TissueFormer: a neural network for labeling tissue from grouped single-cell RNA profiles

Ari S. Benjamin[1] **and Anthony Zador**[1]

***Correspondence:**
benjami@cshl.edu
(AB)

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724

Abstract Single-cell RNA sequencing technologies have enabled unprecedented insights into gene expression and are poised to transform clinical diagnostics. At present, most computational approaches for interpreting single-cell data operate at the level of individual cells, predicting labels or properties based on isolated transcriptomic profiles. This approach overlooks a key class of signals: the composition of cells within a sample or defined population. Such signals are often critical for inferring tissue identity, disease state, or other sample-level phenotypes. To address this limitation, we introduce TissueFormer, a Transformer-based neural network that analyzes groups of single-cell RNA profiles to infer population-level labels while retaining single-cell resolution. Applied to predict the cortical area of groups of cells sampled from spatial transcriptomic data from mouse brains, TissueFormer outperformed both single-cell foundation models and machine learning methods applied to pseudobulk and cell type composition. This higher performance enables the automated construction of high-resolution brain region maps in individual animals directly from spatial transcriptomic data. More broadly, TissueFormer provides a framework for predicting any population-level phenotypes which are influenced by cellular diversity and tissue-level organization.

Introduction

Many clinically and biologically important features such as disease state, tissue identity, or treatment response are defined at the level of cell populations but are driven by patterns of single-cell phenotypes. Such features are often predictable from the composition and organization of cells in a population. For example, functionally distinct areas in the cerebral cortex can be distinguished by the relative abundance of cell types (***Yao et al., 2021b***). In clinical medicine, methods for detecting imbalances in cell type proportions such as complete blood count (CBC) tests form the basis of key diagnostics. For example, a high neutrophil-to-lymphoctye count is a widely used marker of inflammation (***Templeton et al., 2014***). Despite their clinical utility, readouts depending on cell composition remain limited to a handful of canonical patterns identified through decades of observation rather than discovered systematically through large-scale, data-driven analyses.

With the increasing availability of single-cell RNA sequencing, it is now possible to measure patterns in single-cell properties at high resolution across large numbers of cells per sample (***Svensson et al., 2018***). This rich granularity presents a tradeoff for the analysis of phenotypes across tissues, samples, or individuals, each of which contains many sequenced cells. At one extreme, many computational pipelines focus on single cells in isolation, taking the a cell profile as input and producing a label specific to that cell. This is the approach taken by most recent single-cell 'foundation models' of mRNA transcription data (***Lopez et al., 2018***; ***Connell et al., 2022***; ***Cui et al., 2023***; ***Theodoris et al., 2023***; ***Rosen et al., 2023***;

Yuan et al., 2024; ***Schaar et al., 2024***; ***Ito et al., 2025***; ***Hsieh et al., 2024***), reviewed in (***Szałata et al., 2024***). These models learn rich single-cell representations by first 'pretraining' with an unsupervised task on large datasets before fine-tuning on specific tasks, often resulting in improved performance when adapted for tasks such as cell type classification. However, these models are fundamentally single-cell in design, and as result they are unable to detect signals that emerge only from the *composition* of cells present.

An opposing strategy is to first compute summary statistics about cell populations across tissues or samples, then use these for downstream analysis and classification. One such statistic is the average expression, or psuedobulk expression profile, as is frequently used in differential state analysis (***Crowell et al., 2020***). Alternatively, one may calculate compositional features about a population, such as cell type frequencies, ratios of cell types, putative cell-cell communication networks, and cell type specific pathway scores (***Cao et al., 2022***). While informative, both psuedobulk and cell composition measures obscure single-cell granularity and rely on the researcher to specify which compressed representation of a population of cells is likely to be informative about sample labels.

In order to maximize the potential of single-cell data, it is important that methods combine the transfer-learning capabilities and end-to-end trainability of single-cell foundation models with the ability to compare cells across a population. Certain models exist which are end-to-end trainable and can compare across cells, such as CellCnn (***Arvaniti and Claassen, 2017***), scAGG (***Verlaan et al., 2025***), and ScRAT (***Mao et al., 2024***), but these models are not compatible with existing foundation models. The ability to incorporate knowledge from pretrained models is especially important given the high dimensionality of the inputs when predicting sample-level phenotypes. In each independent sample, the input dimensionality is on the order of the number of cells per sample times the number of genes per cell. By first pretraining on self-supervised tasks, foundation models effectively reduce this dimensionality by establishing a learned prior over the representation of each cell.

To bridge these approaches we present TissueFormer, a neural network architecture that analyzes groups of single-cell RNA profiles collectively while utilizing knowledge embedded in pre-trained single-cell foundation models. The model incorporates a module based on the Geneformer architecture for single-cell analysis (***Theodoris et al., 2023***), which can be initialized with pre-trained weights. During processing, each cell is given a learnable representation with this pretrained module. These cell representations are then processed by several further layers of self-attention, allowing the model to learn to attend to the relevant cells and their interactions. Importantly, the output of Tissueformer is invariant to the order of cells presented as input. These design choices allow TissueFormer to learn arbitrary functions from observations of many cells to sample-level phenotypes while simulateously leveraging pretrained single-cell foundation models.

To validate our approach, we applied TissueFormer to apply Allen Brain Atlas labels to regions of the mouse cortex profiled with spatial transcriptomics. This task is a competitive benchmark for the supervised classification of sampled tissues, though it is important to note that clustering or alignment-style methods have also been developed for tissue annotation (e.g. ***Biancalani et al.*** (***2021***); ***Hu et al.*** (***2021***); ***Lee et al.*** (***2025***)). We found that TissueFormer outperforms both standard supervised methods as well as fine-tuned single-cell foundation models at predicting the labels due to previous brain annotations. Furthermore, the performance of TissueFormer scaled roughly logarithmically with the number of cells input as a group, demonstrating the utility of observing multiple cells. These results highlight the importance of sample- or tissue-level features in transcriptomic analysis and establish a framework for integrating single-cell foundation models into higher-order biological investigations.

Results

Our primary goal is to design and evaluate a model which is generally applicable to the supervised problem of sample or tissue annotation from single-cell data. In what follows, we first describe the particular characteristics of this problem that make it challenging for standard methods.

Multi-cell supervised problem setting

A typical problem in single cell analysis is relating mRNA expression levels in each cell to a cellular property such as cell type identity (***Figure 1***a). If we denote the cell-by-transcript matrix of counts as $\boldsymbol{C}$, where each row of $\boldsymbol{C}_i$ contains the mRNA transcript counts for a single cell, this problem seeks to find a function f that maps each row $\boldsymbol{C}_i$ (a cell) to a particular value y_i corresponding to a label. For example, y_i could represent the cortical area from which a given cell was obtained.

$$f(\boldsymbol{C}_i) = y_i \tag{1}$$

In some problem settings, it is advantageous to use information from many cells. In this multi-cell setting, the characteristic equation relates sets of cells to labels. For spatial transcriptomic data, for example, a set might refer to a group of cells within some spatial distance of one another. Denoting $\boldsymbol{G}_i$ as the ith such group, and $\{c \in \boldsymbol{G}_i\}$ as the set of cells in this group, the classification or regression problem seeks to find a function f that maps each set of cells to its label y_i.

$$f(\{c \in \boldsymbol{G}_i\}) = y_i \tag{2}$$

The fundamental difficulty of the multi-cell problem is its high dimensionality. Relative to the single-cell problem, the dimensionality of input examples increases from the number of genes $\boldsymbol{D}$ to $\boldsymbol{D} \times |\boldsymbol{G}|$, where $|\boldsymbol{G}|$ is the number of cells in a group. Meanwhile, the number of independent labels decreases from the total number of cells to the number of disjoint groups, i.e. from $\boldsymbol{N}$ to $\frac{\boldsymbol{N}}{|\boldsymbol{G}|}$. The ratio of input dimensions to independent examples, a proxy for the difficulty of the problem, thus increases as the group size squared, $|\boldsymbol{G}|^2$.

Most strategies for the multi-cell problem explicitly reducing its dimensionality through aggregation (as in psuedobulk analyses) or summary statistics (e.g. ***Cao et al.*** (***2022***)). However, these strategies destroy information which could be critical in multi-cell problems. Pseudobulk methods obscure the variance in the signal, whereas cell type composition or other such signals are fixed and non-adjustable representations of single-cell profiles and are incapable of observing transcriptional dysregulation within cell types. To take full advantage of single-cell data, a model should ideally be able to attend to any aspect of the full RNA transcriptome in each cell while comparing across the population.

It is nevertheless crucial take steps to constrain the distribution of learned functions so as to circumvent the curse of dimensionality in multi-cell prediction problems. One way to address this is by pre-training a foundation model on a self-supervised task before downstream supervised learning. If this self-supervised task is sufficiently similar to the supervised task at hand, the pretrained model weights will be close in weight space to a good solution for the supervised task. This reduces the number of training steps and data needed in the supervised task.

Recent related work has introduced several foundation models tailored for spatial transcriptomic data. Because these models observe groups of cells, they can be seen as solutions to a subcase of the multi-cell problem in Equation 2. For example, scGPT-spatial is pretrained to predict gene expression in held-out cells in the local neighborhood (***Wang et al., 2025***). Unlike TissueFormer, scGPT-spatial averages the embedding of several cells in the local neighborhood for the inter-cell problem, rather than employing self-attention across cells, and thus cannot compare compositional signals. HEIST (***Madhu et al., 2025***) and CI-FM (***You et al., 2025***) define graph neural networks over a hierarchical graph spanning both nearby

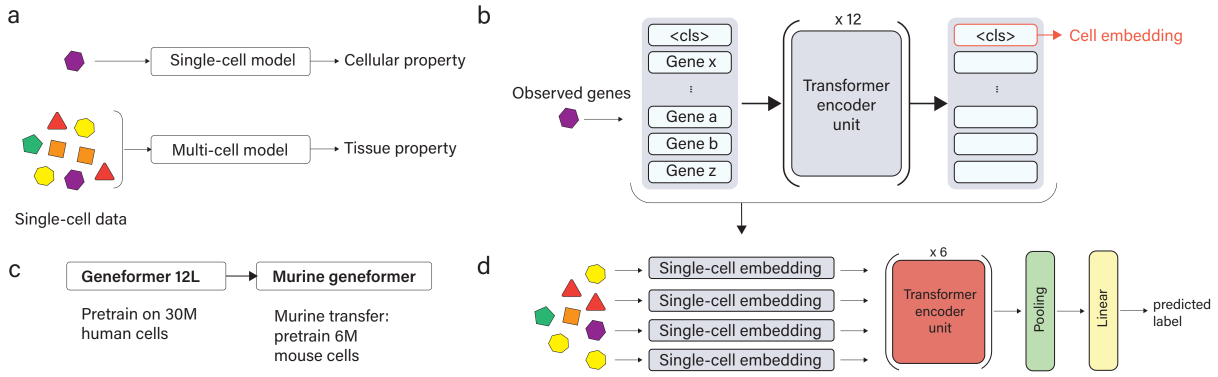

Figure 1. TissueFormer is a Transformer-based neural network to analyze groups of single-cell RNA profiles. a) Single-cell foundation models focus on cellular targets, yet many properties of tissues relate to cellular diversity. b) TissueFormer learns to extract critical information from each cell into a cell embedding vector. The module that extracts these is in architecture identical to Geneformer, a single-cell foundation model. First, cells are processed by a normalized rank-ordering of their genes into a 'sentence' of gene tokens, prefixed by a '<cls>' token to pool gene information, then processed by several transformer layers. c) Experiments in this manuscript use a pretrained single-cell model we call Murine Geneformer, created by adapting the 12-layer Geneformer model (pretrained on human cells) to mouse cells via pairing orthologous genes and further pretraining. d) TissueFormer architecture. Cell embeddings for all cells in a group are fed to 6 Transformer encoding units before average-pooling and a linear readout.

cells and their genes, which in principle enables comparative processing across cells. As explained below, TissueFormer uses Transformer layers and not graph neural networks, making it compatible with single-cell foundation models trained on traditional scRNA-seq data. Unlike these works, the design of TissueFormer is not restricted to spatial transcriptomic data and is applicable to any sample-level phenotype prediction problem.

