signal was distributed
368 broadly across the transcriptome and was enriched for genes regulating external stimulus
369 response, synaptic plasticity, MAP Kinase activity, and circadian rhythm. These predictive
370 genes also function in brain development and cognition. The genes most informative to
371 our model and relevant to these pathways were *Homer1*, *Bdnf,* and *Ntrk2*. These well-
372 known activity response genes did not meet our strict criteria for inclusion as target genes,
373 but their prominent influence on model predictions suggests our model utilizes information
374 from co-regulated genes and pathways to robustly estimate expression of target genes.
375 Notably, we found few genes whose up-regulation was associated with reduced activity
376 score, suggesting model predictions largely rely on positive indicators of activity. Although
377 there is evidence of activity-dependent downregulation of gene expression[26,27], most
378 genes differentially expressed by neuronal activity are transcriptionally upregulated,
379 which is supported by the observation that neuronal activity increases genome-wide
380 chromatin accessibility[1,28,29].
381
382 Prediction of IEG expression, in the absence of further data, is at best a limited
383 proxy of the underlying electrophysiological activity that is presumed to occur in the
384 assayed neurons and brain tissue. To determine whether NEUROeSTIMator predicts
385 actual features of neuronal electrical activity, we investigated a highly unique, single cell
386 Patch-seq data set, which contains both gene expression and electrophysiological
387 readouts from individual GABAergic neurons[20]. The results of this investigation directly
388 link electrophysiological activity of individual neurons with their activity score as generated
389 by NEUROeSTIMator from the corresponding gene expression data. Specifically, we
390 found that higher activity score among all types of GABAergic neurons ascertained
391 corresponded to diminished barriers to depolarization, such as higher resting membrane
392 potential, lower voltage change from rest to action potential threshold, and lower current
393 necessary to elicit depolarization. These findings are consistent with previous work that
394 showed that neurons with higher relative excitability immediately before fear conditioning
395 are more likely to be allocated to the fear memory trace, and more likely to express the
396 IEG *Arc* five minutes after memory recall[30]. Given the results of our investigation, in
397 applications of NEUROeSTIMator to neurons and brain tissue data lacking
398 electrophysiological data, the activity score can be reasonably assumed to tag
399 electrophysiological features related to increased neuronal excitability. Additionally, these
400 analyses demonstrate general utility of NEUROeSTIMator for non-UMI-based
401 sequencing platforms such as SMART-seq.
402
403 With evidence of a robust connection between our model predictions and
404 increased neuronal excitability, we evaluated its performance in identifying signatures of

activity through secondary analyses of several previously published data sets. In a single cell study of chemically induced seizure[8], activity score strongly predicted PTZ exposure in multiple neuron types, in particular L2/3 neurons, which have been previously shown to be especially sensitive to seizures in humans[31]. In a single cell data from rat striatal neurons[22], the activity score was significantly associated with exposure to cocaine, in particular in Drd1+ and Drd3+ neurons. This recapitulation of the study's conclusions is especially noteworthy given that our training set included neither rat nor striatal neurons, demonstrating the generalization of this tool. In human differentiated neurons[23], NEUROeSTIMator detected a sharp increase in activity after 1 hour exposure to KCl that declined in later time points, indicating a greater sensitivity to activity-coupled gene expression than a simple on/off model. These secondary analyses demonstrate our model's ability to identify potent pharmacologically induced neuronal activity, and to generalize across species and cell types.

Detection physiologically relevant activation is a more challenging task than the "hammer" of chemically-induced activity, and NEUROeSTIMator excels in this setting as well. Sensory experience is known to induce activity-dependent gene expression programs in cortical neurons[1]. Predicted activity is markedly higher in visual cortex neurons from dark-reared mice exposed to light, compared to controls. These exposure-dependent changes extended to some non-neuronal cells as well: the activity score can classify visual cortex astrocytes as originating from light-exposed experimental groups with accuracy comparable to neurons. Astrocytes in the visual cortex have been shown to exhibit calcium signaling activity in response to light[32], and this application of the activity score provides further supporting evidence of this light-induced activity. This finding was unexpected, as the model was not trained with any glial or other non-neuronal cell types but provides further evidence for its generalization. We also observed strong increases in predicted activity for endothelial cells in response to light, which is known to increase deoxyhemoglobin in the visual cortex[33,34]. We speculate that the increased predicted activity in endothelial cells may reflect fluid shear stress from blood flow, which has been shown to induce immediate early gene expression[35,36]. As the immediate early gene markers of neuronal activity are, in fact, markers of general stimulus-response in many cell types, even beyond the brain[37], we anticipate that our model may be capable of detecting activation signatures in entirely different cell types such as immune cells, for example. Further research is needed to catalog the breadth of cell types and stimulus responses that can be detected.

Our final demonstration of NEUROeSTIMator's flexibility involves new spatial transcriptomic data from brain sections of mice that were subjected to a spatial learning task, resulting in a spatial map of learning-induced activation. Emerging spatial transcriptomic technologies promise to identify differentially active brain regions following a stimulus such as a training for a learning task. Comparing mice trained in a spatial object recognition task to homecage controls, we found widespread increases in cortical neuron activity, particularly in layers 2/3 of the isocortex. We also observed increased predicted activity in the CA1 region of the dorsal hippocampus, the caudoputamen region of the dorsal striatum, the retrosplenial area, piriform areas, and other regions linked to long-term memory networks[38-48]. The hippocampus is one of the most studied brain regions in

the context of activity-dependent transcription as it directly relates to long-term memory encoding and the concept of the engram ensemble, which is thought to be the physical, synaptic trace of memory. Neuroscientists have long studied engrams, but standard approaches to labeling active neurons typically use a single activity marker to drive expression of GFP or channel rhodopsin for experimental access to putative engram cells[49-53]. Although many immediate early genes are thought to robustly mark activity in neurons, there is also evidence that certain genes do, in fact, exhibit region or stimulus specificity[54-56]. Multigenic approaches to quantifying activity, like NEUROeSTIMator, should be more reliable and generalizable. In our application of NEUROeSTIMator to spatial transcriptomics data after learning, we found that spatial object recognition training induces higher activity in the CA1 subregion of the hippocampus than in the dentate gyrus, and we validated this finding with RNAscope. RNAscope validation also showed gene-specific variability in magnitude of transcriptional induction, which further suggests that individual genes may be less reliable activity markers than multigenic markers. Finally, we show a brain-wide distributed transcriptomic signal of learning, with increased activity in multiple regions involved in memory, navigation, and sensory processing. We speculate this brain-wide map of transcriptional activation may represent a macro-level memory engram. Extending the application of our model to spatial transcriptomics, an entirely different data modality than the training data, we show that the activity score predictions are not confined to use in single cell RNA sequencing datasets. We expect many other distinct brain-wide spatial signatures of transcriptional activation could be identified in relation to other cognitive processes.

Our model uses transcriptional signals to quantify cellular activity response to stimuli. This approach relies on the commonly employed assumption that gene expression is a reliable indicator of activity. We tested the validity of this assumption using a Patch-seq dataset by examining relationships between neuron electrophysiology and transcriptome-based predictions of activity. In the GABAergic neurons we investigated, we found compelling evidence for a link between neuron excitability and transcriptional response to stimulation. However, the lack of broader cell type representation in the Patch-seq dataset is a key limitation of these analyses and increasing cell type representation in Patch-seq datasets will benefit neuroscience research broadly. Another limitation is that our model does not by itself differentiate between types of stimulation (*e.g.*, electrical, chemical, mechanical). Furthermore, the model is sensitive to cellular activation beyond just neurons: we observed stimulus-responsive increases in predicted activity of several non-neuronal cell types, such as astrocytes. However, this is a relatively minor limitation since determination of cell identity is a tractable problem with current single cell techniques, and the meaning of activation can be interpreted within the context of the cell type under investigation. Despite these limitations, we demonstrate utility of NEUROeSTIMator in a wide variety of neuroscience research domains.

NEUROeSTIMator provides the first robust and generalizable means to estimate neuronal activation from transcriptomic data, opening the door to widespread use in molecular neuroscience research. In research involving gene expression, and especially in novel approaches like single cell or spatial transcriptomics, cellular activity state is a variable as fundamental as age, sex, or treatment group. Depending on the goal of an

analysis, it may be a critical covariate, a key grouping variable, or an explanatory variable of central interest that may now be easily estimated using the tools we present here.

METHODS

Dataset for Model Training and Evaluation

To train the model, we utilized publicly available datasets provided by the Allen Institute for Brain Science, including a single-cell RNA-sequencing (scRNAseq) dataset of over a million cells isolated from mouse cortical and hippocampal tissue[17], and a single-nuclei RNA-sequencing (snRNAseq) dataset of 76,000 nuclei isolated from human cortical tissue. Hereafter, these datasets will be referred to as the Allen Mouse and Allen Human datasets. Both datasets used the 10X Genomics Chromium system for droplet capture. The Allen Mouse dataset was prepared using the Chromium Next GEM Single Cell 3' v3 reagent kit, while the Allen Human dataset used v2.

Datasets for Model Application – Publicly Available

We downloaded multiple datasets from Gene Expression Omnibus (GEO) to demonstrate the utility of our model. The following GEO accessions were included in analyses: GSE102827[57], GSE106678[8], GSE136656[23], and GSE137763[22].

Gene Identifier Mapping

We used the R package biomaRt to map gene identifiers from various annotations used in public datasets, and between species, to a common set of reliably mapped genes[58,59]. Ensembl gene identifiers (Ensembl IDs) were used as the primary identifier for mapping genes, and gene symbols were used as secondary identifiers in cases of ambiguous mapping. The Ensembl release 93 archive (July 2018 release) was used for cross-species gene mapping[60]. Genes with one-to-one orthology between mouse and human, as well as mouse and rat, were selected to facilitate cross species utility. All datasets lacking Ensembl ID annotation contained gene symbols, which were then queried against multiple Ensembl archives to determine which archive maximized identifier mapping rate. For instances when a gene symbol mapped to multiple Ensembl IDs, identifiers present in the cross-species mapping table were preferentially selected. We provide a helper function for mapping gene identifiers to the feature set used by our model, and we further demonstrate usage in the associated tutorial.

