SIMVI reveals intrinsic and spatial-induced states in spatial omics data

Mingze Dong[1,2,3], Harriet Kluger[4,5,6], Rong Fan[2,3], Yuval Kluger[1,2,7*]

[1]Interdepartmental Program in Computational Biology & Bioinformatics, Yale University, New Haven, CT, USA.

[2]Department of Pathology, Yale School of Medicine, New Haven, CT, USA.

[3]Department of Biomedical Engineering, Yale University, New Haven, CT, USA.

[4]Department of Medicine, Yale School of Medicine, New Haven, CT, USA.

[5]Yale Cancer Center, Yale School of Medicine, New Haven, CT, USA.

[6]Yale Center for Immuno-Oncology, Yale School of Medicine, New Haven, CT, USA.

[7]Applied Mathematics Program, Yale University, New Haven, CT, USA.

*Corresponding author(s). E-mail(s): yuval.kluger@yale.edu;

Contributing authors: mingze.dong@yale.edu; harriet.kluger@yale.edu; rong.fan@yale.edu;

Abstract

Spatial omics analyze gene expression and interaction dynamics in relation to tissue structure and function. However, existing methods cannot model the intrinsic and spatial-induced variation in spatial omics data, thus failing to identify true spatial interaction effects. Here, we present Spatial Interaction Modeling using Variational Inference (SIMVI), an annotation-free framework that disentangles cell intrinsic and spatial-induced latent variables for modeling gene expression in spatial omics data. SIMVI enables novel downstream analyses, such as clustering and differential expression analysis based on disentangled representations, spatial effect (SE) identification, SE interpretation, and transfer learning on new measurements / modalities. We benchmarked SIMVI on both simulated and real datasets and show that SIMVI uniquely generates highly accurate SE inferences in synthetic datasets and unveils intrinsic variation in complex real datasets. We applied SIMVI to spatial omics data from diverse platforms and tissues (MERFISH human cortex / mouse liver, Slide-seqv2 mouse hippocampus, Spatial-ATAC-RNA-seq) and revealed various region-specific and cell-type-specific spatial interactions. In addition, our experiments on MERFISH human cortex and spatial-ATAC-RNA-seq showcased SIMVI's power in identifying SEs for new samples / modalities. Finally, we applied SIMVI on a newly collected CosMx melanoma dataset. Using SIMVI, we identified immune cells associated with spatial-dependent interactions and revealed the underlying spatial variations associated with patient outcomes.

Main

In recent years, spatial omics technologies have become a prominent tool to explore tissue organization and function at fine-grained resolutions. Emerging imaging-based spatial transcriptomics (ST) technologies, such

as SeqFISH [1], MERFISH [2, 3], CosMx [4] are able to profile hundreds to thousands of genes at a single-cell or even subcellular resolution. Sequencing-based spatial omics technologies, such as DBiT-seq [5, 6], Slide-seqV2 [7], HDST [8], and Stereo-seq [9, 10], provide unbiased gene profiling at a near cellular resolution. These technologies enable researchers to investigate the effects of cellular interactions and niche factors on gene expression phenotypes at the cellular level.

However, a great deal of variability in cellular gene expression may not be attributable to its spatial context. Cells of different types, subtypes or cell cycle stages can display dramatically different gene expression patterns even through they may be in the same tissue niche. Revealation of the quantitative spatial niche effect requires distinguishing two sources of variation in spatial omics data, namely intrinsic variation and spatial-induced variation. Intrinsic variation can be understood as variation independent of neighborhood, such as cell types or phases of cell cycle; while spatial-induced variation can refer to general tissue context induced effects including niche-specific gene expression and the cellular interaction effect. For instance, within an developmental system, the variability across different cell types (such as neurons and fibroblasts) is accounted as the intrinsic variation. Simultaneously, the effect of morphogen gradients on cell transcriptomes should be recognized as spatial-induced variation [11].

Correct disentanglement of intrinsic and spatial-induced variation unveils how one cell with a certain phenotype interacts with its local environment giving rise to gene expression shift. Naturally, the task of disentangling intrinsic and spatial-induced variations is of fundamental importance in spatial omics data analysis. Yet, the task brings forth two significant challenges. One challenge is that cells of the same type often cluster together in space, creating spatial patterns that are not causally related to gene expression. Another challenge is that different cell types may have different responses to their local environment, particularly in the case of cellular interaction. This makes the spatial and intrinsic variations nonlinearly intertwined, significantly diminishing the efficacy of linear factorization methods in meaningful disentanglement (Fig. 1A).

To the best of our knowledge, existing computational approaches do not perform disentanglement of intrinsic and spatial-induced variation from spatial omics data. Existing analysis methods for spatial omics data either aims to identify spatial variable genes [12–17], select genes that mediate local cell interactions [18], predict patient phenotypes [19], or learn gene expression changes associated with local interactions [20]. These methods do not model intrinsic variations of cells or rely on annotated cell type labels assuming their equivalence with intrinsic variation. Two recent works on spatial-aware factor models (MEFISTO [21], NSF [22]) use linear factor models to distinguish spatial and non-spatial variations in spatial omics data. However, they do not address the two aforementioned fundamental challenges and thus cannot achieve a valid disentanglement between the intrinsic and spatial-induced variations.

In this work, we present Spatial Interaction Modeling using Variational Inference (SIMVI), a deep generative model for modeling gene expression in spatial omics data. SIMVI models the underlying gene expression variation of spatial omics data by two sets of latent variables. The first set, intrinsic variation, represents the intrinsic cell state independent of the neighborhood; the second set, spatial-induced variation, represent the variation induced by a cell's spatial context. In order to enforce meaningful disentanglement, SIMVI introduces a novel regularization term based on the mutual information and only explicitly penalize on the spatial-associated latent variables. After training, SIMVI returns the two sets of latent variables enabling further novel downstream analyses. Central to these is the inference of spatial effect (SE) at a single-cell level, using the learned latent variables. Apart from the SE inference, SIMVI enables clustering, visualization and differential expression analysis based on disentangled representations, and interpretation of the SE through either neighborhood enrichment analysis or ligand-receptor relationship prioritization. Finally, SIMVI enables SE identification of new samples or modalities based on the learned mapping from the original count matrix to the intrinsic embedding space.

We comprehensively benchmarked SIMVI in its ability to correctly identify SEs in both synthetic and real datasets. We applied SIMVI in a number of public datasets across different platforms and tissues, including MERFISH human cortex [3], Slide-seqv2 mouse hippocampus [7], MERFISH Vizgen mouse liver, and spatial-ATAC-RNA-seq mouse brain [23]. In these datasets, SIMVI uncovers a range of region-specific and cell-type-specific spatial interactions. Our experiments conducted on MERFISH human cortex and spatial-ATAC-RNA-seq datasets further highlight SIMVI's efficacy in detecting SEs for new samples and modalities. By applying SIMVI on a newly collected CosMx melanoma dataset, we identify macrophage niches associated with spatial dependent interactions and reveal the underlying spatial variations in CD8 T cells that are associated with patient responses.

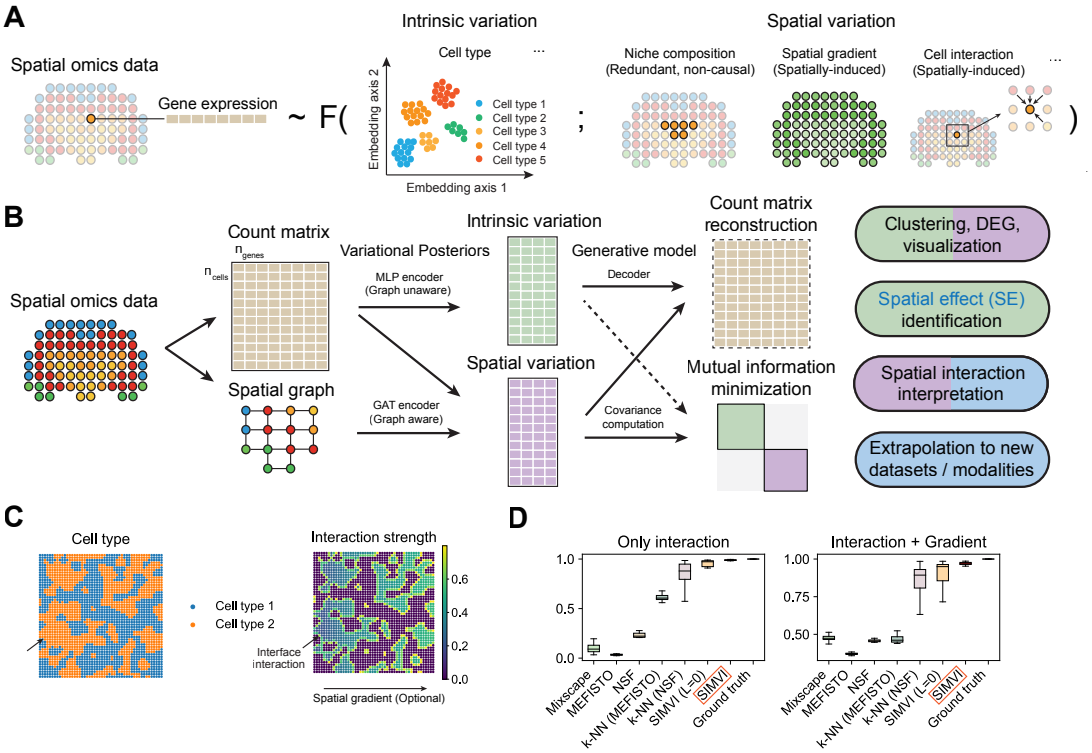

Fig. 1 SIMVI framework and evaluation on simulated data. **A.** Sources of gene expression variation in spatial omics data, including intrinsic factors (e.g., cell type) and spatial factors (e.g., niche composition, spatial gradient, cellular interaction). **B.** SIMVI workflow. Spatial omics data are converted into graph-structured data. Intrinsic and spatial variations are estimated by MLP and GAT variational posteriors, respectively. The loss function consists of the evidence lower bound (ELBO) and the mutual information regularization term. The regularization term is not used for updating intrinsic variation. **C.** Data simulation scheme. Two cell types were simulated on a grid with fixed cell type assignments. Upregulated gene programs were simulated on the cell type interface for cell type 2, and spatial gradients were added in some datasets. **D.** Boxplots of spatial effect (SE) identification performance using PCR-based metric for interaction only and interaction plus gradient scenarios. n=10 for interaction only, n=20 for interaction plus gradient. The top/lower hinge represents the upper/lower quartile and whiskers extend to the largest/smallest value within 1.5× interquartile range from the hinge. The median is shown as the center.

Results

Disentanglement of intrinsic and spatial-induced variations by SIMVI

The SIMVI framework disentangles intrinsic and spatial variations in gene expression. It models the gene expression x_i of each cell i by two sets of low-dimensional latent variables z, s indicating the intrinsic variation and spatial variation, respectively. The intrinsic latent variables z_i are modeled as variational posteriors of the gene expression x_i of cell i: $q(z_i|x_i)$, and the spatial latent variables s are modeled by graph neural network (GNN) variational posteriors, which are functions of neighborhood gene expression: $q(s_i|x_{N(i)})$. The intrinsic variation reflects cell-wise properties such as cell types and cell cycle phases, while the spatial variation captures various effects, such as differential cell composition in the neighborhood, global spatial gradient, and local cellular interaction (Fig. 1A). Among these effects, the latter two indicate meaningful spatial regulations (spatial-induced factors), while the local cell composition enrichment is not causally related to the central cell's expression profile. However with significant local composition enrichment effect, the spatial neighborhood expression would be highly predictive for the central cell. In this case, the GNN variational encoder may learn redundant variations with the intrinsic variation. The redundant variation need to be removed in the spatial latent variables to ensure meaningful disentanglement (Fig. 1A).