TissueFormer architecture

TissueFormer is a neural network architecture tailored for groups of single-cell RNA transcriptomic profiles (***Figure 1***). It is able to attend to information within each cell's relative mRNA transcription profile, yet also look across cells to extract information from the diversity and frequency of mRNA profiles in the group. Architecturally, this is accomplished by first applying a 'single-cell module' to extract information from each cell, then attending across cells in further layers. This combination of intra- and inter-cell expressivity creates a regression problem of high dimensionality, yet this is mitigated by the bottleneck architecture and the easy loading of pre-trained single-cell foundation models into the cell embedding module.

The building block of TissueFormer which performs these capabilities is the Transformer block implementing self-attention (***Vaswani et al., 2017***). Compared to older neural network architectures like Multi-Layer Perceptrons (MLP), Transformers are distinguished by the shape of their inputs. Whereas an MLP observes vector-shaped inputs (i.e. a vector of mRNA transcript counts), Transformers operate on lists of vectors, or equivalently, on matrix-shaped inputs. Self-attention acts to select which vectors in the list are most relevant in the current context.

In order for single-cell data to be compatible with Transformers, each cell's vector of mRNA transcript counts c_i must be expanded into a list of vectors. Multiple strategies for this conversion are possible for single-cell data, as reviewed in ***Szałata et al.*** (***2024***). Here we use a rank-ordering approach as in Geneformer (***Theodoris et al., 2023***). Each cell's expressed genes are ordered by their relative expression normalized by their median expression in a pretraining corpus. Each gene name in this ordered list is then mapped to a high-dimensional embedding vector via a learnable dictionary. While alternatives to rank encoding have been tested in other foundation models, rank encoding is common in bioinformatic pipelines and has generally been found to be robust to common sources of experimental variability (***Ballouz et al., 2015***).

The single-cell module within TissueFormer is designed to compress each cell's unique and relevant information into a *cell embedding* for later processing (***Figure 1***b). This is achieved by prefixing to each cell a special token named the `<cls>` token, following standard design choices (***Devlin et al., 2019***). The `<cls>` token instantiates a working space for the accumulation of information from all genes. During learning, any error-corrective information passed backwards is bottlenecked through the cell embedding, ensuring it collects the necessary information from each cell. In sum, the cell embedding vector which is passed to later layers can be written as a function of the mRNA transcript count vector c_i as:

$$\textbf{CellEmb}[c_i] = \textbf{SingleCellModule[RankEncoding[}c_i\textbf{]]}[\text{<cls>}] \tag{3}$$

The single-cell module shares the Geneformer architecture, meaning that any trained single-cell model can be loaded into this module. This gives TissueFormer the ability to incorporate model weights and biological knowledge from pretrained single-cell foundation models. In this work, all experiments unless otherwise indicated use weights from the Geneformer 12L model, a 12-layer BERT network trained on a masked prediction objective on 30 million human cells (***Theodoris et al., 2023***). We then further pre-trained these weights using an identical objective function but on a collection of datasets of mouse tissue totaling 6 million cells (***Figure 1***c). To facilitate the transfer of knowledge across species, we initialized the gene tokens for mouse genes with their orthologous human genes where available. We call this resulting pretrained single-cell model 'Murine Geneformer'.

TissueFormer attends to information across cells by processing each cell embedding with several additional Transformer layers (***Figure 1***d). At this stage, TissueFormer is invariant to the order of cells presented to the model, seeing them only as a bag or set of cells. The output of any Transformer is inherently equivariant with respect to the order of the input embedding vectors (i.e. rearranging the inputs causes the outputs to be rearranged in the same order). In order to attain invariance to cell order, such that rearrangements to the order have no effect on the output, the output of the Transformer layers is pooled by averaging over cells. A final step is a linear classification layer for classification problems. Collectively, these design choices allow TissueFormer to attend to any relevant information in each cell's transcriptome in a context-dependent manner based on the diversity of cells present and the problem at hand.

Taken together, the equation representing TissueFormer can be summarized as:

$$y_i = \textbf{LinearClassifier[AvgPool[Transformers[}\{\textbf{CellEmb}[c_j] \text{ for } c_j \in G_i\}\textbf{]]]} \tag{4}$$

Cortical annotation

The mammalian brain is often partitioned into regions with distinct functional roles, connectivity, and cellular composition. In any particular brain, variations in neuroanatomy may arise due to genetic factors as well as due to the animal's environment. These variations underlie the evolution of innate behavior and reflect the aptitude for adaptation within a lifetime.

Current methods for the annotation of single brains have certain well-known limitations. One common approach is to register a brain into the average-brain template of the Common Coordinate Framework (CCF), a 3D reference coordinate system constructed from over 1,000 reference mouse brains (***Wang et al., 2020***). Once transformed into CCF coordinates, one can visualize the area boundaries of the Allen Brain Atlas, which again reflect a canonical, average brain. The drawback of this approach is that it obscures individual variation, especially the relative size of brain areas to one another. An alternative approach which enables individual variability is to employ experimental methods that interrogate only a small brain region in each animal—for example, focal viral projection tracing (***Xu et al., 2020***) or two-photon functional imaging for a particular functional modality like visual responses (***Kalatsky and Stryker, 2003***). While accurate, these methods' limited spatial scope prevents a holistic assessment of anatomy across the entire cortex.

An alternative strategy that balances whole-brain coverage with single-animal individuality is to perform *in situ* single-cell RNA sequencing across a single brain. This method modernizes a classic approach of defining neuroanatomy through differences in cell type composition, an idea that goes back at least to Brodmann's brain atlas (***Brodmann, 1909***; ***Zilles and Amunts, 2010***). Currently, the technical efficiency of spatial transcriptomics is now high enough to allow the profiling of cells across an entire mouse brain, and furthermore cheaply enough to cover multiple brains in single studies (***Chen et al., 2024***). Given such datasets, the remaining challenge in a robust pipeline for labeling brain areas is computational in nature.

To investigate whether TissueFormer could serve this purpose, we applied TissueFormer to predict the brain region of cells from their transcriptomes as characterized in a recent dataset of whole-brain spatial transcriptomics of the brains of several animals (***Chen et al., 2024***). Over 1 million cells across the brain, largely in the left hemisphere, were profiled in each of eight animals, four of which were raised in normal rearing conditions. Each cell profile contains counts from a panel of 104 genes selected for their ability to distinguish cell type clusters in excitatory cells. With multiple separate brains and whole-cortex coverage in one hemisphere, this dataset enables a test of whether individual differences in neuroanatomy can be resolved with spatial transcriptomics.

In addition to mRNA counts, this dataset also contains labels of the brain area of each cell as established through a registration into the Common Coordinate Framework (CCF). Each brain in ***Chen et al.*** (***2024***) was previously registered to the CCF using standard software that defines a non-rigid smooth deformation from raw slice coordinates into the CCF. This step normalizes differences in gross anatomy and brain size. Importantly, once transformed into CCF coordinates, one may associate the topography of a new brain with the Allen Brain Atlas to obtain brain area annotations (***Wang et al., 2020***). These CCF annotations provide the labels we use for training.

It is important to note that CCF boundaries may not reflect ground truth neuroanatomy. While CCF registration accounts for overall brain shape, it does not account for more fine-grained differences such as the relative size of brain areas or a shift in a single boundary in one animal versus another. Nevertheless, these labels are sufficiently accurate to allow a comparison of computational methods. It is possible as well that the 'errors' of a model trained on such data will in fact represent a more accurate annotation than the CCF, a possibility we will return in ***Figure 3*** in which we examine the predictions of TissueFormer on held-out brains.

Single-cell profiles are not informative of cortical area

The utility of attending across cells can be made clear by comparing with the performance of machine learning methods which are trained to predict the area given only a single cell. To investigate this, we first split the data into training data (90% of cells in 3 brains), validation data (the remaining 10% of the same 3 brains), and a held-out test brain (***Figure 2***b). We then trained a range of methods on the task of predicting the cortical area of a single cell from its transcriptome.

The methods tested ranged in complexity from simple heuristics to modern single-cell foundation models (***Figure 2***c). We included three standard machine learning benchmarks, namely, a logistic regression model, a random forest model, and k-neighbors classifier. These models map the vector of mRNA transcript counts to the label after a $\log(1+x)$ transformation and z-scoring, and their hyperparameters were tuned on the validation set. We also trained two models which make use of the cell types of the cell in question. These cell types were previously constructed at three levels of granularity (***Chen et al.*** (***2024***)); we selected the finest level. The 'cell type model' predicts the area of a cell based on the most common area of cells of the same type in the training set. The 'cell type k-neighbors' model labels the target cell based on the area of the k-closest neighbors with the same cell type in the training set. Finally, we fine-tuned Murine Geneformer to predict the area of each cell.

Despite the diversity of modeling approaches, all six models performed poorly, with classification accuracies ranging from 20–31%. This highlights the challenge of predicting cortical area from single-cell transcriptomes, consistent with previous findings that many cell types are broadly distributed across the cortex (***Tasic et al., 2018***; ***Yao et al., 2021b***; ***Chen et al., 2024***). While some types are more localized and thus were predicted with higher accuracy (for example, RSP/ACA or Retrosplenial/Anterior Cingulate excitatory cells), overall performance was low. Among the models tested, the Transformer-based Murine Geneformer achieved the highest accuracy, modestly outperforming the others. This is consistent with recent results showing strong performance of large pretrained models on downstream tasks (***Theodoris et al., 2023***). These results motivated our turn toward models that integrate information across cells.

Annotating cell groups with TissueFormer

To apply TissueFormer, we first created a pipeline to group cell sharing a label. Specifically, we grouped cells into cylindrical columns which fall inwards from the surface of the cortex (***Figure 2***d-e). Note that this strategy can easily be generalized to other shapes for spatial data from other tissues, or to other sampling strategies for non-spatial data. The cerebral cortex can be viewed as a layered cake draped over the exterior of the mouse brain, with functional areas as slices (***Figure 2***a). Thus, cortical columns contain cells in the same functional area. Our pipeline programmatically selects a single cell, then grows a cylinder centered on that cell up to the minimum radius needed to contain N total cells. Each columnar group of N cells is labeled with the most common CCF annotation of cells in that group.

