

# Representation learning for multi-modal spatially resolved transcriptomics data

Kalin Nonchev<sup>1, 2</sup>, Sonali Andani<sup>1,2,3\*</sup>, Joanna Ficek-Pascual<sup>1,2\*</sup>,  
Marta Nowak<sup>3</sup>, Bettina Sobottka<sup>3</sup>, Tumor Profiler Consortium,  
Viktor H Koelzer<sup>3,4,5†</sup>, Gunnar Rätsch<sup>1,2,6,7†</sup>

<sup>1</sup>Department of Computer Science, ETH Zurich, Zurich Switzerland.

<sup>2</sup>Swiss Institute of Bioinformatics, Zurich Switzerland.

<sup>3</sup>Computational and Translational Pathology Group, Department of Pathology and Molecular Pathology, University Hospital Zurich, University of Zürich, Zurich Switzerland.

<sup>4</sup>Department of Oncology, University of Oxford, Oxford UK.

<sup>5</sup>Institute of Medical Genetics and Pathology, University Hospital Basel, Basel Switzerland.

<sup>6</sup>AI Center, ETH Zurich, Zurich Switzerland.

<sup>7</sup>Medical Informatics Unit, University Hospital Zurich, Zurich Switzerland.

## Abstract

Spatial transcriptomics enables in-depth molecular characterization of samples on a morphology and RNA level while preserving spatial location. Integrating the resulting multi-modal data is an unsolved problem, and developing new solutions in precision medicine depends on improved methodologies. Here, we introduce *AESTETIK*, a convolutional deep learning model that jointly integrates spatial, transcriptomics, and morphology information to learn accurate spot representations. *AESTETIK* yielded substantially improved cluster assignments on widely adopted technology platforms (e.g., 10x Genomics™, NanoString™) across multiple datasets. We achieved performance enhancement on structured tissues (e.g., brain) with a 21% increase in median ARI over previous state-of-the-art methods. Notably, *AESTETIK* also demonstrated superior performance on cancer tissues with heterogeneous cell populations, showing a two-fold increase in breast cancer, 79% in melanoma, and 21% in liver cancer. We expect that these advances will enable a multi-modal understanding of key biological processes.

---

†Corresponding authors. E-mails: [viktor.koelzer@usz.ch](mailto:viktor.koelzer@usz.ch); [gunnar.raetsch@inf.ethz.ch](mailto:gunnar.raetsch@inf.ethz.ch);

\*Equal contribution.

**Keywords:** machine learning, representation learning, spatial transcriptomics, multi-modal data integration

# 1 Introduction

In multicellular organisms, cells are organized into tissues, groups of cells exhibiting common characteristics related to the biological function [1, 2]. Recent advances in spatial transcriptomics enable in-depth molecular characterization of samples, capturing their morphology and RNA composition while retaining the spatial location (Fig. 1A). The gene expression profiles are usually available per spot, e.g., a 55 $\mu$ m tissue region (10x Genomics<sup>TM</sup>, Visium) covering the whole transcriptome [3], or at a single-cell resolution but with a limited number of captured genes (CosMx NanoString<sup>TM</sup>). More recent spatial transcriptomics technologies provide whole-transcriptome coverage along with higher resolution (e.g., 2 $\mu$ m, 10x Genomics<sup>TM</sup>, Visium HD). Spatially aligning cell types by molecular phenotypes and morphology is important for understanding tissue-specific properties (e.g., neural organization in the brain [4]) in a physiological state and in the context of disease progression and treatment [5–8]. Nevertheless, spatial transcriptomics analysis demands manual annotation of multi-modal data, representing a laborious and resource-intensive process. Achieving reliable automation and overcoming limitations in cross-modal expertise will lead to more accurate annotations, offering a comprehensive, multi-modal perspective on biological mechanisms and interactions [9].

Despite recent progress, computational data analysis that integrates all available data modalities i.e., spatial information, transcriptomics, and morphology, remains challenging. Most existing methods either fall short in effectively integrating all modalities, especially those adapted from single-cell analysis or are computationally expensive [9]. For example, *BayesSpace* employs a Bayesian approach with a prior giving higher weight to physically close spots [10]; *MUSE* relies on a multi-view autoencoder to learn a latent space from transcriptomics and morphology [11]; *stLearn* quantifies morphological distance through histology image features and incorporates these distances with spatial neighbors to refine gene expression [12]. Furthermore, alternative methods suggested a different perspective on modeling the spatial transcriptomics data by employing graph neural networks (GNN) [13–15]. However, the expression profiles often suffer from biological variability (e.g., cell-cycle stage) [16, 17] or technical noise [17, 18]. GNNs’ inherent susceptibility to noise can undermine their robustness and performance in downstream applications [19, 20]. Therefore, a new and reliable integration approach is needed to overcome the aforementioned challenges and improve spatial transcriptomics analysis, ensuring adaptability across spatial transcriptomics technologies.

To this end, we developed *AESTETIK*, a model that jointly integrates spatial, transcriptomics, and morphology information to learn accurate spot representations. We compared its performance against previous state-of-the-art methods on multiple datasets and widely adopted technology platforms: Brain tissue [21], breast cancer [22] and new and yet unreleased metastatic melanoma samples sequenced using Visium from 10x Genomics<sup>TM</sup>; liver from normal and cancer patients using CosMx from NanoString<sup>TM</sup>. We substantially improved the clustering accuracy across all datasets which

yielded spatial domains with coherent expression and morphology. Through an ablation study, we showed the enhanced value of utilizing all available data modalities given the specifics of the analyzed tissue. Further, we validated the learned representation by identifying the main biological drivers and characterizing clusters based on morphology and cell-type composition.

## 2 Results

### 2.1 *AESTETIK* integrates spatial, transcriptomics, and morphology information

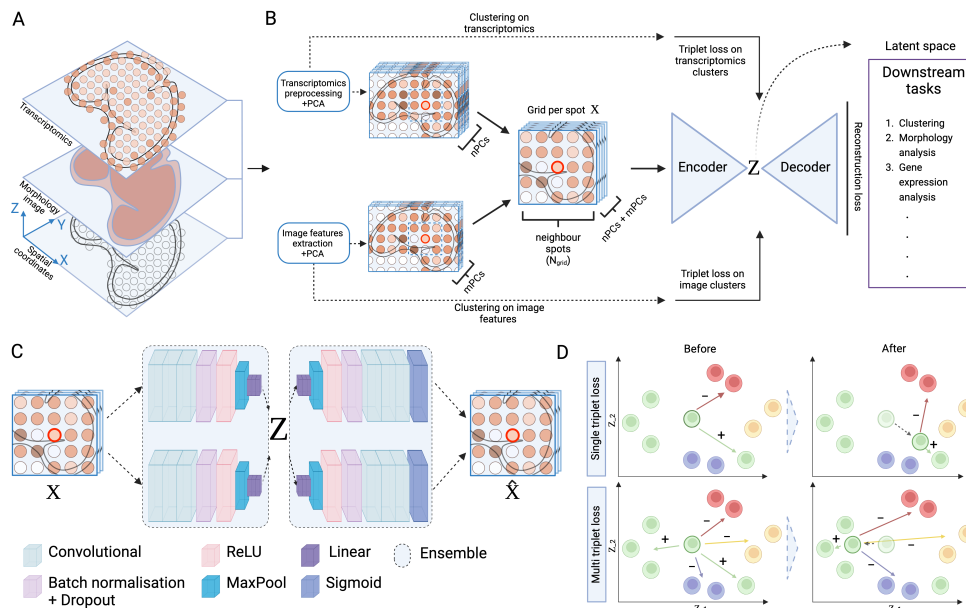
We introduce *AESTETIK* (**A**uto**E**ncoder for **S**patial **T**ranscriptomics **E**xpression with **T**opology and **I**mage **K**nowledge), a convolutional autoencoder model (Fig. 1B). It jointly integrates transcriptomics and morphology information at a spot level and topology at a neighborhood level to learn accurate spot representations that capture biological complexity. Firstly, we preprocess the transcriptomics profiles and apply principal component analysis (PCA) [23]. Simultaneously, the pre-trained on *Imagenet* [24] deep-learning model, *Inception v3* [25], is employed to extract morphology spot features, followed by PCA. After computing clusters separately to preserve the modality-specific structure, we concatenate the top principal components (PC) from both modalities. Next, we construct a square grid for each spot that includes spatially neighboring spots. This grid per spot, along with the precomputed clusters, serves as an input to *AESTETIK*. The model relies on a convolutional encoder-decoder architecture (Fig. 1C) to learn accurate spatial-, transcriptomics-, and morphology-informed spot representations. Ultimately, the learned representations can be leveraged for various downstream applications, including but not limited to clustering, gene expression, morphology, and pathway analysis.