In order to remove the redundant variations in spatial latent variables, SIMVI incorporates a novel mutual information regularization term $I(s, z)$ in the loss term. The mutual information regularization penalizes the dependence between intrinsic and spatial latent variables. In practice, the mutual information can be effectively

estimated from data through empirical covariance matrices (See Methods, Fig. 1B). However, the vanilla mutual information regularization equally penalize intrinsic and spatial embeddings. In our setting, we want to attribute the cell type variation solely to the intrinsic variation z, and penalize only the spatial variation s to eliminate the redundancy. To achieve this, we optimize only the spatial latent variable s with the regularization term, while we update the intrinsic variation z based on the original VAE loss term (See Methods).

The SIMVI model is an variational autoencoder framework implemented with the scvi-tools [24]. SIMVI can be efficiently optimized and easily scaled to most existing spatial omics datasets. After training, SIMVI learns two embeddings that indicate the intrinsic variation and spatial variation in the spatial omics data, respectively. This disentangled representation allows characterization of cell intrinsic heterogeneity and spatial heterogeneity, via clustering, differential gene expression analysis, and visualization.

Using the disentangled representations, SIMVI can model the quantitative effect of spatial neighborhood on individual gene expression. We define the spatial effect (SE) for each cell as the difference between its observed expression and its "spatial null", which has the same intrinsic variation but no spatial variation. We construct the spatial null as the average of k neighboring cells in the intrinsic variation space. Our proposed k-NN estimator of SE is robust to model fitting and can be applied to other modalities (see Methods). We can cluster cells based on their SEs, or further divide them into SE subclusters according to pre-annotated labels (e.g., cell types). In practice, the spatial effect is often dominated by a null cluster that represents homogeneous transcriptomic states across different locations, which may include multiple cell types. By contrast, the "outlier" clusters represent cells with significant spatial effects. We can interpret SEs by either performing differential expression analysis of SE subclusters, or by examining their spatial neighborhoods. One option for the latter is neighborhood enrichment analysis, which is implemented as a workflow in Squidpy [25]. This approach identifies positive and negative co-occurrence of SE subclusters within neighborhoods using permutation tests. Another option is to use a ligand-receptor database to prioritize important ligand-receptor pairs that are strongly correlated with SEs or spatial variation. This option is particularly useful, as existing methods mainly use the average expression of ligand-receptor pairs as an indicator of cellular interaction [26–30]. While these expression-based methods can detect transcriptionally active ligand-receptor pairs, SEs are likely to be regulated by predictive, rather than highly expressed, ligand-receptor pairs. The prioritization of crucial ligand-receptor pairs potentially involved in spatial regulation is a key factor in understanding cellular interactions and represents a distinctive feature of SIMVI.

Evaluation of SIMVI

To demonstrate SIMVI's ability to identify SEs, we first compared SIMVI with alternative methods on simulated data, where we knew the ground truth. We simulated two common mechanisms of spatial interaction: the spatial gradient and local cellular interaction, as collective gene regulatory programs (Fig. 1C, see Methods for details). To evaluate different methods' performance in SE identification, we introduced a novel principal component regression-based score that measures the preservation of spatial variation in the embedding (see Methods). This metric is suitable for our case as it can handle both categorical and continuous covariates and it operates in the latent space, enabling comparison with decomposition-based methods (MEFISTO, NSF). We also used the Adjusted Rand Index (ARI) and the Normalized Mutual Information score (NMI) to assess the accuracy of intrinsic variation identification via cell type separation.

We compared SIMVI with a spatially-unaware baseline that used k-NN in the original PCA space to estimate perturbation effects from scRNA-seq data (Mixscape [31]), MEFISTO, NSF, and a SIMVI model without mutual information regularization (L=0). For MEFISTO and NSF, we also included the k-NN SE based on each cell's expression difference from its neighbors in the non-spatial variation space. We selected the optimal hyperparameters for MEFISTO, NSF, and SIMVI through parameter-sweep analysis (see Methods and Supplementary Fig. 1). SIMVI outperformed all other methods by a large margin, being the only method with comparable performance to the ground truth SE (Fig. 1D). The UMAP visualization of an example dataset further supported the accuracy of SIMVI in identifying intrinsic variation, spatial-induced variation, and SE (Supplementary Fig. 2). The baseline SIMVI model without mutual information regularization (L=0) ranked second, whose lower performance than the full model is conceptually due to the unidentifiability (Fig. 1A). The unidentifiability was also evident from the visualization, where the cell type separation was still present in the spatial variation (Supplementary Fig. 2). Interestingly, the k-NN SE based on non-spatial representations for MEFISTO and NSF performed better than their counterparts using the spatial variation, validating the effectiveness of the k-NN-based construction for SE computation. However, the k-NN MEFISTO and NSF SEs

still fell far behind the SIMVI models. The Mixscape baseline also failed to identify SEs effectively, because it matched cells with similar interaction states and eliminated true SEs (Fig. 1D). Finally, due to the simplicity of the simulated datasets, most models achieved reasonably high performance in cell type separation in the non-spatial variation space (Supplementary Fig. 1).

We also compared SIMVI with MEFISTO and NSF in terms of scalability. For both MEFISTO and NSF, the running time depends on the number of inducing points. Therefore, we tested MEFISTO and NSF with two different settings of inducing points. SIMVI performed much faster and used less memory than MEFISTO and NSF when the cell number was larger than 5,000, even in their favorable settings with fewer inducing points (Supplementary Fig. 3). SIMVI could process 20,000 cells with 1,000 genes in under 10 minutes, with a peak memory usage of around 10GB. This indicates that it can be effectively run on modern laptop computers. Whereas MEFISTO and NSF, due to their slower speeds, were not benchmarked for larger scale datasets, SIMVI performed all experiments in the paper, accommodating up to 60,000 cells, in less than an hour on a Linux server.

SIMVI reveals spatial dependencies in MERFISH human cortex

MERFISH is an imaging-based spatial transcriptomics (ST) technology that enables spatial profiling of gene expression at a subcellular resolution [2, 3]. While imaging-based ST approaches generally suffers from limited numbers of profiled genes, in a recent published MERFISH dataset, the researchers have achieved profiling of 4,000 genes in human cortex [3], including both middle temporal gyrus (MTG) and superior temporal gyrus (STG). The dataset provides an ideal case for application of SIMVI in disentangling different sources of variation and identifying spatial effects (SEs).

In this work, we analyzed two MERFISH slices sampled from human MTG and STG respectively (Fig. 2A, Supplementary Fig. 4A). In both datasets, SIMVI identified cell-type associated variation in intrinsic embeddings and layered spatial patterns (Fig. 2B, Supplementary Fig. 4B) [3]. After obtaining the disentangled variations from SIMVI, we then performed SE identification for each cell in both datasets. In both cell populations, we found various "outlier" clusters (indicating non-zero SEs) in exhibitory (EXC) and inhibitory neurons (INC), and a subcluster of endothelial cells (oENDO) and mural cells (oMURAL) (Fig. 2C-D, Supplementary Fig. 4C-D). Among these, the excitatory and inhibitory neuronal SE subclusters were characterized by distinct KEGG pathways and marker genes, such as *LAMP5*, *ERBB4*, and *KIT* (Supplementary Fig. 5A-D). The endothelial and mural cells from the SE subcluster were spatially co-located, suggesting vascular structures, and had higher expression of *MYH11* and enrichment of a vascular-associated pathway (Vascular smooth muscle contraction, Supplementary Fig. 5A-D) [32]. In contrast, this pathway was not enriched in the non-vascular counterpart mural cell subcluster (oMural_6 cluster in Supplementary Fig. 5C). Overall, The SIMVI SE clusters were consistent with the fine-grained annotation that incorporates the cortex spatial structure [3], while SIMVI uniquely highlighted the SE subcluster resulted from the cellular interaction in the vascular interface (Supplementary Fig. 5A-B), consistent with the previous conclusion that also used the imaging modality [3]. In contrast, the spatial effect derived from the cell type annotation (with the same configuration as SIMVI SE) successfully separated the spatial subclusters, yet due to the heterogeneity within cell types, the resulting embedding was clearly biased by cell types hence did not reflect the quantitative spatial effect (Supplementary Fig. 5E-F). We further explored the spatial dependencies in this dataset using ligand-receptor (LR) prioritization analysis. In MTG, we found LR pairs that were expressed by SE clusters, such as *ERBB4*, a marker gene for a SE subcluster, and *SLIT3*, which was expressed by vascular-associated endothelial and mural cells (Fig. 2E and Supplementary Fig. 5G). In STG, we found various genes that were involved in LR pairs and expressed by spatial-associated subclusters, such as *SORBS1*, *ERBB4*, *AGT*, and *FGF13*, indicating potential spatial regulations (Supplementary Fig. 4E and Supplementary Fig. 5H).

Finally, we evaluated SIMVI with alternative methods, including the PCA baseline, MEFISTO, and NSF. Because we did not have a definitive ground truth SE for this dataset, the assessment of different methods' performances on SE inference was conceptually impossible. Therefore, here we benchmarked the intrinsic (non-spatial) variation returned by each method. The intrinsic variation serves as an intermediate step in spatial effect computation and directly determines quality of spatial effect inference. We used two types of metrics to evaluate the intrinsic variation: cell type preservation (NMI: normalized mutual information score, ARI: adjusted Rand index) and removal of spatial variation (PCR: modified principal component regression score, ASW: average silhouette width per batch, see Methods) (Fig. 2F). We found that SIMVI outperformed both NSF and MEFISTO in terms of cell type preservation and spatial variation removal. The superior performance