We next trained TissueFormer to predict the group's label from the set of transcriptomic profiles within the group. We again cross-validated by training on 90% of cells from three brains, validating on the remaining 10% of cells, and testing on one held-out brain. As we varied the number of cells in each group from $N = 1$ to $N = 256$, we observed that the validation accuracy of TissueFormer increased from around 35% to nearly 80% accuracy (***Figure 2***f). On test brains, the accuracy Tissueformer also increased markedly with group size from 30% to over 60% accuracy (***Figure 2***g). Providing the cortical depth of each cell as additional input to the model did not improve results, suggesting that depth represents redundant information (***Figure 2—figure Supplement 1***f). Note that random guessing on this 1-of-42 classification task yields 8% accuracy due to imbalanced classes. Similar relative accuracies were seen when training with a class-balancing weighted objective (***Figure 2—figure Supplement 1***c). These results demonstrate that TissueFormer is able to dramatically outperform single-cell models at predicting tissue identity.

As the group size increased from $N = 1$ to $N = 256$, the accuracy increased smoothly and roughly logarithmically with group size, indicating a predictable scaling law with more cells. Performance saturated at $N = 128$ cells, possibly because at this scale and at this density of cells per brain the groups sometimes straddle straddle boundaries, at which point group identity begins to lose meaning. Indeed, at $N = 256$ over 70% of groups contained a boundary (***Figure 2—figure Supplement 1***g). Higher sampling densities would enable larger group sizes in smaller locales, and thus possibly an even higher saturating accuracy. Due to this problem of spatial density, it is not possible to verify the maximum group size at which cells contain no further information about area. Overall, the considerable increase of performance with group size highlights the importance of the ability to integrate and compare information across single cells.

The cell embedding module in the above experiment was previously pretrained on a masked mRNA

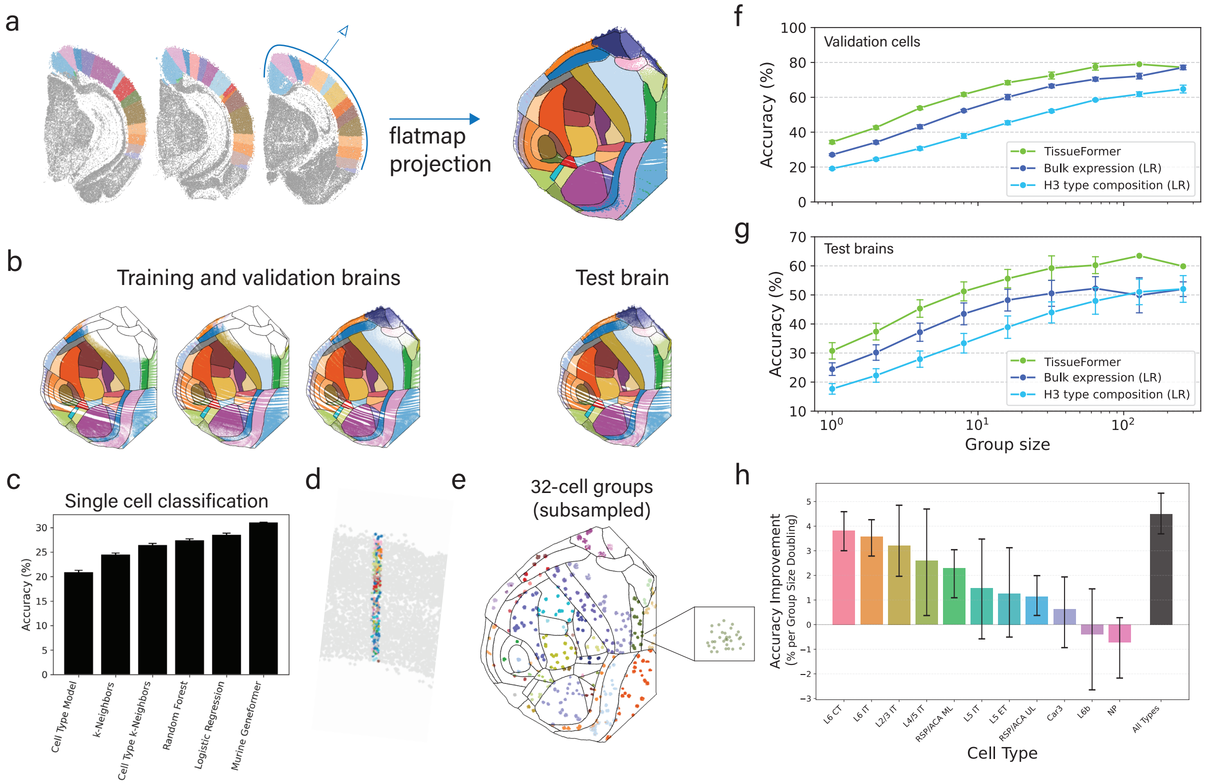

Figure 2. Comparison of accuracy at predicting brain regions shows that TissueFormer outperforms other methods. **a)** Cortical cells from coronal slices of one hemisphere (left, colored by area) are visualized as a 'flatmap', a top-down projection. Color legend is in ***Figure 3***.**b)** Models were evaluated using 4-fold crossvalidation with 3 or 4 brains used for training and validation labels, 1 of 4 brains held out for testing. Shown here are true labels for each cell. **c)** Comparison of accuracy when predicting the area of a single cell in a test brain from its transcriptome shows that all methods perform poorly, though with Murine Geneformer performing best. **d)** An example cell group ($N = 256$) plotted in slice coordinates, colored by cell type. **e)** A sample of groups ($N = 32$) plotted in flatmap coordinates, here colored by the modal area of the group. **f)** Validation accuracy as a function of group size shows that TissueFormer outperforms logistic regression (LR) given either pseudobulk transcription or a histogram of cell types. **g)** Test brain accuracy. Error bars in f-g represent standard deviation across 4 folds. **h)** The average rate of test accuracy improvement with group size for the above TissueFormer (black bar) compared to the accuracy curve slopes of several TissueFormer models trained on homogeneous groups containing only a single cell type. Accuracy curves are visible in supplement 1.

Figure 2—figure supplement 1. Effect of pretraining on TissueFormer, performance of random forest benchmarks, and controls for unequal area size and sampling density.

transcript prediction task on several other datasets containing tens of millions of human and mouse cells. In principle, this pretraining step could help, hinder, or have little effect on our current supervised task of predicting brain area. To investigate its impact, we also trained a TissueFormer model from a random initialization without any pretraining. Surprisingly, the performance of this randomly initialized model was indistinguishable from the pretrained model for all group sizes (***Figure 2—figure Supplement 1***d). This is likely due to the large number of cells in our labeled dataset. To confirm this, we varied the number of training cells used to train both a pretrained and a randomly initialized TissueFormer from one thousand cells to the full dataset of over two million. We found that pretraining indeed offered an advantage, but only in the intermediate range of less than one million cells (***Figure 2—figure Supplement 1***e). Thus, this spatial transcriptomic dataset is large enough that transfer learning from Murine Geneformer offers no advantage.

We next compared the performance of TissueFormer with two standard methods for processing transcriptomic data of groups of single cells. We first examined pseudobulk analyses, which average the single-cell profiles in the group to approximate the measurements of a traditional bulk sequencing experiment. Logistic regression trained on pseudobulk vectors underperformed TissueFormer, yet still showed a nearly logarithmic increase of accuracy with group size (***Figure 2***f). A random forest model trained on the same data underperformed logistic regression (***Figure 2—figure Supplement 1***b). Next, we trained logistic regression and random forests to predict area from the cell type composition of a group, which was represented as a histogram of cell types. We used the finest-grain categorization of cell types from ***Chen et al.*** (***2024***) for this analysis. Overall, type composition was less informative of area than pseudobulk expression (***Figure 2***f). Intriguingly, each method displayed a logarithmic increase in accuracy with group size despite the differences in their representations of RNA transcription.

The logarithmic increase in accuracy with group size across all methodologies could in principle be driven by two effects. One possibility is that the diversity and particular composition of cells provided a new signal not available to single cell models. An alternative possibility is that cells contained independent measurement noise which can be averaged away. Some evidence to the first possibility is that benchmarks based on cell type composition alone showed such an increase (***Figure 2***f-g). To further distinguish these factors within TissueFormer, we reasoned that we could artificially restrict the diversity of cells in each group and then re-test a method's performance. If the scaling were due to measurement noise alone, then this would have little effect and performance would likewise increase with a similar slope on a log-linear plot. We therefore constructed a new method of constructing groups such that they contain only a single cell type from the second level of granularity in the cell type hierarchy ("H2" types). Note that there is still significant diversity within each H2 type, albeit much less than the general population. After training a TissueFormer model from scratch on these homogeneous groups, we found that the test set accuracy given a group of homogeneous type scaled more slowly with increasing group size N, on average (***Figure 2***h). Groups of cells of certain types such as NP cells showed little improvement with larger size. Other types, such as L6 IT cells, nevertheless show comparable increases, likely indicating higher and more informative diversity in that population. The areas of some cell types are more easily predictable than other types due regional localization (***Figure 2—figure Supplement 1***h). These analyses confirmed that the increase in accuracy with group size was at least partially driven by comparative signals across a cell group rather than solely by averaging away technical noise.

The decrease in accuracy between validation and test brain accuracy could have arisen due to multiple reasons. A first reason is the differences in spatial coverage of data from each brain, visible in ***Figure 2***b, which may cause test brains to contain areas not in the training data. However, correcting for this effect by only testing data with high density in the training set yielded improvements in accuracy of at most a few percentage points (***Figure 2—figure Supplement 1***a). Secondly, this could have reflected other batch effects across brains affecting single-cell measurements, effectively making this test brain an out-of-distribution test for the methods. Finally, it is also possible that 'errors' in classification were due to actual changes in neuroanatomy which were correctly predicted by the model but were mislabeled by the pipeline of registering each brain to the CCF. We investigate this possibility in the following section.

Predicted cortical maps

TissueFormer can be used as an automated pipeline to create maps of cortical anatomy from single-cell transcription after being trained on reference brains. Note that annotating from reference brains is a supervised task, not to be confused with the unsupervised task of spatially clustering single-cell data (see Discussion). In ***Figure 3***a, we demonstrate this capability and display the predicted cortical maps for each of 4 animals treated as a held-out test set. Pixels in these maps were colored by a weighted average of nearby cells' predictions (see Methods). For visualization we used the predictions of TissueFormer trained with a class-size-balancing objective ensuring that the prior probability was uniform across all areas regardless of area size. We found that all four maps were coarsely similar to the reference annotation (see ***Figure 2***a), supporting the finding in ***Figure 2*** that TissueFormer is able to predict brain area more accurately than other approaches.

A close inspection of these cortical maps revealed several notable differences in the predicted anatomy of each animal. For example, the boundary between the somatosensory cortex and the motor cortex is consistently shifted in the posterior-lateral direction in brain 1 relative to the Common Coordinate Framework (CCF) annotations, yet is consistently shifted in the opposite (anterior-medial) direction in brain 2 (***Figure 3***b). Additionally, the primary visual area is shifted medially in brain 2, yet laterally in brains 3 and 4. Likewise, the primary auditory area is shifted medially in brain 2 yet is expanded and shifted laterally in brain 4. These shifts are a predominant cause of why test brain 'accuracy' in ***Figure 2***g is lower than validation accuracy.

