TOWARDS A RELIABLE SPATIAL ANALYSIS OF MISSING FEATURES VIA SPATIALLY-REGULARIZED IMPUTATION

A PREPRINT

Chen Qiao
School of Biomedical Sciences
University of Hong Kong
Pokfulam, Hong Kong SAR, China
cqiao@connect.hku.hk

Yuanhua Huang
School of Biomedical Sciences &
Department of Statistis and Actuarial Science
University of Hong Kong
Pokfulam, Hong Kong SAR, China
yuanhua@hku.hk

January 21, 2023

ABSTRACT

Recent spatial transcriptomic (ST) technologies offer a lens for observing the spatial distribution of RNA transcripts in tissues, yet achieving a whole-genome-level spatial landscape remains technically challenging. Multiple computational methods hence have been proposed to impute missing genes from a single-cell reference dataset, while they lack mechanisms of explicitly encoding spatial patterns in the modeling. To fill the research gaps, we introduce a computational model, TransImp, that leverages a spatial auto-correlation metric as a regularization for imputing missing features in ST. Evaluation results from multiple platforms demonstrate that TransImp remarkably preserves the spatial patterns, hence substantially improving the accuracy of downstream analysis in detecting spatially highly variable genes and spatial interactions. Therefore, TransImp offers a way towards a reliable spatial analysis of missing features for both matched and unseen modalities, e.g., nascent RNAs.

Keywords Spatial transcriptomics · Imputation · Spatial regularization

1 Introduction

A variety of biological processes are modulated through the spatial organization of cells, including how different cell types are distributed in a microenvironment and how cells communicate to perform a certain biological function. Prominent examples include localization of cell types in mouse organogenesis [1] and human thymus development [2], as well as inter-cellular communications in squamous cell carcinoma [3] and during intestinal development [4].

In recent years, the rapid development of spatial transcriptomics (ST) technologies makes it further accessible for dissecting the spatial mixture of cells in a wide range of biomedical research. The main two streams of technologies are sequencing-based and imaging-based (via in situ hybridization or in situ sequencing) [5]. The former in principle can cover the whole transcriptome but has a limited resolution of cells (e.g., around 10 cells per spot) while the latter can have a cell-level resolution but is generally limited to probing dozens of pre-selected genes [6]. Recently, breakthroughs on both platforms, e.g., seqFISH+ [7] and Stereo-seq [8], are addressing these limitations in different aspects, while the RNA capture efficiency is still far from perfect in sequencing-based methods and laborious designing of the candidate gene probes is required in the imaging-based methods.

Therefore, computational methods for feature imputation are highly demanded in analyzing ST data, particularly by leveraging the rich single-cell RNA-seq data as a reference, including imputing unseen genes in the imaging-based data or imputing highly sparse genes in sequencing data. In general, modality integration methods can be applied for the task of missing feature imputation, e.g., Liger [9] and Seurat v3 [10]. Recently, multiple tailored methods have also been proposed to address this challenge with improved performance reported. For example, SpaGE was introduced to address this task by averaging k-nearest neighbors (kNN) from the reference scRNA-seq data after projecting both

ST and scRNA-seq datasets into a common low dimensional space which is spanned by adapted principal vectors[11]. Similar kNN-based aggregation strategies were also applied in a joint representation space produced by either shared principal components [1] or the latent encodings of an auto-encoder [12]. Tangram was another appealing method that directly learns a mapping matrix for cells from scRNA-seq to spots in ST by minimizing the cosine distances at both feature and sample levels [13] between imputed and true ST expressions. Moreover, in addition to certain molecular features, mapping meta information of cells, e.g., cell type labels, is also a task that shares the same principles of feature imputation but is usually treated as a standalone task, e.g., in RCTD [14] and Cell2location [15].

However, multiple challenges in ST imputation remain less touched. First, most of the feature imputation methods do not explicitly consider the spatial variability patterns during the imputation, hence downstream analysis may be largely affected by (under- or) over-estimated spatial auto-correlations and/or spatial ligand-receptor interactions. Second, existing methods lack comprehensive assessments regarding spatial patterns, including many downstream tasks. It would be helpful for the community to understand the limitations of different methods in tackling these challenging scenarios. Third, given the rapidly increasing number of cells in the ST data, the efficiency of a method is another demanding feature.

To address these challenges, we introduce a generic framework, TransImp, to transform information from a scRNA-seq reference to the ST context. In particular, we focus on a few challenging tasks, including the detection of spatially-variable features or pairs of features: two very common downstream spatial analyses. Besides, we also consider the prediction of the dominant proportion of missing features in the image-based ST datasets. Finally, we briefly showcase that this method can also be applied to predict the unspliced RNAs, hence enabling trajectory analysis of cell differentiation at the ST level.

2 Results

2.1 The TransImp model

In TransImp model, we aim to learn a mapping (i.e., translation) function to translate the scRNA-seq reference to ST data. Related to the Tangram model [13], our overall translation framework is to learn a linear mapping matrix W from the reference to the ST data. One can view it as a multivariate regression model (i.e., multiple outcome variables), by treating gene as sample and cell as dimension (see the difference of this dual problem in Supp. Fig. S1). Therefore, spatial regularization can be posed for enforcing a constraint on the multiple output variables (i.e., spots in ST). Different from existing models, our TransImp has innovations at two folds (Fig.1.a). First, it defines the mapping matrix in a low-rank manner (by default), which can perform feature imputation not only more robustly but also more efficiently. A special case is fixing the V matrix as the cell type identity matrix, which is equivalent to full-rank on the cell type profiles. Second, we introduce a critical regularizer for preserving spatial auto-correlation (Moran's I) in the auxiliary loss function on top of prediction loss measured by cosine distance between observed and predicted spatial expression (see Section Method). Therefore, the predicted features across space may retain a similar level of spatial auto-correlation to the observed values. After the translation function is learned, Fig.1.b shows how it is used to impute missing genes in spots. To better understand the effect of spatial regularization, we observe the two modes of translation function with or without spatial regularization, hence generating four model configurations shown in Table 1.

To thoroughly evaluate the imputation performance, direct and downstream analysis tasks are conducted on the imputed expression matrices. The performances are compared with existing high-performing methods. Fig.1.c lists seven tasks conducted on the imputed count matrices via five-fold cross training and prediction, details can be found in Section Method.

2.2 TransImp preserves spatial patterns of imputed genes for spatially highly variable gene and spatial ligand-receptor pair detection.