The motivation for the grid construction is to form an image-like representation, with grid encoding for spatial neighborhood and channels for both transcriptomics and morphology modalities. We frame the machine-learning problem as image pattern recognition and compression, with convolutional autoencoders being the state-of-the-art architecture for addressing these challenges [26, 27]. The bottleneck layer serves as a constriction for information flow, forcing the model to capture the biological signal. Moreover, *AESTETIK*'s loss function (Eq. 4) is designed to optimize multiple objectives simultaneously by combining reconstruction loss for accurate latent representation and multi-triplet loss (Fig. 1D, Eq. 3) for structure preservation across modalities. This dual optimization ensures a comprehensive and informative representation of each data modality.

### 2.2 *AESTETIK* improves the identification of spatial domains

We benchmarked *AESTETIK* performance on multiple datasets with available ground truth annotations (Fig. 2A). In line with the methodology of [10–15], we adopted the Adjusted Rand Index (ARI) to measure the similarity between predicted cluster labels and ground truth, with the number of clusters set to match that in



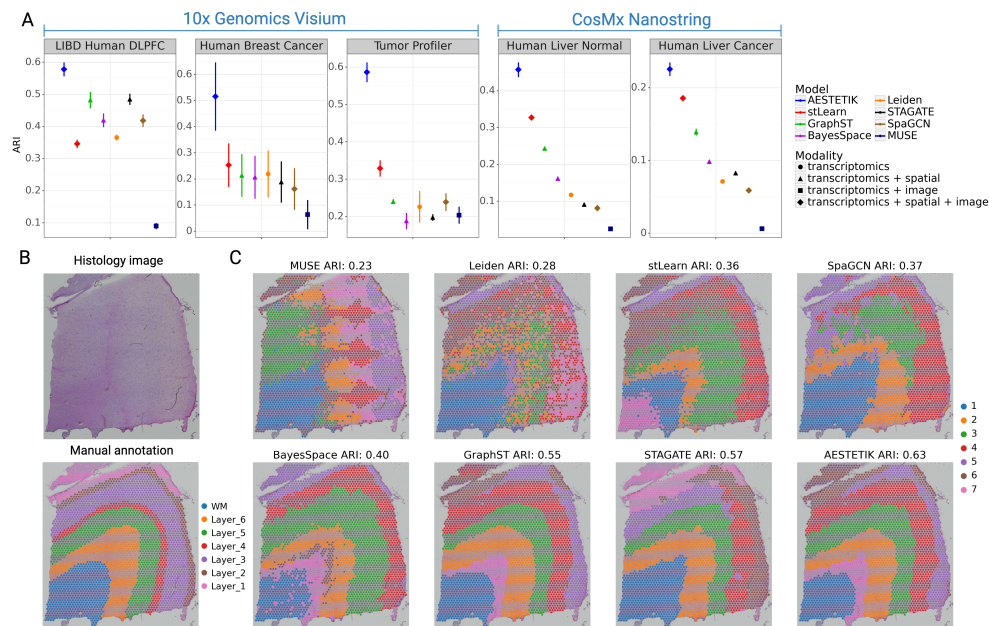


**Fig. 1** *AESTETIK* integrates spatial, transcriptomics, and morphology information to learn accurate spot representations.

**A:** Spatial transcriptomics enables in-depth molecular characterization of samples on a morphology and RNA level while preserving spatial location. **B:** Workflow of *AESTETIK*. Initially, the transcriptomics and morphology spot representations are preprocessed. Next, a dimensionality reduction technique (e.g., PCA) is applied. Subsequently, the processed spot representations are clustered separately to acquire labels required for the multi-triplet loss. Afterwards, the modality-specific representations are fused through concatenation and the grid per spot is built. This is used as an input for the autoencoder. Lastly, the spatial-, transcriptomics-, and morphology-informed spot representations are obtained and used for downstream tasks such as clustering, morphology analysis, etc. **C:** *AESTETIK* relies on a convolutional encoder-decoder architecture to learn accurate spot representations from the spatial transcriptomics data. **D:** Employing a multi-triplet loss, instead of a single triplet loss adds extra positive and negative instances per class around the anchor point, improving the placement of the anchor in the latent space.

the ground truth. To avoid hyperparameter tuning on the samples used for testing, we introduced reversed leave-one-out cross-validation. More specifically, we used a single sample and its replicates to select hyperparameters to maximize the median ARI. Then, the optimal hyperparameters were applied to the remaining test samples. This process was iterated over all folds, and the resulting median ARI, along with the standard error, is reported (Fig. 2A).

*AESTETIK* consistently yielded substantially improved cluster assignments closer to the ground truth annotations over previous state-of-the-art methods across all datasets (Fig. 2A). For example on the LIBD Human DLPFC dataset [21],



**Fig. 2** *AESTETIK* improves the identification of spatial domains with coherent expression and morphology.

**A:** Benchmark of *AESTETIK* and previous state-of-the-art methods in spatial transcriptomics on 5 datasets across 2 technology platforms. The y-axis represents the ARI between the ground truth and the predicted labels. Models are ordered based on their relative rank across the datasets. The shape represents the modalities the model integrates. **B:** Histology image and manual annotation of slice 151676 from the LIBD human DLPFC dataset [21] and **C:** Comparison of cluster assignments for the same slice.

*AESTETIK* achieved the highest ARI of  $0.58 \pm 0.02$ , significantly surpassing the second best model - *GraphST* - by 21%. The LIBD Human DLPFC dataset comprises 12 tissue slices obtained from the dorsolateral prefrontal cortex (DLPFC) brain region, sequenced using Visium from 10x Genomics™, together with curated manual annotations based on brain cytoarchitecture and known marker genes (Fig. 2B). This improvement highlights the superior performance of *AESTETIK* in effectively integrating the spatial modality and generating accurate cluster assignments in structured brain tissue. *STAGATE* and *GraphST* demonstrated lower performance, achieving ARI of  $0.48 \pm 0.02$  and  $0.48 \pm 0.03$ , respectively.

To qualitatively illustrate the cluster assignments, we compare them for slice 151676 (Fig. 2C), using the closest annotations to the ground truth across folds. *MUSE* (ARI 0.23), *Leiden* (ARI 0.28), *stLearn* (ARI 0.36) and *SpaGCN* (ARI 0.37) mixed the brain layers, accompanied by noise along the boundaries. *BayesSpace* (ARI 0.40) partitioned the white matter (WM) and layer 6 into multiple groups. While

*GraphST* (ARI 0.55) and *STAGATE* (ARI 0.57) generated mostly well-defined clusters, layers 1, 2 and 3 were inconsistent. Notably, *AESTETIK* (ARI 0.63) identified the brain architecture, and its clusters displayed clearer definitions at the boundaries, leading to superior performance (Fig. 2C, S1).

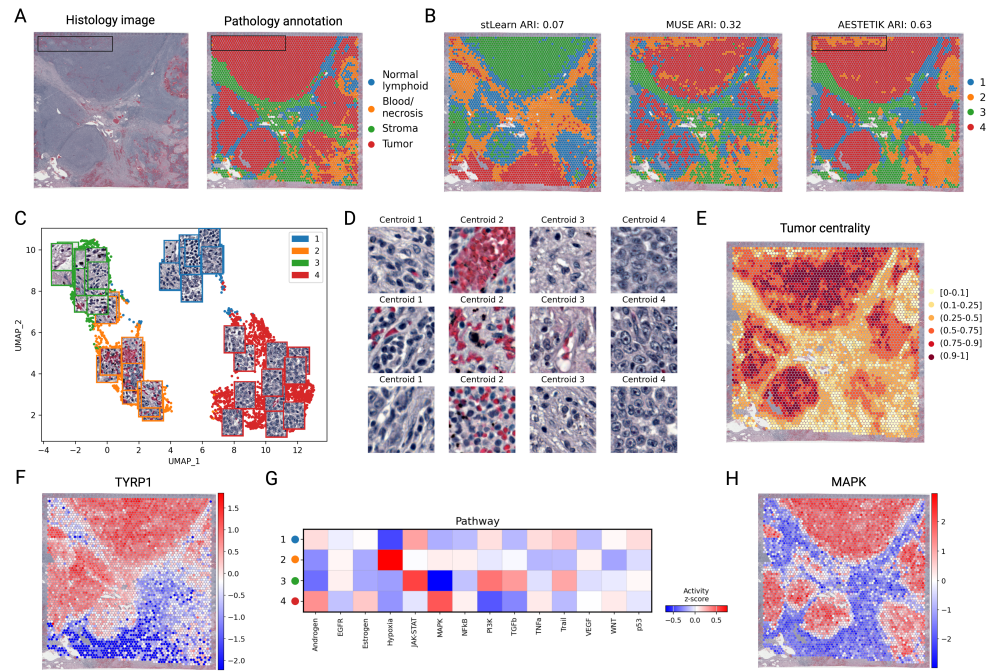
Next, we investigate the methods' performance on the Human Breast Cancer dataset, which comprises 5 tissue slices sequenced using Visium from 10x Genomics™ and annotated independently in two different labs [22]. This dataset presents unique challenges, primarily stemming from the considerable inter- and intra-sample heterogeneity, including variations in the cancer cell population. *AESTETIK* achieved the closest clusters to the ground truth labels with an ARI of  $0.51 \pm 0.13$  (Fig. 2A), indicating a two-fold increase over the second best model, *stLearn* (ARI  $0.25 \pm 0.08$ ). Despite wide standard error intervals observed in all models, *AESTETIK* exhibited heightened robustness, surpassing the challenges posed by the complexity of breast cancer tissue (Fig. S2).