While it is possible that the discrepancies between the CCF annotation and the predicted annotation are 'errors' of the model, it is also possible that this reflects true differences in anatomy between animals or errors in registration with the CCF. The CCF represents the average anatomy of over 1,000 mouse brains, obscuring potential individual variability between brains. To investigate this possibility, we examined the somatosensory-motor boundary in more detail. This boundary can be canonically identified by differences in cell type composition. In particular, motor cortex is classically considered to lack a Layer 4 (***Brodmann, 1909***), and thus should show a large decrease in the density of Layer 4/5 IT excitatory cells relative to Layer 5 IT cells. This easily verifiable boundary was qualitatively visible in coronal slices colored by cell type (***Figure 3***c). To confirm its location, we examined the density of these two cell types along slices taken from brain 1 and brain 2 which intersect the sensory/motor boundary (***Figure 3***d). We found that the CCF boundary was inconsistent across brains relative to the ratio of cell types, but TissueFormer's predicted maps showed a consistent 1:1. Furthermore, in the CCF boundaries, we observed a large shift between animals in the ratio for large distances on either side (***Figure 3***e). In contrast, the density ratios were consistent in their alignment to the predicted boundaries. Thus, a closer inspection of cell type distributions around this verifiable boundary revealed a closer alignment with our predicted atlas than with the boundaries due to CCF alignment.

To provide further verification of the predicted cortical maps, we examined the spatial distributions of individual genes known to correlate with area identity (***Figure 3—figure Supplement 2***). For example, Tshz2 forms a striking medial-to-lateral gradient with high abundance in anterior- and posterior-cingulate/retrosplenial areas with steep drops at the boundaries to motor, somatosensory and visual fields (***Yao et al., 2021b***). Across all four brains, the inter-animal differences in Tshz2 aligned better with the predicted boundary than the CCF area boundaries. Other genes which are also selectively expressed in retrosplenial areas, such as as Coro6 and Zfpm2 (***Chen et al., 2022***), shared this pattern, as did genes with selectively expressed in the dorsal but not the ventral retrosplenial area, such as Nell1 and Zmat4 (***Hashikawa et al.,***

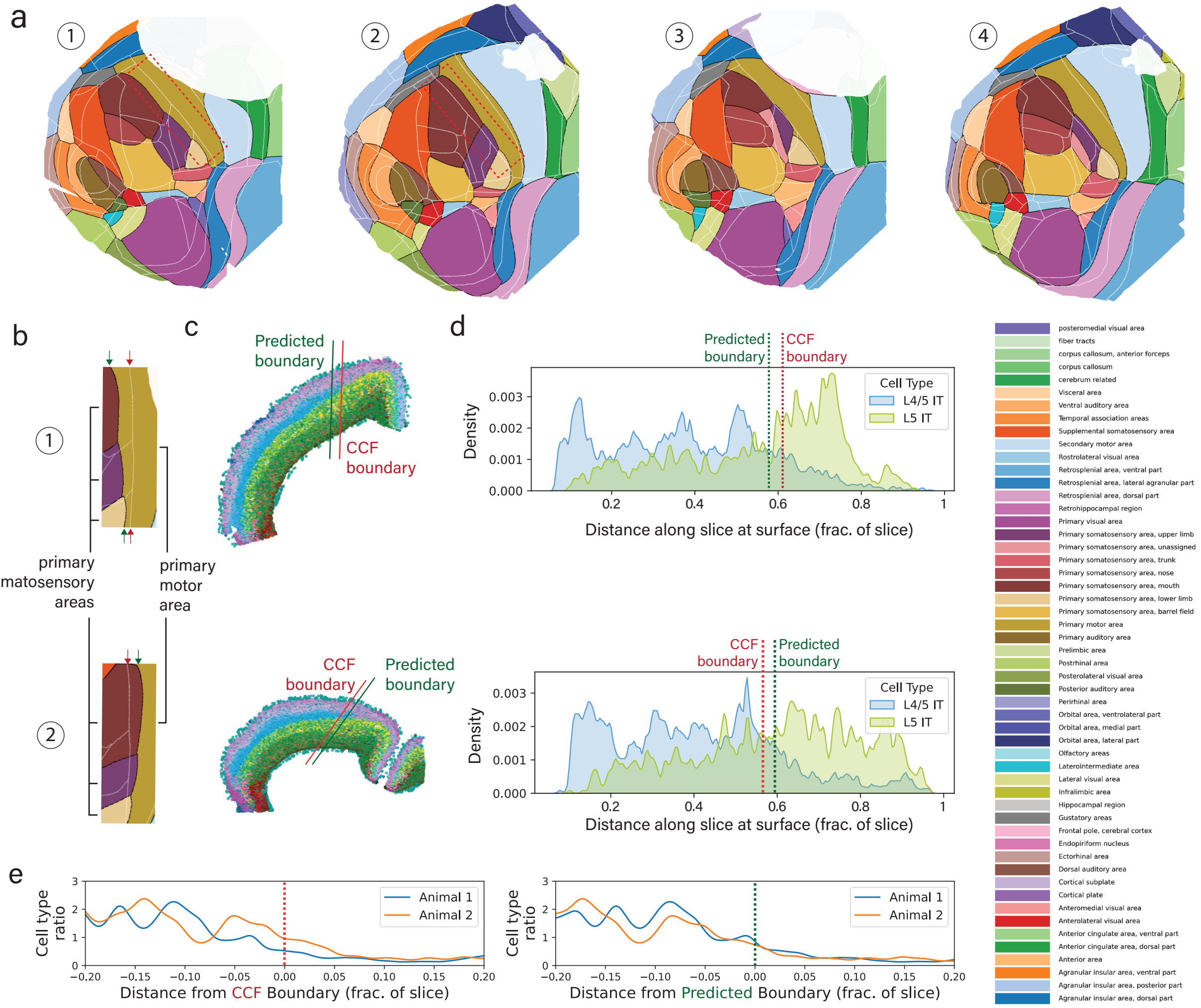

Figure 3. Creation of cortical maps from spatial transcriptomics using TissueFormer. **a)** Predicted cortical areas for each brain when held-out as a test brain. Predicted boundaries are in black, and the reference Common Coordinate Framework (CCF) boundaries are overlaid in white. Areas with low spatial density are masked. **b)** In brains 1 (top) and 2 (bottom), the somatosensory-motor boundary is displaced relative to the CCF, but in opposite directions. **c)** Example slices from brains 1 and 2 which contain the somatosensory-motor boundary boundary. Points are cells, colored by cell type with similar cell types assigned similar colors (see Methods). **d)** Spatial density along the slices of two cell types, L4/5 IT and L5 IT, in the same colors and slices as (c). **e)** The cell type ratio (L4/5 IT / L5 IT) aligned to the CCF boundary (left) or predicted boundary (right), showing an inter-animal shift in CCF vs. inter-animal consistency in the predicted maps.

Figure 3—figure supplement 1. Predicted areas of single cells, highlighting discrepancies ('errors') with CCF labels.

Figure 3—figure supplement 2. Comparison to maps of mRNA transcription of single genes.

2020). Meanwhile, the sensory/motor border analyzed in ***Figure 3*** was correlated to the expression density of Rorb1, Brinp3, and Rcan2 (***Cederquist et al., 2013***), all of which showed consistent shifts across animals with the predicted maps but not with CCF boundaries. These examples of genes with clear spatial boundaries in their patterns of expression lend support to the predicted maps' veracity.

Discussion

We developed TissueFormer, a neural network architecture that attends to groups of single-cell transcriptomic profiles in order to predict a label or annotation common to that group. In tests of predicting the brain area of a group of nearby cells in a multi-brain spatial transcriptomics dataset, TissueFormer achieved much higher accuracy rates compared to methods which predict the area of a single cell from its transcriptomic profile. For example, on held-out test brains TissueFormer achieved 60% accuracy with groups of 64 cells compared to 30% for a matched single-cell model. Furthermore, accuracy steadily increased with group size. TissueFormer also outperformed machine learning methods trained on the pseudobulk expression or cell type composition of the same group of cells. These results highlight the importance of attending across cells when predicting tissue-level properties.

Because TissueFormer accepts arbitrary cell groupings, it can flexibly model phenotypes defined by anatomical contiguity (as in spatial transcriptomics), by patient sample (e.g., blood draws, biopsies), or by dynamic windows in longitudinal sampling. These applications include a wide range of use cases, from immunomonitoring in inflammatory disease to potentially the detection of cancer states (***Weinkauff et al., 1999***; ***Sun et al., 2025***). As increasingly comprehensive healthy and disease cell atlases become available, population-aware models like TissueFormer offer a principled route to integrate those resources into predictive diagnostics that bridge cellular resolution and clinical decision-making (***Dann et al., 2023***).

TissueFormer provides a promising tool for automated brain mapping from spatial transcriptomics. Compared to the labels due to Common Coordinate Framework (CCF) registration, TissueFormer's predicted labels displayed individual variability that better correlated with inter-animal differences in local gene expression and cell type distributions, despite being trained to predict CCF labels. This tool could thus expedite the study of the individual differences in the size and location of transcriptomically-defined regions in the brain.

Our approach to brain mapping is a supervised approach in which reference brains supply ground truth labels. Note that this is distinct from the approach of aligning tissue without regard to area labels, as in Tangram (***Biancalani et al., 2021***) and CAST (***Tang et al., 2024***). An important caveat of the supervised approach is the lack of perfect training data. No datasets of spatial transcriptomics are yet available which have animal-by-animal annotations of neuroanatomy verified by additional modalities such as projection tracing or functional imaging. Without access to such datasets, we instead trained to predict the imperfect CCF labels. In general, there is no formal guarantee that the predictions on held-out brains should reflect a better assessment than the CCF itself. While we are compelled by our empirical analysis to trust the model, this must be done with caution until a model can be trained on a dataset with multiple brains individually annotated with complementing experimental modalities.

In the current study, we did not take up the question of whether different boundaries than the Allen Brain Atlas would be in any sense better, as in recent spatial clustering methods. Current boundaries are widely used and important for shared nomenclature. However, the automated clustering and identification of spatially homogeneous regions within spatial transcriptomic dataset is an interesting and open research question. Several algorithms have already been developed for this purpose (***Dries et al., 2021***; ***Zhao et al., 2021***; ***Chitra et al., 2025***; ***Hu et al., 2021***; ***Dong and Zhang, 2022***; ***Singhal et al., 2024***; ***Jackson et al., 2024***). When algorithms from this family are applied to the mouse cortex, the resulting regions correspond more to cortical layers than to functional regions (***Ortiz et al., 2020***; ***Partel et al., 2020***; ***Lee et al., 2025***). This is consistent with there being larger differences in gene expression across layers than across cortical areas. In contrast, the functional specialization of neurons in cortex varies more across area than across layers within an area; neurons within a cortical column typically mediate the same cognitive tasks and furthermore respond to similar features within that task (***Mountcastle, 1997***; ***Callaway, 1998***). Amid this dichotomy between transcriptomic and functional contiguity in space, we chose in this manuscript to classify functional areas for practical utility and for the reason of it being a benchmark task that leverages transcriptomic diversity within cellular ensembles. Nevertheless, it would be interesting in future work to train TissueFormer with self-supervised or contrastive objectives so as to discover brain areas *de novo*.

Methods

TissueFormer architecture

TissueFormer contains two main components: a single-cell module and an aggregation module. In this sense it can be viewed as an end-to-end trainable hierarchical model above single-cell foundation models.