To evaluate the performance of TransImp, we first applied it to Visium-based ST datasets. Due to Visium's capacity of sequencing almost the whole genome, far more genes can be captured, and we hence can obtain enough positive and negative observations for spatially highly variable gene and spatial ligand-receptor pair detection. Fig.2 reports the various evaluation results, with more detailed benchmarks reported in the supplementary tables S1-S6 and figures S2-S4.

First, by conducting a 5-fold cross-validation, we assessed the performance of our TransImp and three existing methods SpaGE [11], stPlus [12] and Tangram [13], in terms of both entry-level accuracy and spatial pattern preservation. Fig.2.a shows that TransImp with spatial regularization (i.e., TransImpSpa) achieved a comparable imputation performance to the existing best-performing approaches, with 82.3% its imputed genes to have cosine similarities greater than 0.2 to ground truth, surpassing spaGE and stPlus, and is close to the best method Tangram. The better imputation performance

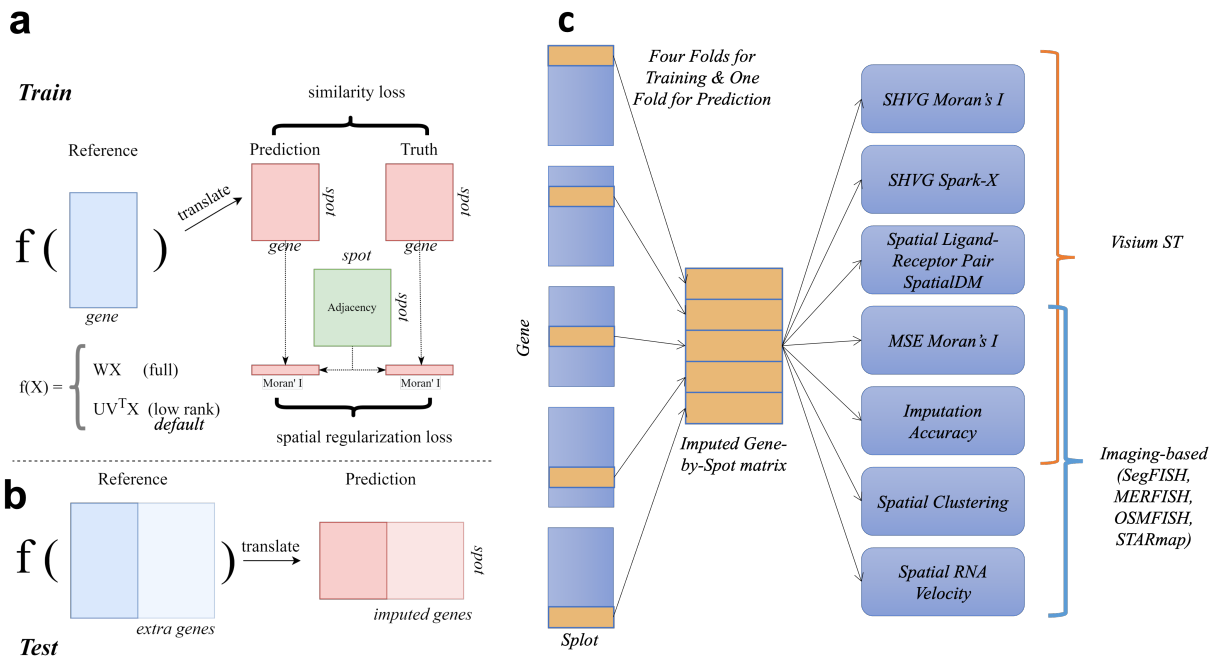

Figure 1: Computational framework and evaluation tasks. a. Training the translation function parameterized as linear full or low-rank (default) matrices with a similarity loss measuring the consistency of prediction against ground-truth, and a spatial regularization loss based on Moran's I; b. Imputation of missing genes after the translation function is learned; c. Five-fold cross imputation of genes at the spatial level, and the direct and downstream evaluation tasks. For Visium Spatial Transcriptomic (ST) datasets, these include the imputation similarity with ground truth, mean squared error (MSE) of Morans' I indices between imputed and ground-truth expressions, spatially-highly-variable (SHVG) gene detection using Moran's I and Spark-X tests, spatially-significant ligand-receptor pair test using SpatialDM are conducted; For imaging-based ST datasets including SeqFISH, MERFISH, OsmFISH and STARmap, evaluation includes the imputation similarity with ground truth, MSE of Moran's I indices between imputed and ground-truth expressions, agglomerative clustering using spatial neighborhood graph as connectivity graph (spatial clustering), as well as Spatial RNA velocity analysis.

of TransImpSpa, unlike Tangram, does not sacrifice much spatial pattern information, as indicated in Fig.2.b with all the imputed genes having mean squared errors (MSE) of Moran's Is below 0.05, much better than Tangram's only 39.4% imputed genes having MSE errors below this threshold.

Second, Fig.2.c-d demonstrate the better spatial pattern preservation performance of the proposed method in downstream analysis. Overall, the tasks are extremely challenging for all the methods, yet TransImpSpa continues to obtain the best results. Specifically, for spatially highly variable gene detection (Fig.2.c and d), TranImpSpa achieved the highest area under precision-recall curve (AUPRC) in both Moran's I and Spark-X testing methods, corroborating its effectiveness of retaining spatial patterns for individual genes. Moreover, for spatial ligand-receptor pair detection using SpatialDM (Fig.2.e), TransImpSpa also outperforms all the other methods by a large margin, showing its good spatial-pattern-preserving performance even in the inter-gene interaction context.

Finally, Fig.2.f visualizes the combined overall scores (combining those shown in Fig.2.a-e) of different methods on Visium datasets of different species, organs, and conditions. It is noticeable that the proposed TransImpSpa has the highest overall scores across all the different datasets.