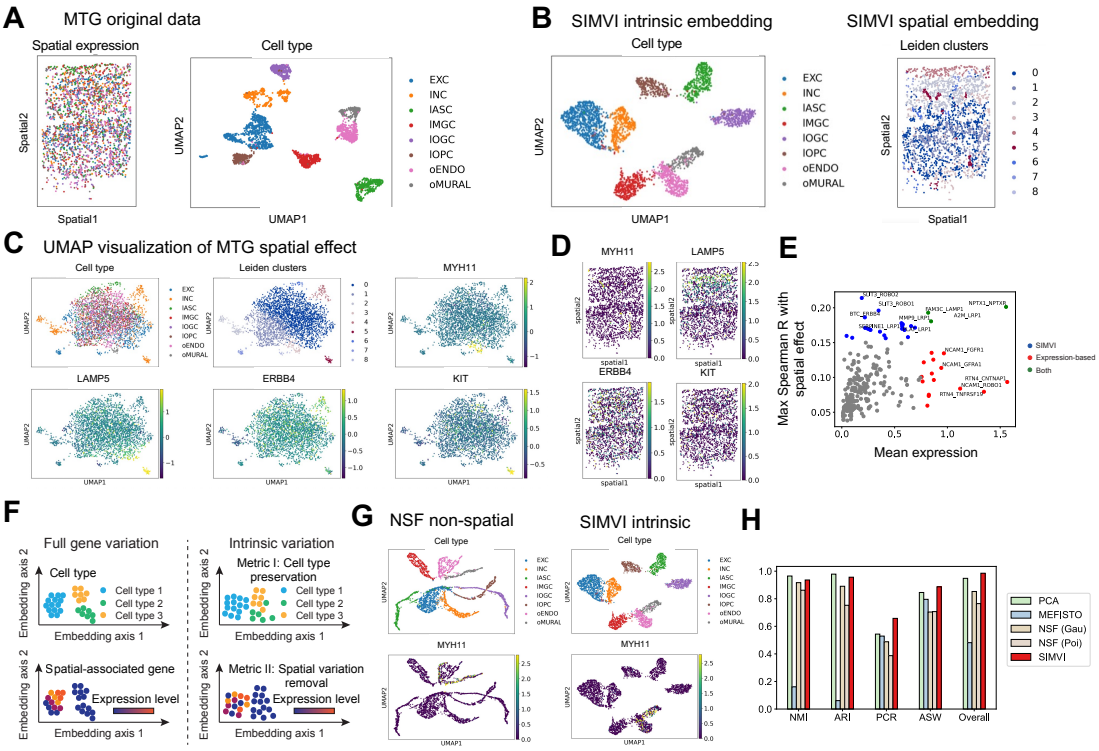

Fig. 2 SIMVI identifies cell-type specific and spatial variations in MERFISH human cortex data. **A.** Spatial (Left) and UMAP (Right) visualization of MERFISH MTG data, colored by annotated cell types. **B.** UMAP visualization of SIMVI intrinsic embedding colored by cell types (left), and spatial visualization of clusters from SIMVI spatial embedding (right). **C.** UMAP visualization of MTG spatial effect (SE) estimated by SIMVI, colored by cell types, clustering, and expression of spatial-associated genes identified by SIMVI (*MYH11*, *LAMP5*, *ERBB4*, *KIT*). **D.** Spatial visualization of genes *MYH11*, *LAMP5*, *ERBB4*, *KIT*. **E.** Ligand-receptor pair prioritization scatter plot for MERFISH MTG data. Grey: not prioritized. Blue: prioritized by high correlation with SIMVI SE. Orange: prioritized by high expression. Green: prioritized by both criteria. **F.** Schematic overview of SIMVI evaluation for real data. Evaluations are based on cell type preservation metrics and spatial variation removal metrics. **G.** UMAP visualization of NSF non-spatial variation and SIMVI intrinsic variation, colored by cell types and *MYH11* expression. **H.** Bar plot of average performance of different methods on MERFISH datasets, in terms of quantitative metrics (Metric I: normalized mutual information score (NMI), adjusted Rand index (ARI); Metric II: principal component regression score (PCR), average silhouette width per batch (ASW)). NSF(Gau) and NSF(Poi) represent two likelihood settings for the NSF model.

of SIMVI was also evident from the UMAP visualizations, where SIMVI balanced cell type preservation and spatial variation removal of the spatial-associated gene *MYH11* (Fig. 2G, Supplementary Fig. 6A-D). The cell type preservation performance of SIMVI was comparable to PCA, but its spatial variation removal performance was better, resulting in the highest overall performance (Fig. 2H, Supplementary Fig. 6E).

SIMVI identifies region-specific gene expression in Slide-seqV2 mouse hippocampus

Slide-seqV2 is a sequencing-based spatial transcriptomics technology that enables unbiased spatial profiling of genes at near-cellular resolution [7]. We applied SIMVI to the widely used Slide-seqV2 mouse hippocampus data [7] (Fig. 3A). Different cell types were well separated in the space of intrinsic variation, wheras the spatial-induced variation included a “null” cluster with multiple cell types and several outlier clusters, such as two subclusters of Endothelial tip cells (Fig. 3B). Further differential expression analysis revealed that these two subclusters were enriched in distinct spatial locations, marked by the genes *Ptgds* and *Ttr*, consistent with previous findings (Fig. 3C) [7, 22].

We observed that the SE clusters did not show significant separation in the UMAP space, but still represented spatial enrichment (Fig. 3D). For instance, different cell types within subcluster 6 tended to colocalize, as confirmed by neighborhood enrichment analysis (Fig. 3E) and spatial visualization (Fig. 3F). Further analysis showed that these localized cells express high levels of the gene *Ttr*, regardless of cell types, when compared

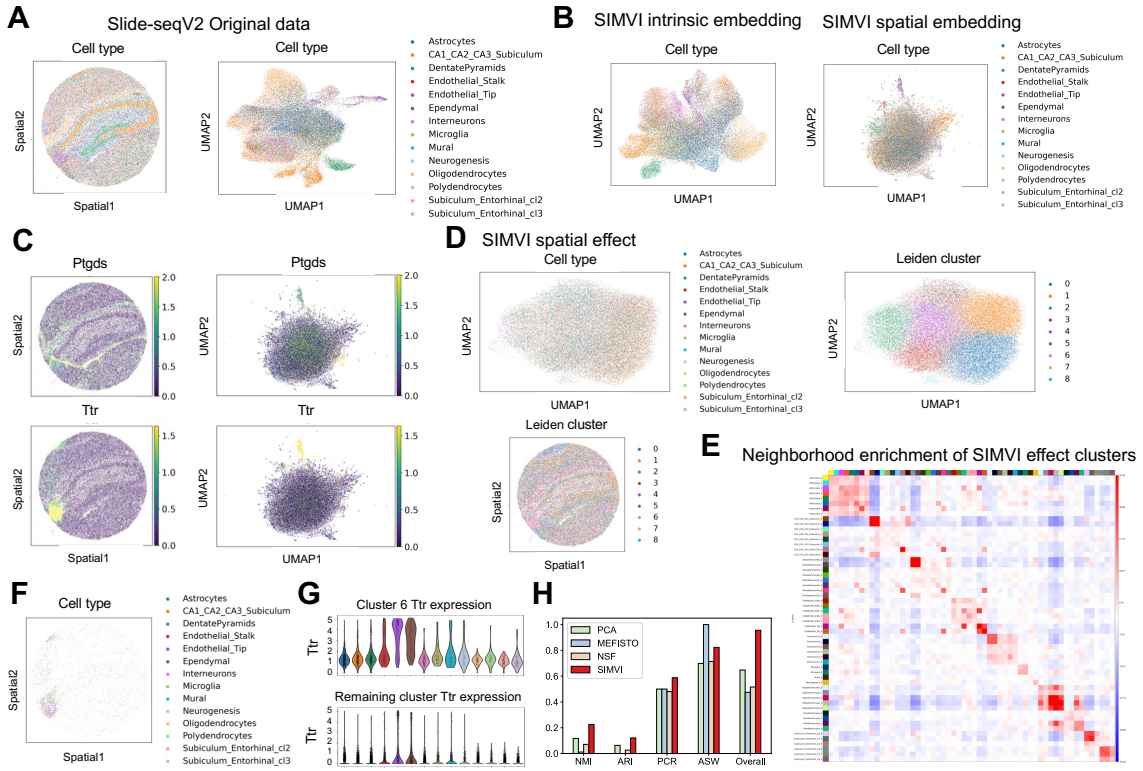

Fig. 3 SIMVI identifies region-specific variations in Slide-seqV2 mouse hippocampus data. **A.** Spatial (left) and UMAP (right) visualization of Slide-seqV2 mouse hippocampus data, colored by annotated cell types. **B.** UMAP visualization of the SIMVI intrinsic variation (left) and spatial variation (right), colored by annotated cell types. **C.** Spatial visualization and UMAP visualization of the SIMVI spatial variation showing *Ptgds* and *Ttr* expression. **D.** UMAP and spatial visualization of the spatial effect (SE) estimated by SIMVI, colored by cell types (UMAP) and clustering (UMAP, spatial). **E.** Neighborhood enrichment plot of SIMVI SE subclusters. Each row/column represents a SE subcluster by dividing the SE clusters based on the cell type label. Each entry represents the permutation z-score between two SE subclusters. **F.** Spatial visualization of SIMVI effect cluster 6, colored by cell types. **G.** Comparative violin plots of *Ttr* expression for all cell types (with color scheme as in **F**) in SIMVI effect cluster 6 and the remaining clusters. **H.** Bar plot of performance of different methods on the Slide-seqV2 dataset, in terms of the quantitative metrics (Metric I: normalized mutual information score (NMI), adjusted Rand index (ARI); Metric II: Principal component regression score (PCR), average silhouette width per batch (ASW)).

with the remaining clusters (Fig. 3G). This observation signifies a region-specific gene expression pattern in the data, which might be attributed to niche effects and warrants further exploration.

Finally, we benchmarked different methods, including the PCA baseline, MEFISTO, and NSF, to identify intrinsic variations through cell type preservation and spatial effect removal. In this dataset, the SIMVI intrinsic embedding separated cell types better than all other tested methods. While MEFISTO achieved high performance in spatial effect removal in terms of ASW, it failed to separate different cell types as reflected by the low values of NMI and ARI and by UMAP visualizations (Supplementary Fig. 7). SIMVI achieved the highest score in the first three metrics (NMI, ARI, and PCR), providing the highest overall performance (Fig. 3H), which was also supported by qualitative UMAP visualizations (Supplementary Fig. 7).

SIMVI captures spatial subclusters in Vizgen liver

While the MERFISH cortex data and the Slide-seqV2 mouse hippocampus data are ideal datasets for SIMVI due to the high gene number and (near) single-cell level resolution, typical imaging-based spatial transcriptomics (ST) datasets include a low number of genes. In order to showcase SIMVI's power in general imaging-based ST data analysis, we applied SIMVI in MERFISH Vizgen mouse liver data, in which 385 genes are profiled.

In this dataset, SIMVI captured cell-type specific variations in the intrinsic embedding and spatial environment difference in vessel (cluster 2) and non-vessel (cluster 0,1) spatial regions (Fig. 4B). More detailed visualization of marker genes *Cyp1a2* and *G6pc* suggested that cluster 0 and 1 corresponded approximately to peri-central

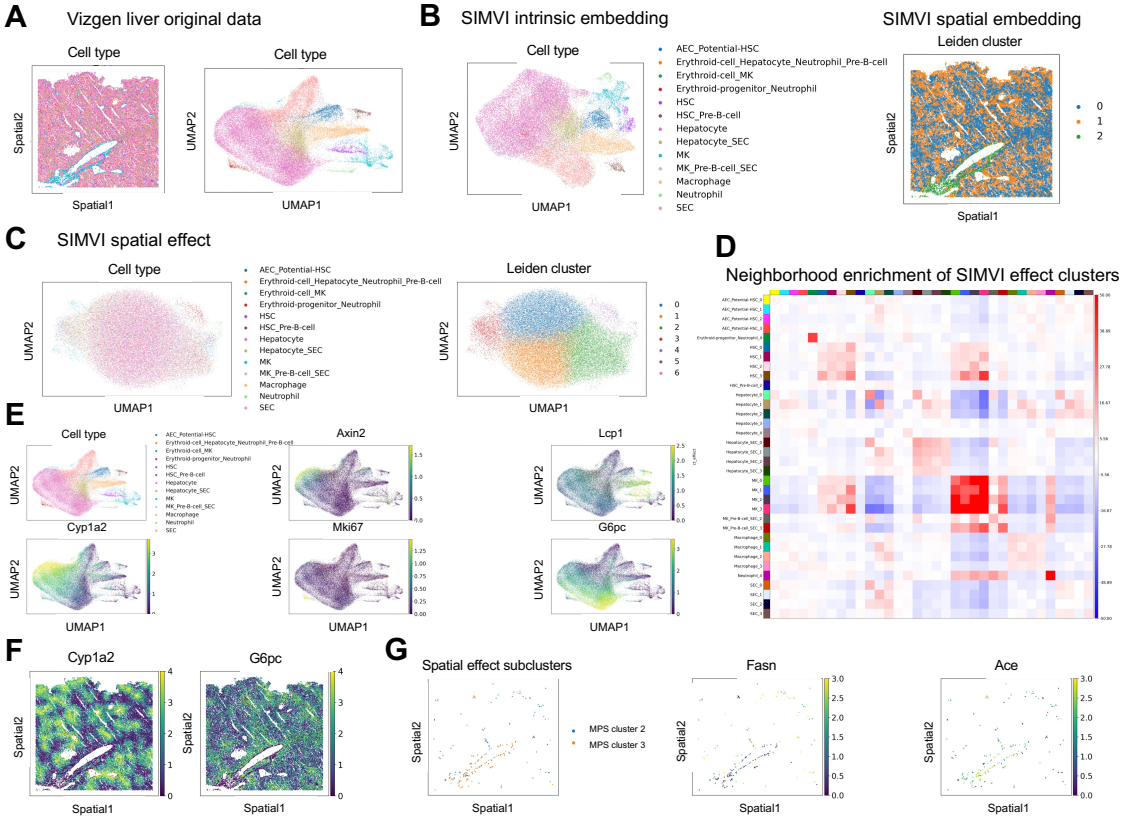

Fig. 4 SIMVI identifies spatial cell-type specific variations in Vizgen liver data. **A.** Spatial (left) and UMAP (right) visualization of Vizgen mouse liver data, colored by annotated cell types (with color scheme as in **B**). **B.** UMAP visualization of the SIMVI intrinsic embedding (Left) colored by cell types and spatial visualization of clusters from the SIMVI spatial embedding (Right). **C.** UMAP visualization and spatial visualization of the spatial effect (SE) estimated by SIMVI, colored by cell types and clustering respectively. **D.** Neighborhood enrichment plot of SIMVI SE subclusters. Each row / column represents a SE subcluster by dividing the SE clusters based on the cell type label. Each entry represents the permutation z-score between two SE subclusters. **E.** UMAP visualization of marker genes selected from SIMVI SE clusters. **F.** Spatial visualization of *Cyp1a2* and *G6pc*. **G.** Spatial visualization of identified SIMVI effect subclusters (MPS cluster 2 and 3) and expression of representative genes (*Fasn* and *Ace*).