Choice of Neural Network Target Genes

To identify robust markers of neuronal activity for use as targets of the neural network, we intersected lists of stimulus-responsive genes from three published RNA-sequencing experiments. Each publication was from a different group of authors, focused on different brain regions, and used different forms of neuronal stimulation (see Table 1). One publication categorized stimulus-responsive genes into three groups – rapid primary response genes (rPRGs), delayed primary response genes (dPRGs), and secondary response genes (SRGs)[61]. As SRGs are thought to demonstrate higher celltype-specificity relative to PRGs[1], only rPRGs and dPRGs were considered from this publication. In another publication, approximately 600 genes upregulated in response to kainic acid treatment in the hippocampus were considered[29].

In a third publication, two sets of KCl-responsive genes were available, one from a brain region enriched in glutamatergic neurons and one enriched for GABAergic neurons[62]. For this study, we sought to intersect the results of both the glutamatergic and GABAergic analyses into one list of genes. Published p-value distributions suggested differences in statistical power between these two analyses, and only statistically significant results were published. To expand the list of genes overlapping between these analyses, we reanalyzed the data using GEO2R to obtain two sets of transcriptome-wide statistics. Using significance rankings from the GEO2R reanalysis, we jointly determined p-value thresholds for each analysis based on rank-rank hypergeometric overlap (RRHO, see **Supplemental Figure 5**) and identified genes with p-values below these thresholds in both sets with concordant direction of effect[63]. Because this approach used unconventional p-value thresholding, we additionally required intersecting genes to have an estimated fold change greater than or equal to 0.5 in both analyses.

Finally, these three lists were intersected, and a set of 41 stimulus-responsive genes were selected as output targets on the basis of being differentially expressed in all three lists.

Sample Filtering, Downsampling, and Partitioning

Cells with less than 3,500 total counts or greater than 30,000 total counts were removed. Non-neuronal cells were removed and imbalances among species, sex, neuron type, quality control metrics and naively-estimated activity were alleviated by weighted random sampling. The R package *groupdata2* was used to create five training folds (88.7%, 886,844 samples) and one test split (11.3%, 113,156 samples) in a way that retains training set diversity while maximizing representation of neuron subclasses in the test split.

Model Input/Output Feature Selection

Input and target features were selected based on mean expression and detection rate in the training data. Input features were required to have a detection rate greater than zero and log mean expression greater than -2 in both the Allen mouse and human samples used for training. The rationale behind removing weakly expressed genes was that the Allen datasets were sequenced deeper than typical datasets, and genes with low detection at high sequencing depth would likely be unreliably detectable at lower depths. The remaining 10,017 genes were used as input features. Of the 32 remaining candidate output target genes, we selected a final set of 20 targets based on consistency of coexpression patterns across datasets and broad cell classes. Gene-gene Pearson correlations were calculated for four cell sets (mouse glutamatergic, mouse GABAergic, human glutamatergic, human GABAergic). In each set, we ranked each gene based on the average correlation to all other candidate genes. These coexpression ranks were averaged across the four cell sets and the top 20 genes were selected as final the final set of output target genes. The targets include *Arc*, *Btg2*, *Crem*, *Dusp1*, *Egr1*, *Egr2*, *Egr3*, *Fbxo33*, *Fos*, *Fosb*, *Fosl2*, *Grasp*, *Junb*, *Npas4*, *Nr4a1*, *Nr4a2*, *Nr4a3*, *Per1*, *Rgs2*, and *Tiparp*.

Dataset Augmentation

The Allen Cell Types Database has higher average total counts per cell compared to many publicly available datasets. To increase representation of lower quality cells, we created augmented samples by synthetically downsampling raw counts using the R package scater[64]. For each combination of species and neuron subclass, an equal number of cells were randomly assigned a value of either 10^3, $10^{3.25}$, or $10^{3.5}$ total counts to be downsampled to. All genes were considered for downsampling.

Feature Normalization and Preprocessing

Log-normalization, as implemented in Seurat, was used to normalize input gene expression. Total counts for each cell were calculated by summing only the 10,017 features used by the model. Normalized expression levels for each gene were centered and scaled based on mean and standard deviation estimated from the training data. For cross validation, mean and standard deviation were estimated without the held-out fold.

Model Architecture

The architecture of the model was adapted from DCA[16]. Briefly, input gene expression is supplied to an encoder; a series of three fully connected dense layers with ELU activations and batch normalization. ELU activations were chosen over RELU to increase learning speed and performance. The first three hidden layers of the encoders contained 16, 8, and 4 units. The encoder then connects to the information bottleneck, a single-unit dense layer with sigmoid activation. To restrict the bottleneck from capturing cell type biases in the training data, we directly supplied neuron subclass to a small encoder of three hidden units, which was then concatenated with the bottleneck activation to form the penultimate decoder layer. The decoder outputs the estimated mean parameter **μ** of the zero-inflated negative binomial (ZINB) model. Two additional encoder branches output estimates of the dispersion and dropout parameters. All output activations were implemented as in DCA, the only difference being the size of the output. The ZINB loss function was used, as implemented in DCA. To impose a prior expected distribution of bottleneck values, we applied a Kullbeck-Leibler divergence loss on the single-unit bottleneck activation using a beta distribution with alpha and beta shape parameters 0.25 and 1, respectively. The alpha parameter was chosen based on prior expectation that the unstimulated Allen Cell Types Database would have very few samples exhibiting a strong transcriptional response to activity. For model applications, the model outputs are not used, but the sigmoidal bottleneck activation value is the metric extracted to estimate neuronal activity.

Model Training

We trained the model using keras, as implemented in keras R package, version 2.3.0.0. Training proceeded for 10 epochs using the ADAM optimizer. Gaussian dropout was applied to input expression to simulate uncertainty in measurements. Augmented samples were given the same output as the original data to curtail the learning of depth-dependent information, and augmented samples appeared in the same training batches as the samples they were derived from. Sample losses were weighted to improve representation of rare cell subclasses based on prevalence in the training set, but weights were clamped to be no more than five times greater than they would be with an unweighted scheme, 1/n. We used basic hyperparameter tuning to evaluate the impact

of pre-bottleneck encoder layer size and units, dropout, and training epochs. Model performance and feature attributions were evaluated using a test split of approximately 50,000 samples that were withheld from model training. Distributions of bottleneck activity and loss were compared across species, sex, neuron class and subclass, and quality control metrics.

Evaluating Feature Importance

To evaluate relative importance of each gene on predicted activity, we implemented a strategy adapted from the integrated gradients[18], a popular feature attribution approach for image classification. Briefly, model gradients were calculated from input genes to bottleneck activation for the test set. For each gene we fit a generalized additive model (GAM) with smoothing between the sample gradients and bottleneck activation. We estimated global feature importance by approximating the integral of the fit. We took the summation of the GAM output for the full range of possible bottleneck activation values [0,1], with a step size of 1×10^{-3}. Gene set enrichment analysis was run using the fgsea implementation in R. We examined all gene ontology terms with size between 100-200, and collapsing redundant pathways using default parameters.

Comparison of Patch-seq and SMART-seq Datasets

The SMART-seq and Patch-seq datasets were integrated, clustered, and visualized using the R package Seurat[21]. The sctransform 'v2' method was used for normalization[65]. Cell clustering results were cross-referenced with provided neuron type annotations to identify neuron types common to both datasets and harmonize labels. We used the Kolmogorov-Smirnov test (KS test) to assess distributional differences in predicted activity of samples from the SMART-seq and Patch-seq datasets at the neuron type level. Only neuron types with at least one hundred samples in each dataset were considered. To derive empirical p-values and determine whether dataset-specific biases is transcriptome-wide expression levels contributed to observed differences in predicted activity, we ran 1,000 permutations of the KS test using activity score predictions from within-sample shuffled gene expression values. Empirical p-values were calculated from the frequency of permutation D statistics that were more extreme than those observed in the actual data.

Modeling Activity Score with Electrophysiological Features

The whole-cell Patch-seq electrophysiology recordings from neurons of mouse visual cortex provided by the Allen Institute for Brain Sciences were downloaded as NWB files from the DANDI (ID: 000020, September 2021 update)[20]. The dataset included electrophysiological recordings of 4,284 specimens from 1,040 mouse subjects. The Python package IPFX was used to perform quality control and compute electrophysiological features from the NWB files. We trained a random forest model to predict the activity score output of NEUROeSTIMator using only electrophysiological features. Cells included in training belonged to major neuron types (*e.g.*, Vip, Sst) with at least 250 cells, and minor neuron types (*e.g.*, Sst Rxfp1 Prdm8) with at least 25 cells. Features with missing values in five hundred or more cells were removed, and cells with any missing values for the remaining features were discarded. We engineered additional features from the IPFX output, including time and voltage changes between various landmarks of action potentials (*e.g.*, voltage change between action potential threshold

and peak). Minor neuron type was regressed out of the activity score and electrophysiological features. To identify outliers, we performed single value decomposition and computed Mahalanobis distance for sample loadings on the components cumulatively explaining at least 90% of the dataset variance. Samples with a Mahalanobis distance greater than four median absolute deviations were considered outliers and removed. The random forest model was trained using the randomForest R package with 1,500 trees, a terminal node size of five, and three variables randomly sampled as candidates at each split (e.g., `mtry=3`). Model performance is reported using out-of-bag error. To facilitate model interpretation, we trained a single decision tree using the same input features to predict the random forest predictions, and examined the features used at each split. To identify active and inactive cells whose activity score could be predicted using electrophysiological features, we identified two groups of cells in the top or bottom of NEUROeSTIMator output and random forest predictions of NEUROeSTIMator output. The KS test was used to determine if the most important feature in the random forest model could significantly separate active and inactive cells.

Preparation of Application Datasets

Application datasets were downloaded from the Gene Expression Omnibus (GEO). In datasets without provided labels, cell types were identified by running the standard Seurat pipeline, identifying cell clusters, and cross-referencing cluster marker genes to those used in the source publications or the Allen Cell Types Database. To prepare datasets for input to NEUROeSTIMator, gene expression data was log-normalized sample-wise and scaled gene-wise using mean and standard deviation values from the original training dataset. Samples with at least 1000 counts were then input to the model and activity score was the 1-unit bottleneck activation layer output.

Testing Differences in Predicted Activity

For all datasets analyzed in figures 2 and 3, we used the KS test to evaluate differences in activity score distributions. A linear model was used to test for differences (i.e., using t-statistic of the regression slope) in predicted activity of the spatial transcriptomic clusters in figure 5.