2.3 TransImp contributes to error-reduced spatial patterns in imputed genes while retaining comparable imputation performance to state-of-the-art methods

We next applied TransImp to a dataset generated with seqFISH platform on mouse organogenesis, where 351 genes were probed in 57,536 spots [1]. We again conducted a 5-fold cross-validation and compared the performances of different methods shown in Fig.3.a-b). In Fig.3.a, the Cosine similarity score (CSS) indicates that the proposed method achieved the best performance regarding imputation consistency with ground-truth (medain CSS= 0.500, 0.487 for

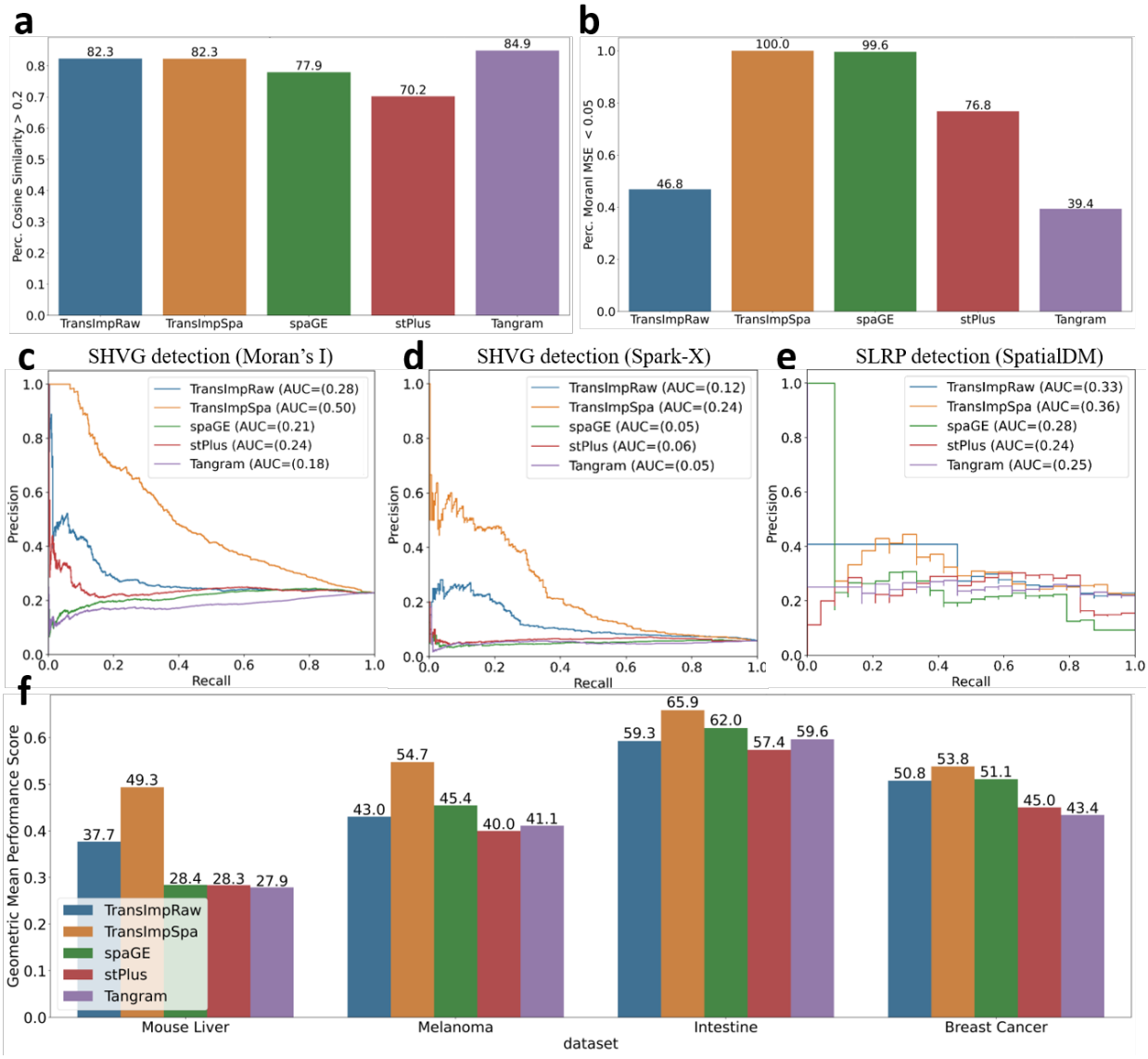

Figure 2: Evaluation results of Visium ST datasets: detailed results of mouse liver (a-e), and combined results for all the four Visium ST datasets (f). a. Proportion of cosine similarity greater than 0.2 between imputed and ground-truth; b. Proportion of MSE error of Moran's I indices smaller than 0.05 on imputed v.s. ground-truth expressions; c. Precision-recall curves and AUC scores of Moran's I Spatially-highly variable gene (SHVG) detection tests; d. Precision-recall curves and AUC scores of Spark-X SHVG detection tests; e. Precision-recall curves and AUC scores of spatial ligand-receptor pair (SLRP) detection tests; f. combined overall scores of all the methods across different ST datasets.

TransImpRaw and TransImpSpa respectively), outperforming the existing methods at least moderately. We noticed that stPlus and Tangram also work comparably well (median CSS=0.482 and 0.477, respectively), while SpaGE are less accurate on average (median CSS=0.454).

Besides, we noticed two evident observations in Fig.3.b. First, the similarity measurement seems limited in capturing global spatial patterns quantified by Moran's I, resulting in unaligned performance rankings for the methods. Second, the spatial regularization shows successful preservation of spatial auto-correlation patterns in the translation function, where on average 80.5% (46.9-80.5% across datasets) of Mean Squared Errors (MSE) of Moran's I on imputed gene profiles were reduced by TransImpSpa compared to Tangram. This capability of preserving the observed spatial auto-correlation is particularly remarkable, as Moran's Is of these test genes have not been seen during the training.