### 2.3 *AESTETIK* effectively incorporates the morphology modality

We introduce a new and yet unreleased spatial transcriptomics dataset with 9 distinct tissue regions sequenced using Visium from 10x Genomics™ from the Tumor Profiler study [28]. Each region has a replicate resulting in 18 samples of size  $6.5 \times 6.5 \text{mm}^2$ , with data including 10x Genomics™ *Space Ranger v3.0.0* outputs and a corresponding H&E image scanned at a high resolution of  $0.3 \mu\text{m}/\text{pixel}$ . The tissue regions originate from 7 patients with metastatic melanoma each characterized by one of the following immune subtypes: immune desert, immune excluded, or inflamed. The ground truth annotations were obtained using histopathology software (HALO AI™ (Indica Labs, Corrales, NM, USA)), classifying the spots into one of the following categories: tumor, stroma, normal lymphoid, and blood/necrosis. Following this, a pathologist manually reviewed the model predictions (Fig. 3A). We consider this dataset to be a valuable reference benchmark for evaluating the performance of spatial transcriptomics models, particularly in terms of their ability to integrate morphology effectively.

On this dataset, *AESTETIK* achieved a 79% increase in ARI ( $0.59 \pm 0.03$ ) over previous state-of-the-art methods, demonstrating effective use of the morphology modality (Fig. 2A). While both *stLearn* (ARI  $0.33 \pm 0.02$ ) and *MUSE* (ARI  $0.20 \pm 0.02$ ) use the same pre-trained *Inception v3* [25] for extracting morphology features, they fall short in effectively leveraging this information (Fig. 3B). On the other hand, *AESTETIK* not only produced accurate cluster assignments (Fig. 3B), but also identified a hemorrhage region in the upper left of the H&E image, that was overlooked during annotation (black box in Fig. 3A, B, S3).

Next, we qualitatively explore the latent representations and the identified spatial domains by focusing on slice MACEGEJ-2-2 (Fig. 3A). We visualized the latent space using *UMAP* [30] with randomly sampled morphology spot representations (Fig. 3C).



**Fig. 3** *AESTETIK* effectively incorporates the morphology modality revealing biologically relevant spatial organization of cancer tissue.

**A:** Histology image and pathology annotation of slice MACEGEJ-2-2 from the Tumor Profiler dataset. **B:** Comparison of cluster assignments for slice MACEGEJ-2-2. **C:** UMAP plot of the *AESTETIK*'s latent space with randomly sampled spot images. **D:** Most representative cluster spots based on the obtained representations. **E:** Euclidean distance in latent space of each spot to the tumor centroid plotted in spatial space. Most representative spots are located in the middle of the tumor formations. **F:** Spatial marker gene expression of TYRP1. **G:** Pathway analysis of the identified clusters using *decoupler* [29]. **H** Spatial activation of MAPK pathway.

We observe an aggregation of tumor spots (cluster 4) on the bottom-right side, showcasing similarities in their characteristics. On the lower left side, there are areas with blood and necrosis. While clusters 1 and 3, representing normal lymphoid tissue and stromal cells, are positioned in the upper part, a closer inspection reveals discernible differences in their underlying structures. Besides, for enhanced explainability of the spot representations, we selected the most representative spots per cluster (Fig. 3D). Visually, tumor cells within cluster 4 exhibit distinct characteristics; they appear significantly larger, displaying irregular shapes, and possessing enlarged nuclei. Stromal cells (cluster 3) have an elongated morphology and are noticeably more scattered [31] compared to the normal lymphoid cells (cluster 1), which are generally smaller and denser [32].

Furthermore, to illustrate the effect of encoding spatial information in latent space,

we computed the Euclidean distance of each spot to the tumor centroid and visualized it spatially (Fig. 3E). While stromal cells are the furthest, we observed that tumor cells close in latent space are clustered spatially, with the most representative spots located in the middle of the tumor formations.

To provide additional insights, we found TYRP1 and TKTL1 among the top tumor marker genes in the tumor cluster, confirming the model predictions for the spatially resolved identification of melanoma cells (Fig. 3F, S4). TYRP1 gene is involved in melanocyte pigmentation, associated with melanoma progression and is a target for oncological immunotherapy [33–35]. The second highly upregulated gene, TKTL1, is implicated in the progression of melanoma and contributes to the increased invasion of melanoma cells [36]. Further, we performed a pathway analysis of the tumor clusters using *decoupler* [29] (Fig. 3G) which revealed increased activity of the MAPK pathway in the cancer cluster (Fig. 3H), known for promoting cell proliferation, invasion, metastasis, migration, survival, and angiogenesis [37–39]. Furthermore, we observed that hypoxia signatures were predominant in areas of necrosis and hemorrhage and JAK-STAT inflammatory signaling was predominant in the cancer microenvironment clusters 1 and 3. These pathway-level analyses underline the robust associations of the spatially resolved clustering results achieved by *AESTETIK* and support interpretation in the context of the underlying biology.

## 2.4 *AESTETIK* improves cluster assignment in single-cell spatial transcriptomics

CosMx NanoString™ released a liver dataset with single-cell resolution, encompassing two tissue regions from normal and cancer patients and capturing 1000 genes. It offers valuable insights into liver biology and cancer characteristics. More specifically, using these datasets, we assess the models' effectiveness at single-cell resolution by comparing the clusters they produce with the cell types reported by NanoString™. In both normal and cancer liver tissue, *AESTETIK* exhibited outstanding performance, substantially outperforming the other models by 39% and 21%, with ARI of  $0.46 \pm 0.02$  and  $0.23 \pm 0.00$ , respectively (Fig. 2A, S5, S5). The second best model, *stLearn*, attained a score of  $0.33 \pm 0.00$  and  $0.19 \pm 0.00$ , followed by *GraphST* with  $0.24 \pm 0.00$  and  $0.14 \pm 0.00$ . Overall, the clustering accuracy on the cancer tissue is lower compared to the normal sample. However, the relative trend in the ranking of the models remained consistent.

## 2.5 Joint integration of multi-modal data enhances computational analysis

To pinpoint the benefit of the spatial modality in the LIBD Human DLPFC and Tumor Profiler datasets, we systematically varied the grid's window size, ranging from 1 (w/o spatial information) to 11, and measured the change in ARI (Fig. 4A). The grid size determines the number of spatially adjacent spots to consider. Local spatial information proved important, preserving local details and spot-to-spot variability. However, incorporating a more extensive global context through a larger window size (e.g., 11) introduced noise and hampered performance, which was likely due to signal



over-smoothing and the loss of spot-specific details.

