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

Ethan Bahl[1,2], Snehajyoti Chatterjee[3,4], Muhammad Elsadany[1,2], Yann Vanrobaeys[2,3], Li-Chun Lin[3,4], 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.
* 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 signals of activation distributed throughout the broader transcriptome to estimate neuronal activation in a way that 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 single cell and time course studies of activity-induced gene expression. Further, using spatial transcriptomic techniques, we demonstrate the model's ability to identify patterns of learning-induced activation. In conclusion, NEUROeSTIMator is 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 controls synaptic plasticity and is dysregulated in many disorders of the nervous system[1-6]. Currently, a major focus in neuroscience research is aimed at understanding tissue and cell type specificity of activity-dependent transcription factors and the corresponding 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 significant barriers to characterizing and quantifying gene expression signatures of neuronal activity.

Upregulated expression levels of several transcription factors, such as *Fos* and *Egr1,* are commonly used as markers of activity-dependent transcriptional response in RNA-sequencing experiments[1,4]. However, single cell RNA-sequencing data exhibits sparsity and variability in gene expression measurements that can be attributed 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 in the absence of stimulation or activity[1]. The consequence of 1) heterogeneous within-cell type responses to stimulation (*e.g.*, memory-associated engram populations[13]), where 2) subsets of responsive cells cannot be confidently identified results in a drastic reduction of power in e.g., differential expression analyses between experimental conditions at the cell type level, as signal from the experimental group is diluted by non-responsive cells. Therefore, it is not only crucial to identify subsets of responsive cells for between-sample comparisons to a control group, but it also opens further avenues for higher powered within-sample comparisons. 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 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)[14], and uniform manifold projection (UMAP)[15], similarly aim 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[16]. DCA[17] 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 expression patterns 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. Applying our approach to a diverse collection of datasets, we demonstrate that this 1-dimensional bottleneck, hereafter referred to as the 'activity score', can identify individual neuronal activation caused by seizure, cocaine administration, and sensory experience. We demonstrate the use of the activity score to classify cells by experimental manipulation, expose genes involved in these predictions, and identify neuron type-specific expression signatures of activity. Furthermore, we demonstrate generalizability of our approach to new spatial transcriptomic data from brain slices following learning. To enable the neuroscience community to take advantage of these efforts, we have developed NEUROeSTIMator, an R package with an accompanying tutorial that demonstrates an application 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[18], 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 approximately 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[19] 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*. In general, target genes with lower mean expression contributed less to predictions. 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 also found several cell type markers exhibiting influence on predictions such as *Gad2*, *Sst*, and *Pde10a*. We observed that many influential genes had higher mean expression levels, although lower abundance genes also exerted influence. For example, the gene *Cyr61*, known to regulate dendritic arborization[20], 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 several annotation sets (**Fig. 1c**). Among the most highly enriched gene sets were BDNF/NTRK signaling ($p_{adjusted} < 0.05$ for 3 largely redundant annotations), circadian rhythm ($p_{adjusted} = 0.0084$), and nuclear receptor metapathways ($p_{adjusted} = 0.0319$). Many influential genes were members of multiple significant gene sets, with *Homer1, Bdnf, Ntrk2, Jun,* and *Sst* showing high degrees of centrality within the significant gene set membership network (**Fig. 1d**).

To demonstrate the utility and generalizability of our model for external applications, we applied it to five datasets (**Table 1**).

Detecting Pharmacological Activation of Neurons

We first 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 medial amygdala (MeA) 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 increases in predicted activity for several neuronal subtypes, including GABAergic subtypes N2-N4 as well as glutamatergic subtypes N10-N12 (Wilcoxon test, $p_{adjusted} < 0.001$) (**Fig. 2a**).

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[21]. We found neuron subtype-specific increases in activity score (**Fig. 2b**). D1 and D3-type medium spiny neurons (MSN) were the most profoundly affected neuron subtypes (Wilcoxon test, D1: $p_{adjusted} < 2.22x10^{-16}$, D2: $p_{adjusted} = 0.0018$). 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[22] (**Fig. 2c**). 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 increases in predicted activity were observed for all cell types ($p_{adjusted} < 0.05$), the activity score of post-mitotic neurons displayed a stronger response to KCl compared to NES+ neural progenitor clusters. Among the most responsive neuron types was the Tbr1+ pallial glutamatergic cluster (cluster 6) ($p_{adjusted} < 2.22x10^{-16}$ for all time points compared to baseline). Notably, all neuron 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 levels 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 stimulation for 0, 1, or 4 hours (**Fig. 3a**). To determine whether the ability of our model to detect activity signatures is restricted to neurons, we additionally examined predicted activity in non-

neuronal cell types. 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. 3b**). 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. Surprisingly, despite the model being trained on purely neuronal cell type populations, the activity score was able to separate stimulation groups for astrocytes just as accurately as it could for neurons. Astrocytes in the visual cortex have been shown to reliably respond to light[23]. Not only does this application of the activity score provide further evidence of astrocytic responsiveness to light, it also directly suggests that light induces transcriptional changes in astrocytes.