and peri-portal regions, respectively (Fig. 4F). Using SIMVI SE analysis, we identified different SE subclusters within hepatocytes (Fig. 4C). To show the meaning of these SE subclusters, we selected representative marker genes for each SE cluster and showed their expression in the original gene expression UMAP space (Fig. 4E). We found that all selected genes exhibited expression patterns that were not fully determined by their underlying cell types. In particular, *Akin2* was selected as the marker gene with a leading score for SE cluster 1, which was highlighted as a marker for peri-central hepatocytes [33]. On the other hand, *G6pc*, the leading marker gene for SE cluster 6, was enriched in complementary peri-portal hepatocyte populations. These findings are consistent with prior knowledge and the clustering result of the SIMVI spatial embedding (Fig. 4B). To further explore cell-type specific SE patterns, we performed neighborhood enrichment analysis using cell-type fine-grained SE subclusters (Fig. 4D). Through the analysis, we identified two subclusters within the MK_Pre-B_SEC (MPS) cell type that have distinct neighborhoods. Further investigation revealed that the subclusters are associated with vessel and non-vessel regions both spatially and functionally (Fig. 4G, Supplementary Fig. 8).

SIMVI enables transfer learning across measurements and modalities

The SIMVI model learns two variational posteriors, with one mapping from each cell's gene expression to the intrinsic variation space, the other mapping the cell neighborhood profile to the spatial-induced variation. Consequently, the two mappings can be utilized to extract variation from other measurements of the same tissue, presuming the generative process of the data is consistent. Moreover, as our definition of spatial effect (SE) depends on only the original count matrix and the learned intrinsic variation, the quantitative SE can be

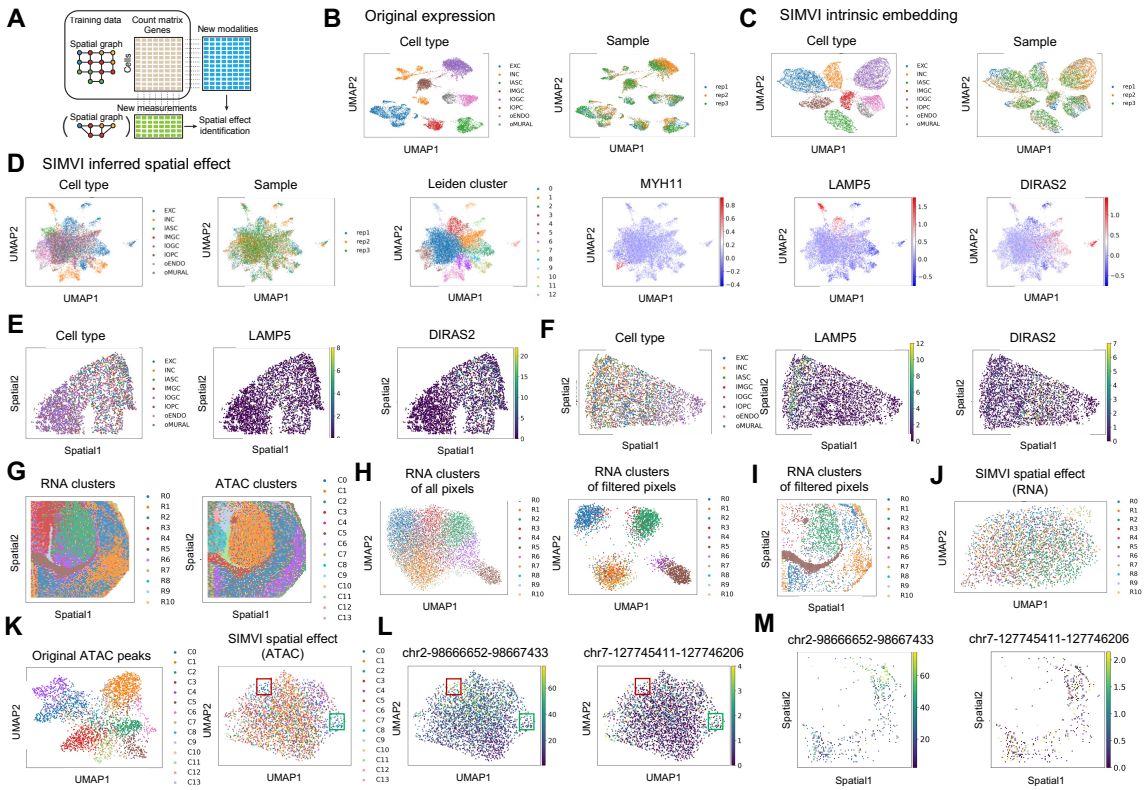

Fig. 5 SIMVI enables spatial effect (SE) identification for new measurements / modalities. **A.** Illustration of SIMVI SE identification for new measurements or modalities. **B.** UMAP visualization of the MERFISH human MTG data, colored by annotated cell types and replicate label. **C.** UMAP visualization of the SIMVI intrinsic embedding colored by cell types and replicate label. **D.** UMAP visualization of the SE estimated by SIMVI, colored by cell types, replicate label, and representative gene (*MYH11*, *LAMP5*, *DIRAS5*) effect. **E.** Spatial visualization of the replicate 2, colored by cell types and expression of *LAMP5* and *DIRAS5*. **F.** Spatial visualization of the replicate 3, colored by cell types and expression of *LAMP5* and *DIRAS5*. **G.** Spatial visualization of the spatial-RNA-ATAC mouse brain data, colored by RNA cluster labels and ATAC cluster labels from [23]. **H.** UMAP visualization the RNA expression for all pixels and filtered pixels. **I.** Spatial visualization of filtered pixels colored by annotated RNA clusters. **J.** UMAP visualization of SIMVI SE, colored by annotated RNA clusters. **K.** UMAP visualization of TF-IDF normalized ATAC peaks and the ATAC SE, colored by annotated ATAC clusters. **L.** UMAP visualization of the ATAC SE, colored by two representative peaks. **M.** Spatial visualization of the filtered ATAC C0 cluster colored by two representative peaks.

directly learned from new data that is generated with a non-spatial measurement. Moreover, in spatial multi-omics data [23, 34], the SE from another modality that is not used for SIMVI can be directly obtained by the neighborhood in the RNA intrinsic variation space. Both the task of learning SE from datasets that do not have spatial information and learning SE from arbitrary additional modalities are unique features of SIMVI (Fig. 5A).

In order to showcase SIMVI's power in extrapolations to both new measurements and new modalities, we performed experiments on MERFISH human MTG cortex data replicates (Fig. 5B) and spatial-ATAC-RNA-seq data [23] (Fig. 5G). Due to the lack of paired scRNA-seq measurements with single-cell level ST datasets, we ignored the MERFISH replicates' spatial coordinates, using their count matrices to imitate measurements without spatial information. We used the trained SIMVI model on one MTG replicate (described in the previous section) to infer SE on the remaining replicates. SIMVI identified the shared cell type variation in MERFISH MTG data (Fig. 5C) and shared SEs highlighted by representative genes (Fig. 5D), without using the spatial information of the replicates. The relevance of the selected genes was confirmed through their spatial distributions in both replicates (Fig. 5E-F).

In the Spatial-ATAC-RNA-seq dataset, the pixel resolution ($20\mu m$) is larger than the cell size. Therefore, each pixel may represent a mixture of cells resulting in unidentifiability between spatial gradient and continuous cell type composition change along space. As a result, we first applied a strict thresholding scheme to select the pixels with high likelihood to contain pure cell types from the dataset through unsupervised archetypal

analysis (See Methods, Supplementary Fig. 9A). After filtering, the subset of pixels formed 4 well separated clusters in the UMAP space, rather than a continuum for the original pixels, yet still covering most regions in space (Fig. 5H-I). By applying SIMVI to the RNA modality of the filtered multi-modal data, we identified two subclusters with non-zero SEs, corresponding to a subcluster of R3 cells and the cells from RNA cluster R7 respectively (Fig. 5J). The two SEs both indicate meaningful spatial patterns indicated by representative genes *Meg*, *Mbp*, and that of *Tpm1* respectively (Supplementary Fig. 9D-E).

Based on the intrinsic representation, the spatial effect (SE) of ATAC can be established by the k-NN estimator as previously described. However, we discovered that this estimator failed to consider the unique variations solely exhibited in the ATAC modality, leading to a dominant ATAC cluster effect overshadowing the spatial effect (Supplementary Fig. 9F). To overcome this, we concatenated the one-hot representation of the ATAC cluster label with the SIMVI intrinsic embedding, serving as a soft regularization for the k-NN SE to more robustly match pixels from the same ATAC cluster. The resulting ATAC SE is shown in Fig. 5K. For an ablation test, we also evaluated the k-NN SE of the same configuration, derived solely from the ATAC cluster label. We observed similar ATAC cluster-specific effects (Supplementary Fig. 9F), which further suggested the essence of the SIMVI intrinsic embedding in estimating the ATAC SE. To showcase the validity of the computed spatial effects, we focused on the differential SE behavior observed in ATAC cluster C0. The UMAP visualization shows two subclusters (marked by red and green squares). The upper-left subcluster (the red square) was marked by the ATAC peak chr2-98666652-98667433, while another marker peak chr7-127745411-127746206 was shared across the two subclusters (Fig. 5L). Further spatial visualization revealed that the first peak was indeed enriched spatially in the upper-right area, while the latter peak had dispersed spatial distribution (Fig. 5M), confirming the validity of the two subclusters.

SIMVI captures spatial interactions in cohort Melanoma data

Melanoma is the most common cause of skin cancer related fatalities, resulting in over 7000 deaths per year in the United States [35]. In recent years, the death rate has dropped dramatically due to the advent of immune checkpoint inhibitors which reverse T cell exhaustion and disinhibit effector T cells, among other functions [36]. However, not all patients respond to these therapies, and extensive efforts are ongoing to understand mechanisms of resistance or sensitivity. Powerful new technologies, such as the Nanostring CosMx platform, provide gene expression profiles in situ with single cell resolution and preservation of spatial information [4]. Here we employed SIMVI to explore the intricate spatial interactions between tumor and immune cells in CosMx samples from 25 melanoma patients treated with immune checkpoint inhibitors with various outcomes (Fig. 6A-C).

Tumor cells are expected to be highly heterogenous when comparing across patients, due to the variability in mutations in melanoma, whereas the heterogeneity in cells in the tumor microenvironment was less pronounced. Due to these fundamental differences, we ran the SIMVI model on all cells while focusing on the non-tumor cell subsets for downstream analysis and interpretation (Fig. 6D). In non-tumor cells, SIMVI identified cell-type specific variation as shown by the separated cell type clusters in the UMAP spaces (Fig. 6E, Supplementary Fig. 10D), while the spatial variation was not primarily driven by cell type variation but showed separation with respect to the likelihood of response to treatment (Fig. 6F, Supplementary Fig. 10E).