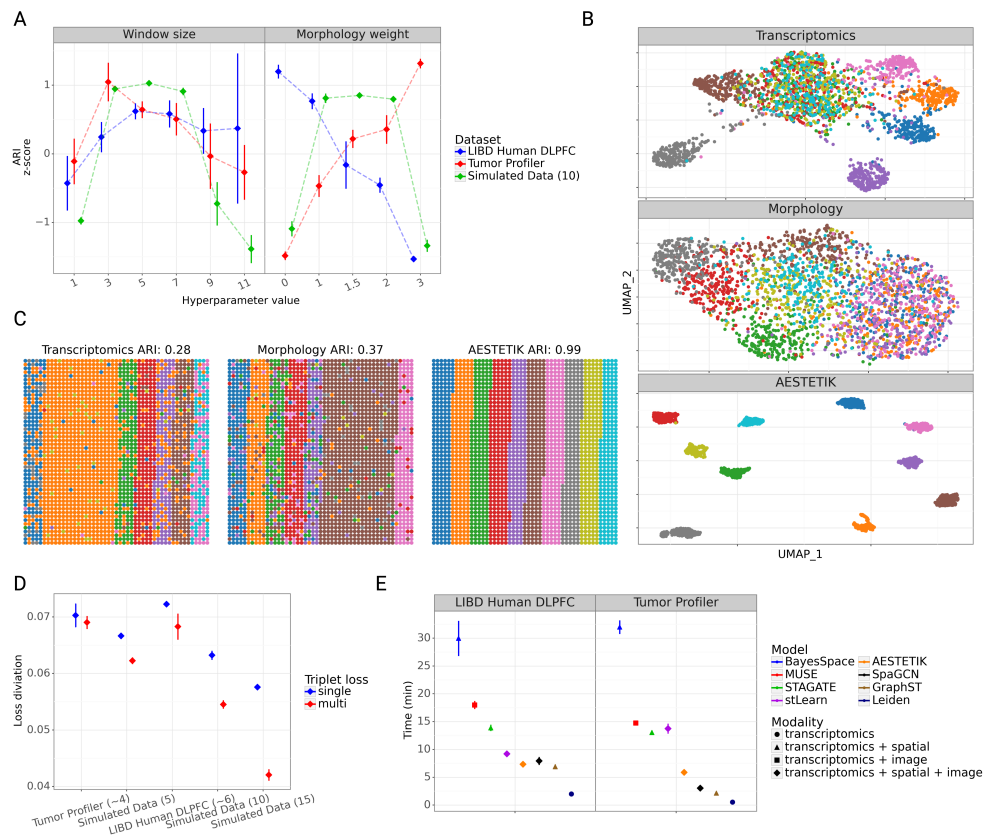
Similarly, we studied the contribution of each modality by varying the morphology weight (0 - no morphology; 1.5 - equal weight between transcriptomics and morphology; 3 - only morphology). As expected, we found that the transcriptomics modality in the brain dataset is informative given the provided ground truth annotations, relying on known cytoarchitecture and marker genes [21]. In contrast, the ground truth for the Tumor Profiler, derived from histopathology software, was morphology driven (Fig. 4A). Furthermore, to underscore the significance of methods incorporating all data modalities, we present a scenario illustrating the necessity of both modalities to reveal ground truth annotations (Fig. 4B). Following the approach of [11], we simulated data where both modalities are essential for accurate cluster identification. Our ablation study demonstrated that the optimal ARI was achieved when accounting for both modalities, thus emphasizing the critical significance of multi-modal data integration (Fig. 4C).

Additionally, the multi-triplet loss demonstrated an enhancement in loss stability during training compared to the single triplet loss (Fig. 4D). The refined positioning of clusters in latent space, considering multiple positive and negative spots, becomes crucial, especially when dealing with datasets containing numerous clusters. Lastly, the runtime per tissue slice for our model, incorporating all three modalities (~8 min), was either lower or comparable to that of other models (Fig. 4E). For example, *BayesSpace*, *MUSE*, and *STAGATE*, incorporating only two modalities, required ~28 min, ~17 min, and ~13 min, respectively. Moreover, we demonstrate that *AESTETIK* is well-suited for analyzing large spatial transcriptomics datasets, scaling to millions of spots (Fig. S7).

### 3 Discussion

In this work, we propose *AESTETIK*, a method that jointly integrates spatial, transcriptomics, and morphology information to learn accurate spot representations. Our results consistently showed superior performance to state-of-the-art methods across structured tissues (e.g., brain) and cancer tissues with heterogeneous cell populations (e.g., breast, melanoma, liver) across widely adopted spatial transcriptomics technologies (e.g., 10x Genomics™, Visium, CosMx NanoString™). We systematically demonstrated the significance of jointly integrating multi-modal data to improve spatial transcriptomics analysis and yield more precise spot annotations.

This improvement in spot representation resulted from modeling the spatial transcriptomics modalities as a grid encoding the spatial spot neighborhood and channels as transcriptomics and morphology modalities. Our approach framed the machine-learning problem as image pattern recognition and compression, where convolution filters jointly learn the importance of neighboring spots and channels. This proved beneficial in both structured and heterogeneous tissues. In contrast, the GNNs, employed by *SpaGCN*, *GraphST* and *STAGATE* demonstrated variations in



**Fig. 4 Joint integration of multi-modal data enhances computational analysis**

**A:** Ablation study on the influence of window size and morphology weight on the ARI. The y-axis represents the ARI, normalized by dataset. **B:** UMAP visualization of single (transcriptomics, morphology) and combined (AESTETIK) modality representations on a simulated tissue slice, colored by the 10 ground truth annotations. **C:** Cluster assignments based on only a single modality (transcriptomics, morphology) and AESTETIK's joint representations. **D:** Comparing the stability of the single and multi-triplet loss on the loss function of AESTETIK across datasets. The number, following the dataset name, is the median number of clusters present. The y-axis represents the standard deviation computed on the loss difference over successive training epochs. **E:** Runtime for the evaluated clustering methods. The y-axis represents the time in minutes. Models are ordered based on their relative rank across the datasets.

their performance relative to the other methods across tissue types and spatial transcriptomics technologies. This could be attributed to the inherent susceptibility of GNNs to noise [19, 20]. The graph structure ensures connectivity among neighboring spots, which is useful in structured tissues (e.g., brain) with coherent spatial patterns. However, it presents challenges in samples of lower sequencing quality or tissues with higher heterogeneity (e.g., cancer cell populations), where noise, introduced

through node perturbations and edge alterations, might affect the graph structure. Consequently, this undermines the robustness and performance of current GNNs in downstream applications.

Further, in our ablation study on the brain dataset, we quantitatively demonstrated the significance of the spatial modality in identifying the brain layer structure. We discovered that a relatively small grid's window size (5-7) sufficiently captures the desired spatial signal. Opting for a larger neighborhood (e.g., window size 11) offers no extra value. Unlike the global tissue context, the local environment better preserves spot-specific signals and nearby variability. Ultimately, our ablation results underscore **1)** the importance of jointly integrating the available spatial transcriptomics data modalities for accurate spot representation, and **2)** the necessity for external knowledge to prioritize the signal of interest, depending on the particular research question at hand.

Several paths to further improve model accuracy appear promising. **1)** We employed the pre-trained *Inception v3* [25] to extract morphology features. However, adopting a model tailored to a specific task (e.g., cell nuclei segmentation and classification) would likely yield more informative spot features, potentially leading to improved overall performance. **2)** *AESTETIK* randomly selects the positive and negative pairs for each anchor point during training. We believe this process can be improved by utilizing a *smarter* strategy for triplet mining, which would eventually improve the performance, and robustness to noise.

In the future, *AESTETIK* could be effectively applied to fine-map cell populations in spatial transcriptomics datasets [40–42], to systematically analyze the interplay between different modalities by varying their contribution and to gain a multi-modal understanding of key biological processes. To foster these downstream applications, we have released the code for *AESTETIK* along with examples demonstrating its usage. Moreover, we anticipate that upon its release, the 10x Genomics™, Visium dataset from the Tumor Profiler study will serve as a valuable reference benchmark for assessing spatial transcriptomics model performance and explainability. Thus, we hope that our model, together with this dataset, will stimulate further improvements in computational spatial transcriptomics analysis.

## 4 Methods & Materials

### 4.1 Data preprocessing

*AESTETIK* takes in spatial, transcriptomics, and morphology information. We apply the same preprocessing pipeline across datasets and sequencing technologies. For simplicity, we refer to both spot and cell as a spot (*a single spot can contain 1 cell*).



#### 4.1.1 Transcriptomics modality

Starting with raw counts, genes expressed in fewer than 10 spots are removed. Then, the *scanpy* function *highly variable genes* computes normalized variance in *Seurat v3* style, removing genes with variance below 1 [23, 43]. Each spot undergoes normalization by total counts over all genes, followed by  $\log_{1p}$  transformation and scaling. Subsequently, PCA is applied to the preprocessed counts, extracting the first 15 PCs [10].