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 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 and clustered all spots into anatomical regions. The 23 resulting clusters were annotated with brain region names from the Allen Mouse Brain Atlas (**Fig. 4a**). At baseline, we noted a weak activation signature in HC slices, primarily covering cortical layers of the isocortex and subregions of the hippocampus (**Fig. 4b, left**). To identify a spatial activation signature of SOR, we tested for differences in predicted activity for each brain region cluster (**Fig. 4c**). We observed significant increases in predicted activity for several cortical and subcortical regions. Multiple layers of the isocortex and the retrosplenial area showed large increases in activity following SOR (**Fig. 4d**), with the greatest increase observed in layers 2/3. We also found a significant activation of the caudoputamen area of the dorsal striatum, also known as the tail of the striatum. The amygdala, hippocampus, and the

olfactory/piriform areas also showed significant increases in activity, of comparable magnitude. Subregions of the hippocampus were variably activated by SOR, with the strongest increase 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.

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 to stimulation. NEUROeSTIMator can be used to rapidly identify and prioritize subsets of neurons showing transcriptional evidence of a stimulus response. In tests of predictive performance and generalizability, we demonstrate that the neuronal activity score can robustly detect signatures of activation from multiple types of stimulation, neuron subtypes, species, and sequencing technologies, including spatial transcriptomics.

To gain an understanding of the genes most influential in the model, we systematically perturbed expression of input genes and evaluated the effect on predicted activity. We found broadly distributed signal across the transcriptome, enriched for genes related to BDNF/NTRK signaling, circadian rhythm, and nuclear receptor pathways. The genes most informative to our model and relevant to these gene sets were *Homer1*, *Bdnf, Ntrk2, and Jun*. These well-known activity response genes were not included as model targets based on our selection criteria, but their prominent influence on model predictions suggests our model utilizes information from coregulated genes and pathways to robustly estimate expression of target genes. Notably, we found few genes associated with lower activity score relative to genes whose expression was associated with a higher score, suggesting model predictions largely rely on positive indicators of activity. Although there is evidence of activity-dependent downregulation of gene expression[25,26], most genes differentially expressed by neuronal activity are transcriptionally upregulated, which is supported by findings that neuronal activity increases genome-wide chromatin accessibility[1,27,28].

We observed a positive relationship between mean expression levels and gene influence. The most influential target genes, *Egr1* and *Nr4a1*, were the most highly expressed targets. We also observed that predicted activity is influenced by several known cell-type markers, which tend to be highly expressed. We suspect these observations are driven by reliability of gene detection at lower sequencing depths. Highly expressed markers of activity or cell type have greater detection rate, at both deep and shallow sequencing depths, than weakly expressed genes and therefore, are more reliable markers. We did not initially expect cell-type markers to strongly influence model predictions, and we suspect that cell type markers exert influence on predicted activity by reliably providing the model with information about cell identity, thereby allowing the model to establish cell-type specific intercepts for target gene expression that represent basal expression. Notably, the within cell-type heterogeneity of predicted

activity that we observe suggests that these cell-type markers are not sufficient by themselves to lead to predictions of neuronal activation.

We examined gene set annotations related to the top genes influencing predicted activity and identified circadian rhythm and BDNF signaling as key pathways in predicting neuronal activation. *Per1*, a target of our model, is a circadian regulator gene upregulated by neuronal activity[29]. It has been demonstrated that disrupted activity-dependent binding of CREB to CREB-binding protein (CBP) impairs long-term memory in mice and blunts the transcriptional upregulation of immediate early genes and circadian rhythm genes[30]. Together with our results, this suggests a subset of the activity response is allocated to a group of genes regulating circadian rhythm, and our model extracts this information from the transcriptome to predict expression levels of activity-dependent genes. Multiple significant gene sets were related to BDNF/NTRK signaling. *Bdnf* is a well-established activity response gene and one of the most extensively studied regulators of synaptic plasticity[31-34]. One of the most influential genes with membership in several significant gene sets was *Homer1*. Synaptic plasticity induced by neuronal activity has been shown to remodel synaptic scaffolding proteins[35], in part through regulation of *Homer1*[36]. Together, these findings suggest our model predicts activity, in part, by leveraging gene coexpression networks that interact with the immediate early gene activity markers.