Next we aimed to further investigate the SE of non-tumor cells, particularly immune cells, in tumor microenvironment. We found that macrophages exhibited heterogeneous SEs (Supplementary Fig. 10F). As a result, we focused on the SE of macrophages. The SE clustering of macrophages showed two distinct clusters marked by high (Leiden cluster 3) versus low (Leiden clusters 0-2) expression levels of *SPP1*, a marker for tumor-infiltrating macrophages. The low-*SPP1* macrophage group can be further characterized by the state of the complement pathway (*C1QA*, *C1QB*, *C1QC*) [38] and expression of inflammatory mediators (*LYZ*, *CXCL9*), creating 3 subclusters with different combinations of complement activation level and inflammatory marker level (Fig. 6G). Further pathway enrichment analysis confirms the three distinct subclusters of low *SPP1* macrophages demonstrate varying immunological states, as depicted through differential enrichment of KEGG immunological processes (Fig. 6G) [32]. Our identified macrophage subclusters were consistent with the existing knowledge of tumor-associated macrophages (TAMs) [39]. The validity of the computed SE was supported by plotting of the original gene expression in the SE space (Supplementary Fig. 11B). Moreover, the macrophage SE embedding effectively harmonized over the patient samples providing more interpretable clustering, while the individual patient variability played a prominent role on the original gene expression space (Supplementary Fig. 11A). The tumor infiltrating role of the high *SPP1* cluster was supported by the spatial

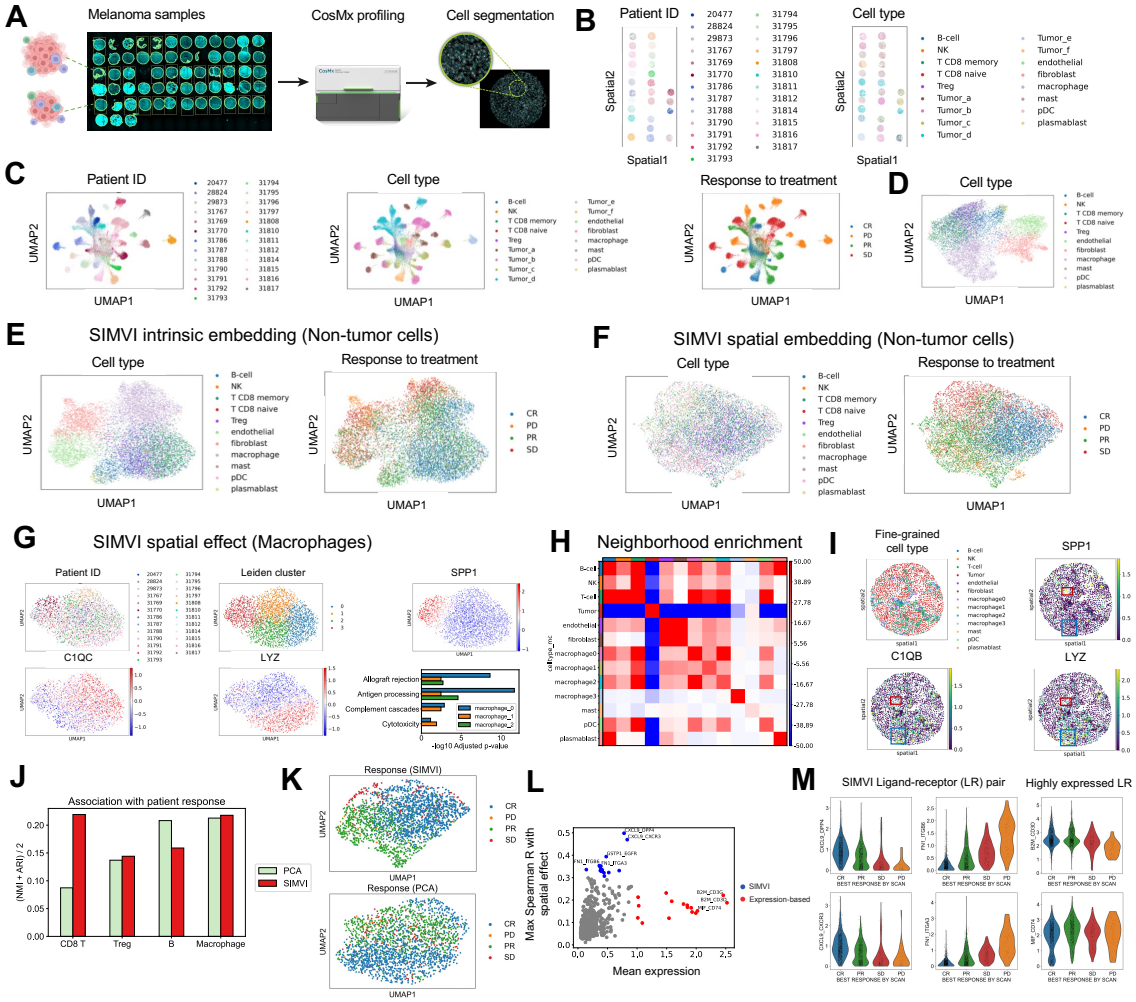

Fig. 6 SIMVI identifies macrophage populations and T cell spatial variations in melanoma samples profiled by CosMx. **A.** Schematic overview of the experimental pipeline. **B.** Spatial visualization of the data, colored by annotated patient ID and cell types. **C.** UMAP visualization of the data gene expression, colored by patient ID, cell type, and response to treatment. The response to treatment was defined as disease progression (PD), stable disease (SD), partial or complete response (PR, CR), as described in prior publications [37]. **D.** UMAP visualization of the non-tumor cell subset, colored by cell type. **E.** UMAP visualization of the SIMVI intrinsic embedding colored by cell type and best radiographic response. **F.** UMAP visualization of the SIMVI spatial embedding colored by cell type and best radiographic response. **G.** The first five panels show UMAP visualization of the macrophage SE estimated by SIMVI, colored by cell types, Leiden clusters and differentially expressed genes. The last panel shows enrichment of selected KEGG immunological pathways for low *SPP1* macrophage subclusters, including allograft rejection, antigen processing, complement cascades and NK-mediated cytotoxicity. **H.** Neighborhood enrichment plot of SIMVI fine-grained cell types. Macrophages are divided into 4 clusters according to the SE analysis. Other cell subtypes from each cell type were aggregated. **I.** Spatial visualization of a selected patient sample (31788) with different macrophage niches, colored by fine-grained cell type and expressions of *SPP1*, *C1QB*, *LYZ*. **J.** Bar plot of different immune cell types' average value of NMI (normalized mutual information score) and ARI (adjusted Rand index) using PCA and SIMVI spatial embedding. **K.** UMAP visualization of CD8 T cells using the SIMVI spatial embedding (upper) and PC embedding (lower), colored by patient response. Here the patient response was separated in the SIMVI UMAP space of CD8 T cells, but mixed in the PCA UMAP space of these cells. **L.** Ligand-receptor pair prioritization scatter plot for CD8 T cells. Grey: Not prioritized. Blue: Prioritized by high correlation with SIMVI SE. Orange: Prioritized by high expression. **M.** Violin plots of selected SIMVI prioritized ligand-receptor (LR) pairs and highly expressed LR pairs, grouped by patient response.

neighborhood enrichment plot (Fig. 6H), where most non-tumor cell types, including macrophage clusters 0-2 show strong negative enrichment with respect to tumor while macrophage cluster 3 does not. By additional spatial visualization of patient tumor examples, we found that the high *SPP1* macrophage population had a different spatial niche enrichment (red square) than macrophages enriched with the inflammatory markers (Blue square, Fig. 6I).

As immune checkpoint inhibitors function by activating exhausted or dysfunctional effector T cells, we expect that the immunotherapy modulates the interaction dynamics between effector T cells and tumor cells. The efficacy of the immunotherapy is directly reflected by tumor response and shrinkage, and should be associated with spatially-induced variation in effector T cells [40, 41]. Quantifying this association revealed that in CD8 T cells, the SIMVI spatial embedding had a higher association score than the PC embedding. Interestingly, this pattern was absent in other immune cells, including the transcriptomically akin Treg cells (Fig. 6J, Supplementary Fig. 11C). This indicates that SIMVI selectively prioritized the spatial-induced variation of CD8 T cells consistent with the underlying biology. The stronger association between SIMVI spatial embedding and patient outcome was also directly verified by the UMAP visualization (Fig. 6K). Further, our SIMVI ligand-receptor (LR) prioritization analysis identified a number of LR interactions mainly including *CXCL9* and *FN1* mediated LR interactions, which were associated with improved / worsened patient outcomes respectively (Fig. 6L-M). This aligns with previous studies on other cancer types indicating that *FN1* overexpression correlates with an adverse prognosis [42, 43] and the involvement of *CXCL9* in both immune activation and tumor metastasis [44, 45]. In contrast, the LR pairs identified by mere high expression did not show clear associations with response to therapy (Fig. 6M). Finally, we note that the patient outcome information was not accessed by SIMVI throughout the analysis, which highlights the power of SIMVI for detecting biologically meaningful spatial regulations in an unsupervised manner.

Discussion

We introduced SIMVI (Spatial Interaction Modeling using Variational Inference), a powerful approach to disassociate intrinsic and spatial-induced variations in spatial omics data. To the best of our knowledge, SIMVI is the first model that correctly attributes spatial variation and considers the nonlinear effect induced by cell-type specific spatial effect, leading to meaningful spatial effect identification. SIMVI enables various novel analyses, covering downstream analysis using the disentangled variations, spatial effect computation and interpretation, and learning spatial effects for new measurements / modalities. With our benchmarking study in simulated data and real datasets, we showed that SIMVI outperforms alternative methods in terms of various quantitative metrics and qualitative comparisons. We applied SIMVI to five real datasets from different tissues and platforms, including MERFISH human cortex, Slide-seqV2 mouse hippocampus, MERFISH Vizgen mouse liver, spatial-ATAC-RNA-seq mouse brain, and CosMx melanoma. SIMVI revealed new biological insights into all analyzed datasets. Given the rapid development of high-resolution spatial omics, we anticipate SIMVI to be of immediate interest to the spatial omics community.

SIMVI was designed to handle spatial omics data with near-cellular resolution, such as imaging-based spatial omics technologies and high-resolution sequencing-based spatial omics data like Slide-seqV2 [7] and Stereo-seq [9]. For spatial omics datasets with lower resolution, SIMVI may face some limitations that can be overcome by using additional methods. One limitation is that these datasets have more than one cell in each pixel, which may obscure the cellular interactions and make the spatial gradient and gradual shifts in cell composition difficult to distinguish. Another limitation is that these technologies may have non-negligible gaps between pixels, which restricts the interpretation of local interactions between observed pixels. We addressed the first limitation partially by selecting pure cell type pixels in the spatial-ATAC-RNA-seq data analysis. To address these limitations more systematically, advanced deconvolution methods such as Tangram [46], CARD [47], and DestVI [48] could help reveal the single-cell profile within each pixel by using scRNA-seq references. Moreover, computational techniques that model the spatial image and spatial transcriptomics datasets, such as XFuse [49] and TESLA [50], may be extended to provide imputations for cells not covered by pixels.

SIMVI uses a single-layer graph neural network to model spatial interaction, with the graph being built by either k nearest neighbors or Delaunay triangulation. Therefore, SIMVI may have difficulty identifying cell communication over long spatial distances, which could potentially be detected with additional ligand-receptor information in the SIMVI model. The SIMVI model could be improved by incorporating long-range ligand-receptor interaction [26–30], possibly through collective optimal transport as suggested in COMMOT [51], to enhance SIMVI's potential for interpretable inferences of spatial effects.