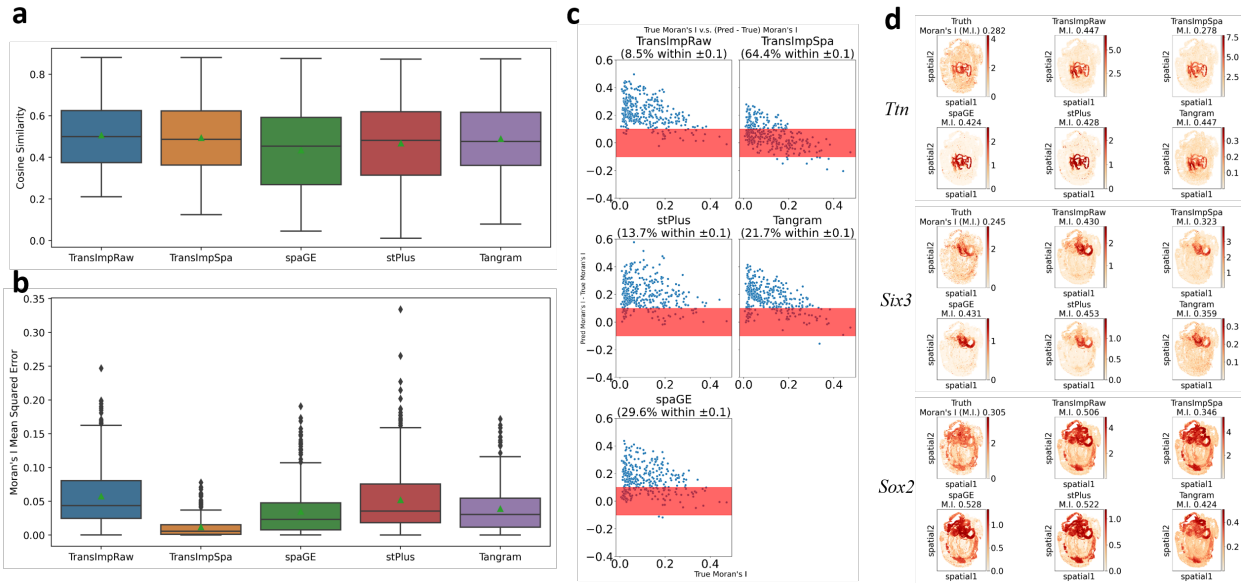

Figure 3: Results on seqFISH ST dataset. a. Box plots of the Cosine Similarity Scores (CSS) among different imputation methods; b. Box Plots of the Moran's I mean squared error (MSE); c. Deviance of Moran's Is computed on imputed genes from ground truth; d. Imputation results for example marker genes across different methods.

To further investigate the behaviors of different methods in preserving spatial patterns, scatter plots of true Moran's Is (x-axis) by predicted minus true Moran's Is (y-axis) are listed in Fig.3.c, where positive values on y-axis may imply false positive, i.e. predicted Moran's Is are greater than the true values. Without spatial regularization, it is evident from the plots that the imputed gene profiles of all the methods tend to incur overestimated Moran's Is: the majority of the points (70-91%) are located above the upper border of the ±0.1 error bar area. By contrast, spatial regularization helps reduce such bias by moving the majority points towards the ±0.1 interval area (64%), reducing the MSEs for Moran's Is.

In Fig.3.d, we visualize the true and imputed expressions of three marker genes. Although all the methods obtained comparable CSS values, we find only our method with spatial regularization can relieve the overestimation trend of Moran's I.

2.4 TransImp is efficient and robust

Next, we conducted evaluation experiments on three more spatial transcriptomic datasets generated using different technologies. In Fig.4a and b, we summarized the CSS of gene profiles and MSE of Moran's Is across datasets and methods. It can be found that the proposed framework not only exhibited comparable performance to those state-of-the-art methods in cosine similarity scores but also on average much better spatial preservation powers with spatial regularization.

Moreover, downstream clustering analysis was conducted on the imputed genes from the 5-fold test sets. To involve spatial information in the clustering, we adopt Agglomerative Clustering with adjacency matrices calculated using the spatial coordinates. We compare the clusters of imputed versus true expressions after applying the same Agglomerative Clustering procedure. The averaged clustering metrics (covering multiple scores, e.g., Adjusted Rand Index; see Method) are visualized as Fig.4c, demonstrating relatively good and robust performances of the proposed method across datasets.

In addition, we recorded the training run-time for each method. As shown in Fig. 4.d, benefited from GPU acceleration, Tangram and our proposed methods are much more efficient than stPlus and SpaGE, particularly on larger datasets (MERFISH and seqFISH). While the training run-time for Tangram and none-spatially-regularized translation models are close on larger datasets, that of spatially-regularized models increases slightly the computing time, due to the need of computing global Morans' I in each training iteration.

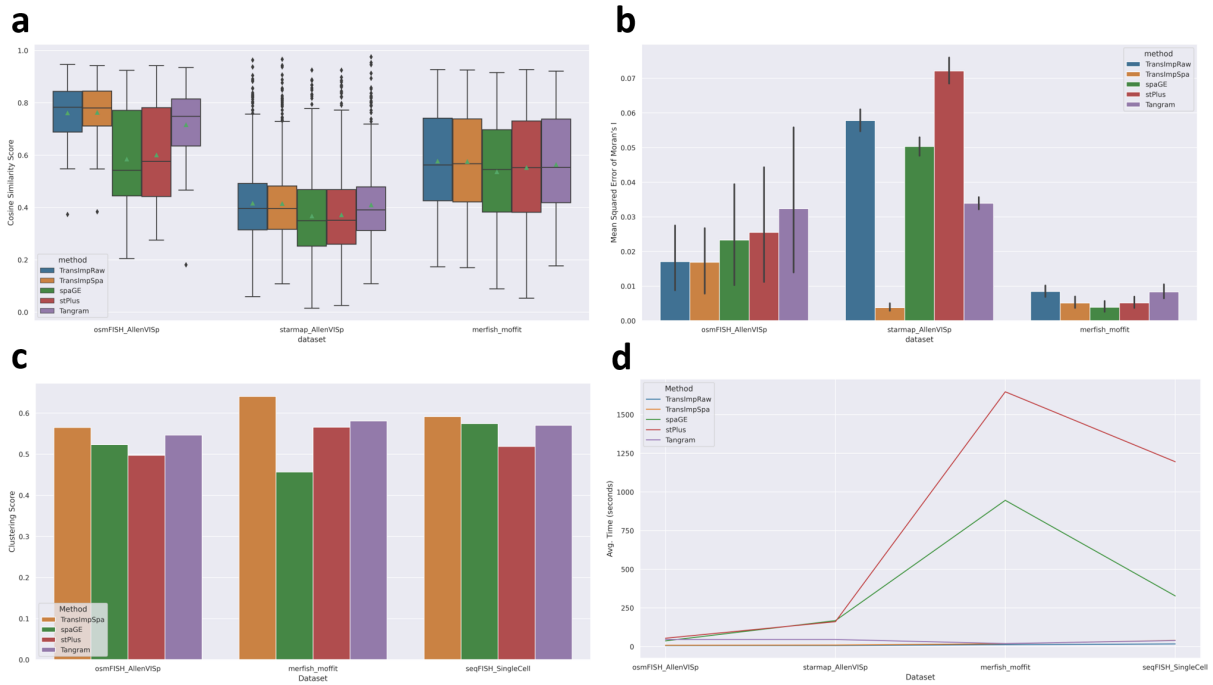

Figure 4: Evaluation of imputation methods. a. Cosine similarity scores over osmFISH, starmap and MERFISH spatial transcriptomic datasets; b. Mean Squared Error of Moran's I statistics; c. Averaged Clustering scores from multiple metrics (Methods); d. Averaged computational runtime of different methods.