We applied our model to a single cell dataset containing neurons subjected to seizure in the medial amygdala and demonstrated the ability of our model to predict increased neuronal activation in response to PTZ in multiple cell types. We next applied our model to a single cell experiment treating the rat striatum with either saline or cocaine. Our estimates of activity recapitulate a key finding from the source study, that Drd1+ and Drd3+ medium spiny neurons display a strong activation response to cocaine treatment. This finding is particularly noteworthy given that our model was trained using only mouse and human cortical and hippocampal neurons, none of which were medium spiny neurons.

Although we included human single nuclei data in the model training process, several neuronal subclasses from human samples in our test set showed low levels of predicted activity compared to the corresponding subclass in mouse samples. Although we reasoned this could be due to the nature of transcriptional machinery shutting down and RNA degradation in post-mortem neurons, it was not clear whether our model had inappropriately learned to equate human gene expression signatures to low neuronal activity. We applied our model to experimental data from human cell lines exposed to a time course of depolarizing KCl treatment. Our model detected a sharp increase in activity following 1 hour of treatment across multiple cell types, suggesting the model is indeed capable of identifying human signatures of neuronal activity. Further, our model predicted gradually declining levels of activity after 1 hour of KCl treatment, suggesting it is capable of discerning activity signatures beyond a simple on-off model of transcriptional activation. Not only do these results demonstrate our model can identify potent pharmacologically induced forms of neuronal activity, but it can also robustly do

so across species and cell types and discern gradual temporal changes over a time course of treatment.

Sensory experience is known to induce activity-dependent gene expression programs in cortical neurons. We demonstrated that the predicted activity is markedly higher in visual cortex neurons from mice exposed to light, as compared to controls. Further, we identified strong increases in predicted activity for non-neuronal cell types responding to light exposures. We specifically demonstrated that the activity score can classify visual cortex astrocytes as originating from light-exposed experimental groups with accuracy comparable to neurons. This finding was unexpected, as the model was not trained with any glial or other non-neuronal cell types. However, as the immediate early gene markers of activity are, in fact, markers of activity in many cell types, even beyond the brain, we anticipate that our model may be capable of detecting activation signatures in entirely different cell types such as immune cells, for example. Not only does this analysis provide a further line of evidence for astrocytic responsiveness to light in the visual cortex, it also demonstrates the ability of our model to detect such a response via transcriptomic data.

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 spatial object recognition 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 hippocampus, the caudoputamen region of the dorsal striatum, the retrosplenial area, and piriform areas. The CA1 region of the hippocampus has been shown to play a role in long term spatial memory in rodents[37-39], and the caudate nucleus has been demonstrated to play a role in spatial working memory in both monkeys and humans. These regions, particularly CA1, retrosplenial area and the caudoputamen have known involvement in spatial learning and working memory. The piriform area is involved in olfaction, which may reflect sensory processing involved in long term memory encoding. 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 activation could be identified in relation to other cognitive processes.

NEUROeSTIMator provides the first robust and generalizable means to quantify neuronal activation from gene expression data, opening the door to widespread inclusion in molecular neuroscience research. In neuroscience research involving gene expression, and especially in novel approaches like single cell or spatial transcriptomics, neuronal 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 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[18], 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[40], GSE103976[7], GSE136656[22], and GSE137763[21].

Datasets for Model Application – Spatial Transcriptomics

We generated a novel spatial transcriptomic dataset examining the effects of spatial object recognition (SOR) training in mice. The dataset contains spatial RNA-sequencing of whole brain slices from 1 hour after SOR training or home cage controls. SOR training was performed as previously described[24]. Mouse brain section per mouse was cut at 10 µm thickness and mounted onto each Visium slide capture area. After H&E staining, each bright-field image was taken as described in the spatial transcriptomics protocol. Tissue permeabilization was performed for 18 minutes, as established in the tissue optimization assay. The Visium Spatial Gene Expression Slide & Reagent kit (10x Genomics) was used to generate sequencing libraries for Visium samples. Libraries were constructed according to the 10x Visium library construction protocol and sequenced by Illumina NovaSeq6000. Raw data was then processed using the 10x Genomics Space Ranger analysis pipeline. See **Supplemental Figure S2** for images of predicted activity for each replicate.

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[41,42]. 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[43]. 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)[44]. 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[28].

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[45]. 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 S1**) and identified genes with p-values below these thresholds in both sets with concordant direction of effect[46]. 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

Raw counts were downsampled using the R package scater[47]. 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[17]. Briefly, input gene expression is supplied to an encoder; a series of three fully connected dense layers with ELU activations and batch normalization. The first layer contained 16 units and each successive layer halved the units of the previous layer. The encoder then connects to the information bottleneck, a single-unit dense layer with sigmoid activation. The bottleneck then connects to the first output, the estimated mean parameter **μ** of the zero-inflated negative binomial model. Two additional, independent, encoder branches output estimates of the dispersion and dropout parameters theta and pi, respectively. The zero-inflated negative binomial (ZINB) loss function was used, as implemented in DCA. 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, which were synthetically downsampled to simulate lower sequencing depths, were given the same output as the original data to curtail the learning of depth-dependent information. Sample losses were weighted to improve representation of rare cell subclasses but were limited to be no more than five times greater than they would be in an equally weighted scheme.