Comparing Performance to a Scaled Additive Approach

We compared the performance of NEUROeSTIMator to a scaled additive approach to quantify activity. The scaled additive measure of activity is calculated by first scaling log-normalized target expression values to a range between 0 and 1, and then summing the scaled targets in each cell or sample. Performance was defined by the ability to separate cells based on the stimulation status (*i.e.*, experimental group) using area under the curve (AUC). For GSE136656 and GSE102827, which contained multiple time points after stimulation, the first time point after stimulation (one hour) was chosen as the stimulated group.

Behavioral training for Spatial Transcriptomics

We generated a novel spatial transcriptomic dataset examining the effects of spatial learning in mice. The dataset contains spatial RNA-sequencing of whole brain slices from 1 hour after SOR training or home cage controls (HC). 10-12 weeks old C57BL/6J male

726 mice were trained in SOR task as described previously[24]. Briefly, mice were handled for
727 5 consecutive days prior to the SOR training. On the day of the training, animals were
728 individually placed in an open field with a spatial cue and was allowed to habituate for 6
729 mins. They were returned to their home cage and the arena was wiped clean with ethanol.
730 The animals were again placed back into the open field with three objects at specific
731 spatial locations. The animals were allowed to freely explore the objects in the arena for
732 6 mins. This training with objects comprised of three trials of 6 mins each. One hour after
733 the completion of the training, mice were euthanized and whole brains were flash frozen
734 in isopentane placed on dry ice.
735
736 **Spatial transcriptomics**
737 Mouse brain section per mouse was cut at 10 µm thickness and mounted onto each
738 Visium slide capture area. After H&E staining, each bright-field image was taken as
739 described in the spatial transcriptomics protocol. Tissue permeabilization was performed
740 for 18 minutes, as established in the tissue optimization assay. The Visium Spatial Gene
741 Expression Slide & Reagent kit (10x Genomics) was used to generate sequencing
742 libraries for Visium samples. Libraries were constructed according to the 10x Visium
743 library construction protocol and sequenced by Illumina NovaSeq6000. Raw data was
744 then processed using the 10x Genomics Space Ranger analysis pipeline. See
745 **Supplemental Figure S2** for images of predicted activity for each replicate.
746
747 ***In situ* hybridization**
748 *In situ* hybridization was performed on 20 µm coronal brain sections from OR-trained and
749 HC mice using RNAscope reagents (Advanced Cell Diagnostics). Briefly, fixed frozen
750 brains were sectioned on a microtome and mounted on Superfrost™ Plus microscope
751 slides (Fisherbrand). Slides then underwent serial dehydration in Ethanol, followed by
752 Hydrogen Peroxide treatment, target retrieval, and Protease III treatment – all of which
753 were done according to the manufacturer's protocol. Hybridization of probes was done at
754 40°C for 2 hours in an HybEZ oven using a 50:1:1 cocktail of Egr1, Nr4a1, and Egr3
755 respectively. The probe signals were amplified with Pre-amplifier 1 (Amp 1-FL) and
756 counterstained with OPAL dyes (Akoya Biosciences) corresponding to different
757 excitation-emission wavelengths. Finally, the slides were mounted with Vectashield®
758 Antifade Mounting Medium with DAPI (Vector Laboratories). Slides were stored in 4^0C
759 until they were imaged.
760
761 **Confocal imaging and analysis**
762 After performing the RNAscope assay, brain sections were imaged using Olympus
763 FV3000 confocal microscope with a 10X NA = 0.4 objective at 800 × 800-pixel resolution.
764 High magnification images of the hippocampus were obtained using a 100X NA = 1.45 oil
765 immersion objective at 800 × 800-pixel resolution and 1.25 optical zoom. All images (8
766 bit) were acquired with identical settings for laser power, detector gain and pinhole
767 diameter for each experiment and between experiments. Images from the different
768 channels were stacked and projected at maximum intensity using ImageJ (NIH).
769 Hippocampal subregion-specific Mean Fluorescence Intensity (MFI) of each RNAscope
770 probe was evaluated using plugins in ImageJ. The average MFI for all the 3 probes

corresponding to each hippocampal subregion was compared between the two experimental groups.

Statistics for RNAscope analysis

Average MFI corresponding to hippocampal subregions illustrated between Homecage and Learning groups. Two-way ANOVA: significant group x session interaction ($F(1, 6) = 8.015$, $p=0.0299$), Sidak's multiple comparisons test: **$p= 0.0075$ (subregion CA1, Homecage versus Learning, n=4), $p= 0.5553$ (subregion DG, Homecage versus Learning, n=4).

DATA AVAILABILITY

Spatial RNA-sequencing data, including gene expression measurements, tissue images, spot coordinates, and raw FASTQ files have been deposited in the Gene Expression Omnibus repository under the reference series ID GSE201610.

CODE AVAILABILITY

NEUROeSTIMator is available at https://research-git.uiowa.edu/michaelson-lab-public/neuroestimator/ as a free R package with installation instructions and a tutorial.

ACKNOWLEDGEMENTS

Funding: This work was supported by NIH grant R01 MH 087463 to T.A., NIH grant R01 DC014489 to J.J.M., NIH grant K99 AG 068306 and the Nellie Ball Trust to S.C., and The University of Iowa Hawkeye Intellectual and Developmental Disabilities Research Center (HAWK-IDDRC) P50 HD103556 to T.A. and Lane Strathearn. T.A. and J.J.M. are also supported by the Roy J. Carver Charitable Trust.

AUTHOR CONTRIBUTIONS

EB, SC, TA, and JM conceived the study. EB performed data curation, model training, analyses, and visualizations. SC and TA designed the molecular biology experiments with input from JM and KPG. SC performed behavioral experiments, stereotactic surgeries, and single nuclei sequencing. LC performed Visium spatial transcriptomic sequencing experiments. YV and ME processed Visium data. JR and MM performed RNAscope experiments. UM performed statistical analysis and created visualization pertaining to the RNAscope data. EB and JM wrote the article.

Data Sources

We thank the Allen Institute for Brain Sciences for their valuable datasets we used to train our model. We thank the creators and authors of DCA, whose work inspired the approach we implemented in this paper. Finally, we thank the investigators who make their data publicly available on repositories like GEO. Specifically, we would like to thank the authors providing data for GEO series GSE111899, GSE125068, GSE55591, GSE106678, GSE137763, GSE136656, and GSE102827.

Competing interests: The authors declare that they have no competing interests.

We also thank Mahesh Shetty for his valuable contributions to discussions about this project.

dataset	use	species	stimulation	tissue
Allen Cell Types Database: mouse	model training	mouse	-	cortex, hippocampus
Allen Cell Types Database: human	model training	human	-	cortex
GSE111899	target selection	mouse	Sensory experience	cortex (visual)
GSE125068	target selection	mouse	PTZ	hippocampus
GSE55591	target selection	mouse	KCl	cortex (neuron culture)
Allen Patch-Sequencing Dataset	model application	mouse	Patch-sequencing protocol	cortex
Allen Smart-Sequencing Dataset	model application	mouse	NA	cortex
GSE106678	model application	mouse	PTZ	cortex
GSE137767	model application	rat	cocaine	nucleus accumbens
GSE136656	model application	human	KCl	neuron culture
GSE102827	model application	mouse	Sensory experience	cortex (visual)
SOR - Visium	model application	mouse	Spatial object recognition training	whole brain slice
SOR – Single Nuclei	model application	mouse	Spatial object recognition training	dorsal hippocampus

Table 1. Datasets used.

References

1 Yap, E. L. & Greenberg, M. E. Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron* **100**, 330-348 (2018). https://doi.org:10.1016/j.neuron.2018.10.013

2 Cohen, S. & Greenberg, M. E. Communication between the synapse and the nucleus in neuronal development, plasticity, and disease. *Annu Rev Cell Dev Biol* **24**, 183-209 (2008). https://doi.org:10.1146/annurev.cellbio.24.110707.175235

3 West, A. E. & Greenberg, M. E. Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb Perspect Biol* **3** (2011). https://doi.org:10.1101/cshperspect.a005744

4 Gallo, F. T., Katche, C., Morici, J. F., Medina, J. H. & Weisstaub, N. V. Immediate Early Genes, Memory and Psychiatric Disorders: Focus on c-Fos, Egr1 and Arc. *Front Behav Neurosci* **12**, 79 (2018). https://doi.org:10.3389/fnbeh.2018.00079

5 Mews, P. *et al.* From Circuits to Chromatin: The Emerging Role of Epigenetics in Mental Health. *J Neurosci* **41**, 873-882 (2021). https://doi.org:10.1523/JNEUROSCI.1649-20.2020

6 Nido, G. S., Ryan, M. M., Benuskova, L. & Williams, J. M. Dynamical properties of gene regulatory networks involved in long-term potentiation. *Front Mol Neurosci* **8**, 42 (2015). https://doi.org:10.3389/fnmol.2015.00042

7 Wu, Y. E., Pan, L., Zuo, Y., Li, X. & Hong, W. Detecting Activated Cell Populations Using Single-Cell RNA-Seq. *Neuron* **96**, 313-329 e316 (2017). https://doi.org:10.1016/j.neuron.2017.09.026

8 Hu, P. *et al.* Dissecting Cell-Type Composition and Activity-Dependent Transcriptional State in Mammalian Brains by Massively Parallel Single-Nucleus RNA-Seq. *Mol Cell* **68**, 1006-1015 e1007 (2017). https://doi.org:10.1016/j.molcel.2017.11.017

9 Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S. & Marioni, J. C. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat Methods* **14**, 565-571 (2017). https://doi.org:10.1038/nmeth.4292

10 Hicks, S. C., Townes, F. W., Teng, M. & Irizarry, R. A. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* **19**, 562-578 (2018). https://doi.org:10.1093/biostatistics/kxx053

11 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, 296 (2019). https://doi.org:10.1186/s13059-019-1874-1

12 Lahnemann, D. *et al.* Eleven grand challenges in single-cell data science. *Genome Biol* **21**, 31 (2020). https://doi.org:10.1186/s13059-020-1926-6

13 van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579-2605 (2008).

14 McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv:1802.03426 (2018). <https://ui.adsabs.harvard.edu/abs/2018arXiv180203426M>.