2.5 TransImp facilitates spatial RNA velocity analysis by imputing unspliced and spliced RNAs in spots

Finally, as a generic translation framework, TransImp may have the capability of translating the unseen feature modality in the ST data from the reference scRNA-seq. Here, we specifically assess how the unspliced RNA (and the spliced RNA) abundance can be translated from scRNA-seq data to ST data. To explore, we trained the TransImpSpa framework on the Chicken Heart and Mouse Brain datasets with corresponding reference datasets retrieved from [16]. On each dataset, the model was trained on the anchor genes (shared between scRNA-seq and ST) and then used to predict the unspliced and spliced expression matrices into the spatial space, followed by RNA velocity analysis with scVelo stochastic mode [17].

On the Mouse Brain dataset, we first performed the clustering on the scRNA-seq data and translated the cell type to the ST data (consistently between Tangram). As there lacks ground truth of the unspliced and spliced RNA abundance, we evaluated the performance by comparing the consistency of the predicted differentiation directional between ST and scRNA-seq. In Fig. 5a, we found that the neuron differentiation from neuroblast cells is well captured in both scRNA-seq and ST data (Supp. Fig. S7 and S8).

By performing a similar analysis on the chicken heart dataset where the cell types were provided, we found the overall transition trends terminating at Fibroblast cells are consistent with the pseudotime analysis for the Epicardial lineage in [18]. Taken together, these results demonstrate the potential of spatial RNA Velocity analysis translating unseen modalities using the proposed framework.

3 Discussion

To summarize, we introduced a framework for imputing missing features from reference scRNA-seq data. With various datasets from different platforms, we demonstrated that the proposed framework has a good balance to both achieve state-of-the-art prediction accuracy and preserve the spatial auto-correlation of the imputed features. Also, the wide applicability in common downstream analyses and computational efficiency of our method may further accelerate the analysis of increasingly popular ST data.

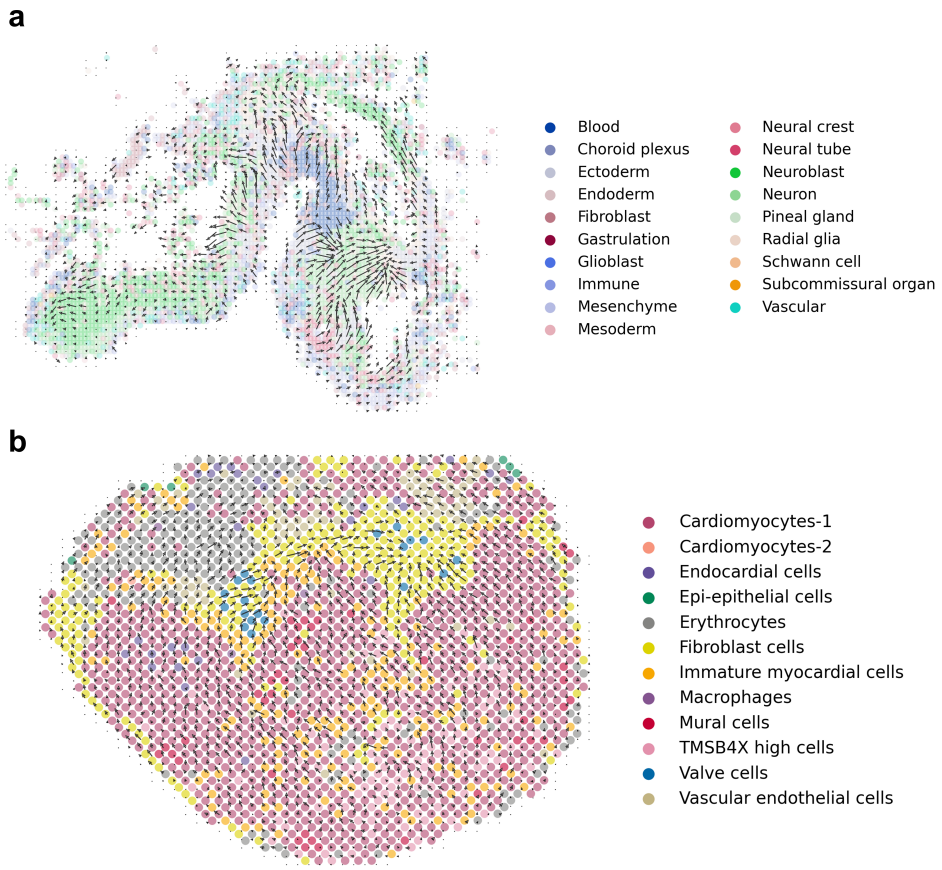

Figure 5: Spatial RNA velocity analysis based on imputed nascent and mature RNA counts, spot cell types are mapped from single cell annotation. a. Spot-level transition grid map on the Mouse Brain ST dataset, sub-regions of the map show transitional directions from Neuralblast to Neuron captured in single cell level; b. Spot-level transition grid map on the Chicken Heart ST dataset, the overall trends of terminating at Fibroblast cells are consistent with prior work.

On the other hand, it still remains an open challenge how to better structure imputation models with the location information available from spatial transcriptomics. The first step might be to re-evaluate spatial metrics that play an important role in quantifying spatial patterns of gene expressions, based on which the discrepancies between prediction and truth can be measured. It should be noted that the properties of a spatial metric affect downstream analysis. In this initial study, we leveraged Global Moran's I as a proxy for quantifying spatial patterns. However, studies have argued for more powerful metrics such as SpatialDE [19] and SparkX [20] for mining spatial patterns. Besides, applying local spatial metrics such as local Moran's I and Geary's C can be achieved by enabling spot-centered mini-batch optimization for translation functions, instead of fully batched training resulted from calculating global spatial metrics. This allows scaling the spatially-regularized frameworks to even larger datasets. Moreover, from Fig.2.a-b and Fig.3.a-b we noticed the unaligned performance rankings between the cosine similarity scores and spatial metric errors, which indicates the practical requirement for a more competent spatial metric.