#### 4.1.2 Morphology modality

The raw RGB image for each tissue slice is divided into tiles, each representing a spot and its defined neighborhood based on the spot diameter. Following the default preprocessing steps of *Inception v3* [25], the tiles are resized to 299, their center is cropped and the RGB channels are normalized. Morphology features are then extracted from the last network layer (with 2,048 dimensions) of the pre-trained on *Imagenet* [24] deep-learning model *Inception v3*. Finally, PCA reduces the feature dimension from 2,048 to 15.

#### 4.1.3 Grid construction

To begin, each spot is represented by two vectors containing the first  $n_{d1}$  and  $n_{d2}$  PCs obtained via PCA from the preprocessed transcriptomics and morphology modalities, along with their spatial coordinates. For simplicity, we assume  $n_{d1} = n_{d2} = n_{pca}$ , but the following workflow holds also for  $n_{d1} \neq n_{d2}$ . These vectors are concatenated and scaled in the range  $[0, 1]$ . Then, a square grid for each spot is constructed with the number of spatial neighbors,  $N_{grid}$ , chosen as an odd number to ensure the center position of the selected spot in the window. This results in a tensor of size:

$$\text{spot}_i = (N_{grid}, N_{grid}, 2 * n_{pca}) \quad (1)$$

which can be interpreted as  $N_{grid} \times N_{grid}$  image with  $2 * n_{pca}$  channels. For missing or located on the borders spots, we apply padding by taking the median expression over each channel in  $\text{spot}_i$ .

#### 4.1.4 Clustering

The default clustering algorithm is *Bayesian Gaussian Mixture* with a diagonal covariance matrix from the *sklearn* package, but we also support *K-Means*, *Leiden*, and *Lowvain*. Once the cluster labels are obtained, an additional preprocessing step can be applied. A K-Neighbors Classifier is fitted using spatial coordinates and the already obtained clusters to refine the cluster assignments in spatial space through majority voting.

## 4.2 Model architecture

*AESTETIK* utilizes a convolutional deep-learning autoencoder with a standard encoder-decoder architecture and a bottleneck layer. The encoder comprises a convolutional layer, max-pooling layer, batch normalization, *ReLU* activation, and linear layer. Default hyperparameters include 64 convolutional kernels (size 7), dropout ( $p = 0.3$ ), max-pooling (stride 3), and a linear layer (size 16). The decoder follows a mirrored architecture, concluding with a sigmoid function to constrain output values in the range  $[0, 1]$ . *AESTETIK* is a *Python* package implemented in *PyTorch*.

### 4.2.1 Autoencoder ensemble

To improve the network stability, we employ ensemble architectures for both the encoder and decoder, utilizing random *LeCun* [44] initialization. The ensemble's output is determined by taking the median over predictions. We train an ensemble with 3 encoders and decoders. The final representation is computed by dropout sampling 1,000 times and taking the median value.

### 4.2.2 Reconstruction loss

We employ a reconstruction loss to ensure that the latent space effectively captures the biological complexity of the morphology and transcriptomics modalities. We define it as:

$$\mathbb{L}_{rec} = \alpha * \mathbb{L}_{rec_m} + (3 - \alpha) * \mathbb{L}_{rec_{tr}}, \quad (2)$$

where  $\alpha \in [0, 3]$  is a hyperparameter for the morphology weight.  $\mathbb{L}_{rec_{[m, tr]}}$  is the standard L1 reconstruction loss with  $m$  for morphology and  $tr$  for transcriptomics modality. We use L1 loss due to the input-output range being  $[0, 1]$ .

### 4.2.3 Multi-triplet loss

We apply triplet loss to preserve the structure across modalities. Its primary objective is to learn a spot representation in which similar instances are closer together, while dissimilar instances are farther apart. Define an anchor point  $A$  with label  $l_i$ , then we draw at random a positive point  $P$  with label  $l_i$  and a negative point  $N$  with label  $l_j$  such as  $l_i \neq l_j$ , then the single triplet loss is defined as:

$$\mathbb{L}_{triplet}(A, P, N) = \max(0, \|f(A) - f(P)\|^2 - \|f(A) - f(N)\|^2 + margin)$$

In spatial transcriptomics, multiple classes and high noise ratios are typical. Using just one positive and negative point can lead to unstable representation and increase the training time due to alternations. To improve the spot representation robustness, we propose the multi-triplet loss motivated by [45]. Let  $L$  be the number of unique labels in the dataset. Define an anchor point  $A$  with label  $l_i$ , then we draw, with replacement,  $L - 1$  positive points  $\{P_1, P_2, \dots, P_{L-1}\}$  with label  $l_i$ . Additionally, for each label  $l_j$  where  $j \neq i$ , we draw a single negative point, resulting in  $\{N_1, N_2, \dots, N_{L-1}\}$ . Then

the multi-triplet loss for a single modality can be defined as:

$$\mathbb{L}_{multi\_triplet}(A) = \frac{1}{L-1} \sum_{k=1}^{L-1} \mathbb{L}_{triplet}(A, P_k, N_k)$$

which when extended to all spots:

$$\mathbb{L}_{multi\_triplet} = \frac{1}{N_{spots}} \sum_{l=1}^{N_{spots}} \mathbb{L}_{multi\_triplet}(A_l)$$

The modality-weighted multi-triplet loss is defined as:

$$\mathbb{L}_{multi\_triplet} = \alpha * \mathbb{L}_{multi\_triplet_m} + (3 - \alpha) * \mathbb{L}_{multi\_triplet_{tr}} \quad (3)$$

with  $\alpha$  defined as in equation 2.

#### 4.2.4 Loss function

The overall loss function for training combines reconstruction loss to ensure accurate latent representation and multi-triplet loss for preserving structure across modalities. Formally, it is defined as:

$$\mathbb{L}_{\text{AESTETIK}} = \mathbb{L}_{rec} + \mathbb{L}_{multi\_triplet},$$

which can be rewritten as:

$$\mathbb{L}_{\text{AESTETIK}} = \alpha * (\mathbb{L}_{rec_m} + \mathbb{L}_{multi\_triplet_m}) + (3 - \alpha) * (\mathbb{L}_{rec_{tr}} + \mathbb{L}_{multi\_triplet_{tr}}) \quad (4)$$

#### 4.2.5 Training details

The model is trained for 100 epochs using *Adam* [46] with a weight decay of 1e-6, a learning rate of 1e-3, and a batch size corresponding to the number of spots in a tissue slice. The run time is approximately 8 minutes on a GPU, with inference time under a minute. Computational data analysis was performed at Leonhard Med (<https://sis.id.ethz.ch/services/sensitiveresearchdata/>) secure trusted research environment at ETH Zurich.

#### 4.2.6 Evaluation

We propose reversed leave-one-out cross-validation for model evaluation to avoid hyperparameter tuning on test samples. We utilize a single sample and its replicates to select hyperparameters through a grid search, aiming to maximize the median ARI. Subsequently, the optimal hyperparameters are applied to the remaining test samples. For state-of-the-art methods, we consider hyperparameter values suggested by the authors, as well as those discussed in the corresponding paper. Hyperparameter values are provided in the supplement. To ensure comparable conditions across models, the

number of clusters is pre-defined based on the provided ground truth. Performance is assessed using the ARI between ground-truth labels and cluster assignments. We bootstrap 10,000 times from the median ARI across the test folds and report the resulting median ARI and its standard error. For most datasets, we generated all possible sample combinations (folds). However, for the larger CosMx NanoString™ Liver dataset we use each FOV to select the hyperparameters, but we evaluate on randomly selected 20 FOVs, not adjacent to the FOVs used for optimization.

#### 4.2.7 Ablation study

In the ablation study on *AESTETIK*, we adhere to the procedure described in 4.2.6, where we fix the value for the hyperparameter of interest and assess its impact on the ARI. To ensure comparability across datasets, we compute an ARI z-score.

### 4.3 Downstream applications

#### 4.3.1 Marker genes

For marker genes, we employ the *rank genes groups* function from *scanpy* using the Wilcoxon signed-rank test. Significant marker genes (adjusted p-value < 0.05) are selected and sorted by their average log-fold change. The top 15 marker genes per cluster are reported.

#### 4.3.2 Pathway analysis

For pathway analysis, we utilize the multivariate linear model from the *decoupler* package [29] to compute regulatory pathway activities from the *PROGENy* database [47].

#### 4.3.3 Cluster centroids in latent space

To determine the centroid for each cluster in the latent space, we employed a method minimizing the sum of Euclidean distances among all samples within that class. Subsequently, we computed the top N spots near each cluster centroid.