15 Bao, S. *et al.* Deep learning-based advances and applications for single-cell RNA-sequencing data analysis. *Brief Bioinform* **23** (2022). https://doi.org:10.1093/bib/bbab473

16 Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S. & Theis, F. J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat Commun* **10**, 390 (2019). https://doi.org:10.1038/s41467-018-07931-2

17 Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241 e3226 (2021). https://doi.org:10.1016/j.cell.2021.04.021

18 Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. arXiv:1703.01365 (2017). <https://ui.adsabs.harvard.edu/abs/2017arXiv170301365S>.

19 Malik, A. R. *et al.* Cyr61, a matricellular protein, is needed for dendritic arborization of hippocampal neurons. *J Biol Chem* **288**, 8544-8559 (2013). https://doi.org:10.1074/jbc.M112.411629

20 Gouwens, N. W. *et al.* Integrated Morphoelectric and Transcriptomic Classification of Cortical GABAergic Cells. *Cell* **183**, 935-953 e919 (2020). https://doi.org:10.1016/j.cell.2020.09.057

21 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). https://doi.org:10.1016/j.cell.2021.04.048

22 Savell, K. E. *et al.* A dopamine-induced gene expression signature regulates neuronal function and cocaine response. *Sci Adv* **6**, eaba4221 (2020). https://doi.org:10.1126/sciadv.aba4221

23 Boulting, G. L. *et al.* Activity-dependent regulome of human GABAergic neurons reveals new patterns of gene regulation and neurological disease heritability. *Nat Neurosci* **24**, 437-448 (2021). https://doi.org:10.1038/s41593-020-00786-1

24 Chatterjee, S. *et al.* Endoplasmic reticulum chaperone genes encode effectors of long-term memory. *Sci Adv* **8**, eabm6063 (2022). https://doi.org:10.1126/sciadv.abm6063

25 Wang, F. *et al.* RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J Mol Diagn* **14**, 22-29 (2012). https://doi.org:10.1016/j.jmoldx.2011.08.002

26 Iijima, T., Emi, K. & Yuzaki, M. Activity-dependent repression of Cbln1 expression: mechanism for developmental and homeostatic regulation of synapses in the cerebellum. *J Neurosci* **29**, 5425-5434 (2009). https://doi.org:10.1523/JNEUROSCI.4473-08.2009

27 Itoh, K., Stevens, B., Schachner, M. & Fields, R. D. Regulated expression of the neural cell adhesion molecule L1 by specific patterns of neural impulses. *Science* **270**, 1369-1372 (1995). https://doi.org:10.1126/science.270.5240.1369

28 Koberstein, J. N. *et al.* Learning-dependent chromatin remodeling highlights noncoding regulatory regions linked to autism. *Sci Signal* **11** (2018). https://doi.org:10.1126/scisignal.aan6500

29 Fernandez-Albert, J. *et al.* Immediate and deferred epigenomic signatures of in vivo neuronal activation in mouse hippocampus. *Nat Neurosci* **22**, 1718-1730 (2019). https://doi.org:10.1038/s41593-019-0476-2

30 Yiu, A. P. *et al.* Neurons are recruited to a memory trace based on relative neuronal excitability immediately before training. *Neuron* **83**, 722-735 (2014). https://doi.org:10.1016/j.neuron.2014.07.017

31 Beaumont, T. L., Yao, B., Shah, A., Kapatos, G. & Loeb, J. A. Layer-specific CREB target gene induction in human neocortical epilepsy. *J Neurosci* **32**, 14389-14401 (2012). https://doi.org:10.1523/JNEUROSCI.3408-12.2012

32 Sonoda, K., Matsui, T., Bito, H. & Ohki, K. Astrocytes in the mouse visual cortex reliably respond to visual stimulation. *Biochem Biophys Res Commun* **505**, 1216-1222 (2018). https://doi.org:10.1016/j.bbrc.2018.10.027

33 Scholkmann, F., Zohdi, H. & Wolf, U. The Role of Systemic Physiology in Individual Hemodynamic Responses Measured on the Head Due to Long-Term Stimulation Involving Colored Light Exposure and a Cognitive Task: An SPA-fNIRS Study. *Brain Sci* **12** (2022). https://doi.org:10.3390/brainsci12050597

34 Toronov, V. Y., Zhang, X. & Webb, A. G. A spatial and temporal comparison of hemodynamic signals measured using optical and functional magnetic resonance imaging during activation in the human primary visual cortex. *Neuroimage* **34**, 1136-1148 (2007). https://doi.org:10.1016/j.neuroimage.2006.08.048

35 Khachigian, L. M. *et al.* Egr-1 is activated in endothelial cells exposed to fluid shear stress and interacts with a novel shear-stress-response element in the PDGF A-chain promoter. *Arterioscler Thromb Vasc Biol* **17**, 2280-2286 (1997). https://doi.org:10.1161/01.atv.17.10.2280

36 Schwachtgen, J. L., Houston, P., Campbell, C., Sukhatme, V. & Braddock, M. Fluid shear stress activation of egr-1 transcription in cultured human endothelial and epithelial cells is mediated via the extracellular signal-related kinase 1/2 mitogen-activated protein kinase pathway. *J Clin Invest* **101**, 2540-2549 (1998). https://doi.org:10.1172/JCI1404

37 Bahrami, S. & Drablos, F. Gene regulation in the immediate-early response process. *Adv Biol Regul* **62**, 37-49 (2016). https://doi.org:10.1016/j.jbior.2016.05.001

38 Iaria, G., Petrides, M., Dagher, A., Pike, B. & Bohbot, V. D. Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *J Neurosci* **23**, 5945-5952 (2003).

39 Postle, B. R. & D'Esposito, M. Dissociation of human caudate nucleus activity in spatial and nonspatial working memory: an event-related fMRI study. *Brain Res Cogn Brain Res* **8**, 107-115 (1999). https://doi.org:10.1016/s0926-6410(99)00010-5

40 Woolley, D. G. *et al.* Virtual water maze learning in human increases functional connectivity between posterior hippocampus and dorsal caudate. *Hum Brain Mapp* **36**, 1265-1277 (2015). https://doi.org:10.1002/hbm.22700

41 Blazing, R. M. & Franks, K. M. Odor coding in piriform cortex: mechanistic insights into distributed coding. *Curr Opin Neurobiol* **64**, 96-102 (2020). https://doi.org:10.1016/j.conb.2020.03.001

42 Tsien, J. Z., Huerta, P. T. & Tonegawa, S. The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell* **87**, 1327-1338 (1996). https://doi.org:10.1016/s0092-8674(00)81827-9

43 Tanimizu, T., Kono, K. & Kida, S. Brain networks activated to form object recognition memory. *Brain Res Bull* **141**, 27-34 (2018). https://doi.org:10.1016/j.brainresbull.2017.05.017

44 Haettig, J., Sun, Y., Wood, M. A. & Xu, X. Cell-type specific inactivation of hippocampal CA1 disrupts location-dependent object recognition in the mouse. *Learn Mem* **20**, 139-146 (2013). https://doi.org:10.1101/lm.027847.112

45 Bridi, M. S. & Abel, T. The NR4A orphan nuclear receptors mediate transcription-dependent hippocampal synaptic plasticity. *Neurobiol Learn Mem* **105**, 151-158 (2013). https://doi.org:10.1016/j.nlm.2013.06.020

46 Pittenger, C. *et al.* Reversible inhibition of CREB/ATF transcription factors in region CA1 of the dorsal hippocampus disrupts hippocampus-dependent spatial memory. *Neuron* **34**, 447-462 (2002). https://doi.org:10.1016/s0896-6273(02)00684-0

47 Danielson, N. B. *et al.* Sublayer-Specific Coding Dynamics during Spatial Navigation and Learning in Hippocampal Area CA1. *Neuron* **91**, 652-665 (2016). https://doi.org:10.1016/j.neuron.2016.06.020

48 Muzzio, I. A., Kentros, C. & Kandel, E. What is remembered? Role of attention on the encoding and retrieval of hippocampal representations. *J Physiol* **587**, 2837-2854 (2009). https://doi.org:10.1113/jphysiol.2009.172445

49 Roy, D. S. *et al.* Brain-wide mapping reveals that engrams for a single memory are distributed across multiple brain regions. *Nat Commun* **13**, 1799 (2022). https://doi.org:10.1038/s41467-022-29384-4

50 Josselyn, S. A., Kohler, S. & Frankland, P. W. Finding the engram. *Nat Rev Neurosci* **16**, 521-534 (2015). https://doi.org:10.1038/nrn4000

51 Liu, X., Ramirez, S., Redondo, R. L. & Tonegawa, S. Identification and Manipulation of Memory Engram Cells. *Cold Spring Harb Symp Quant Biol* **79**, 59-65 (2014). https://doi.org:10.1101/sqb.2014.79.024901

52 DeNardo, L. A. *et al.* Temporal evolution of cortical ensembles promoting remote memory retrieval. *Nat Neurosci* **22**, 460-469 (2019). https://doi.org:10.1038/s41593-018-0318-7

53 Wheeler, A. L. *et al.* Identification of a functional connectome for long-term fear memory in mice. *PLoS Comput Biol* **9**, e1002853 (2013). https://doi.org:10.1371/journal.pcbi.1002853

54 Nguyen, T. V., Kosofsky, B. E., Birnbaum, R., Cohen, B. M. & Hyman, S. E. Differential expression of c-fos and zif268 in rat striatum after haloperidol, clozapine, and amphetamine. *Proc Natl Acad Sci U S A* **89**, 4270-4274 (1992). https://doi.org:10.1073/pnas.89.10.4270

55 Polston, E. K. & Erskine, M. S. Patterns of induction of the immediate-early genes c-fos and egr-1 in the female rat brain following differential amounts of mating stimulation. *Neuroendocrinology* **62**, 370-384 (1995). https://doi.org:10.1159/000127027

56 Slattery, D. A. *et al.* Comparison of alterations in c-fos and Egr-1 (zif268) expression throughout the rat brain following acute administration of different

classes of antidepressant compounds. *Neuropsychopharmacology* **30**, 1278-1287 (2005). https://doi.org:10.1038/sj.npp.1300717