The full translation framework working with cluster-aggregated gene signature matrix (TransImpCls and TransImpClsSpa, in Supp. Tables S1-S6, Figures S5 and S6) can be viewed as a special case of the low-rank framework with a cell-by-gene reference matrix, in that the low-dimensional matrix $\mathbf{V}$ is fixed to be a binary matrix of shape cell-by-cluster. Each row of the matrix is a one-hot vector turned on at the dimension corresponding to this cell's cluster type. The low-rank setting hence offers an additional interface for injecting prior knowledge (e.g., cell types) into the translation function, either via explicit regularization on V and/or U or through a Bayesian manner. Moreover, beyond a linear setting for the low-rank framework, the property of none-linearity may also be added to the translation function via, e.g., none-linear activation functions after the dot product with $\mathbf{V}$ and/or $\mathbf{U}$, increasing model capacity for more complex mapping.

Both modes of the proposed method: TransImpSpa and TransClsSpa obtained decreased spatial MSE errors at the cost of receiving slightly worse CSS scores when regularized by spatial losses, which urges us to be cautious when interpreting the imputation evaluation results: A high CSS score may not necessarily indicate a similarly good performance of spatial pattern preservation. Consequently, spatial analysis on such imputations may be questionable, as indicated by the examples shown in Fig.3.c: many overestimated Moran's I scores on unregularized imputations. On the other hand, the strength of regularization may affect CSS scores, which advocates tuning a proper weight (λ) for balanced performance although we set all $\lambda = 1.0$ in benchmarking for simplicity.

4 Method

As shown in Fig.1.a, a standard translation framework translates an input reference gene profile or signature into the target gene profile on spots, where we treat genes rather than cells/spots as an input instance. The translation function entails the simplest configuration where it is linear without bias terms (assuming both datasets are normalized to the same scale), reducing to an alignment/mapping matrix as inferred by Tangram [13]. Nevertheless, a translation function could be more flexible as being low-rank or even non-linear. Moreover, we may add spatial regularizations in fitting translation functions, so that they could better preserve spatial patterns.

Translation function A translation function $f(\cdot)$ takes as input a reference gene profile $\mathbf{x}$ and outputs the target spatial profile $\mathbf{y}$: $\mathbf{y} = f(\mathbf{x})$. This work considers simple settings of it being full or low-rank linear mapping:

$$f(\mathbf{x}) = \begin{cases} \mathbf{W}\mathbf{x} & \text{full} \\ \mathbf{U} \cdot \mathbf{V}^T\mathbf{x} & \text{low-rank, default} \end{cases}$$

Where non-negative matrix $\mathbf{W} \in R^{N_s \times N_c}$, N_s and N_c are the numbers of spots and cells, respectively. In the low-rank setting, $\mathbf{W}$ is approximated by the matrix multiplication of two low-dimensional non-negative matrices $\mathbf{U} \in R^{N_s \times K}$ and $\mathbf{V} \in R^{N_c \times K}$, where K is the hyper-parameter specifying the dimensionality. We constrain $\mathbf{V}$ to be non-negative by taking the element-wise square of an unconstrained matrix $\tilde{\mathbf{V}}$, denoted as $\mathbf{V} := \tilde{\mathbf{V}}^2$. $\mathbf{U}$ is constrained by taking the softmax of an unconstrained matrix $\tilde{\mathbf{U}}$ over the K latent dimensions. The design motivation is to first implicitly generate latent cluster centers (in space R^K) and then construct spots by weighted combinations of these centers. Besides, we consider two modes of the input $\mathbf{x}$: 1) Cluster mode, aggregating reference matrix by summing over clusters provided by e.g., Leiden method; and 2) Cell mode, the whole gene profile vector. For efficiency of computation, we apply low-rank mapping for the cell mode (as default) and full mapping for the cluster mode (a special case of the low-rank cell mode).

Translation loss Shared genes between reference and spatial transcriptomic datasets are used for supervised training of the translation function. We denote the output of the translation function for all the genes as matrix $\hat{\mathbf{Y}}$ which is compared with the ground-truth matrix $\mathbf{Y}$. The translation loss is computed based on the cosine similarity between the rows and columns of the two matrices, capturing both the spot-wise and gene-wise expression distributions:

$$\ell_{trans} = \frac{1}{N_s}\sum_{i=1}^{N_s}\left(1 - \cos(\hat{\mathbf{y}}_{i,:}, \mathbf{y}_{i,:})\right) + \frac{1}{N_g}\sum_{j=1}^{N_g}\left(1 - \cos(\hat{\mathbf{y}}_{:,j}, \mathbf{y}_{:,j})\right)$$

where $\mathbf{y}_{i,:}$ and $\mathbf{y}_{:,j}$ index the ith row and jth column from matrix $\mathbf{Y}$, respectively. N_g is the total number of shared genes.

Spatial regularization loss To explicitly encode spatial patterns into the training procedure, we adopt global Moran's I [21], a well-studied spatial auto-correlation metric, as the quantitative measurement, and compare I values on predicted and true expressions of each gene using mean squared error (MSE) loss:

$$\ell_{spa} = \frac{1}{N_g}\sum_{i=1}^{N_g}(\hat{I}_i - I_i)^2$$

where $\hat{I}_i$ and I_i are Moran's I computed on predicted and true expressions of gene i respectively.

Model configuration and training Four settings of the proposed framework are studied and evaluated on different datasets. As shown in Table 1, configurations denoted as TransImpClsSpa and TransImpCls are Cluster-based full mapping frameworks with and without spatial regularization respectively. Likewise, TransImpSpa and TransImpRaw

Table 1: Model configurations

	w/o Spa.Reg	w/ Spa.Reg
Low-Rank (cell mode)	TransImpRaw	TransImpSpa
Full (cluster mode)	TransImpCls	TransImpClsSpa

are cell-based low-rank settings of the translation framework with and without spatial regularization. Note that the evaluation results of the cluster-based full mapping framework are provided as additional materials in supplementary tables S1-S6 and figures S5 and S6 for manuscript space conservation.

For configurations with spatial regularization, a hyper-parameter λ is used to balance the spatial regularization strength in the total loss:

$$\ell_{total} = \ell_{trans} + \lambda \cdot \ell_{spa}$$

Only the translation loss ℓ_{trans} is used for configurations without spatial regularization.

All the models are implemented using pytorch [22] and trained with AdamW optimizer [23] on GPU.