Evaluating Model Performance

Model performance was evaluated using a test split that was entirely shielded from model training or selection. 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 the integrated gradients[19] approach. Integrated gradients were averaged for each of the 4 species-by-class groups, and then averaged again to allow the gradients from each species and cell class contribute equally to the final importance metric.

Testing Differences in Predicted Activity

For all datasets analyzed in figure 2, we used the Wilcox test. 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 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.

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, GSE103976, GSE137763, GSE136656, and GSE102827.

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

We also thank Mahesh Shetty and Utsav Mukherjee for their valuable contributions to discussions about this project.

FIGURES

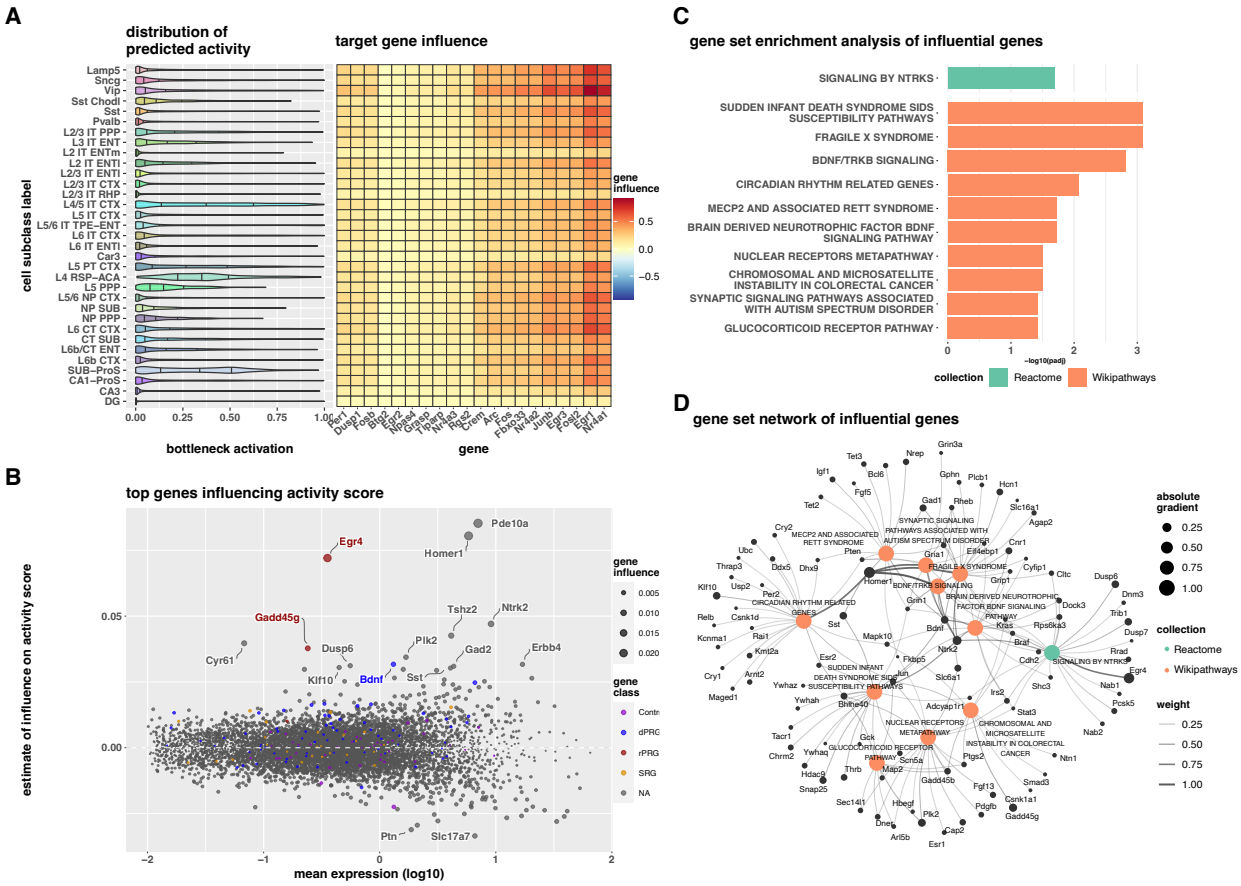


Figure 1 - **Gene-wise contributions to the activity score and its distribution within cell types of the training data.** The training data for the model was assembled from stimulus-naïve single-cell and single-nucleus experiments, and the proportion of active cells among each neuronal subclass varied, as indicated by the