Data availability

All datasets analyzed in this paper from previous publications are publicly available, with downloading and preprocessing instructions available in the Methods section. The newly generated dataset (the CosMx melanoma dataset) will be released soon.

Code availability

We have made SIMVI available as a public open-source Python package, which can be accessed at https://github.com/KlugerLab/SIMVI.

Acknowledgements

R.F. and Y.K. disclose support for the research of this work from NIH [U54AG076043, U54AG079759]. H.K. and Y.K. disclose support for the research of this work from NIH [P50CA121974]. Y.K. also discloses support for the research of this work from NIH [R01GM131642, UM1DA051410, and U01DA053628].

Author contributions

M.D. conceived the study. Y.K. provided overall supervision of the study. H.K. and R.F. provided additional supervision. M.D. developed SIMVI. H.K. provided melanoma tumor samples and clinical data. Y.K. and H.K. provided the CosMx melanoma dataset. M.D. performed the computational analysis. M.D., H.K., and Y.K. wrote the manuscript.

Competing interests

R.F. is co-founder of and scientific advisor to IsoPlexis, Singleton Biotechnologies, and AtlasXomics with financial interest. The remaining authors declare no competing interests.

Methods

The SIMVI model

Here we present the SIMVI model in more detail. We use $X \in \mathbb{R}^{n \times p}$ to denote the count matrix of a spatial omics dataset, where n is the number of cells / points and p is the number of genes. We use $C \in \mathbb{R}^{n \times 2}$ to denote the coordinate matrix of cells / points, where C_i is the (x, y) coordinate of cell i. We preprocess the spatial information to generate a graph $G = (V, E)$, where $V = \{1, ..., n\}$ is the vertex set and E is the edge set of ordered pairs (i, j), where $i, j \in V$. The neighbors of cell i are $N(i) = \{j | (i, j) \in E\}$. We then describe the generative model and inference procedure of SIMVI using these notations.

Generative model of SIMVI

SIMVI assumes the following generative process for modeling the distribution of entries x_{ig} in the count matrix of spatial omics data:

$$\begin{aligned}
[z_i, s_i] &\sim \text{Normal}(0, I) \\
l_i &\sim \text{log normal}(l_\mu^{b_i}, l_\sigma^{b_i}) \\
\rho_i &= f(z_i, s_i, b_i) \\
w_{ig} &\sim \text{Gamma}(\rho_{ig}, \theta_g) \\
y_{ig} &\sim \text{Poisson}(l_i w_{ig}) \\
h_{ig} &\sim \text{Bernoulli}(l_i w_{ig}) \\
x_{ig} &= h_{ig} y_{ig}
\end{aligned} \tag{1}$$

In the generative process, z_i and s_i are the two sets of latent variables representing the intrinsic variation and spatial variation respectively. The marginal distribution of $[z_i, s_i]$ is modeled as Gaussian distribution, meaning the z_i and s_i are independent joint Gaussian distributions. In the model, b_i represents experimental covariates such as batch label. The library size is modeled as a latent variable following log normal distribution with parameters $l_\mu^{b_i}, l_\sigma^{b_i}$. In practice, we have found the statistical estimation of library size l_i is usually sufficient, especially for the datasets that do not have more than one batch. In this case, the generative process of library size l_i in (1) can be straightforwardly replaced with $l_i = \sum_g x_{ig}$. $\rho_i \in \mathbb{R}^p$ represents the mean of gene expression output by neural network decoder f. The function f is constrained to let $\sum_{g=1}^{p} \rho_{ig} \equiv 1$ for all i through the use of a soft-max layer. $\theta \in \mathbb{R}^p$ specifies the gene-specific shape parameter of the Gamma distribution. Taken together, y_{ig} follows a negative binomial (NB) distribution, which can be seen through the

Gamma-Poisson mixture representation of NB distribution. h_{ig} is a parameter that represents the zero inflation level. Together, x_{ig} is modeled as a sample from a zero-inflated negative binomial (ZINB) distribution. The generative process of SIMVI mostly follows the scvi framework [24, 52], with the additional modeling of the spatial variation represented by s_i.

Approximate posterior inference of SIMVI

In order to infer the parameters in the SIMVI model, we approximate the posterior distribution via variational inference, a standard framework for inferring parameters for deep probabilistic models. The posterior distribution is approximated to be factorized as follows:

$$q_\Phi(z_i, s_i, l_i | x_i, x_{N(i)}, b_i) = q_{\Phi_z}(z_i | x_i, b_i) q_{\Phi_s}(s_i | x_{N(i)}, b_i) q_{\Phi_l}(l_i | x_i, b_i) \tag{2}$$

Here Φ denotes neural network weights that determine the variational posterior. In SIMVI, different from the original scvi model [52], we additionally model the spatial variation s_i's approximated variational posterior as the function of cell i's neighborhood $N(i)$ gene expression. We use a one-layer graph attention network (GAT) with dynamic attention [53] to model the variational posterior $p(s_i | x_{N(i)})$. After the factorization, the evidence lower bound (ELBO) can be derived via straightforward computation.

$$\begin{aligned}
\log p(x_i|b_i) &= \log \int p(x_i, z_i, s_i, l_i | b_i) dz_i ds_i dl_i \\
&= \log \int \frac{p(x_i, z_i, s_i, l_i | b_i) q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i)}{q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i)} dz_i ds_i dl_i \\
&\geq \int q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i) \log \frac{p(x_i, z_i, s_i, l_i | b_i)}{q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i)} dz_i ds_i dl_i \\
&= \int q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i) \log \frac{p(x_i | z_i, s_i, l_i, b_i) p(z_i, s_i, l_i | b_i)}{q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i)} dz_i ds_i dl_i \\
&= \mathbb{E}_{q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i)} \log p(x_i | z_i, s_i, l_i, b_i) \\
&\quad - D_{\mathrm{KL}}(q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i) || p(z_i, s_i, l_i | b_i)) \\
&= \mathbb{E}_{q(z_i, s_i, l_i | x_i, x_{N(i)}, b_i)} \log p(x_i | z_i, s_i, l_i, b_i) - D_{\mathrm{KL}}(q(z_i | x_i, b_i) || p(z_i | b_i)) \\
&\quad - D_{\mathrm{KL}}(q(s_i | x_{N(i)}, b_i)) || p(s_i | b_i)) - D_{\mathrm{KL}}(q(l_i | x_i, b_i) || p(l_i | b_i)). \quad \text{(ELBO)}
\end{aligned} \tag{3}$$

In our derivation, Jensen inequality is used in the third step and the independence between z, s, l leads to the KL divergence decomposition in the last step. While the ELBO loss can be potentially used as the objective function, optimization over only the ELBO loss results in correlated s and z, meaning the model fails to disentangle intrinsic variation and spatial variation. The reason is that the ELBO loss does not enforce the independence of the optimized latent variables as they are computed separately. In order to further enforce the disentanglement, we regularize the mutual information between s and z ($I(s, z)$). Assuming $[s, z]$ follows joint Gaussian distribution, which is a necessary condition for the generative process of the SIMVI, the mutual information term can be derived analytically:

$$I(s, z) = \frac{1}{2} \log \frac{\det \Sigma_s \det \Sigma_z}{\det \Sigma_{[s,z]}}, \tag{4}$$

where $\Sigma_s, \Sigma_z, \Sigma_{[s,z]}$ denote the covariance matrices for $s, z, [s, z]$ respectively. In practice, the covariance matrices can be estimated from sample covariance matrices thus the mutual information can be also effectively estimated.

We note a similar regularization term has been explored in InfoVAE [54], which proposes a VAE framework with a mutual information regularization in order to improve the performance. However, our definitions differ conceptually from the InfoVAE model. In InfoVAE, the mutual information between the original variables x and the latent space z is maximized to achieve improvement in performance. The mutual information term in InfoVAE is transformed into an additional term $MMD(q_\phi(z)|p(z))$ in the loss term, while the coefficient of KL divergence term can be arbitrarily selected. This additional term enforces the variational distribution to the marginal distribution, rather than enforcing independence between two variational distributions. Our proposed

mutual information term can be also seen as a form of the Hilbert-Schmidt independence criterion (HSIC) [55] with a statistically meaningful kernel selection under the distributional assumption.

Finally, denote L as the regularization strength, the loss function of SIMVI is formulated as

$$\text{Loss} = -\text{ELBO} + L \cdot I(s, z). \tag{5}$$

The SIMVI model is trained to minimize the loss term through stochastic gradient descent, enabled by the reparametrization trick. Since some of the variations are both intrinsic and spatial (e.g. cells of the same cell type in a niche), we have designed a procedure to attribute the redundant variation to intrinsic variation. Specifically, during the optimization, the intrinsic variation network Φ_z is updated only based on the minus ELBO part of the loss, while the spatial variation Φ_s is updated based on the full loss term. The Adam optimizer [56] is used to update the model parameters. Overall, the regularization term is designed to attribute the redundant variations as intrinsic variation, enforcing the identifiability of the VAE-based SIMVI model [57–59].

We finalize by noting that the train / validation set construction for the SIMVI model is different from most scvi based models [52, 60]. The reason is that for any train / validation set separation, there can be edges across the two sets. In this work, in order to define train and validation sets, we adopt the semi-supervised node classification framework for graph neural networks [61]. Specifically, the full dataset is fed into the SIMVI model to compute the intermediate cell-level outputs (embeddings) for each cell, including intrinsic and spatial variation parameters and library sizes. Then the cell-level outputs are masked to generate train set and validation set. During training, only the embedding outputs from cells in train sets are used to compute the loss function.

Spatial effect identification with SIMVI

While SIMVI disentangles intrinsic variation and spatial variation, the spatial effect itself is usually determined by both intrinsic and spatial variations. For example, different types of cells can have different responses to neighboring cells even in the same niche. Therefore, in order to model the spatial effect, we consider the following estimand for estimating cell i's spatial effect E_i:

$$E_i = \mathbb{E}[\rho_i - \mathbb{E}_{z_i}\rho_i] = \mathbb{E}[f(z_i, s_i, b_i) - \mathbb{E}_{z_i} f(z_i, s_i, b_i)]. \tag{6}$$

Here ρ is the latent variable representing normalized gene expression as defined in (1). Estimation of (6) can be performed in various ways. In this work, we consider the following k-NN estimator for computing the expectation term:

$$E_i^{\text{k-NN}} = \mathbb{E}[\rho_i - \sum_{j \in N_z(i)} \rho_j], \tag{7}$$

where $N_z(i)$ is the set of k nearest neighbors defined on $\mathbb{E}z \in \mathbb{R}^{n \times d}$ with the same covariate label b.

As ρ is a random latent variable, the estimation of $\mathbb{E}\rho_i$ requires substantial Monte-Carlo sampling and is not robust to fitting quality. In this work, we note that the k-NN based estimator can be also defined by log-normalized count matrix, which is deterministic and enables robust and efficient spatial effect analysis. Denoting the log-normalize function as h, then the estimator is defined by

$$\tilde{E}_i^{\text{k-NN}} = \mathbb{E}[h(x_i) - \sum_{j \in N_z(i)} h(x_j)]. \tag{8}$$

Benchmarking on simulated data

In order to validate SIMVI's ability to disentangle variations and identifying spatial effect, we generated simulated spatial omics datasets with ground truth spatial interactions. In this simulation, two mechanisms, including spatial gradient and cellular interactions, were implemented for generating spatial patterns. In all simulations, 2500 cells times 1000 genes scRNA-seq count matrices with two cell types of equal probabilities were first simulated by the Python package scSim [62]. In order to simulate the spatial patterns, we employed a kinetic Ising model using the TDQL equation:

$$\frac{\partial \phi}{\partial t} = \phi - \phi^3 + \nabla^2 \phi \tag{9}$$

The TDGL equation was simulated using 5-point Laplacian discretization form. The initial condition was set as i.i.d samples from normal distributions $N(0, 1)$ on a 50×50 spatial grid. The system state at $T = 50$ was discretized by the quantile corresponding to the simulated cell type ratio to generate the cell type labels.