### 4.4 Data availability

The LIBD Human DLPFC dataset is available at <https://github.com/LieberInstitute/HumanPilot> and <http://research.libd.org/spatialLIBD>; Human Breast Cancer - Zenodo <https://doi.org/10.5281/zenodo.4739739>, Human Liver Normal and Cancer - <https://nanosttring.com/products/cosmx-spatial-molecular-imager/human-liver-rna-ffpe-dataset/>. The metastatic melanoma dataset with 18 tissue slices from Tumor Profiler samples sequenced using Visium from 10x Genomics™ will be made available upon acceptance of publication.

### 4.5 Code availability

The open-source implementation of *AESTETIK* along with a tutorial is available at:

- [www.github.com/ratschlab/aestetik](https://www.github.com/ratschlab/aestetik)

The Snakemake pipeline for reproducing the results is available at:

- [www.github.com/ratschlab/st-rep](https://www.github.com/ratschlab/st-rep)

**Acknowledgments.** This work was supported by the Swiss Federal Institutes of Technology (strategic focus area of personalized health and related technologies; 2021–367). The 10x spatial transcriptomics sequencing of the Tumor Profiler samples was made possible through a technology access program by 10x Genomics™, with special acknowledgments to Jacob Stern, James Chell, Rudi Schläfli, Laura Lipka, Mario Werner, Nikhil Rao, and Scott Brouillette for their invaluable contributions. The Tumor Profiler study was supported by a public-private partnership involving Roche Holding AG, ETH Zurich, University of Zurich, University Hospital Zurich, and University Hospital Basel. V.H.K. gratefully acknowledges additional funding by the Promedica Foundation (F-87701-41-01).

**TUMOR PROFILER CONSORTIUM.** Rudolf Aebersold<sup>5</sup>, Melike Ak<sup>33</sup>, Faisal S Al-Quaddoomi<sup>12,22</sup>, Silvana I Albert<sup>10</sup>, Jonas Albinus<sup>10</sup>, Ilaria Alborelli<sup>29</sup>, Sonali Andani<sup>9,22,31,36</sup>, Per-Olof Attinger<sup>14</sup>, Marina Bacac<sup>21</sup>, Daniel Baumhoer<sup>29</sup>, Beatrice Beck-Schimmer<sup>44</sup>, Niko Beerenwinkel<sup>7,22</sup>, Christian Beisel<sup>7</sup>, Lara Bernasconi<sup>32</sup>, Anne Bertolini<sup>12,22</sup>, Bernd Bodenmiller<sup>11,40</sup>, Ximena Bonilla<sup>9</sup>, Lars Bosshard<sup>12,22</sup>, Byron Calgua<sup>29</sup>, Ruben Casanova<sup>40</sup>, Stéphane Chevrier<sup>40</sup>, Natalia Chicherova<sup>12,22</sup>, Ricardo Coelho<sup>23</sup>, Maya D’Costa<sup>13</sup>, Esther Danenberg<sup>42</sup>, Natalie R Davidson<sup>9</sup>, Monica-Andreea Dragan<sup>7</sup>, Reinhard Dummer<sup>33</sup>, Stefanie Engler<sup>40</sup>, Martin Erkens<sup>19</sup>, Katja Eschbach<sup>7</sup>, Cinzia Esposito<sup>42</sup>, André Fedier<sup>23</sup>, Pedro F Ferreira<sup>7</sup>, Joanna Ficek-Pascual<sup>1,9,16,22,31</sup>, Anja L Frei<sup>36</sup>, Bruno Frey<sup>18</sup>, Sandra Goetze<sup>10</sup>, Linda Grob<sup>12,22</sup>, Gabriele Gut<sup>42</sup>, Detlef Günther<sup>8</sup>, Pirmin Haeuptle<sup>3</sup>, Viola Heinzlmann-Schwarz<sup>23,28</sup>, Sylvia Herter<sup>21</sup>, Rene Holtackers<sup>42</sup>, Tamara Huesser<sup>21</sup>, Alexander Immer<sup>9,17</sup>, Anja Irmisch<sup>33</sup>, Francis Jacob<sup>23</sup>, Andrea Jacobs<sup>40</sup>, Tim M Jaeger<sup>14</sup>, Katharina Jahn<sup>7</sup>, Alva R James<sup>9,22,31</sup>, Philip M Jermann<sup>29</sup>, André Kahles<sup>9,22,31</sup>, Abdullah Kahraman<sup>22,36</sup>, Viktor H Koelzer<sup>36,41</sup>, Werner Kuebler<sup>30</sup>, Jack Kuipers<sup>7,22</sup>, Christian P Kunze<sup>27</sup>, Christian Kurzeder<sup>26</sup>, Kjong-Van Lehmann<sup>2,4,9,15</sup>, Mitchell Levesque<sup>33</sup>, Ulrike Lischetti<sup>23</sup>, Flavio C Lombardo<sup>23</sup>, Sebastian Lugert<sup>13</sup>, Gerd Maass<sup>18</sup>, Markus G Manz<sup>35</sup>, Philipp Markolin<sup>9</sup>, Martin Mehnert<sup>10</sup>, Julien Mena<sup>5</sup>, Julian M Metzler<sup>34</sup>, Nicola Miglino<sup>35,41</sup>, Emanuela S Milani<sup>10</sup>, Holger Moch<sup>36</sup>, Simone Muenst<sup>29</sup>, Riccardo Murri<sup>43</sup>, Charlotte KY Ng<sup>29,39</sup>, Stefan Nicolet<sup>29</sup>, Marta Nowak<sup>36</sup>, Monica Nunez Lopez<sup>23</sup>, Patrick GA Pedrioli<sup>6</sup>, Lucas Pelkmans<sup>42</sup>, Salvatore Piscuoglio<sup>23,29</sup>, Michael Prummer<sup>12,22</sup>, Prélot, Laurie<sup>9,22,31</sup>, Natalie Rimmer<sup>23</sup>, Mathilde Ritter<sup>23</sup>, Christian Rommel<sup>19</sup>, María L Rosano-González<sup>12,22</sup>, Gunnar Rättsch<sup>1,6,9,22,31</sup>, Natascha Santacroce<sup>7</sup>, Jacobo Sarabia del Castillo<sup>42</sup>, Ramona Schlenker<sup>20</sup>, Petra C Schwalie<sup>19</sup>, Severin Schwan<sup>14</sup>, Tobias Schär<sup>7</sup>, Gabriela Senti<sup>32</sup>, Wenguang Shao<sup>10</sup>, Franziska Singer<sup>12,22</sup>, Sujana Sivapatham<sup>40</sup>, Berend Snijder<sup>5,22</sup>, Bettina Sobottka<sup>36</sup>, Vipin T Sreedharan<sup>12,22</sup>, Stefan Stark<sup>9,22,31</sup>, Daniel J Stekhoven<sup>12,22</sup>, Tanmay Tanna<sup>7,9</sup>, Alexandre PA Theocharides<sup>35</sup>, Tinu M Thomas<sup>9,22,31</sup>, Markus Tolnay<sup>29</sup>, Vinko Tosevski<sup>21</sup>, Nora C Toussaint<sup>12,22</sup>, Mustafa A Tuncel<sup>7,22</sup>, Marina Tusup<sup>33</sup>,

Audrey Van Drogen<sup>10</sup>, Marcus Vetter<sup>25</sup>, Tatjana Vljajnic<sup>29</sup>, Sandra Weber<sup>32</sup>, Walter P Weber<sup>24</sup>, Rebekka Wegmann<sup>5</sup>, Michael Weller<sup>38</sup>, Fabian Wendt<sup>10</sup>, Norbert Wey<sup>36</sup>, Andreas Wicki<sup>35,41</sup>, Mattheus HE Wildschut<sup>5,35</sup>, Bernd Wollscheid<sup>10</sup>, Shuqing Yu<sup>12,22</sup>, Johanna Ziegler<sup>33</sup>, Marc Zimmermann<sup>9</sup>, Martin Zoche<sup>36</sup>, Gregor Zuend<sup>37</sup>