plotted distributions of bottleneck activation (i.e., the activity score, **A**). To understand the contribution of individual genes to the activity score, we examined gradient values from the DNN model (**B**). Higher positive gene gradient values indicate that increased gene expression is linked to increased bottleneck activation. To identify the strongest contributors to the activity score, we compared overall gene influence on predicted activity (X-axis) versus influence on model predictions (Y-axis) (**C**). Genes are colored by class as reported by Tyssowski, et. al, 2018: Rapid primary response genes (rPRGs, red), delayed primary response genes (dPRGs, blue), and secondary response genes (sRG, yellow).

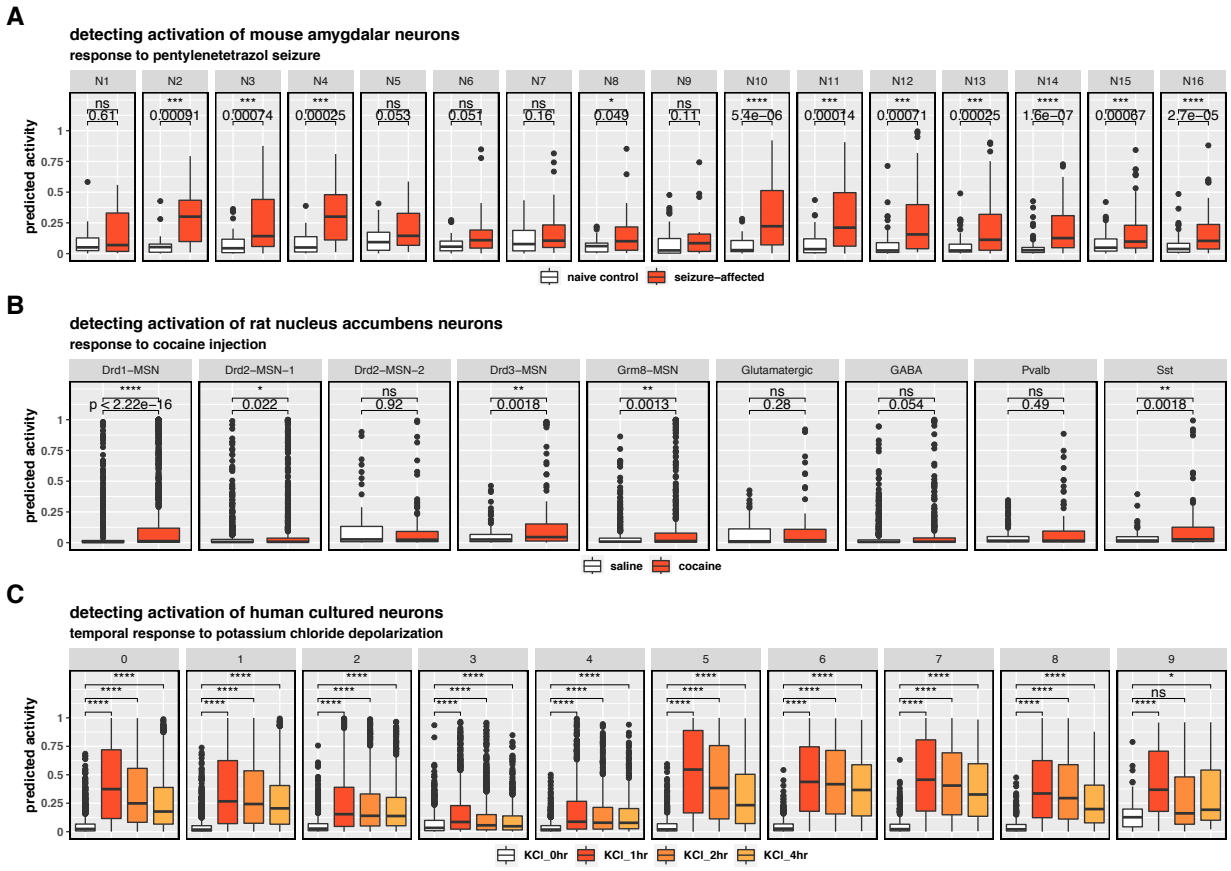


Figure 2 – Multi-species generalization of neuronal activity score applied to previously published chemical induction studies. Predicted activity for various amygdala neuron subtypes (mouse) stimulated with PTZ (red) or controls (gray) (**A**). Cell-type specific activation predicted for rat neurons of the nucleus accumbens treated with either cocaine (red) or saline (gray) (**B**). Time series of predicted activation of human GABAergic-like iPSCs treated with depolarizing potassium chloride at 0 hours, 1 hour, 2 hours, and 4 hours (**C**).