57 Hrvatin, S. *et al.* Single-cell analysis of experience-dependent transcriptomic states in the mouse visual cortex. *Nat Neurosci* **21**, 120-129 (2018). https://doi.org:10.1038/s41593-017-0029-5

58 Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439-3440 (2005). https://doi.org:10.1093/bioinformatics/bti525

59 Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009). https://doi.org:10.1038/nprot.2009.97

60 Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Res* **46**, D754-D761 (2018). https://doi.org:10.1093/nar/gkx1098

61 Tyssowski, K. M. *et al.* Different Neuronal Activity Patterns Induce Different Gene Expression Programs. *Neuron* **98**, 530-546 e511 (2018). https://doi.org:10.1016/j.neuron.2018.04.001

62 Spiegel, I. *et al.* Npas4 regulates excitatory-inhibitory balance within neural circuits through cell-type-specific gene programs. *Cell* **157**, 1216-1229 (2014). https://doi.org:10.1016/j.cell.2014.03.058

63 Plaisier, S. B., Taschereau, R., Wong, J. A. & Graeber, T. G. Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res* **38**, e169 (2010). https://doi.org:10.1093/nar/gkq636

64 McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186 (2017). https://doi.org:10.1093/bioinformatics/btw777

65 Choudhary, S. & Satija, R. Comparison and evaluation of statistical error models for scRNA-seq. *Genome Biol* **23**, 27 (2022). https://doi.org:10.1186/s13059-021-02584-9

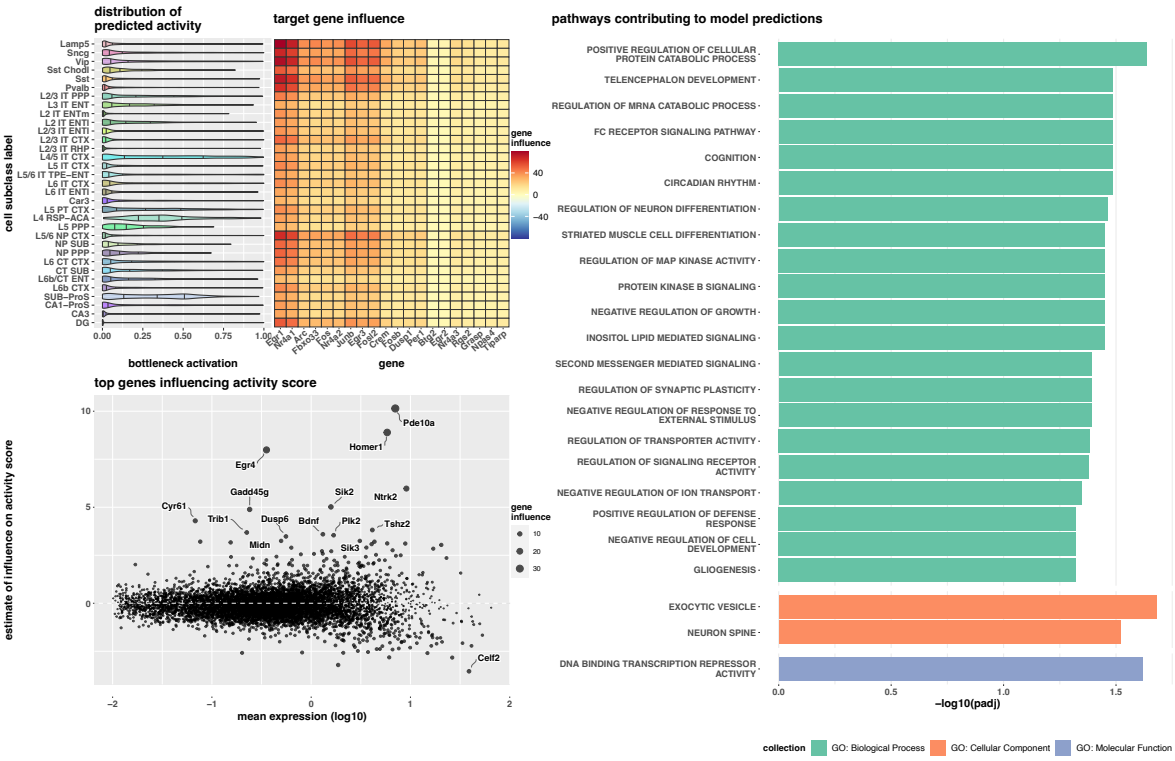


Figure 1 - **Gene-wise contributions to the activity score and its distribution within cell types of the training data. A)** The proportion of active cells among each neuronal subclass is shown by the plotted distributions of bottleneck activation (i.e., the activity score), **left**. Subclass-specific contribution of target genes to the activity score prediction is shown on the **right**. Gene influence is defined by mean integrated gradient values from the model. Higher positive gene gradient values indicate a positive relationship between gene expression input to the model and activity score. **B**) To identify the strongest non-target contributors to the activity score, we compared overall gene influence on predicted activity (X-axis) versus influence on model predictions (Y-axis). (**C**) Pathways and gene sets significantly enriched for genes contributing to activity score predictions. Significance of gene set enrichment test with multiple testing correction. Bars are grouped and colored by annotation source.

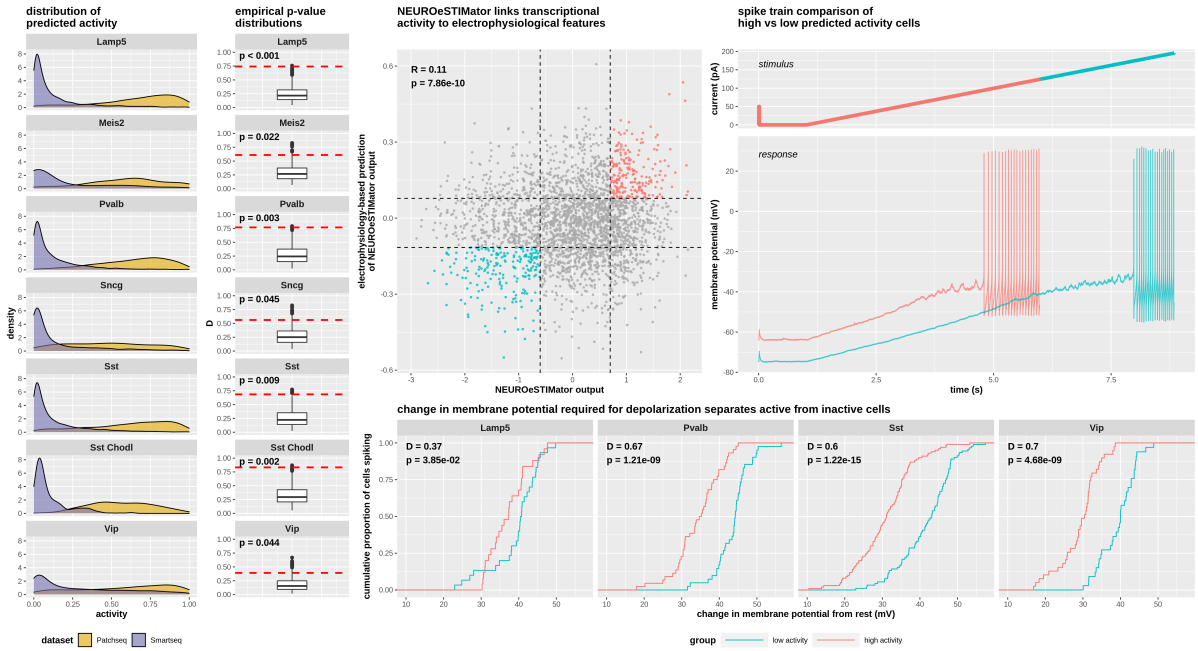


Figure 2 – Electrophysiological features of neuronal stimulation captured by transcriptome-based activity score. (**A**) Activity score distributions compared between stimulated Patch-seq neurons (gold) and unstimulated SMART-seq neurons (violet). (**B**) Observed D statistics (red line) and empirical distribution from Kolmogorov-Smirnov tests between activity score distributions of Patch-seq and SMART-seq datasets, using one thousand permutations of shuffled expression values . Observed NEUROeSTIMator (x-axis) versus electrophysiology-based random forest prediction of NEUROeSTIMator output (y-axis). Cells in the same quartile of observed and predicted contain neurons whose transcriptome-based activity score can be predicted electrophysiologically (high activity – red, low activity – blue) (**C**). Spike train comparison of two Sst Rxfp1 Prdm8 neurons from the high (red) and low (blue) activity groups which demonstrate electrophysiological features predictive of NEUROeSTIMator output (ramp current injected - top panel, membrane potential - bottom panel) (**D**). Cumulative proportion of cells spiking for as difference in resting and threshold potential increases. Annotated statistics from Kolmogorov-Smirnov test (**E**).