The single-cell module is based on the BERT architecture, an encoder-only Transformer-based neural network (***Devlin et al., 2019***). We use identical hyperparameter settings as used by Geneformer 12L (***Theodoris et al., 2023***), allowing pretrained weights to be readily employed. Specifically, we use a 12-layer BERT network with ReLU activations, hidden layer sizes of 512, and a context window of up to 2048 genes. Each Transformer uses multi-head attention with 8 heads, and the inputs are appended with sinusoidal positional embeddings.

The aggregation model pools across single-cell embeddings. As a first step, the first token from the single-cell module for each cell, which is always the `<cls>` token in the input token list, is extracted. This group is then fed as input to a stack of Transformer layers (without positional embedding so as to preserve order equivariance). Each layer employs LayerNorm normalization and GELU nonlinearities. Finally, the output is averaged over cells, resulting in a single vector for the group of cells. This is given as input to a single linear classifier layer, outputting unnormalized log probabilities of each of the potential labels. All classification tasks use a cross-entropy objective.

In the present manuscript, we used the following hyperparameters. The aggregation module had 6 Transformer layers, each with a hidden size of 768 units. All training runs on the classification task used the Adam optimizer with a learning rate of 0.001 and a linear learning rate decay following a learning rate warmup with warmup ratio 0.1. Our implementation relies on the Hugging Face Transformers library (***Wolf et al., 2020***), based on Pytorch (***Paszke et al., 2019***). We used Hydra for configuration management (***Yadan, 2019***).

Single-cell pretraining

The single-cell module can be pretrained on single-cell data using a self-supervised objective. In our experiments we use a version of Geneformer adapted for mouse brain tissue. We first downloaded the Geneformer 12L model, which was trained to predict the identity of randomly masked genes (15% of total genes per cell) on a corpus of 30 million human cells. We then adapted this to data mouse tissue by repeating this methodology on a hand-curated dataset containing over 6 million mouse cells. Using the CellXGene Census explorer (***Program et al., 2025***), we aggregated data from several publications totaling over 6 million cells (***Yao et al., 2021b***; ***Govek et al., 2022***; ***Kozareva et al., 2021***; ***Steuernagel et al., 2022***; ***Yao et al., 2021a***). After tokenizing (see below), we then trained the Geneformer architecture on this dataset with an identical masked gene prediction objective. Models were trained for 3 epochs using AdamW, a learning rate of 0.001, and a batch size of 32 cells on two Nvidia V100 cards. We tested training on the murine dataset from scratch as well as finetuning the human model. In order to finetune the human model, we gave mouse genes the same initial embedding vector as their human orthologs, where available, and gave unmatched genes random initializations. Orthologs were obtained from Ensembl (***Harrison et al., 2024***), with the first selected if multiple orthologous genes existed. This transfer from human genes significantly improved the model compared to training on mouse genes alone as measured with the cross-entropy loss on 100,000 held-out cells from the mouse brain datasets. We call this finetuned Geneformer pretrained model 'Murine Geneformer'.

Data and tokenization

Data

The data published in ***Chen et al.*** (***2024***) contain spatial transcriptomic data obtained via BARseq. BARseq is a probe-based *in site* sequencing technology based on Illumina chemistry. BARseq establishes mRNA transcript counts in cells based on the identification of probe sequences at spatially resolved locations, followed by cell segmentation and assignment of genes to cells. In this dataset, probes were selected to resolve 104 marker genes chosen to resolve excitatory cells in cortex. After quality control, the overall dataset contains 10.3 million cells across nine brains, including a pilot brain, four brains of mice raised in normal rearing conditions, and four brains of mice raised following monocular enucleation.

After downloading spatial transcriptomic data were downloaded from ***Chen et al.*** (***2024***), we first converted all files into AnnData format (***Virshup et al., 2023***) with a custom script. We selected the four animals raised in control conditions, then selected only cortical cells, resulting in over 1.6 million cortical cells across 4 animals. Several metadata annotations for each cell were computed in the original publication. These include cell type annotations at 3 hierarchical levels, and cellular location within each brain slice in 'slice coordinates' as well as in CCF coordinates after each brain was registered to the Allen Brain Atlas. This dataset also contains the location within warped 'flatmap' coordinates, which use the butterfly projection from the `ccf-streamlines` python package to place cells in a 3D coordinate space in which cortical depth is the third axis (***Wang et al., 2020***).

For plots in which cells were colored by cell type (e.g. ***Figure 3***), we assigned to each cell type a color so that similar cell types would have similar colors. Specifically, the perceptual distance between any two cell types' colors was chosen to match the correlation between the cell types' pseudobulk expression vectors. Unlike standard color maps, this ensures that perceptual saliency aligns with biological meaning, with no salient colors (e.g. red) arbitrarily assigned. Technically, this involves computing the cell type similarity matrix, embedding this matrix into 3D via multidimensional scaling (MDS) to best preserve distances, then interpreting this 3D space as the LUV axes of the perceptually uniform LUV color space. We released a Python package which computes such colormaps called `ColorMyCell` to accompany this manuscript (***Benjamin, 2025***).

This dataset also contains labels of the brain area of each cell. These are inherited from the CCF coordinate space. We used these annotations as training data when predicting area from cellular transcription. As downloaded, these areas were at the finest level of the hierarchical tree of brain annotations and distinguished between cortical layers. In order to obtain the cortical area without regard to layers, we ascended one level up the hierarchical 'ancestor' tree of annotation layers in the Allen Brain Atlas. The resulting areas, displayed in e.g. ***Figure 3***, comprise 42 functionally distinct cortical areas.

Tokenization

All cells' mRNA transcript vectors were 'tokenized', or converted into a format readable by Transformer models, with the following procedure. We used an identical strategy as performed by Geneformer, but wrote custom code to accept data in anndata format (***Virshup et al., 2023***). First, each cell was normalized by the total counts in each cell. All mRNA counts were then normalized by a fixed vector representing the median per-gene counts in a reference data corpus. When fine-tuning Geneformer, it was empirically better to use the same median count vector used to train Geneformer (i.e. on human data) rather than re-establish median counts in mouse data. These normalized counts were then rank sorted, and the indices of the sort (i.e. the output of argsort) was used as the model input. For example, a model input of $[\text{gene } 5, \text{gene } 2, ...]$ describes that the gene 5 had highest relative transcription, followed by gene 2, and so on. In a modification of the Geneformer pipeline, we ensured genes with zero counts were given a special token `<pad>`. We also modified the Geneformer pipeline to prepend a `<cls>` token. The 'sentences' seen by the model were thus of the form $[\text{<cls>}, \text{gene } 5, \text{gene } 2, ..., \text{<pad>}]$. The tokens are mapped to high-dimensional embedding vector via a dictionary learned within TissueFormer.

Group construction strategy

In order to train and predict with TissueFormer, it is necessary to deliver to it groups of cells which share a label. This requires code infrastructure to appropriately group cells according to the proper criteria. Since single-cell datasets are already quite large, this would ideally occur without requiring the duplication of data for cells that occur in multiple potential groups. Thus, we took an online strategy which loaded the single-cell dataset (after tokenization) and then in an online manner constructed groups to deliver to TissueFormer in batches. To accomplish this we designed a custom sampler and data collator within the Hugging Face code ecosystem for training TissueFormer.

Our strategy was to select cells in groups which are approximately cortical columns. In order to cover the cortical surface with a roughly equal density, our pipeline first selected a seed cell at random from the training data, which potentially covers multiple brains. We next extracted the N cells in the same brain which were closest to that cell in the XY plane of the flatmap coordinate system. This resulted in a group of N cells which lie within a column. The width of the column differed in each group depending on the local density of cells. Note that because each brain contains a comparable number of cells, the column size in the test brain was comparable to the column size during training despite there being 3 training brains and 3 times the overall number of cells.

Predicting area from single cells

To provide a baseline for the multi-cell methods, we trained several machine learning models to predict area from single-cell data. Benchmark models either saw cell type (H3 types), raw transcriptomic data, or both in the case of the k-nearest neighbor cell type model. To finetune Murine Geneformer, we removed the last layer and replaced it with a linear classifier of area. We found best performance with a progressive training strategy popular in supervised finetuning. For one epoch, we froze all but the last layer, then each epoch progressively unfroze one Transformer layer in Murine Geneformer. We used a batch size of 32 and learning rate 0.0005. Results of training this model from scratch without pretraining are visible in ***Figure 2—figure Supplement 1***e.

TissueFormer training

We trained TissueFormer on an NVIDIA H100 card with a batch size of 4096 total cells divided into N groups, each with $|G| = 4096/N$ cells. A single model trains to completion for this dataset set (1.4 million cells) in 10 epochs in less than an hour. The most efficient training strategy involves saturating the GPU memory, and this is largely driven by the overall number of cells rather than the number of distinct groups. We trained all models with varying group size N to convergence to ensure a fair comparison. However, as convergence can be difficult to verify, we attempted to further match the training resources for each model by comparing models with an equal number of training steps, which is proportional to the number of cell groups seen.

Due to the imbalance in the size of brain areas, some classes have many more cells than others. A Bayes optimal model trained on this data would learn an unequal prior over classes, adding a slight bias to classify cell groups as belonging to the largest areas. This may not be desirable in certain applications of brain mapping. For example, cells on the boundary between a large area and a small area which have zero evidence towards one or the other will be classified into the large area due to the prior. This would cause an inflation of large areas. In ***Figure 3*** we display the results of a model trained with an objective designed to eliminate this bias and enforce an equal prior. Specifically we down-weighted examples from large areas with the factor as computed by the `compute_class_weight` function in scikit-learn (***Pedregosa et al., 2011***). For M total cells, C classes, and M_i cells in class i, this factor is $\frac{M}{C*M_i}$. This ensures that the magnitude of the loss function equally reflects examples from all areas.

Pseudobulk and type composition benchmarks

To ensure a fair comparison with TissueFormer, we constructed train and test sets using the same data loading pipeline with the same random seed. For the pseudobulk results, we then averaged the mRNA transcript counts within each group of cells, then trained our benchmark methods. For the cell type composition models, we extracted the cell type of each cell in the group, the constructed a normalized histogram reflecting the relative composition of cell types in that group. For both data modalities, our benchmark methods were logistic regression and random forests. Specifically, we trained logistic regression with the `lbfgs` solver. We used nested cross-validation to determine the regularization parameter, and found that no regularization was optimal at this dataset size. We also used a random forest classifier with 200 estimators, max depth of 15, 33% of features seen per tree, and a minimum samples per split and 5 and per leaf at 2. We used the scikit-learn implementation of both classifiers (***Pedregosa et al., 2011***).

Brain map visualization

To create ***Figure 3***a, we first trained 4 models in a cross-validated scheme, with 1 brain held out each fold. To achieve a high-resolution coverage, we then predicted on the test brain using every cell in the test brain once as a center seed of group selection. The prediction on each cell was stored as its label. These are visualized in ***Figure 3—figure Supplement 1***. We then constructed a method to present smooth brain maps with clear boundaries between areas. Namely, we trained a support vector classifier (SVC) with a radial basis function kernel on the cells and their predicted labels, then visualized the predictions on pixels. We used the cuML implementation, a GPU-accelerated method, and used $\gamma = 0.00001$ and $C = 1$ as hyperparameters of the SVC (***Raschka et al., 2020***). Intuitively, this method acts to color each pixel in ***Figure 3***a by a weighted average of the TissueFormer-predicted labels of the cells surrounding that pixel. The weights are selected due to a combination of proximity to the pixel and a factor determined by the algorithm to minimize the rate at which cells are misclassified, i.e. located on a pixel with a different color (area) than that cell's TissueFormer-predicted area.