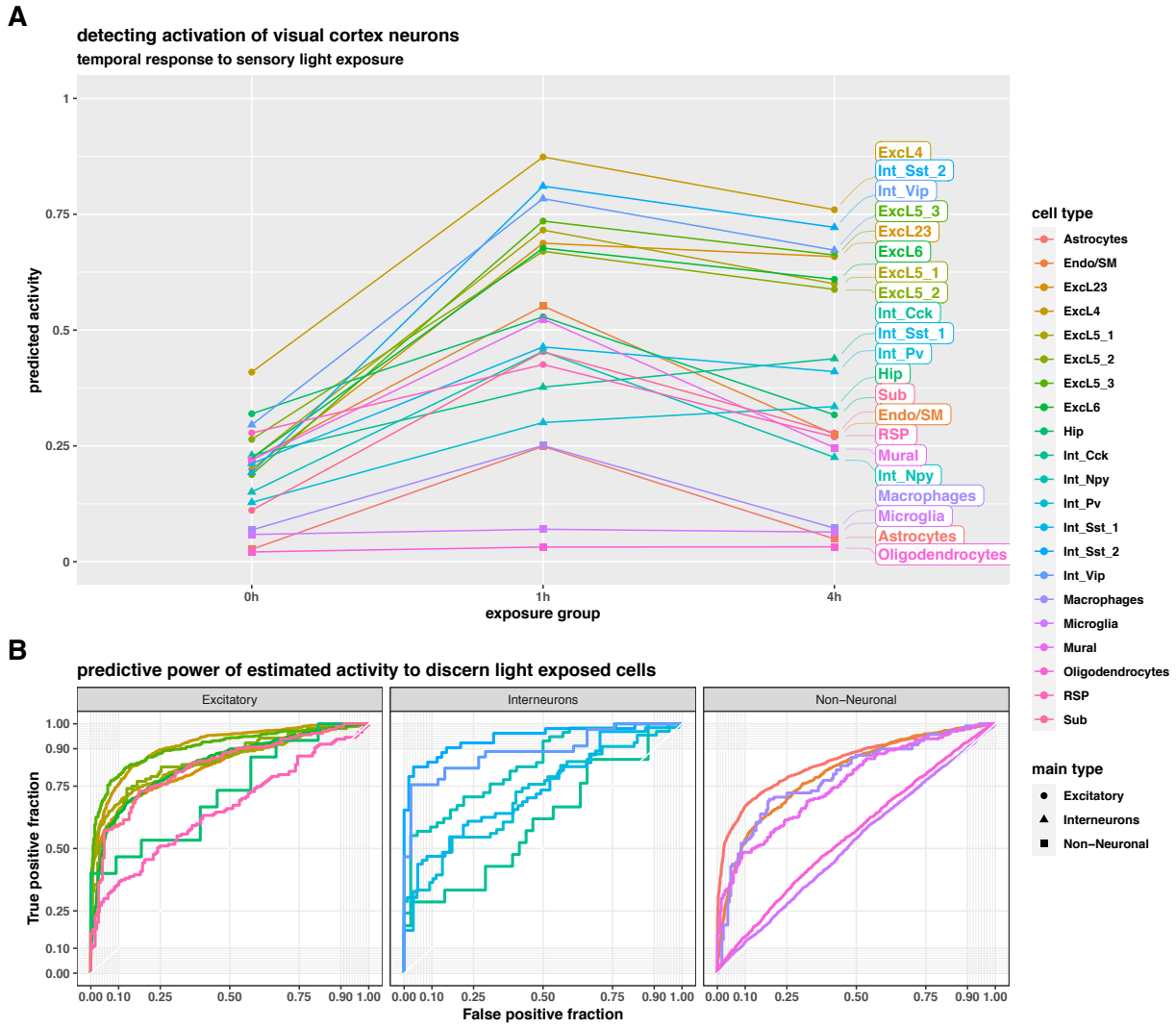
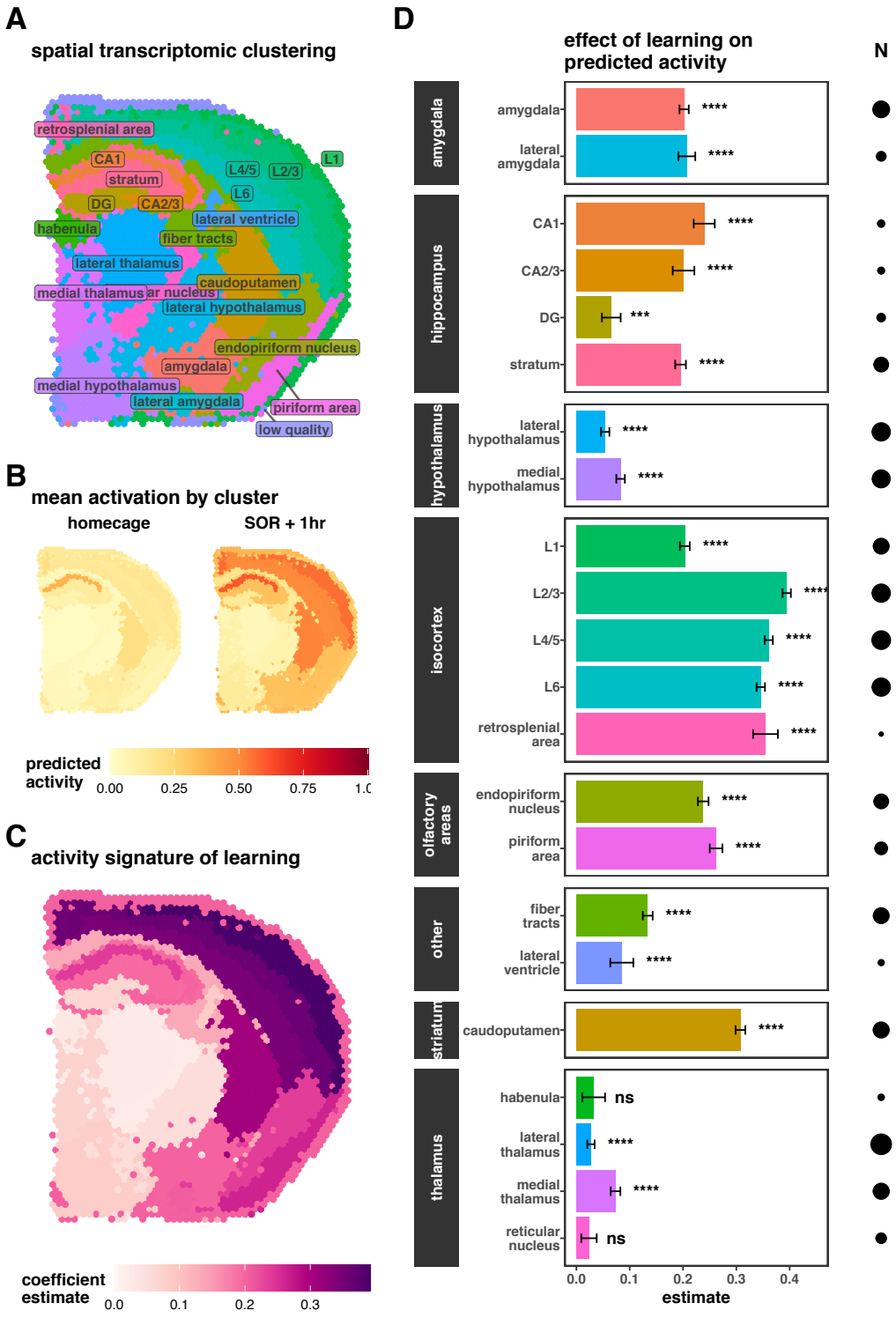