For simulating both the spatial gradient and cellular interaction, we additional modeled a gene regulatory program in the simulated count matrix. We arranged the rows so that the gene activity changed gradually from left to right, by sorting `program_usage` parameter in the scSim model. Then to simulate cellular interaction, for each cell of cell type 2, we computed the neighbor number with cell type 1. Then the neighbor number is used to further sort `program_usage` parameter. We repeated this process 20 times to get 20 different simulations. Additionally, we simulated 10 datasets with only interface interaction, by not preserving the row coordinate order while sorting according to the `program_usage` parameter.

For benchmarking the Mixscape baseline, we computed each cell's difference with its 20 nearest neighbors' (k-NN graph obtained by leading 50 principal components) average expression following descriptions in the paper [31]. For benchmarking of both MEFISTO and NSF, we have used the implementation in the NSF paper [22]: https://github.com/willtownes/nsf-paper that provides wrappers for both MEFISTO and NSF models. In order to model both spatial and non-spatial variations, in all of our experiments, we used the NSFH model for NSF and select the numbers of spatial factors and non-spatial factors both 10. For the MEFISTO model, it does not explicitly provide spatial / non-spatial components. Instead, it returns a parameter ζ for determining the smooth level of each factor. Therfore we trained MEFISTO models with 10 factors and select the factors with zeta above a certain quantile threshold as intrinsic factors and factors below median as spatial factors. All the other settings, including selections of kernel functions and stopping criteria, were consistent with the default in both MEFISTO and NSF. As the validation set is not used in the default stopping criteria for both methods, here we took the training set to be the full dataset for convenience in outputting desired embeddings.

Both MEFISTO and NSF give us two sets of latent variables that represent the spatial and non-spatial aspects of the data. To compare how well these two methods can capture the spatial effect, we used two different ways to generate the spatial effect. One way was to use the latent variables that show the spatial variation directly. The other way was to use the k-NN based method as previously described using log normalized count matrices.

For benchmarking SIMVI, the model was trained with 90% of the full data points as training set. The learning rate and weight decay for Adam optimizer were set to be 1e-3 and 1e-4 respectively. All the models were trained for 300 epochs. We also included the baseline SIMVI model with the regularization equal to zero. All the remaining settings of the baseline SIMVI model is consistent with the full SIMVI model.

To evaluate the non-spatial / intrinsic variation, we applied cell type preservation metrics including normalized mutual information score (NMI) and adjusted rand index (ARI) as implemented by `scib.metrics.metrics(_nmi=True,_ari=True)`. To evaluate the spatial effect identification, here we used the ratio of explained variance by spatial label (being the "program_usage" in interaction plus gradient experiment, and "neighborhood_group" in the only interaction experiment) and the sum of explained variance by cell type and by spatial label as the PCR-based metric for measuring spatial effect preservation level relative to cell type removal level.

In order to investigate the effect of hyperparameter settings on different methods respectively, we performed parameter-sweep analysis for all tested methods on the key model settings. The sweeped hyperparameters for all methods are summarized as follows. The selected hyperparameter settings are highlighted in bold. For the experiment with interaction plus gradient:

MEFISTO: The number of inducing points: [100,**500**]; The quantile threshold: [**0.25**,0.5,0.75].

NSF: The number of inducing points: [100,**500**]; The model likelihood: ['poi',**'gau'**] (For the Gaussian likelihood setting, the data is log-normalized first).

SIMVI: The regularization strength L: [0,1,**5**,20,100]. The neighbor number for the spatial graph k: [4,**8**,20].

For the experiment with only interaction:

MEFISTO: The number of inducing points: [50,100,**500**]; The quantile threshold: [**0.25**,0.5,0.75].

NSF: The number of inducing points: [100,**500**]; The model likelihood: ['poi',**'gau'**] (For the Gaussian likelihood setting, the data is log-normalized first).

SIMVI: The regularization strength L: [0,1,5,**20**,100]. As our previous parameter sweep analysis show little difference with respect to the neighbor number, here we fix $k = 10$ for SIMVI.

The detailed parameter-sweep analysis results are listed in Supplementary Fig. 1.

MERFISH human cortex dataset

We used a dataset of gene expression from two brain regions: MTG and STG. The dataset was converted to AnnData format and we downloaded it from https://datadryad.org/stash/dataset/doi:10.5061/dryad.x3ffbg7mw. We used the following samples from the dataset: MTG sample: H18.06.006.MTG.4000, with 3044 cells and 4000 genes; STG sample: H19.30.001.STG.4000, with 4085 cells and 4000 genes. We analyzed the samples with SIMVI to separate the gene expression into two parts: intrinsic variation and spatial-induced variation. We used the same settings as in our benchmarking study with neighbor number $k = 10$ and regularization strength L set to 5. We trained SIMVI for 300 epochs on each of the two datasets, during which the validation losses eventually stabilized. The intrinsic variation and spatial variation are extracted as the variational posterior expectations of s, z respectively. We used the k-NN estimator on log normalized count matrices to estimate spatial effect for each cell in the MTG (k=50) and STG (k=200) dataset respectively. We then clustered and visualized the cells based on their intrinsic variation, spatial variation, and k-NN spatial effect. To do this, we used some functions from Scanpy, a tool for single-cell analysis including: `sc.pp.pca`: to do principal component analysis (PCA), which reduces the dimensionality of the data; `sc.pp.neighbors`: to find the neighbors of each cell based on their distance; `sc.tl.leiden`: to cluster the cells into groups based on their similarity. We also looked for marker genes that were differentially expressed in each cluster of spatial variation. We used the function `sc.tl.rank_gene_groups` to identify most signficantly upregulated genes in each cluster by Wilcoxon test. We further divided the clusters into subclusters based on their cell types. We only kept the subclusters that had enough cells (more than 40 for MTG and more than 20 for STG). The differential expression analysis for these subclusters was performed by one-side Wilcoxon signed-rank test on spatial effect minus 0.05 to filter out genes with small effect sizes. We then performed multi-hypothesis correction via `statsmodels.stats.multitest.fdrcorrection` and selected the genes with adjusted p-values smaller than 0.05 to be the differential expressed genes. Then the pathway analysis was performed by GSEAPy [3] with KEGG pathway gene set [32] and the background set to be the full profiled gene list. For the ligand-receptor prioritization analysis, we used the CellTalkDB [30] for extracting the ligand-receptor pairs. For each ligand-receptor pair, we computed the sum of the average neighborhood normalized ligand expression (k=10) and the central cell receptor normalized receptor expression as the ligand-receptor strength for each cell. For each pair, we calculated a score that combines the expression of the ligand in the neighboring cells and the expression of the receptor in the central cell. For each pair, its maximum Spearman correlation between PC components of the spatial effect matrix was computed.

For benchmarking, we additionally applied MEFISTO and NSF on the MTG and STG datasets. For both methods, we first selected 1,000 highly variable genes consistent with the implementation in [22], and used 500 inducing points. We used the setting of quantile threshold = 0.5 for MEFISTO and both Gaussian and Poisson likelihood settings for NSF. We used the NMI and ARI for benchmarking cell type preservation of the intrinsic embedding, and used a modified version of PCR to measure the removal of spatial differentially expressed gene signal (*MYH11*) in endothelial and mural cells. The modified PCR is defined as the average of the original PCR [63] that measures signal removal level, and the ratio of the cell type label regression score of the new embedding and the original PC embedding. The latter term is truncated to have max 1. Compared to the original version of PCR, the modified version of PCR evaluates batch removal while penalizing uninformative embeddings that merges different cell types. Finally, we used a threshold (log2) of spatial differentially expressed gene (*MYH11*) value to separate different states in the corresponding cell types (endothelial and mural cells) and used the Silhouette score to measure the overlap level of the states. All the metrics were implemented by the scib Python package [63].

For the replicate experiment, we additionally used MTG samples H18.06.006.MTG.4000 (rep2, rep3). The count matrices of the two datasets are fed into the MLP encoder of the trained SIMVI model to return intrinsic embeddings. Then we used our k-NN (k=50) SE to compute each dataset's spatial effect and concatenated them as the final spatial effect output.

Slide-seqV2 mouse hippocampus dataset

We acquired the annotated dataset in AnnData format from Squidpy [25]. As the dataset does not provide raw counts, we additionally accessed the raw count matrix from the Broad single-cell portal: https://singlecell.broadinstitute.org/single_cell/study/SCP815/highly-sensitive-spatial-transcriptomics-at-near-cellular-resolution-with-slide-seqv2#study-summary. The preprocessed dataset contains 41786 cells and 4000 highly variable genes. For SIMVI analysis, we used all the 4,000 highly variable genes and the same graph construction and hyperparameter settings as those used in benchmarking study with neighbor number $k = 10$ and regularization strength L set to 5. The SIMVI model was trained for 300 epochs for the Slide-seqV2 dataset. The intrinsic variation and spatial variation were extracted as the variational posterior expectations of s, z respectively. We used the k-NN estimator on log normalized count matrices to estimate spatial effect for each cell in the dataset (k=200). The downstream analysis settings are consistent with the analysis on MERFISH cortex data. Additionally, we performed neighborhood enrichment analysis to explore the spatial relationship between different spatial effect clusters. The neighborhood enrichment analysis was performed on SE clusters that are further segragated by cell types, and were thresholded so that only the clusters that have more than 200 cells were retained. Then the neighborhood enrichment score between clusters was computed by functions `sq.gr.spatial_neighbors` and `sq.gr.nhood_enrichment(n_perms=1000)` in Squidpy. For NSF and MEFISTO, the first 1,000 highly variable genes were used, consistent with the implementation in [22]. 1000 inducing points are used for both MEFISTO and NSF models. We used quantile threshold = 0.5 for MEFISTO and Poisson likelihood for NSF, consistent with [22]. We utilized both NMI and ARI to benchmark the preservation of cell type in the intrinsic embedding, while the modified PCR metric same as described in the previous section was employed to assess the removal of spatially differentially expressed gene signals (*Ttr*, *Ptgds*) in Endothelial Tip cells. Additionally, we applied a threshold (2) to the values of the spatially differentially expressed genes (*Ttr*, *Ptgds*) to segregate different spatial states within the corresponding cell types (Endothelial Tip cells). The Silhouette score was then used to evaluate the degree of overlap between these states. All the metrics were implemented by the scib Python package [63].

Vizgen mouse liver dataset

The dataset is available at https://info.vizgen.com/mouse-liver-access. We preprocessed the raw data and performed annotation according to the vignette provided in Squidpy (https://squidpy.readthedocs.io/en/stable/notebooks/tutorials/tutorial_vizgen_mouse_liver.html) [25]. The full dataset after preprocessing comprises 367335 cells and 385 genes, with 13 cell types in total including the main hepatocyte phenotypes and immune cells. For the SIMVI analysis, we specifically selected a spatial range of cells located in the upper middle portion of the image that covers liver biological structures. The preprocessed subset of the dataset comprises 42,872 cells and 385 genes. We used the same graph construction and hyperparameter settings for the SIMVI analysis as those utilized in the benchmarking study with neighbor number $k = 10$ and regularization strength L set to 5. Training of the SIMVI model was performed for 500 epochs on this dataset. We used the k-NN estimator on log normalized count matrices to estimate spatial effect for each cell in the dataset (k=500). The setup for the neighborhood enrichment analysis was the same as that employed in the Slide-seqV2 data analysis. The marker genes for each cluster were selected using the `sc.tl.rank_gene_groups` function via the Wilcoxon test.

Spatial-ATAC-RNA-seq mouse brain data

The dataset is available at https://cells.ucsc.edu/?ds=brain-spatial-omics+p22-atac. The RNA raw counts were downloaded from https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM6753043. The data was converted into AnnData format. The raw RNA count was used for SIMVI analysis, while the spatial effect identification was performed using the log-normalized RNA expression and the TF-IDF normalized matrix of ATAC peaks.