Data Availability

All code to train TissueFormer as well as to create the figures in the manuscript is available at https://github.com/ZadorL
annotation.

Acknowledgements

This project was funded with support from NIH grants 1RF1 MH126883-01A1, 5U19NS123716-04, and 7U01NS132363-02. Computational resources were provided with support from NIH grant S10OD028632-01.

References

Arvaniti E, Claassen M. Sensitive detection of rare disease-associated cell subsets via representation learning. Nature Communications. 2017 Apr; 8(1):14825. https://www.nature.com/articles/ncomms14825, doi: 10.1038/ncomms14825, publisher: Nature Publishing Group.

Ballouz S, Verleyen W, Gillis J. Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. Bioinformatics. 2015; 31(13):2123–2130. Publisher: Oxford University Press.

Benjamin A, ColorMyCells: A Python package for biologically faithful colormaps for cell type visualization. Zenodo; 2025. https://doi.org/10.5281/zenodo.15595324, doi: 10.5281/zenodo.15595324.

Biancalani T, Scalia G, Buffoni L, Avasthi R, Lu Z, Sanger A, Tokcan N, Vanderburg CR, Segerstolpe A, Zhang M, Avraham-Davidi I, Vickovic S, Nitzan M, Ma S, Subramanian A, Lipinski M, Buenrostro J, Brown NB, Fanelli D, Zhuang X, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. Nature Methods. 2021 Nov; 18(11):1352–1362. https://www.nature.com/articles/s41592-021-01264-7, doi: 10.1038/s41592-021-01264-7, publisher: Nature Publishing Group.

Brodmann K. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. Barth; 1909.

Callaway EM. Local circuits in primary visual cortex of the macaque monkey. Annual review of neuroscience. 1998; 21(1):47–74. Publisher: Annual Reviews 4139 El Camino Way, PO Box 10139, Palo Alto, CA 94303-0139, USA.

Cao Y, Lin Y, Patrick E, Yang P, Yang JYH. scFeatures: multi-view representations of single-cell and spatial data for disease outcome prediction. Bioinformatics. 2022 Oct; 38(20):4745–4753. https://doi.org/10.1093/bioinformatics/btac590, doi: 10.1093/bioinformatics/btac590.

Cederquist GY, Azim E, Shnider SJ, Padmanabhan H, Macklis JD. Lmo4 Establishes Rostral Motor Cortex Projection Neuron Subtype Diversity. The Journal of Neuroscience. 2013 Apr; 33(15):6321–6332. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3698850/, doi: 10.1523/JNEUROSCI.5140-12.2013.

Chen SQ, Chen CH, Xiang XJ, Zhang SY, Ding SL. Chemoarchitecture of area prostriata in adult and developing mice: Comparison with presubiculum and parasubiculum. The Journal of Comparative Neurology. 2022 Oct; 530(14):2486–2517. doi: 10.1002/cne.25346.

Chen X, Fischer S, Rue MCP, Zhang A, Mukherjee D, Kanold PO, Gillis J, Zador AM. Whole-cortex in situ sequencing reveals input-dependent area identity. Nature. 2024 Apr; https://www.nature.com/articles/s41586-024-07221-6, doi: 10.1038/s41586-024-07221-6.

Chitra U, Arnold BJ, Sarkar H, Sanno K, Ma C, Lopez-Darwin S, Raphael BJ. Mapping the topography of spatial gene expression with interpretable deep learning. Nature Methods. 2025 Feb; 22(2):298–309. https://doi.org/10.1038/s41592-024-02503-3, doi: 10.1038/s41592-024-02503-3.

Connell W, Khan U, Keiser MJ, A single-cell gene expression language model. arXiv; 2022. http://arxiv.org/abs/2210.14330, doi: 10.48550/arXiv.2210.14330, arXiv:2210.14330 [cs, q-bio] version: 1.

Crowell HL, Soneson C, Germain PL, Calini D, Collin L, Raposo C, Malhotra D, Robinson MD. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. Nature Communications. 2020 Nov; 11(1):6077. https://www.nature.com/articles/s41467-020-19894-4, doi: 10.1038/s41467-020-19894-4, publisher: Nature Publishing Group.

Cui H, Wang C, Maan H, Pang K, Luo F, Wang B, scGPT: Towards Building a Foundation Model for Single-Cell Multi-omics Using Generative AI. bioRxiv; 2023. https://www.biorxiv.org/content/10.1101/2023.04.30.538439v2, doi: 10.1101/2023.04.30.538439, pages: 2023.04.30.538439 Section: New Results.

Dann E, Cujba AM, Oliver AJ, Meyer KB, Teichmann SA, Marioni JC. Precise identification of cell states altered in disease using healthy single-cell references. Nature genetics. 2023; 55(11):1998–2008. Publisher: Nature Publishing Group US New York.

Devlin J, Chang MW, Lee K, Toutanova K, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv; 2019. http://arxiv.org/abs/1810.04805, doi: 10.48550/arXiv.1810.04805, arXiv:1810.04805 [cs].

Dong K, Zhang S. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nature Communications. 2022 Apr; 13(1):1739. https://doi.org/10.1038/s41467-022-29439-6, doi: 10.1038/s41467-022-29439-6.

Dries R, Zhu Q, Dong R, Eng CHL, Li H, Liu K, Fu Y, Zhao T, Sarkar A, Bao F, George RE, Pierson N, Cai L, Yuan GC. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. Genome Biology. 2021 Mar; 22(1):78. https://doi.org/10.1186/s13059-021-02286-2, doi: 10.1186/s13059-021-02286-2.

Govek KW, Chen S, Sgourdou P, Yao Y, Woodhouse S, Chen T, Fuccillo MV, Epstein DJ, Camara PG. Developmental trajectories of thalamic progenitors revealed by single-cell transcriptome profiling and Shh perturbation. Cell reports. 2022; 41(10). Publisher: Elsevier.

Harrison PW, Amode MR, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Becker A, Bennett R, Berry A, Bhai J, others. Ensembl 2024. Nucleic acids research. 2024; 52(D1):D891–D899. Publisher: Oxford University Press.

Hashikawa Y, Hashikawa K, Rossi MA, Basiri ML, Liu Y, Johnston NL, Ahmad OR, Stuber GD. Transcriptional and Spatial Resolution of Cell Types in the Mammalian Habenula. Neuron. 2020 Jun; 106(5):743–758.e5. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7285796/, doi: 10.1016/j.neuron.2020.03.011.

Hsieh KL, Chu Y, Li X, Pilié PG, Dai Y, scEMB: Learning context representation of genes based on large-scale single-cell transcriptomics. bioRxiv; 2024. https://www.biorxiv.org/content/10.1101/2024.09.24.614685v1, doi: 10.1101/2024.09.24.614685, pages: 2024.09.24.614685 Section: New Results.

Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, Lee EB, Shinohara RT, Li M. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nature Methods. 2021 Nov; 18(11):1342–1351. https://doi.org/10.1038/s41592-021-01255-8, doi: 10.1038/s41592-021-01255-8.

Ito K, Hirakawa T, Shigenobu S, Fujiyoshi H, Yamashita T. Mouse-Geneformer: A deep learning model for mouse single-cell transcriptome and its cross-species utility. PLOS Genetics. 2025 Mar; 21(3):e1011420. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1011420, doi: 10.1371/journal.pgen.1011420, publisher: Public Library of Science.

Jackson KC, Booeshaghi AS, Galvez-Merchan A, Moses L, Chari T, Kim A, Pachter L, Identification of spatial homogeneous regions in tissues with concordex. bioRxiv; 2024. https://www.biorxiv.org/content/10.1101/2023.06.28.546949v2, doi: 10.1101/2023.06.28.546949, pages: 2023.06.28.546949 Section: New Results.

Kalatsky VA, Stryker MP. New paradigm for optical imaging: temporally encoded maps of intrinsic signal. Neuron. 2003 May; 38(4):529–545. doi: 10.1016/s0896-6273(03)00286-1.

Kozareva V, Martin C, Osorno T, Rudolph S, Guo C, Vanderburg C, Nadaf N, Regev A, Regehr WG, Macosko E. A transcriptomic atlas of mouse cerebellar cortex comprehensively defines cell types. Nature. 2021 Oct; 598(7879):214–219. https://www.nature.com/articles/s41586-021-03220-z, doi: 10.1038/s41586-021-03220-z, number: 7879 Publisher: Nature Publishing Group.

Lee AJ, Dubuc A, Kunst M, Yao S, Lusk N, Ng L, Zeng H, Tasic B, Abbasi-Asl R. Data-driven fine-grained region discovery in the mouse brain with transformers. bioRxiv. 2025; https://www.biorxiv.org/content/early/2025/02/26/2024.05.05.592608, doi: 10.1101/2024.05.05.592608, publisher: Cold Spring Harbor Laboratory _eprint: https://www.biorxiv.org/content/early/2025/02/26/2024.05.05.592608.full.pdf.

Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nature Methods. 2018 Dec; 15(12):1053–1058. https://www.nature.com/articles/s41592-018-0229-2, doi: 10.1038/s41592-018-0229-2, publisher: Nature Publishing Group.

Madhu H, Rocha JF, Huang T, Viswanath S, Krishnaswamy S, Ying R. HEIST: A Graph Foundation Model for Spatial Transcriptomics and Proteomics Data. arXiv preprint arXiv:250611152. 2025; .

Mao Y, Lin YY, Wong NKY, Volik S, Sar F, Collins C, Ester M. Phenotype prediction from single-cell RNA-seq data using attention-based neural networks. Bioinformatics. 2024 Feb; 40(2):btae067. https://doi.org/10.1093/bioinformatics/btae067, doi: 10.1093/bioinformatics/btae067.

Mountcastle V. The columnar organization of the neocortex. Brain. 1997 Apr; 120(4):701–722. https://academic.oup.com/brain/article-lookup/doi/10.1093/brain/120.4.701, doi: 10.1093/brain/120.4.701.

Ortiz C, Navarro JF, Jurek A, Märtin A, Lundeberg J, Meletis K. Molecular atlas of the adult mouse brain. Science Advances. 2020; 6(26):eabb3446. https://www.science.org/doi/abs/10.1126/sciadv.abb3446, doi: 10.1126/sciadv.abb3446, _eprint: https://www.science.org/doi/pdf/10.1126/sciadv.abb3446.

Partel G, Hilscher MM, Milli G, Solorzano L, Klemm AH, Nilsson M, Wählby C. Automated identification of the mouse brain's spatial compartments from in situ sequencing data. BMC Biology. 2020 Dec; 18(1):144. https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-020-00874-5, doi: 10.1186/s12915-020-00874-5.

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc Fd, Fox E, Garnett R, editors. *Advances in Neural Information Processing Systems*, vol. 32 Curran Associates, Inc.; 2019. https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011; 12:2825–2830.

Program CCS, Abdulla S, Aevermann B, Assis P, Badajoz S, Bell SM, Bezzi E, Cakir B, Chaffer J, Chambers S, others. CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data. Nucleic Acids Research. 2025; 53(D1):D886–D900. Publisher: Oxford University Press.

Raschka S, Patterson J, Nolet C. Machine Learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence. arXiv preprint arXiv:200204803. 2020; .