Figure 3 - Temporal patterns and classification of *in vivo* sensory activation. Cell type activity predictions of visual cortex neurons in freely behaving mice exposed to light for 0, 1, or 4 hours (**A**). ROC plots indicate the ability of predicted activity to separate various cell types into 0h vs 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.



615

616 **Figure 4 – Spatial transcriptomic patterns of neuronal activation after spatial learning.**
617 Spatial anatomical clustering of RNA-sequencing spots (**A**). Regions were labeled by comparing
618 transcriptionally-defined clusters to the Allen coronal mouse brain atlas. Activity score per spot,
619 averaged across experimental groups (**B**) of home cage controls (left) and 1 hour after spatial

object recognition (SOR) training (right). Using the spatial anatomical clustering derived from the expression data, we were able to group individual spots into clusters to test the significance of activity induction, here indicated by the cluster-wise coefficient estimate. Brain regions differentially activated by SOR training (**C**). Cluster-wise differential activity statistics are summarized in (**D**) and provided in **Supplementary Table S1**. Bar length represents estimated effect of SOR on activity score, based on linear models. Brackets indicate standard error and circle size represents the number of spots per comparison, which indexes statistical power. Non-significant regions with an adjusted p-value > 0.05 are colored grey.

629
630

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)
GSE103976	model application	mouse	PTZ	amygdala
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

Table 1. Datasets used.

631

References

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

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, doi:10.1146/annurev.cellbio.24.110707.175235 (2008).

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

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, doi:10.3389/fnbeh.2018.00079 (2018).