Datasets and configuration Two categories of spatial transcriptomic (ST) datasets are used for evaluating the imputation performance of the proposed methods, as shown in Table 2. The top four rows summarize the information of the Visium-based ST datasets. The preprocessed mouse liver ST c1 sample and its reference are from [24]. The breast cancer ST sample 1142243F and single-cell reference are from [25] and [26], respectively. We achieved the preprocessed human melanoma [27] ST dataset and its reference as well as the intestine [28] ST A1 sample from the spatialDM authors [29], and the intestine reference from [30]. For the visium-based datasets, we train all the models with 2000 epochs, a learning rate 0.01, weight decay 0.01, and the latent dimension for low-rank modes being 256. We clip the elements of $\mathbf{V}$ to be within 0.5 for the intestine and melanoma experiments to prevent overfitting, while keeping the default threshold of 10 for the mouse liver and breast cancer experiments. The bottom four rows summarize the imaging-based ST datasets. We obtained the preprocessed STARmap, MERFISH and OsmFISH, as well as the corresponding references AllenVISp and Moffit from [11], while the SeqFISH dataset with its single-cell reference is obtained from [1]. For all the datasets, we used the same learning rate 0.01 and weight decay 0.01. and set the low-rank latent dimensions to 128. We clipped the elements of $\mathbf{V}$ to be within 5 for low-rank models on MERFISH dataset to counteract overfitting, while keeping the default 10 on all the other ST datasets. The models were trained in 1000 epochs across all the datasets, except for TransImpSpa on the SeqFISH dataset, where 2000 epochs of training were conducted for sufficient model fitting.

Table 2: Dataset information

Dataset	# Spots	# Cells	# Spa.genes	# Ref.genes
Visium Mouse Liver	2,110	4,759	16,225	19,355
Visium Human Intestine	2,649	14,537	33,538	19,525
Visium Human Breast Cancer	4,784	35,276	28,402	33,745
Visium Human Melanoma	293	4,645	5,779	21,118
osmFISH_AllenVISp	3,405	14,249	33	34,617
starmap_AllenVISp	1,549	14,249	1,020	34,617
Merfish_Moffit	64,373	31,299	155	18,646
seqFISH_Single Cell	57,536	32,844	351	29,452

For RNA velocity analysis, the preprocessed versions of the two spatial transcriptomic datasets with the corresponding single-cell references, Day 14 Chicken Heart [18] and Developing MouseBrain Atlas [31] are achieved from [16].

Evaluation of imputation We compare our method and previous methods including stPlus [12], SpaGE[11] and Tangram [13] on 5-fold cross-validataion results over different ST datasets. To measure the similarity between predicted and true gene profiles, cosine similarity score (CSS) is calculated for each gene and aggregated either by the proportion of genes passing some threshold (e.g., > 0.2, in Fig.2.a) or median (Fig.3.a) within each dataset. Likewise, we aggregate MSE values by either thresholding a proportion (e.g., < 0.05, in Fig.2.b) or computing the median (Fig.3.b) for Moran's Is on predicted and true gene profiles for each dataset.

Evaluation on spatially highly variable gene detection To further evaluate the imputation methods, we assess the downstream task of detecting spatially highly variable genes from the imputed expression matrices. This evaluation can

only be conducted on Visium-based ST datasets, due to their almost whole genome-wide sequencing capacity that can capture enough positive and negative genes for analysis. We adopted both the classical Moran's I test [21] and the more recent non-parametric Spark-X test [20], and set the significance level to 0.01 for both methods. Viewing significant and non-significant results as binary classification, we may draw precision-recall curves (PRC) (in Fig.2.c and d, and Supp. Fig.S2 and S3) and summarize the performances of different methods as the area under the curve (AUC), which is a better metric than the area under the Receiver-Operating Characteristic curve in scenarios of label imbalance.

Evaluation on spatial ligand-receptor pair detection Beyond individual spatial patterns, we evaluate methods in a more challenging task that tries to identify spatially interactive ligand-receptor pairs. The recently developed method SpatialDM [29] leverages a bivariant Moran's statistic to detect spatial co-expression patterns of ligand and receptor pairs, and is used as the assessment tool in our evaluation. We set the significance level of FDR to 0.01, and after running the test for the ground truth and all the imputed expression matrices, we could also plot PRC and calculate the summary AUC for model comparison, as shown in Fig.2.e, and Supp. Fig.S4.

Spatial clustering evaluation We further evaluated our methods on downstream clustering analysis. The analysis was conducted on OsmFISH, MERFISH and seqFISH, where the number of cell types is available. Agglomerative clustering structured by a spatial adjacency matrix was applied to each dataset, including both the true and the predicted expression matrices. We then compare the clustering results of predicted and true expressions with averaged clustering indices of Adjusted Rand Score (ARS), Adjusted Mutual Information Score (AMIS), Homogeneity Score (HOMO), and Normalized Mutual Information Score (NMI).

Spatial RNA velocity exploration To explore the potential application of the proposed method on Spatial RNA velocity analysis, we fit TransImpSpa on reference datasets and translate its unspliced and spliced mRNA count matrices into the spatial space, where RNA velocity was inferred using scVelo [17].

Code availability The TransImp is implemented as an open-source Python package and freely at `https://github.com/qiaochen/tranSpa`. For reproducibility, all analysis notebooks are also available in this repository with links to the pre-processed datasets.

Acknowledgement

We thank Miss Zhuoxuan Li for kindly sharing the pre-processed Melanoma and Intestine datasets. We acknowledge support from the University of Hong Kong and its Li Ka Shing Faculty of Medicine as well as Centre for Translational Stem Cell Biology (CTSCB).

References

[1] T Lohoff, S Ghazanfar, A Missarova, N Koulena, N Pierson, JA Griffiths, ES Bardot, C-HL Eng, RCV Tyser, R Argelaguet, et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nature biotechnology*, 40(1):74–85, 2022.

[2] Jong-Eun Park, Rachel A Botting, Cecilia Domínguez Conde, Dorin-Mirel Popescu, Marieke Lavaert, Daniel J Kunz, Issac Goh, Emily Stephenson, Roberta Ragazzini, Elizabeth Tuck, et al. A cell atlas of human thymic development defines T cell repertoire formation. *Science*, 367(6480):eaay3224, 2020.

[3] Andrew L Ji, Adam J Rubin, Kim Thrane, Sizun Jiang, David L Reynolds, Robin M Meyers, Margaret G Guo, Benson M George, Annelie Mollbrink, Joseph Bergenstråhle, et al. Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma. *Cell*, 182(2):497–514, 2020.