<sup>1</sup>AI Center at ETH Zurich, Andreasstrasse 5, 8092 Zurich, Switzerland, <sup>2</sup>Cancer Research Center Cologne-Essen, University Hospital Cologne, Cologne, Germany, <sup>3</sup>Cantonal Hospital Baselland, Medical University Clinic, Rheinstrasse 26, 4410 Liestal, Switzerland, <sup>4</sup>Center for Integrated Oncology Aachen (CIO-A), Aachen, Germany, <sup>5</sup>ETH Zurich, Department of Biology, Institute of Molecular Systems Biology, Otto-Stern-Weg 3, 8093 Zurich, Switzerland, <sup>6</sup>ETH Zurich, Department of Biology, Wolfgang-Pauli-Strasse 27, 8093 Zurich, Switzerland, <sup>7</sup>ETH Zurich, Department of Biosystems Science and Engineering, Mattenstrasse 26, 4058 Basel, Switzerland, <sup>8</sup>ETH Zurich, Department of Chemistry and Applied Biosciences, Vladimir-Prelog-Weg 1-5/10, 8093 Zurich, Switzerland, <sup>9</sup>ETH Zurich, Department of Computer Science, Institute of Machine Learning, Universitätsstrasse 6, 8092 Zurich, Switzerland, <sup>10</sup>ETH Zurich, Department of Health Sciences and Technology, Otto-Stern-Weg 3, 8093 Zurich, Switzerland, <sup>11</sup>ETH Zurich, Institute of Molecular Health Sciences, Otto-Stern-Weg 7, 8093 Zurich, Switzerland, <sup>12</sup>ETH Zurich, NEXUS Personalized Health Technologies, Wagistrasse 18, 8952 Zurich, Switzerland, <sup>13</sup>F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, 4070 Basel, Switzerland, <sup>14</sup>F. Hoffmann-La Roche Ltd, Grenzacherstrasse 124, 4070 Basel, Switzerland, <sup>15</sup>Joint Research Center Computational Biomedicine, University Hospital RWTH Aachen, Aachen, Germany, <sup>16</sup>Life Science Zurich Graduate School, Biomedicine PhD Program, Winterthurerstrasse 190, 8057 Zurich, Switzerland, <sup>17</sup>Max Planck ETH Center for Learning Systems, <sup>18</sup>Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany, <sup>19</sup>Roche Pharmaceutical Research and Early Development, Roche Innovation Center Basel, Grenzacherstrasse 124, 4070 Basel, Switzerland, <sup>20</sup>Roche Pharmaceutical Research and Early Development, Roche Innovation Center Munich, Roche Diagnostics GmbH, Nonnenwald 2, 82377 Penzberg, Germany, <sup>21</sup>Roche Pharmaceutical Research and Early Development, Roche Innovation Center Zurich, Wagistrasse 10, 8952 Schlieren, Switzerland, <sup>22</sup>SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland, <sup>23</sup>University Hospital Basel and University of Basel, Department of Biomedicine, Hebelstrasse 20, 4031 Basel, Switzerland, <sup>24</sup>University Hospital Basel and University of Basel, Department of Surgery, Brustzentrum, Spitalstrasse 21, 4031 Basel, Switzerland, <sup>25</sup>University Hospital Basel, Brustzentrum & Tumorzentrum, Petersgraben 4, 4031 Basel, Switzerland, <sup>26</sup>University Hospital Basel, Brustzentrum, Spitalstrasse 21, 4031 Basel, Switzerland, <sup>27</sup>University Hospital Basel, Department of Information- and Communication Technology, Spitalstrasse 26, 4031 Basel, Switzerland, <sup>28</sup>University Hospital Basel, Gynecological Cancer Center, Spitalstrasse 21, 4031 Basel, Switzerland, <sup>29</sup>University Hospital Basel, Institute of Medical Genetics and Pathology, Schönbeinstrasse 40, 4031 Basel, Switzerland, <sup>30</sup>University Hospital Basel, Spitalstrasse 21/Petersgraben 4, 4031 Basel, Switzerland, <sup>31</sup>University Hospital Zurich, Biomedical Informatics, Schmelzbergstrasse 26, 8006 Zurich, Switzerland, <sup>32</sup>University Hospital Zurich, Clinical Trials Center, Rämistrasse 100, 8091 Zurich, Switzerland, <sup>33</sup>University Hospital Zurich, Department of Dermatology, Gloriastrasse 31, 8091 Zurich, Switzerland, <sup>34</sup>University Hospital Zurich, Department of Gynecology, Frauenklinikstrasse 10, 8091 Zurich, Switzerland, <sup>35</sup>University Hospital Zurich, Department of Medical Oncology and Hematology, Rämistrasse

100, 8091 Zurich, Switzerland, <sup>36</sup>University Hospital Zurich, Department of Pathology and Molecular Pathology, Schmelzbergstrasse 12, 8091 Zurich, Switzerland, <sup>37</sup>University Hospital Zurich, Rämistrasse 100, 8091 Zurich, Switzerland, <sup>38</sup>University Hospital and University of Zurich, Department of Neurology, Frauenklinikstrasse 26, 8091 Zurich, Switzerland, <sup>39</sup>University of Bern, Department of BioMedical Research, Murtenstrasse 35, 3008 Bern, Switzerland, <sup>40</sup>University of Zurich, Department of Quantitative Biomedicine, Winterthurerstrasse 190, 8057 Zurich, Switzerland, <sup>41</sup>University of Zurich, Faculty of Medicine, Zurich, Switzerland, <sup>42</sup>University of Zurich, Institute of Molecular Life Sciences, Winterthurerstrasse 190, 8057 Zurich, Switzerland, <sup>43</sup>University of Zurich, Services and Support for Science IT, Winterthurerstrasse 190, 8057 Zurich, Switzerland, <sup>44</sup>University of Zurich, VP Medicine, Künstlergasse 15, 8001 Zurich, Switzerland

**Consent for publication.** This manuscript has been seen and approved by all listed authors. The figures were created using BioRender.com and exported under a paid subscription.

**Funding.** We gratefully acknowledge funding from the Tumor Profiler Initiative and the Tumor Profiler Center (to V.H.K., G.R.). The Tumor Profiler study is jointly funded by a public-private partnership involving F. Hoffmann-La Roche Ltd., ETH Zurich, University of Zurich, University Hospital Zurich, and University Hospital Basel. We also acknowledge funding of S.A. from the Swiss Federal Institutes of Technology strategic focus area of personalized health and related technologies project 2021-367 (to G.R., V.H.K.), of K.N. by Swiss National Science Foundation grants 220127 (to G.R.) and 201656, and ETH core funding (to G.R.), UZH core funding (to V.H.K) and funding by the Promedica Foundation grant F-87701-41-01 (to V.H.K).

**Conflict of interest/Competing interests.** V.H.K reports being an invited speaker for Sharing Progress in Cancer Care (SPCC) and Indica Labs; advisory board of Takeda; and sponsored research agreements with Roche and IAG, all unrelated to the current study. VHK is a participant in a patent application on the assessment of cancer immunotherapy biomarkers by digital pathology; a patent application on multimodal deep learning for the prediction of recurrence risk in cancer patients, and a patent application on predicting the efficacy of cancer treatment using deep learning. GR is a participant in a patent application on matching cells from different measurement modalities which is not directly related to the current work. Moreover, G.R. is a cofounder of Computomics GmbH, Germany, and one of its shareholders.

## References

- [1] Rao, A., Barkley, D., França, G.S., Yanai, I.: Exploring tissue architecture using spatial transcriptomics. *Nature* **596**(7871), 211–220 (2021)
- [2] Asp, M., Bergenstråhle, J., Lundeberg, J.: Spatially resolved transcriptomes—next generation tools for tissue exploration. *BioEssays* **42**(10), 1900221 (2020)