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

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, doi:10.3389/fnmol.2015.00042 (2015).

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, doi:10.1016/j.neuron.2017.09.026 (2017).

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, doi:10.1016/j.molcel.2017.11.017 (2017).

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, doi:10.1038/nmeth.4292 (2017).

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, doi:10.1093/biostatistics/kxx053 (2018).

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

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

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

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

15 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>.

16 Bao, S. *et al.* Deep learning-based advances and applications for single-cell RNA-sequencing data analysis. *Brief Bioinform* **23**, doi:10.1093/bib/bbab473 (2022).
17 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, doi:10.1038/s41467-018-07931-2 (2019).
18 Yao, Z. *et al.* A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. *Cell* **184**, 3222-3241 e3226, doi:10.1016/j.cell.2021.04.021 (2021).
19 Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. arXiv:1703.01365 (2017). <https://ui.adsabs.harvard.edu/abs/2017arXiv170301365S>.
20 Malik, A. R. *et al.* Cyr61, a matricellular protein, is needed for dendritic arborization of hippocampal neurons. *J Biol Chem* **288**, 8544-8559, doi:10.1074/jbc.M112.411629 (2013).
21 Savell, K. E. *et al.* A dopamine-induced gene expression signature regulates neuronal function and cocaine response. *Sci Adv* **6**, eaba4221, doi:10.1126/sciadv.aba4221 (2020).
22 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, doi:10.1038/s41593-020-00786-1 (2021).
23 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, doi:10.1016/j.bbrc.2018.10.027 (2018).
24 Chatterjee, S. *et al.* Endoplasmic reticulum chaperone genes encode effectors of long-term memory. *Sci Adv* **8**, eabm6063, doi:10.1126/sciadv.abm6063 (2022).
25 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, doi:10.1523/JNEUROSCI.4473-08.2009 (2009).
26 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, doi:10.1126/science.270.5240.1369 (1995).
27 Koberstein, J. N. *et al.* Learning-dependent chromatin remodeling highlights noncoding regulatory regions linked to autism. *Sci Signal* **11**, doi:10.1126/scisignal.aan6500 (2018).
28 Fernandez-Albert, J. *et al.* Immediate and deferred epigenomic signatures of in vivo neuronal activation in mouse hippocampus. *Nat Neurosci* **22**, 1718-1730, doi:10.1038/s41593-019-0476-2 (2019).
29 Kwapis, J. L. *et al.* Epigenetic regulation of the circadian gene Per1 contributes to age-related changes in hippocampal memory. *Nat Commun* **9**, 3323, doi:10.1038/s41467-018-05868-0 (2018).
30 Chatterjee, S. *et al.* The CBP KIX domain regulates long-term memory and circadian activity. *BMC Biol* **18**, 155, doi:10.1186/s12915-020-00886-1 (2020).

31 Kida, S. & Serita, T. Functional roles of CREB as a positive regulator in the formation and enhancement of memory. *Brain Res Bull* **105**, 17-24, doi:10.1016/j.brainresbull.2014.04.011 (2014).

32 Horvath, P. M., Chanaday, N. L., Alten, B., Kavalali, E. T. & Monteggia, L. M. A subthreshold synaptic mechanism regulating BDNF expression and resting synaptic strength. *Cell Rep* **36**, 109467, doi:10.1016/j.celrep.2021.109467 (2021).

33 Castren, E. & Monteggia, L. M. Brain-Derived Neurotrophic Factor Signaling in Depression and Antidepressant Action. *Biol Psychiatry* **90**, 128-136, doi:10.1016/j.biopsych.2021.05.008 (2021).

34 Wang, C. S., Kavalali, E. T. & Monteggia, L. M. BDNF signaling in context: From synaptic regulation to psychiatric disorders. *Cell* **185**, 62-76, doi:10.1016/j.cell.2021.12.003 (2022).

35 Wiesner, T. *et al.* Activity-Dependent Remodeling of Synaptic Protein Organization Revealed by High Throughput Analysis of STED Nanoscopy Images. *Front Neural Circuits* **14**, 57, doi:10.3389/fncir.2020.00057 (2020).

36 Kim, S., Kim, H. & Um, J. W. Synapse development organized by neuronal activity-regulated immediate-early genes. *Exp Mol Med* **50**, 1-7, doi:10.1038/s12276-018-0025-1 (2018).

37 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, doi:10.1016/s0092-8674(00)81827-9 (1996).

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

39 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, doi:10.1101/lm.027847.112 (2013).

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

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

42 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, doi:10.1038/nprot.2009.97 (2009).

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

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

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

46 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, doi:10.1093/nar/gkq636 (2010).

47 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, doi:10.1093/bioinformatics/btw777 (2017).