Rosen Y, Roohani Y, Agarwal A, Samotorčan L, Tabula Sapiens Consortium, Quake SR, Leskovec J, Universal Cell Embeddings: A Foundation Model for Cell Biology; 2023. http://biorxiv.org/lookup/doi/10.1101/2023.11.28.568918, doi: 10.1101/2023.11.28.568918.

Schaar AC, Tejada-Lapuerta A, Palla G, Gutgesell R, Halle L, Minaeva M, Vornholz L, Dony L, Drummer F, Bahrami M, Theis FJ, Nicheformer: a foundation model for single-cell and spatial omics. bioRxiv; 2024. https://www.biorxiv.org/content/10.1101/2024.04.15.589472v1, doi: 10.1101/2024.04.15.589472, pages: 2024.04.15.589472 Section: New Results.

Singhal V, Chou N, Lee J, Yue Y, Liu J, Chock WK, Lin L, Chang YC, Teo EML, Aow J, Lee HK, Chen KH, Prabhakar S. BANKSY unifies cell typing and tissue domain segmentation for scalable spatial omics data analysis. Nature Genetics. 2024 Mar; 56(3):431–441. https://doi.org/10.1038/s41588-024-01664-3, doi: 10.1038/s41588-024-01664-3.

Steuernagel L, Lam BYH, Klemm P, Dowsett GKC, Bauder CA, Tadross JA, Hitschfeld TS, del Rio Martin A, Chen W, de Solis AJ, Fenselau H, Davidsen P, Cimino I, Kohnke SN, Rimmington D, Coll AP, Beyer A, Yeo GSH, Brüning JC. HypoMap—a unified single-cell gene expression atlas of the murine hypothalamus. Nature Metabolism. 2022 Oct; 4(10):1402–1419. https://www.nature.com/articles/s42255-022-00657-y, doi: 10.1038/s42255-022-00657-y, number: 10 Publisher: Nature Publishing Group.

Sun X, Axelrod ML, Waks AG, Fu J, DiLullo M, Van Allen EM, Tolaney SM, Mittendorf EA, Xu Y, Balko JM. Dynamic single-cell systemic immune responses in immunotherapy-treated early-stage HR+ breast cancer patients. npj Breast Cancer. 2025 Jul; 11(1):65. https://doi.org/10.1038/s41523-025-00776-1, doi: 10.1038/s41523-025-00776-1.

Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nature Protocols. 2018 Apr; 13(4):599–604. https://www.nature.com/articles/nprot.2017.149, doi: 10.1038/nprot.2017.149, publisher: Nature Publishing Group.

Szałata A, Hrovatin K, Becker S, Tejada-Lapuerta A, Cui H, Wang B, Theis FJ. Transformers in single-cell omics: a review and new perspectives. Nature Methods. 2024 Aug; 21(8):1430–1443. https://www.nature.com/articles/s41592-024-02353-z, doi: 10.1038/s41592-024-02353-z, publisher: Nature Publishing Group.

Tang Z, Luo S, Zeng H, Huang J, Sui X, Wu M, Wang X. Search and match across spatial omics samples at single-cell resolution. Nature Methods. 2024 Oct; 21(10):1818–1829. https://www.nature.com/articles/s41592-024-02410-7, doi: 10.1038/s41592-024-02410-7, publisher: Nature Publishing Group.

Tasic B, Yao Z, Graybuck LT, Smith KA, Nguyen TN, Bertagnolli D, Goldy J, Garren E, Economo MN, Viswanathan S, others. Shared and distinct transcriptomic cell types across neocortical areas. Nature. 2018; 563(7729):72–78. Publisher: Nature Publishing Group UK London.

Templeton AJ, McNamara MG, Šeruga B, Vera-Badillo FE, Aneja P, Ocaña A, Leibowitz-Amit R, Sonpavde G, Knox JJ, Tran B, Tannock IF, Amir E. Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis. Journal of the National Cancer Institute. 2014 Jun; 106(6):dju124. doi: 10.1093/jnci/dju124.

Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, Mantineo H, Brydon EM, Zeng Z, Liu XS, Ellinor PT. Transfer learning enables predictions in network biology. Nature. 2023 Jun; 618(7965):616–624. https://www.nature.com/articles/s41586-023-06139-9, doi: 10.1038/s41586-023-06139-9.

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. Advances in neural information processing systems. 2017; 30.

Verlaan T, Bouland G, Mahfouz A, Reinders MJT. scAGG: Sample-level embedding and classification of Alzheimer's disease from single-nucleus data. bioRxiv. 2025; https://www.biorxiv.org/content/early/2025/01/30/2025.01.28.635240, doi: 10.1101/2025.01.28.635240, publisher: Cold Spring Harbor Laboratory _eprint: https://www.biorxiv.org/content/early/2025/01/30/2025.01.28.635240.full.pdf.

Virshup I, Bredikhin D, Heumos L, Palla G, Sturm G, Gayoso A, Kats I, Koutrouli M, Berger B, others. The scverse project provides a computational ecosystem for single-cell omics data analysis. Nature biotechnology. 2023; 41(5):604–606. Publisher: Nature Publishing Group US New York.

Wang C, Cui H, Zhang A, Xie R, Goodarzi H, Wang B. scGPT-spatial: Continual Pretraining of Single-Cell Foundation Model for Spatial Transcriptomics. bioRxiv. 2025; https://www.biorxiv.org/content/early/2025/02/08/2025.02.05.636714, doi: 10.1101/2025.02.05.636714, publisher: Cold Spring Harbor Laboratory _eprint: https://www.biorxiv.org/content/early/2025/02/08/2025.02.05.636714.full.pdf.

Wang Q, Ding SL, Li Y, Royall J, Feng D, Lesnar P, Graddis N, Naeemi M, Facer B, Ho A, others. The Allen mouse brain common coordinate framework: a 3D reference atlas. Cell. 2020; 181(4):936–953. Publisher: Elsevier.

Weinkauff R, Estey EH, Starostik P, Hayes K, Huh YO, Hirsch-Ginsber C, Andreeff M, Keating M, Kantarjjan HM, Freireich EJ, others. Use of peripheral blood blasts vs bone marrow blasts for diagnosis of acute leukemia. American journal of clinical pathology. 1999; 111(6):733–740. Publisher: Oxford University Press Oxford, UK.

Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, Cistac P, Rault T, Louf R, Funtowicz M, Davison J, Shleifer S, Platen Pv, Ma C, Jernite Y, Plu J, Xu C, Scao TL, Gugger S, Drame M, et al. Transformers: State-of-the-Art Natural Language Processing. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* Online: Association for Computational Linguistics; 2020. p. 38–45. https://www.aclweb.org/anthology/2020.emnlp-demos.6.

Xu X, Holmes TC, Luo MH, Beier KT, Horwitz GD, Zhao F, Zeng W, Hui M, Semler BL, Sandri-Goldin RM. Viral Vectors for Neural Circuit Mapping and Recent Advances in Trans-synaptic Anterograde Tracers. Neuron. 2020 Sep; 107(6):1029–1047. https://www.cell.com/neuron/abstract/S0896-6273(20)30527-4, doi: 10.1016/j.neuron.2020.07.010, publisher: Elsevier.

Yadan O, Hydra - A framework for elegantly configuring complex applications; 2019. https://github.com/facebookresearch/hydra.

Yao Z, Liu H, Xie F, Fischer S, Adkins RS, Aldridge AI, Ament SA, Bartlett A, Behrens MM, Van den Berge K, Bertagnolli D, de Bézieux HR, Biancalani T, Booeshaghi AS, Bravo HC, Casper T, Colantuoni C, Crabtree J, Creasy H, Crichton K, et al. A transcriptomic and epigenomic cell atlas of the mouse primary motor cortex. Nature. 2021 Oct; 598(7879):103–110. https://www.nature.com/articles/s41586-021-03500-8, doi: 10.1038/s41586-021-03500-8, number: 7879 Publisher: Nature Publishing Group.

Yao Z, van Velthoven CTJ, Nguyen TN, Goldy J, Sedeno-Cortes AE, Baftizadeh F, Bertagnolli D, Casper T, Chiang M, Crichton K, Ding SL, Fong O, Garren E, Glandon A, Gouwens NW, Gray J, Graybuck LT, Hawrylycz MJ, Hirschstein D, Kroll M, et al. A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation. Cell. 2021 Jun; 184(12):3222–3241.e26. https://linkinghub.elsevier.com/retrieve/pii/S0092867421005018, doi: 10.1016/j.cell.2021.04.021.

You Y, Wang ZJ, Fleisher K, Liu R, Thomson M. Building Foundation Models to Characterize Cellular Interactions via Geometric Self-Supervised Learning on Spatial Genomics. bioRxiv. 2025; https://www.biorxiv.org/content/early/2025/01/27/2025.01.25.634867, doi: 10.1101/2025.01.25.634867, publisher: Cold Spring Harbor Laboratory _eprint: https://www.biorxiv.org/content/early/2025/01/27/2025.01.25.634867.full.pdf.

Yuan X, Zhan Z, Zhang Z, Zhou M, Zhao J, Han B, Li Y, Tang J, Cell-ontology guided transcriptome foundation model. arXiv; 2024. http://arxiv.org/abs/2408.12373, doi: 10.48550/arXiv.2408.12373, arXiv:2408.12373 [cs].

Zhao E, Stone MR, Ren X, Guenthoer J, Smythe KS, Pulliam T, Williams SR, Uytingco CR, Taylor SEB, Nghiem P, Bielas JH, Gottardo R. Spatial transcriptomics at subspot resolution with BayesSpace. Nature Biotechnology. 2021 Nov; 39(11):1375–1384. https://doi.org/10.1038/s41587-021-00935-2, doi: 10.1038/s41587-021-00935-2.

Zilles K, Amunts K. Centenary of Brodmann's map — conception and fate. Nature Reviews Neuroscience. 2010 Feb; 11(2):139–145. https://www.nature.com/articles/nrn2776, doi: 10.1038/nrn2776, publisher: Nature Publishing Group.