[4] David Fawkner-Corbett, Agne Antanaviciute, Kaushal Parikh, Marta Jagielowicz, Ana Sousa Gerós, Tarun Gupta, Neil Ashley, Doran Khamis, Darren Fowler, Edward Morrissey, et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell*, 184(3):810–826, 2021.

[5] Anjali Rao, Dalia Barkley, Gustavo S França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, 2021.

[6] Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature Methods*, 19(5):534–546, 2022.

[7] Chee-Huat Linus Eng, Michael Lawson, Qian Zhu, Ruben Dries, Noushin Koulena, Yodai Takei, Jina Yun, Christopher Cronin, Christoph Karp, Guo-Cheng Yuan, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, 568(7751):235–239, 2019.

[8] Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, Xiaojie Qiu, Jin Yang, Jiangshan Xu, Shijie Hao, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022.

[9] Joshua D Welch, Velina Kozareva, Ashley Ferreira, Charles Vanderburg, Carly Martin, and Evan Z Macosko. Single-cell multi-omic integration compares and contrasts features of brain cell identity. *Cell*, 177(7):1873–1887, 2019.

[10] Tim Stuart, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William M Mauck III, Yuhan Hao, Marlon Stoeckius, Peter Smibert, and Rahul Satija. Comprehensive integration of single-cell data. *Cell*, 177(7):1888–1902, 2019.

[11] Tamim Abdelaal, Soufiane Mourragui, Ahmed Mahfouz, and Marcel JT Reinders. SpaGE: spatial gene enhancement using scRNA-seq. *Nucleic acids research*, 48(18):e107–e107, 2020.

[12] Chen Shengquan, Zhang Boheng, Chen Xiaoyang, Zhang Xuegong, and Jiang Rui. stPlus: a reference-based method for the accurate enhancement of spatial transcriptomics. *Bioinformatics*, 37(1):299–307, 07 2021.

[13] Tommaso Biancalani, Gabriele Scalia, Lorenzo Buffoni, Raghav Avasthi, Ziqing Lu, Aman Sanger, Neriman Tokcan, Charles R Vanderburg, Åsa Segerstolpe, Meng Zhang, et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nature methods*, 18(11):1352–1362, 2021.

[14] Dylan M Cable, Evan Murray, Luli S Zou, Aleksandrina Goeva, Evan Z Macosko, Fei Chen, and Rafael A Irizarry. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature Biotechnology*, 40(4):517–526, 2022.

[15] Vitalii Kleshchevnikov, Artem Shmatko, Emma Dann, Alexander Aivazidis, Hamish W King, Tong Li, Rasa Elmentaite, Artem Lomakin, Veronika Kedlian, Adam Gayoso, et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nature biotechnology*, 40(5):661–671, 2022.

[16] Tamim Abdelaal, Boudewijn PF Lelieveldt, Marcel JT Reinders, and Ahmed Mahfouz. SIRV: Spatial inference of RNA velocity at the single-cell resolution. *bioRxiv*, 2021.

[17] Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, and Fabian J Theis. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature biotechnology*, 38(12):1408–1414, 2020.

[18] Madhav Mantri, Gaetano J Scuderi, Roozbeh Abedini-Nassab, Michael FZ Wang, David McKellar, Hao Shi, Benjamin Grodner, Jonathan T Butcher, and Iwijn De Vlaminck. Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis. *Nature communications*, 12(1):1–13, 2021.

[19] Valentine Svensson, Sarah A Teichmann, and Oliver Stegle. SpatialDE: identification of spatially variable genes. *Nature methods*, 15(5):343–346, 2018.

[20] Jiaqiang Zhu, Shiquan Sun, and Xiang Zhou. SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome biology*, 22(1):1–25, 2021.

[21] Patrick AP Moran. Notes on continuous stochastic phenomena. *Biometrika*, 37(1/2):17–23, 1950.

[22] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[23] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

[24] Alexis Vandenbon, Rin Mizuno, Riyo Konishi, Masaya Onishi, Kyoko Masuda, Yuka Kobayashi, Hiroshi Kawamoto, Ayako Suzuki, Masahito Shimizu, Yasuhito Tanaka, Yutaka Suzuki, and Shinpei Kawaoka. Murine breast cancers disorganize the liver transcriptome in zonated manners. *bioRxiv*, 2022.

[25] Sunny Z Wu, Ghamdan Al-Eryani, Daniel Lee Roden, Simon Junankar, Kate Harvey, Alma Andersson, Aatish Thennavan, Chenfei Wang, James R Torpy, Nenad Bartonicek, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nature genetics*, 53(9):1334–1347, 2021.

[26] G Gambardella, G Viscido, B Tumaini, A Isacchi, R Bosotti, and D di Bernardo. A single-cell analysis of breast cancer cell lines to study tumour heterogeneity and drug response. *Nature communications*, 13(1):1–12, 2022.

[27] Kim Thrane, Hanna Eriksson, Jonas Maaskola, Johan Hansson, and Joakim Lundeberg. Spatially resolved transcriptomics enables dissection of genetic heterogeneity in stage iii cutaneous malignant melanoma. *Cancer research*, 78(20):5970–5979, 2018.

[28] David Fawkner-Corbett, Agne Antanaviciute, Kaushal Parikh, Marta Jagielowicz, Ana Sousa Gerós, Tarun Gupta, Neil Ashley, Doran Khamis, Darren Fowler, Edward Morrissey, et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell*, 184(3):810–826, 2021.

[29] LI Zhuoxuan, Tianjia Wang, Pengtao Liu, and Yuanhua Huang. Spatialdm: Rapid identification of spatially co-expressed ligand-receptor reveals cell-cell communication patterns. *bioRxiv*, 2022.

[30] Yalong Wang, Wanlu Song, Jilian Wang, Ting Wang, Xiaochen Xiong, Zhen Qi, Wei Fu, Xuerui Yang, and Ye-Guang Chen. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *Journal of Experimental Medicine*, 217(2), 2020.

[31] Gioele La Manno, Kimberly Siletti, Alessandro Furlan, Daniel Gyllborg, Elin Vinsland, Alejandro Mossi Albiach, Christoffer Mattsson Langseth, Irina Khven, Alex R Lederer, Lisa M Dratva, et al. Molecular architecture of the developing mouse brain. *Nature*, 596(7870):92–96, 2021.