In this dataset, each pixel represents a mixture of cells. To identify the pixels that include only one cell type for SIMVI analysis, we applied a filtering strategy based on archetypal analysis. First, we identified 40 archetypes using the 50 leading principal components obtained from from preprocessed RNA expression. Then the data was clustered using the Leiden algorithm with a pre-specified resolution (2.0). A likelihood score of each cluster can be obtained by counting the total number of cells that are among the 5 nearest neighbors for each of the 40 archetypes. Then the clusters were sorted by the likelihood, only cells from clusters that are passing a quantile threshold (0.35) were counted as pure cells.

As this dataset is of different nature from the other analyzed datasets in this work, we used a modified list of parameters, namely 5 for spatial neighbors and 20 for the regularization strength. The SIMVI model was trained for 300 epoches on the dataset. To construct the k-NN graph for estimating ATAC spatial effect, we first concatenated the SIMVI intrinsic embedding from the RNA modality with the ATAC annotation label. We then computed the k-NN graph for the concatenated embedding in order to adequately filter out the distinct intrinsic variation in the ATAC modality, with $k = 20$. The marker genes and ATAC peaks were selected using the `sc.tl.rank_gene_groups` function via the Wilcoxon test.

CosMx melanoma dataset

Tissue acquisition and retrieval of patient information was approved by an Institutional Review Boards at Yale University. Tissue microarray blocks were constructed as previously described [37], containing tumor samples from 60 patients treated with immunotherapy. Slices from the block were submitted to to NanoString for CosMx spatial profiling. Twenty five samples were randomly selected for analysis. The dataset used contains samples from 16 male patients and 9 female patients, ages ranging from 35-90. Among these patients, 11 patients received ipilimumab and nivolumab in combination (IPI+NIVO), 13 received pembrolizumab (PEMBRO), and one patient received nivolumab (NIVO) alone. CosMx Human Universal Cell Characterization RNA Panel was used as the SMI reagent. This panel included genes for cell typing and mapping (243 genes), cell state and function (269 genes), cell-cell interaction (435 genes), and hormone activities (46 genes). No statistical methods were used to predetermine sample size. The data was divided into 11 categories of non-tumor cells and six subclasses of tumor cells. The preprocessed dataset consisted of 56,761 cells and 960 genes. We analyzed the dataset with SIMVI of the same settings as in our benchmarking study, with neighbor number $k = 10$ and the regularization strength L set to 5. We trained SIMVI for 500 epochs on the dataset. We then splitted the dataset into two subsets: non-tumor cells and tumor cells. We used the k-NN estimator on log normalized count matrices of non-tumor cells to estimate spatial effect for non-tumor cells (k=50). The setup for the neighborhood enrichment analysis is consistent with that employed in the Slide-seqV2 data analysis. The marker genes for each cluster were selected using the `sc.tl.rank_gene_groups` function via the Wilcoxon test. The differential expression analysis for the *SPP1-* macrophages subclusters is performed by one-side Wilcoxon signed-rank test on spatial effect minus 0.05. Then we performed multi-hypothesis correction via `statsmodels.stats.multitest.fdrcorrection`. Genes with adjusted p-values smaller than 0.05 were selected to be the differential expressed genes. Then we performed the pathway analysis by GSEAPy [3] with KEGG pathway gene set [32]. The background was set to be the full profiled gene list. Next we computed the NMI (normalized mutual information score) and ARI (adjusted Rand index) for both SIMVI spatial embedding and the original PC embedding across different immune cell types. The setting for ligand-receptor strength computation was consistent with those adopted in the MERFISH human cortex data analysis. For each ligand-receptor pair, we computed its maximum Spearman correlation with respect to the SIMVI spatial variation components.

References

[1] Lohoff, T. *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nature biotechnology* **40**, 74–85 (2022).

[2] Zhang, M. *et al.* Spatially resolved cell atlas of the mouse primary motor cortex by merfish. *Nature* **598**, 137–143 (2021).

[3] Fang, R. *et al.* Conservation and divergence of cortical cell organization in human and mouse revealed by merfish. *Science* **377**, 56–62 (2022).

[4] He, S. *et al.* High-plex imaging of rna and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. *Nature Biotechnology* **40**, 1794–1806 (2022).

[5] Liu, Y. *et al.* High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681 (2020).

[6] Liu, Y., Enninful, A., Deng, Y. & Fan, R. Spatial transcriptome sequencing of ffpe tissues at the cellular level. *bioRxiv* 2020–10 (2020).

[7] Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqv2. *Nature biotechnology* **39**, 313–319 (2021).

[8] Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nature methods* **16**, 987–990 (2019).

[9] Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).

[10] Wei, X. *et al.* Single-cell stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science* **377**, eabp9444 (2022).

[11] Gilbert, S. F. *Developmental biology.* (sinauer associates, Inc, 2010).

[12] Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome biology* **22**, 1–31 (2021).

[13] Weber, L. M., Saha, A., Datta, A., Hansen, K. D. & Hicks, S. C. nnsvg: scalable identification of spatially variable genes using nearest-neighbor gaussian processes. *bioRxiv* 2022–05 (2022).

[14] Hao, M., Hua, K. & Zhang, X. Somde: a scalable method for identifying spatially variable genes with self-organizing map. *Bioinformatics* **37**, 4392–4398 (2021).

[15] Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nature methods* **17**, 193–200 (2020).

[16] Zhu, J., Sun, S. & Zhou, X. Spark-x: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies. *Genome Biology* **22**, 1–25 (2021).

[17] Svensson, V., Teichmann, S. A. & Stegle, O. Spatialde: identification of spatially variable genes. *Nature methods* **15**, 343–346 (2018).

[18] Dong, M. & Kluger, Y. Geass: Neural causal feature selection for high-dimensional biological data. *The Eleventh International Conference on Learning Representations* (2023).

[19] Wu, Z. *et al.* Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nature Biomedical Engineering* **6**, 1435–1448 (2022).

[20] Fischer, D. S., Schaar, A. C. & Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nature Biotechnology* **41**, 332–336 (2023).

[21] Velten, B. *et al.* Identifying temporal and spatial patterns of variation from multimodal data using mefisto. *Nature methods* **19**, 179–186 (2022).

[22] Townes, F. W. & Engelhardt, B. E. Nonnegative spatial factorization applied to spatial genomics. *Nature Methods* **20**, 229–238 (2023).

[23] Zhang, D. *et al.* Spatial epigenome–transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113–122 (2023).

[24] Gayoso, A. *et al.* A python library for probabilistic analysis of single-cell omics data. *Nature biotechnology* **40**, 163–166 (2022).

[25] Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nature methods* **19**, 171–178 (2022).

[26] Vento-Tormo, R. *et al.* Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).

[27] Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. Cellphonedb: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature protocols* **15**, 1484–1506 (2020).

[28] Garcia-Alonso, L. *et al.* Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nature genetics* **53**, 1698–1711 (2021).

[29] Garcia-Alonso, L. *et al.* Single-cell roadmap of human gonadal development. *Nature* **607**, 540–547 (2022).

[30] Shao, X. *et al.* Celltalkdb: a manually curated database of ligand–receptor interactions in humans and mice. *Briefings in bioinformatics* **22**, bbaa269 (2021).

[31] Papalexi, E. *et al.* Characterizing the molecular regulation of inhibitory immune checkpoints with multimodal single-cell screens. *Nature genetics* **53**, 322–331 (2021).

[32] Kanehisa, M. & Goto, S. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27–30 (2000).

[33] Sun, T. *et al.* Axin2+ pericentral hepatocytes have limited contributions to liver homeostasis and regeneration. *Cell stem cell* **26**, 97–107 (2020).

[34] Liu, Y. *et al.* Spatial-cite-seq: spatially resolved high-plex protein and whole transcriptome co-mapping. *Research Square* (2022).

[35] Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *Ca Cancer J Clin* **73**, 17–48 (2023).

[36] Pichler, A. C. *et al.* Tcr-independent cd137 (4-1bb) signaling promotes cd8+-exhausted t cell proliferation and terminal differentiation. *Immunity* (2023).

[37] Wong, P. F. *et al.* Multiplex quantitative analysis of tumor-infiltrating lymphocytes and immunotherapy outcome in metastatic melanoma. *Clinical Cancer Research* **25**, 2442–2449 (2019).

[38] Zhang, L. *et al.* Single-cell analyses inform mechanisms of myeloid-targeted therapies in colon cancer. *Cell* **181**, 442–459 (2020).

[39] Ma, R.-Y., Black, A. & Qian, B.-Z. Macrophage diversity in cancer revisited in the era of single-cell omics. *Trends in immunology* (2022).

[40] Couzin-Frankel, J. Cancer immunotherapy (2013).

[41] Murphy, K. & Weaver, C. *Janeway's immunobiology* (Garland science, 2016).

[42] Sun, Y. *et al.* High expression of fibronectin 1 indicates poor prognosis in gastric cancer. *Oncology Letters* **19**, 93–102 (2020).

[43] Zhang, X.-X., Luo, J.-H. & Wu, L.-Q. Fn1 overexpression is correlated with unfavorable prognosis and immune infiltrates in breast cancer. *Frontiers in Genetics* **13**, 913659 (2022).

[44] Ding, Q. *et al.* Cxcl9: evidence and contradictions for its role in tumor progression. *Cancer medicine* **5**, 3246–3259 (2016).

[45] Tokunaga, R. *et al.* Cxcl9, cxcl10, cxcl11/cxcr3 axis for immune activation–a target for novel cancer therapy. *Cancer treatment reviews* **63**, 40–47 (2018).

[46] Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram. *Nature methods* **18**, 1352–1362 (2021).

[47] Ma, Y. & Zhou, X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nature biotechnology* **40**, 1349–1359 (2022).

[48] Lopez, R. *et al.* Destvi identifies continuums of cell types in spatial transcriptomics data. *Nature biotechnology* **40**, 1360–1369 (2022).

[49] Bergenstråhle, L. *et al.* Super-resolved spatial transcriptomics by deep data fusion. *Nature biotechnology* **40**, 476–479 (2022).

[50] Hu, J. *et al.* Deciphering tumor ecosystems at super resolution from spatial transcriptomics with tesla. *Cell systems* **14**, 404–417 (2023).

[51] Cang, Z. *et al.* Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nature Methods* **20**, 218–228 (2023).

[52] Xu, C. *et al.* Probabilistic harmonization and annotation of single-cell transcriptomics data with deep generative models. *Molecular systems biology* **17**, e9620 (2021).

[53] Brody, S., Alon, U. & Yahav, E. How attentive are graph attention networks? *arXiv preprint arXiv:2105.14491* (2021).

[54] Zhao, S., Song, J. & Ermon, S. Infovae: Balancing learning and inference in variational autoencoders **33**, 5885–5892 (2019).

[55] Gretton, A., Bousquet, O., Smola, A. & Schölkopf, B. Measuring statistical dependence with hilbert-schmidt norms 63–77 (2005).

[56] Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).

[57] Hyvärinen, A. & Pajunen, P. Nonlinear independent component analysis: Existence and uniqueness results. *Neural networks* **12**, 429–439 (1999).

[58] Khemakhem, I., Kingma, D., Monti, R. & Hyvarinen, A. Variational autoencoders and nonlinear ica: A unifying framework. *International Conference on Artificial Intelligence and Statistics* 2207–2217 (2020).

[59] Dong, M. *et al.* Causal identification of single-cell experimental perturbation effects with cinema-ot. *bioRxiv* 2022–07 (2022).

[60] Weinberger, E., Lin, C. & Lee, S.-I. Isolating salient variations of interest in single-cell data with contrastivevi. *bioRxiv* 2021–12 (2021).

[61] Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[62] Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell rna-seq. *Elife* **8** (2019).

[63] Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nature methods* **19**, 41–50 (2022).