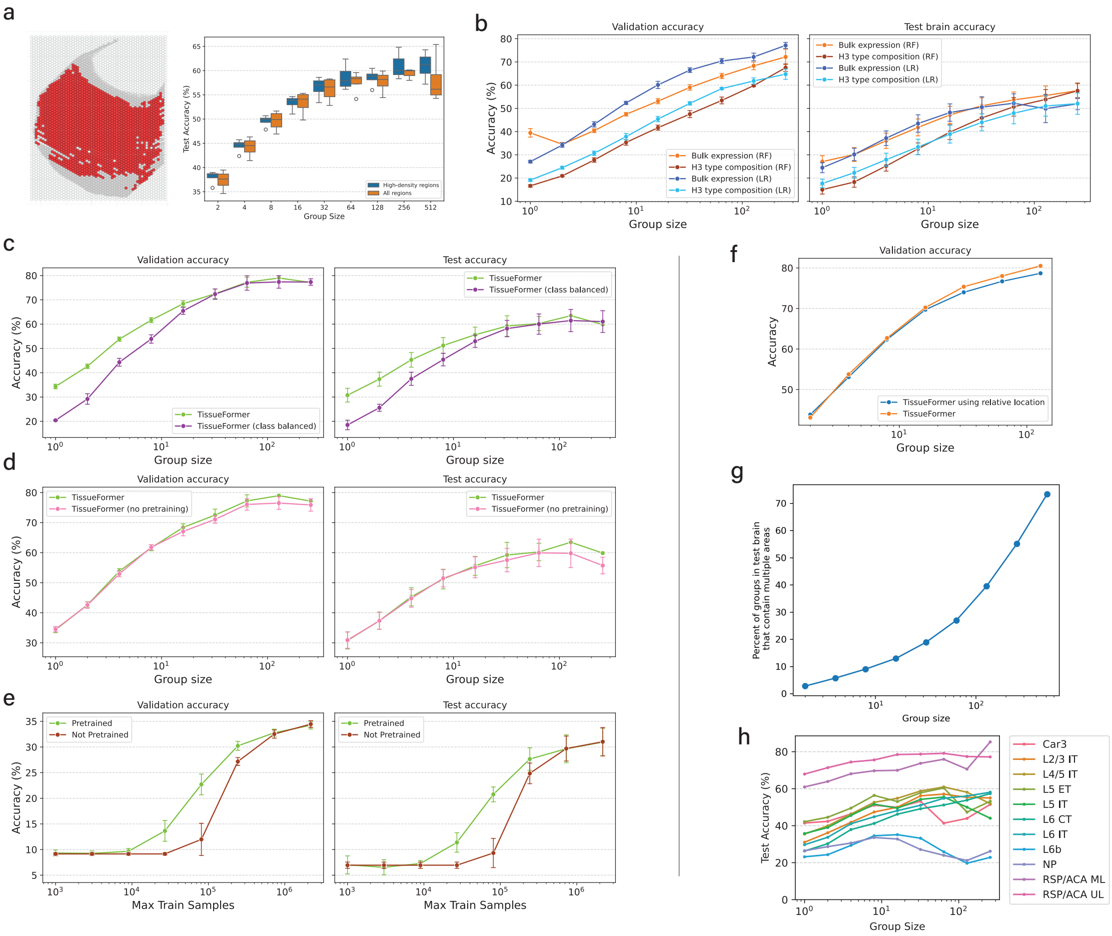

Figure 2—figure supplement 1. a) The drop in test brain accuracy relative to validation accuracy is only partially due to differences in sampling density. We selected hexes in the test brain which had high density in the 3 training brains (example at left), specifically, hexes with at least as many cells as the group size. (right) Test accuracy on data in those hexes was marginally higher than for all points, yet still much below validation accuracy. **b)** Performance of random forests (RF) compared against logistic regression (LR) given identical inputs. **c)** Performance of a class-balanced model which controls for some areas having more cells in training brains with a modified objective that discounts samples from frequent areas. **d)** Effect of initializing TissueFormer with a pretrained Murine Geneformer as single-cell module vs. randomly initialized. **e)** Varying the number of train samples, comparing random initalizations to pretrained initializations, here for $N = 1$ group size. $N = 1$ was chosen to reduce ambiguity and ensure that the number of cells seen equaled the number of groups seen. **f)** Effect of giving TissueFormer the relative spatial location (relative to the center of mass) within each group of each cell. **g)** Number of groups in the test brain that contain an area boundary as a function of the group size. **h)** The accuracy curves of models trained while only ever able to observe groups containing a single cell type. Groups were constructed by selecting a single cell and choosing the nearest N cells of the same time. The high performance of some types, such as RSP/ACA, reflects the fact that these cells are localized in just one area of cortex (retrospenial/anterior cingulate areas). Slopes of these curves are shown in Fig. 2.

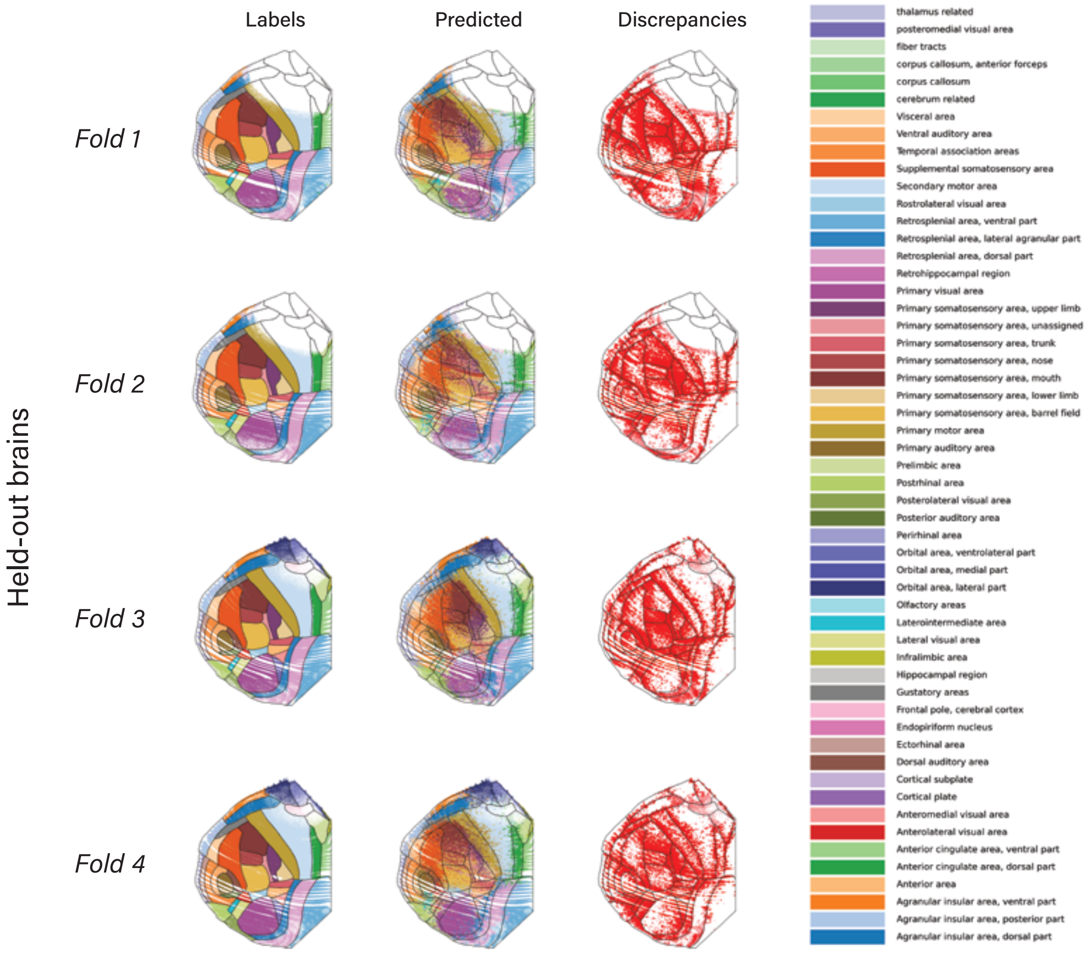

Figure 3—figure supplement 1. Across all 4 brains, we show the CCF labels on single cells (left), the predicted labels (middle) and the location of mismatches between the two. We call these discrepancies rather than errors as it is not clear which is closer to ground truth. Notable systematic differences include a gross shift of the Primary Visual Area (magenta), and the somatosensory-motor boundary. In both cases, a shift appears in the discrepancy map as a large amount of discrepancies on one side of an area but very few on the other side, and a large discontinuity at the CCF boundary.

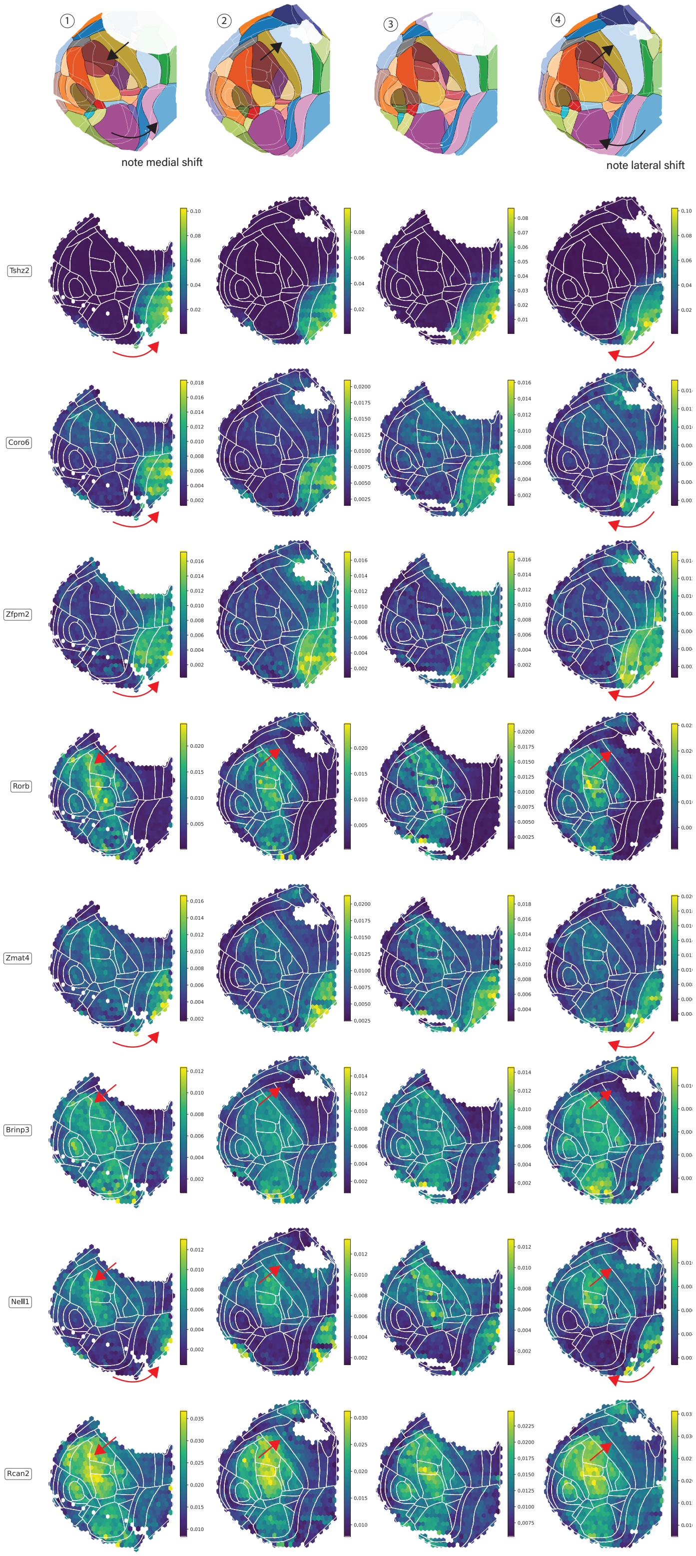

Figure 3—figure supplement 2. Normalized spatial density of select genes in each animal. Arrows highlight the animal-specific shifts of regional anatomy relative to the CCF (white boundaries), and how these are consistent between genes and with the predictions of Tissueformer (top row, reproduced from ***Figure 3***. The top three rows (Tshz2, Coro6, and Zfmp2) recapitulate the medial shift in the boundary to the retrospenial areas in Brain 1 and the lateral shift in this boundary in Brain 4. The bottom 5 rows recapitulate the shift in the somatosensory/motor boundary, the same as analyzed in cell type distributions in Figure 3. To construct normalized gene densities, we first normalized mRNA transcription levels within each cell. We then plot the average normalized transcription within each hex. Normalization was necessary to reduce confounds of unequal spatial sampling.