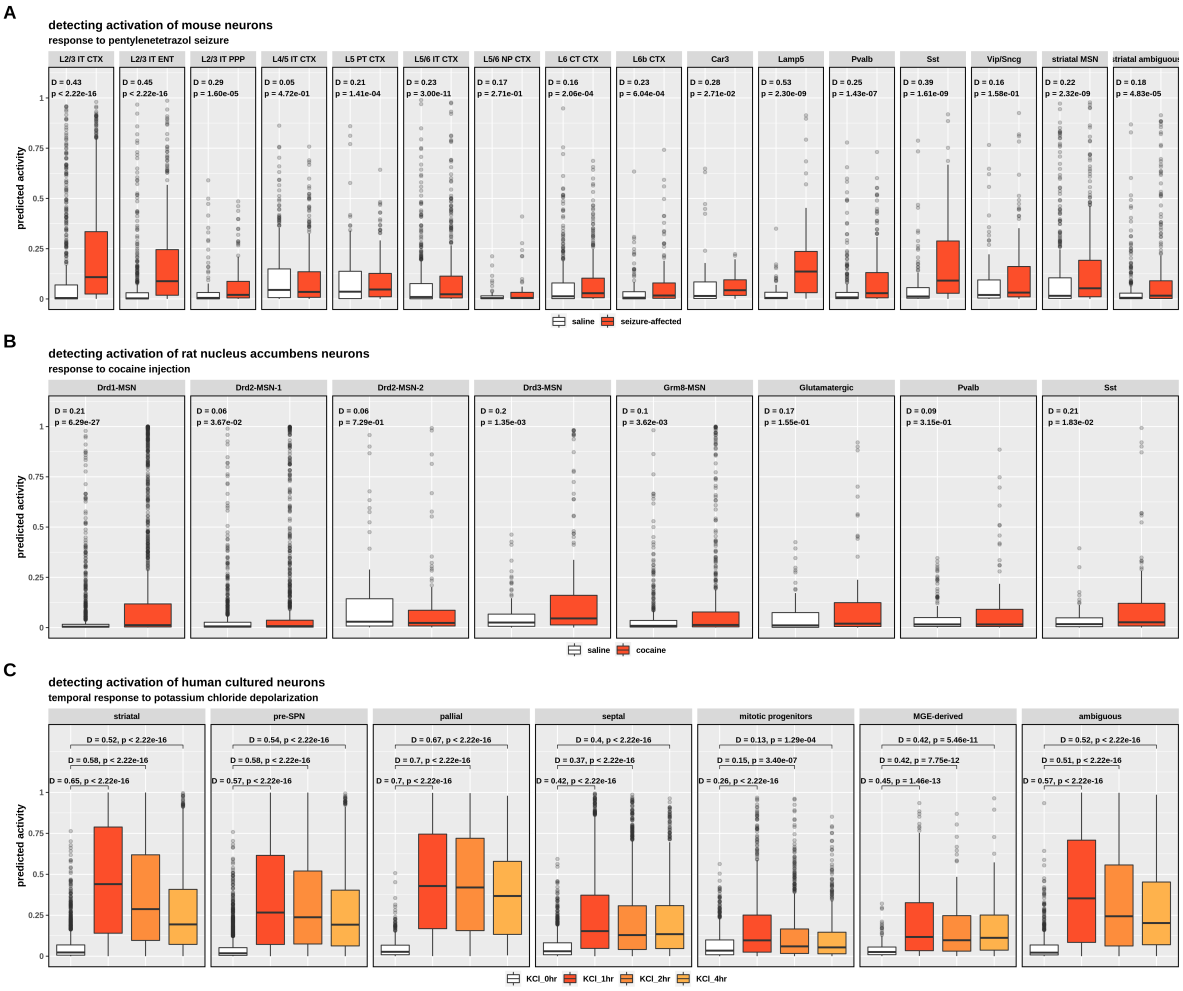


Figure 3 – Multi-species generalization of neuronal activity score applied to previously published chemical induction studies. Predicted activity for various neuron types responding to treatment with potent chemical modulators of neuron activity. (**A**) Response to PTZ-induced seizure (red) and controls (white) in mouse cortical neurons. (**B**) Cocaine treatment (red) or controls (white) in rat nucleus accumbens neurons. (**C**) Time series of predicted activity for human iPSCs treated with depolarizing potassium chloride at 0 hours (control), 1 hour, 2 hours, and 4 hours. Annotated statistics from Kolmogorov-Smirnov test.

1065
1066

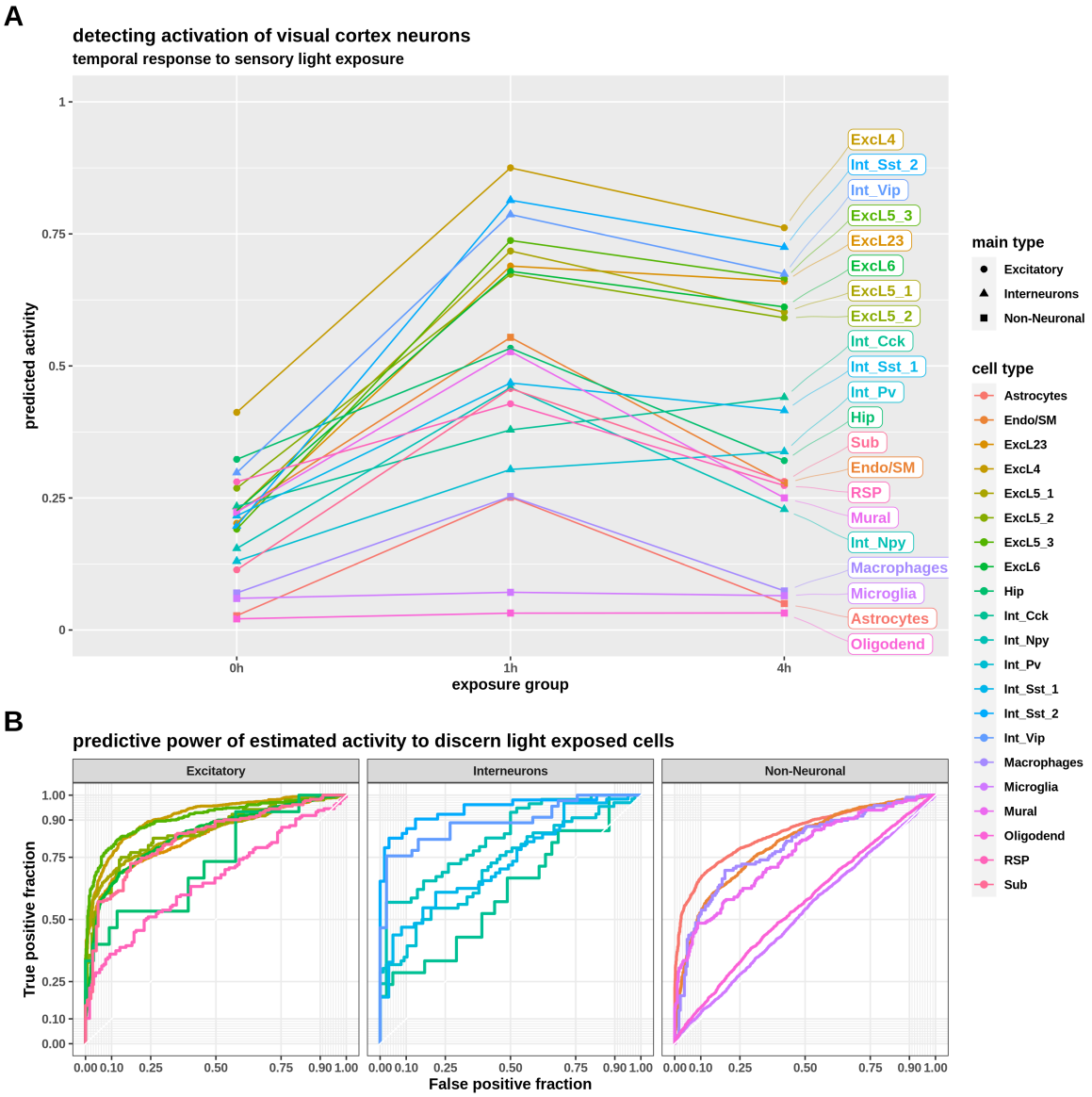


Figure 4 - Temporal patterns and classification of *in vivo* sensory activation. (**A**) Cell type activity predictions of visual cortex neurons in freely behaving mice exposed to light for 0, 1, or 4 hours. Activity predictions summarized by cell type and experimental group. (**B**) ROC plots indicate the ability of predicted activity to distinguish between the 0h and 1h experimental groups. Diagonals from bottom left to top right indicate an accuracy similar to random chance, while lines moving straight vertically, then straight horizontally indicate perfect separation.