- [3] Williams, C.G., Lee, H.J., Asatsuma, T., Vento-Tormo, R., Haque, A.: An introduction to spatial transcriptomics for biomedical research. *Genome Medicine* **14**(1), 1–18 (2022)
- [4] Lein, E., Borm, L.E., Linnarsson, S.: The promise of spatial transcriptomics for neuroscience in the era of molecular cell typing. *Science* **358**(6359), 64–69 (2017)
- [5] Ståhl, P.L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J.F., Magnusson, J., Giacomello, S., Asp, M., Westholm, J.O., Huss, M., *et al.*: Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**(6294), 78–82 (2016)
- [6] Yoosuf, N., Navarro, J.F., Salmén, F., Ståhl, P.L., Daub, C.O.: Identification and transfer of spatial transcriptomics signatures for cancer diagnosis. *Breast Cancer Research* **22**, 1–10 (2020)
- [7] Chen, W.-T., Lu, A., Craessaerts, K., Pavie, B., Frigerio, C.S., Corthout, N., Qian, X., Laláková, J., Kühnemund, M., Voytyuk, I., *et al.*: Spatial transcriptomics and in situ sequencing to study alzheimer’s disease. *Cell* **182**(4), 976–991 (2020)
- [8] Longo, S.K., Guo, M.G., Ji, A.L., Khavari, P.A.: Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics* **22**(10), 627–644 (2021)
- [9] Zeng, Z., Li, Y., Li, Y., Luo, Y.: Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome biology* **23**(1), 1–23 (2022)
- [10] Zhao, E., Stone, M.R., Ren, X., Guenthoer, J., Smythe, K.S., Pulliam, T., Williams, S.R., Uyttingco, C.R., Taylor, S.E., Nghiem, P., *et al.*: Spatial transcriptomics at subspot resolution with bayesspace. *Nature biotechnology* **39**(11), 1375–1384 (2021)
- [11] Bao, F., Deng, Y., Wan, S., Shen, S.Q., Wang, B., Dai, Q., Altschuler, S.J., Wu, L.F.: Integrative spatial analysis of cell morphologies and transcriptional states with muse. *Nature biotechnology* **40**(8), 1200–1209 (2022)
- [12] Pham, D., Tan, X., Xu, J., Grice, L.F., Lam, P.Y., Raghubar, A., Vukovic, J., Ruitenber, M.J., Nguyen, Q.: stlearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *BioRxiv*, 2020–05 (2020)
- [13] Hu, J., Li, X., Coleman, K., Schroeder, A., Ma, N., Irwin, D.J., Lee, E.B., Shinohara, R.T., Li, M.: Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature methods* **18**(11), 1342–1351 (2021)



- [14] Dong, K., Zhang, S.: Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature communications* **13**(1), 1739 (2022)
- [15] Long, Y., Ang, K.S., Li, M., Chong, K.L.K., Sethi, R., Zhong, C., Xu, H., Ong, Z., Sachaphibulkij, K., Chen, A., *et al.*: Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nature Communications* **14**(1), 1155 (2023)
- [16] Eldar, A., Elowitz, M.B.: Functional roles for noise in genetic circuits. *Nature* **467**(7312), 167–173 (2010)
- [17] Kharchenko, P.V., Silberstein, L., Scadden, D.T.: Bayesian approach to single-cell differential expression analysis. *Nature methods* **11**(7), 740–742 (2014)
- [18] Buettner, F., Natarajan, K.N., Casale, F.P., Proserpio, V., Scialdone, A., Theis, F.J., Teichmann, S.A., Marioni, J.C., Stegle, O.: Computational analysis of cell-to-cell heterogeneity in single-cell rna-sequencing data reveals hidden subpopulations of cells. *Nature biotechnology* **33**(2), 155–160 (2015)
- [19] Dai, H., Li, H., Tian, T., Huang, X., Wang, L., Zhu, J., Song, L.: Adversarial attack on graph structured data. In: *International Conference on Machine Learning*, pp. 1115–1124 (2018). PMLR
- [20] Zügner, D., Akbarnejad, A., Günnemann, S.: Adversarial attacks on neural networks for graph data. In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2847–2856 (2018)
- [21] Maynard, K.R., Collado-Torres, L., Weber, L.M., Uytingco, C., Barry, B.K., Williams, S.R., Catallini, J.L., Tran, M.N., Besich, Z., Tippianni, M., *et al.*: Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nature neuroscience* **24**(3), 425–436 (2021)
- [22] Wu, S.Z., Al-Eryani, G., Roden, D.L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J.R., Bartonicek, N., *et al.*: A single-cell and spatially resolved atlas of human breast cancers. *Nature genetics* **53**(9), 1334–1347 (2021)
- [23] Lun, A.T., McCarthy, D.J., Marioni, J.C.: A step-by-step workflow for low-level analysis of single-cell rna-seq data with bioconductor. *F1000Research* **5** (2016)
- [24] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255 (2009). Ieee
- [25] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE Conference on*

Computer Vision and Pattern Recognition, pp. 2818–2826 (2016)

- [26] Pramerdorfer, C., Kampel, M.: Facial expression recognition using convolutional neural networks: state of the art. arXiv preprint arXiv:1612.02903 (2016)
- [27] Voulodimos, A., Doulamis, N., Doulamis, A., Protopapadakis, E., et al.: Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience* **2018** (2018)
- [28] Irmisch, A., Bonilla, X., Chevrier, S., Lehmann, K.-V., Singer, F., Toussaint, N.C., Esposito, C., Mena, J., Milani, E.S., Casanova, R., *et al.*: The tumor profiler study: integrated, multi-omic, functional tumor profiling for clinical decision support. *Cancer Cell* **39**(3), 288–293 (2021)
- [29] Badia-i-Mompel, P., Santiago, J.V., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C.H., Flores, R.O.R., Saez-Rodriguez, J.: decoupler: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances* (2022) <https://doi.org/10.1093/bioadv/vbac016>
- [30] McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426 (2018)
- [31] Manetti, M.: Molecular morphology and function of stromal cells. MDPI (2021)
- [32] Van Der Meer, W., Van Gelder, W., Keijzer, R., Willems, H.: The divergent morphological classification of variant lymphocytes in blood smears. *Journal of clinical pathology* **60**(7), 838–839 (2007)
- [33] Ghanem, G., Fabrice, J.: Tyrosinase related protein 1 (tyrp1/gp75) in human cutaneous melanoma. *Molecular oncology* **5**(2), 150–155 (2011)
- [34] Qiu, C., Li, P., Bi, J., Wu, Q., Lu, L., Qian, G., Jia, R., Jia, R.: Differential expression of tyrp1 in adult human retinal pigment epithelium and uveal melanoma cells. *Oncology Letters* **11**(4), 2379–2383 (2016)
- [35] Journe, F., Boufker, H.I., Van Kempen, L., Galibert, M.-D., Wiedig, M., Salès, F., Theunis, A., Nonclercq, D., Frau, A., Laurent, G., *et al.*: Tyrp1 mRNA expression in melanoma metastases correlates with clinical outcome. *British journal of cancer* **105**(11), 1726–1732 (2011)
- [36] Jayachandran, A., Lo, P.-H., Chueh, A.C., Prithviraj, P., Molania, R., Davalos-Salas, M., Anaka, M., Walkiewicz, M., Cebon, J., Behren, A.: Transketolase-like 1 ectopic expression is associated with dna hypomethylation and induces the warburg effect in melanoma cells. *BMC cancer* **16**, 1–15 (2016)
- [37] Inamdar, G.S., Madhunapantula, S.V., Robertson, G.P.: Targeting the mapk

- pathway in melanoma: why some approaches succeed and other fail. *Biochemical pharmacology* **80**(5), 624–637 (2010)
- [38] Amaral, T., Sinnberg, T., Meier, F., Krepler, C., Levesque, M., Niessner, H., Garbe, C.: The mitogen-activated protein kinase pathway in melanoma part i-activation and primary resistance mechanisms to braf inhibition. *European journal of cancer* **73**, 85–92 (2017)
- [39] Shain, A.H., Yeh, I., Kovalyshyn, I., Sriharan, A., Talevich, E., Gagnon, A., Dummer, R., North, J., Pincus, L., Ruben, B., *et al.*: The genetic evolution of melanoma from precursor lesions. *New England Journal of Medicine* **373**(20), 1926–1936 (2015)
- [40] Fan, Z., Chen, R., Chen, X.: Spatialdb: a database for spatially resolved transcriptomes. *Nucleic acids research* **48**(D1), 233–237 (2020)
- [41] Xu, Z., Wang, W., Yang, T., Chen, J., Huang, Y., Gould, J., Du, W., Yang, F., Li, L., Lai, T., *et al.*: Stomicsdb: a database of spatial transcriptomic data. *bioRxiv*, 2022–03 (2022)
- [42] Zheng, Y., Chen, Y., Ding, X., Wong, K.H., Cheung, E.: Aquila: a spatial omics database and analysis platform. *Nucleic Acids Research* **51**(D1), 827–834 (2023)
- [43] Wolf, F.A., Angerer, P., Theis, F.J.: Scanpy: large-scale single-cell gene expression data analysis. *Genome biology* **19**, 1–5 (2018)
- [44] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D.: Backpropagation applied to handwritten zip code recognition. *Neural computation* **1**(4), 541–551 (1989)
- [45] Sohn, K.: Improved deep metric learning with multi-class n-pair loss objective. *Advances in neural information processing systems* **29** (2016)
- [46] Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014)
- [47] Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M.J., Blüthgen, N., Saez-Rodriguez, J.: Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature communications* **9**(1), 20 (2018)