1067
1068

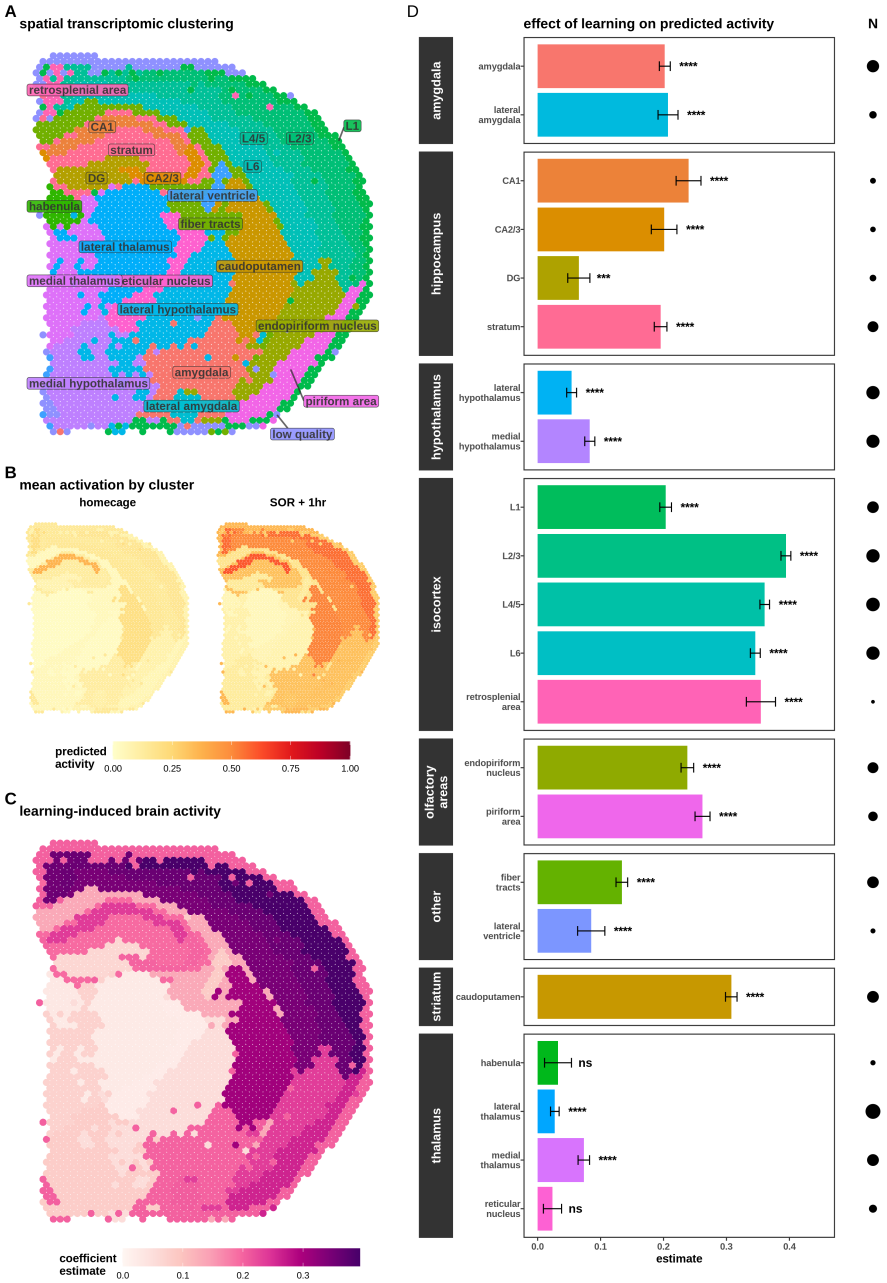
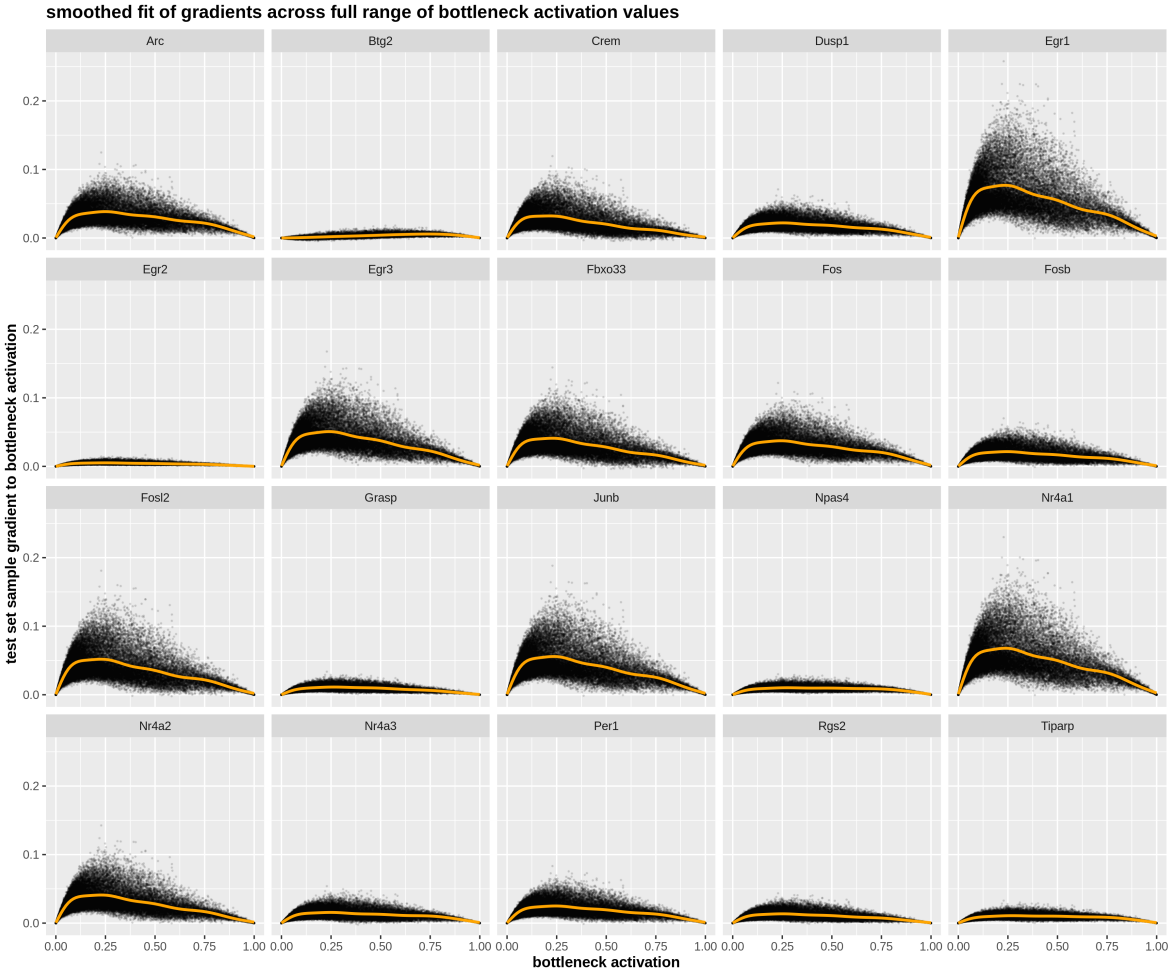


Figure 5 – Spatial transcriptomic patterns of neuronal activation after spatial learning. (**A**) Spatial anatomical clustering of RNA-sequencing spots. (**B**) Activity score per spot, averaged within experimental groups of homecage controls (left) and 1 hour after SOR training (right). Using the spatial anatomical clustering derived from the expression data, we were able to group individual spots into clusters to. (**C**) Brain region-specific induction of activity following SOR training. Differences tested using linear models, colored by the SOR coefficient estimate. Cluster-wise differential activity statistics are summarized in (**D**) and provided in **Supplementary Table S1**. Bar length represents estimated coefficient of SOR on activity score. Brackets indicate standard error and circle size represents the number of spots per comparison, an index of statistical power. Non-significant regions with an adjusted p-value > 0.05 are colored grey.

1069
1070



Supplemental Figure 1 – Integrated gradient approach to feature attribution.
Gradients from target gene inputs to bottleneck activation output (activity score) for test set samples. The orange line is the smoothed fit, and the integral for the range of [0,1] represents the estimate of gene influence.

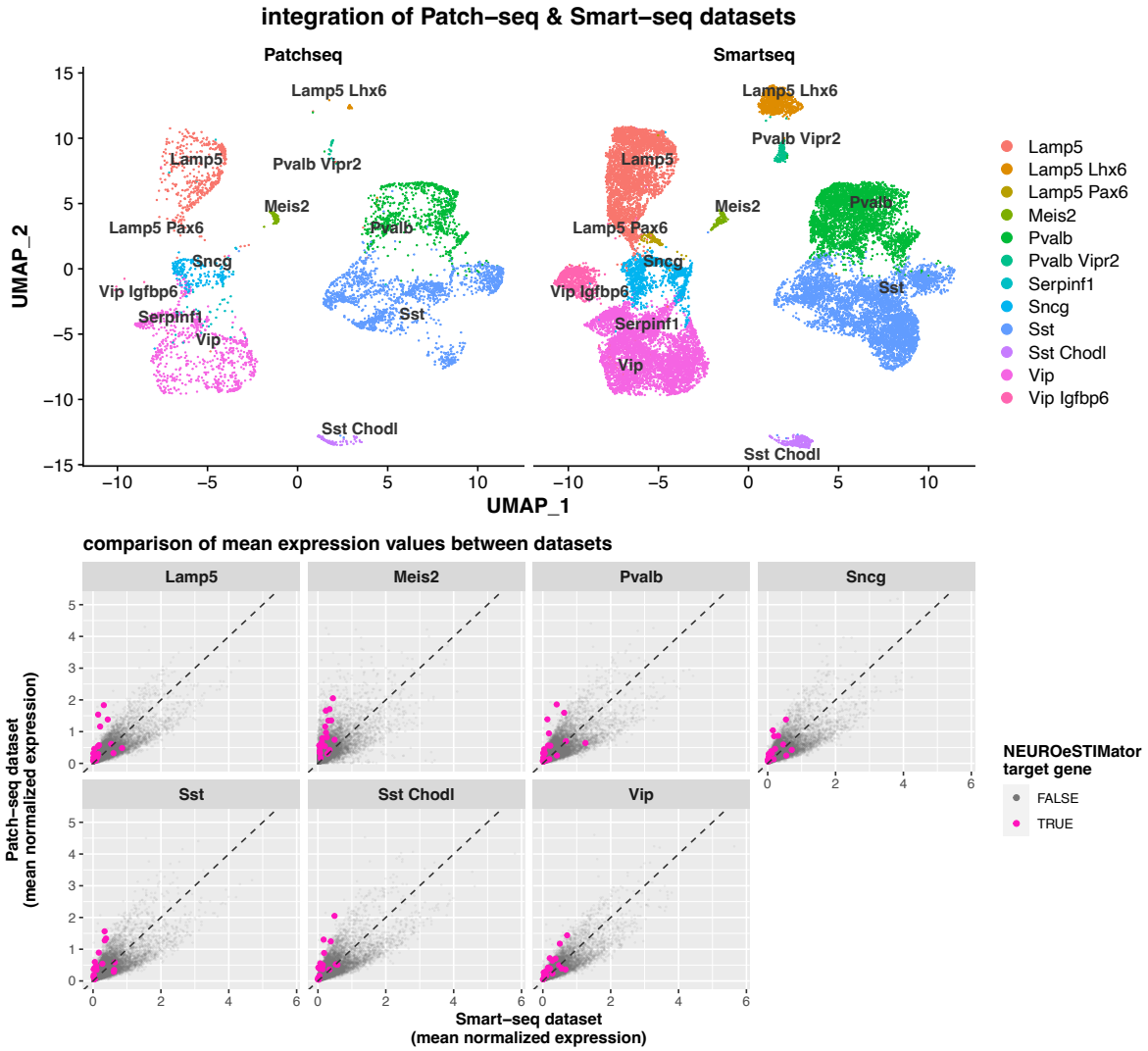


Figure S1 - Integration of Patch-sequencing dataset to GABAergic reference dataset. (**A**) UMAP reduction showing the Seurat integration for the Patch-sequencing dataset (left) and GABAergic reference dataset (right). (**B**) Comparison of average cell-type specific expression per gene. Several model target genes (pink) are amongst the most discrepant genes between datasets, with higher expression in the Patch-sequencing dataset.

1073
1074

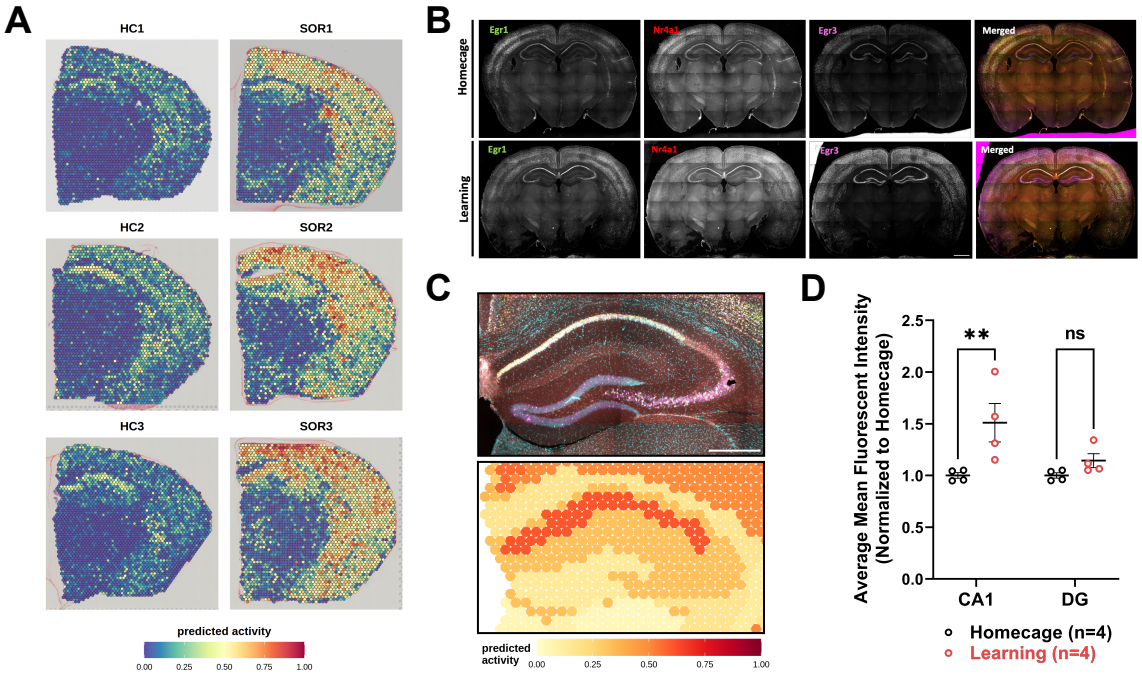


Figure S2 - Spatial transcriptomic techniques reveal region-specific activity in response to spatial learning. (**A**) Predicted activity of individual Visium spots for each sample from a homecage control group (left) and an SOR-trained group (right). (**B**) RNAScope validation of region-specific activation *Egr1*, *Nr4a1*, and *Egr3*. Scale bar represents 1000 um in length. (**C**) Image comparison of learning induced activity as measured by RNAScope (top) and Visium (bottom). Scale bar depicts 500 um in length. (**D**) Dorsal hippocampus subregion-specific induction of transcriptional activity following SOR.

1075
1076

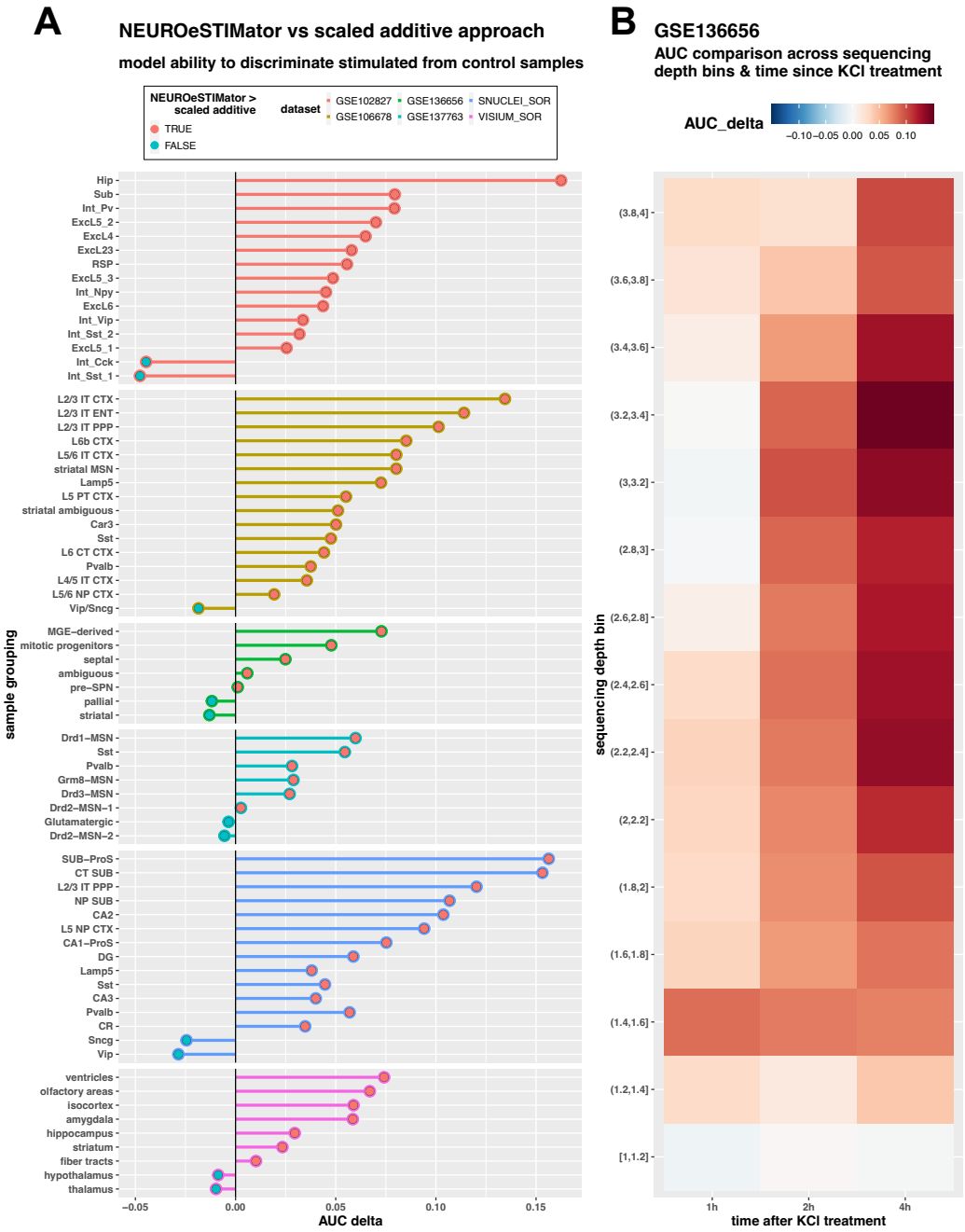


Figure S3 - Generalized Improvement in Classification Accuracy Over Scaled Additive Approach. Comparison of classification accuracy (AUC) of experimental groups for the NEUROeSTIMator model and a scaled additive model. (**A**) Depicts the increase in AUC afforded by NEUROeSTIMator. Datasets are group by color and rows correspond to cell/sample groupings. (**B**) Examines improved classification performance for cells grouped by timepoints (columns, biological effect) and total log counts (rows, technical effect). NEUROeSTIMator improves classification accuracy over a scaled additive model, particularly for distal time points and for cells with lower total UMI counts.

1077
1078

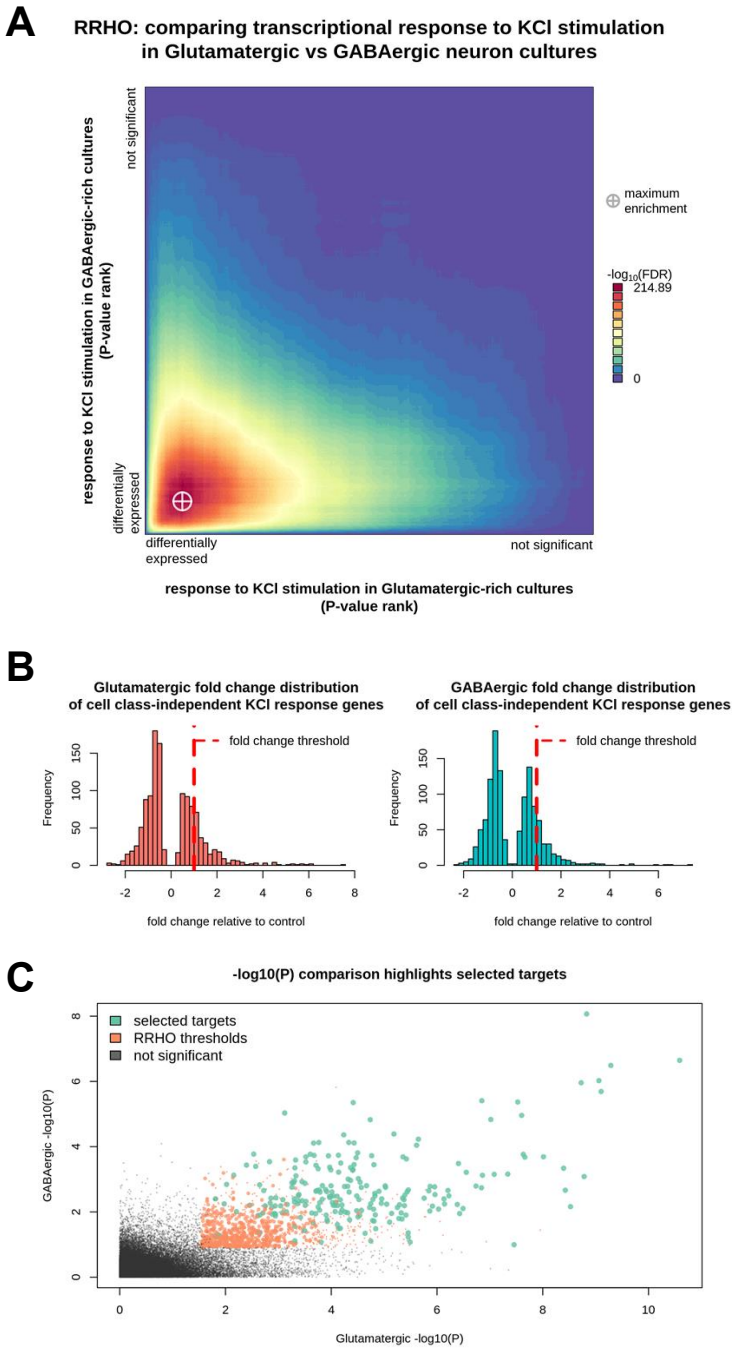


Figure S4 - Candidate Target Gene Selection. (**A**) Rank-rank hypergeometric overlap (RRHO) was used to identify p-value thresholds of maximum enrichment between differential expression lists for potassium chloride depolarization in glutamatergic-enriched and GABAergic-enriched neuron samples. Log fold change thresholds were set at one (**B**) to select the final set of candidate targets from genes below RRHO p-value thresholds in both lists (**C**).

1079