Cell composition inference and identification of layer-specific spatial transcriptional profiles with POLARIS

Jiawen Chen[1], Tianyou Luo[1], Minzhi Jiang[2], Jiandong Liu[3], Gaorav P Gupta[4,5,*], Yun Li[1,6,7,*,#]

[1] Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2] Department of Applied Physical Sciences, University of North Carolina at Chapel Hill, North Carolina, USA

[3] Department of Pathology and Laboratory Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[4] Department of Radiation Oncology, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[5] Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[6] Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[7] Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

* These authors contributed equally.

[#] Corresponding author: Yun Li, Department of Genetics, 120 Mason Farm Road, Campus Box 7264, University North Carolina, Chapel Hill, NC 27599, USA. E-mail: yunli@med.unc.edu

Key words: Spatial transcriptomics, deconvolution, differentially expressed genes, single cell, deep learning

Abstract

Spatial transcriptomics (ST) technology, providing spatially resolved transcriptional profiles, facilitates advanced understanding of key biological processes related to health and disease. Sequencing-based ST technologies provide whole-transcriptome profiles, but are limited by the non-single cell level resolution. Lack of knowledge in the number of cells or cell type composition at each spot can lead to invalid downstream analysis, which is a critical issue recognized in ST data analysis. Methods developed, however, tend to under-utilize histological images, which conceptually provide important and complementary information including anatomical structure and distribution of cells. To fill in the gaps, we present POLARIS, a versatile ST analysis method that can perform cell type deconvolution, identify anatomical or

functional layer-wise differentially expressed (LDE) genes and enable cell composition inference from histology images. Applied to four tissues, POLARIS demonstrates high deconvolution accuracy, accurately predicts cell composition solely from images, and identifies LDE genes that are biologically relevant and meaningful.

Introduction

Molecular analysis of messenger RNA patterns in histological tissue sections is a key component of biomedical research and diagnostics. The development of novel spatial transcriptomic (ST) technologies has advanced dramatically over the last few years. There are two main categories of ST technologies: imaging-based or sequencing-based. Technologies based on imaging directly image individual RNA molecules within single cells [1, 2]. Sequencing-based techniques first label spatial spots on histological tissue sections with unique barcodes to indicate their two-dimensional spatial positions, and utilize RNA-sequencing to provide gene expression quantifications for each spot along with the spatial coordinates [3, 4]. Commonly used methods include MERFISH [1], seqFISH+ [5] in the former category and 10X Genomics' Visium platform [3] in the latter category. More information can be found in recent review papers [6-8]. Some of the sequencing-based techniques (exemplary platforms include Spatial Transcriptomics and Visium) also provide a co-registered hematoxylin and eosin (H&E) stained histology image for the analyzed sample. Empowered by these technologies, we can obtain gene expression profiling with retained spatial information and histological images, which enable researchers and clinicians to gain a new level of insight into complex tissue samples.

In parallel to these technological developments, computational methods to analyze spatial data derived from tissue samples have substantially advanced. For instance, focusing on histology images, multiple machine learning and deep learning methods have been developed to maximally extract information from these images [9-12]. In the presence of pathological annotations, histology images can be used for various purposes including cell segmentation [13, 14], tissue type registration [15], mutation rate inference [16, 17], and gene expression prediction [12]. Most of these tasks, however, require pathologists to fully annotate each cell in the histological image, entailing substantial manual time and human resources. Such pathologist annotation is currently unavailable for the vast majority of publicly available ST data, thus making using traditional cell detection methods to perform ST deconvolution inaccessible. In the field of ST, histology imaging has primarily been used to predict gene expression and

perform tissue registration, where the image data is usually subject to a pre-trained model to extract image features [9, 10, 12]. Several popular pre-trained models, such as convolutional neural networks, stacked sparse autoencoders, and masked autoencoders (MAE), have been employed as a first step to reduce image dimensions and demonstrate advantages in many applications [9, 10, 12, 13, 15]. However, cell composition inference hasn't benefited from these models yet. In recent literature, histology images have been utilized to improve deconvolution accuracy [18, 19] but methods that can predict cell composition *solely* from histology images are currently unavailable.

Besides the histology image, ST data allow for the extraction and revelation of tissue structure through coordinated gene expression. Researchers have developed methods such as SPARK [20] and SpatialDE [21] for identifying genes whose expression varies within a tissue slice, known as spatially differentially expressed (SDE) genes. Gene expression changes spatially across spots within a tissue slice, often reflecting some underlying structured heterogeneity such as anatomical layers, clusters of similar spots, and/or spatial domains. Such structured heterogeneity motivates the development of ST clustering methods including BayesSpace to identify layers/clusters within each ST slice [11, 22, 23]. As aforementioned, the identified layers often correspond to different functions or morphological changes in the tissue [22, 24, 25]. The across-spot variation in expression can be largely attributed to three factors: variation in cell number, variation in cell composition, and true spatially driven variation in gene expression profile (**Fig. 1a**).

When the expression variation is truly driven by difference in spatial coordinates (in contrast to difference in cell number or composition), we can consider having sub-cell types located in different spatial regions (**Fig. 1a**). However, when SDE genes are detected by the aforementioned methods, the identified spatial difference is a result of the interplay of all three factors, and it is difficult to distinguish genes that are truly spatially differentially expressed from those that merely *appear* so due to differential cell number or composition across spatial spots. We would be able to differentiate among the driving factors if we had single-cell resolution data with the entire transcriptome or at least a large number of genes measured. However, in practice, we normally do not have this luxury: we have either data from imaging-based technologies that are single-cell resolution but measure only a small number of genes, or data from sequencing-based methods that provide transcriptome-wide measurement, but are limited in resolution. Sequencing-based ST methods have spatial spots of 2-100 μm in diameter,

implying that each spot can easily contain tens of cells of different cell types. Lacking ideal (i.e., single-cell resolution with many genes measured) data motivates the development of computational methods to infer variation in gene expression profile across layers, while simultaneously estimating and adjusting for the estimated cell number and composition.

As a matter of fact, lack of knowledge in the number of cells at each spot nor the cell type composition itself has been recognized as a critical issue in ST data analysis, because failure to adjust for this accurately can lead to invalid downstream analyses. In order to address this problem, a number of ST deconvolution methods have been developed [18, 26-30]. However, most ST deconvolution methods assume that the gene expression profile for the same cell type is invariant across the entire tissue sample, which is a strong assumption whose violation will result in inaccurate cell composition inference. Methods such as DestVI that assume a continuous or smoothly changing gene expression profile across the tissue, however, have exhibited inconsistent performance across tissue types [6, 31]. Therefore, how to model layer-specific gene expression variation and utilize histological images to infer cell composition is a problem that remains unsolved.

Here we present POLARIS, Probabilistic-based cell cOmposition inference with LAyer infoRmatIon Strategy, to perform cell type deconvolution and infer layer-wise differentially expressed (LDE) genes (**Fig. 1b**). POLARIS integrates single-cell RNA-seq reference and ST data with annotated layer information. By examining histology images and the coordinated expression profile, one can reasonably infer layers or sub-regions that correspond to different biological functions (e.g. cancer vs non-cancer regions in a tumor biopsy, different layers in a brain cortical sample, ventricle and atrium areas in heart). By explicitly allowing and modeling layer-specific gene expression patterns, POLARIS is not only capable of identifying cell type composition with high accuracy, but also could identify LDE genes while simultaneously correcting for differential cell composition. An additional key characteristic of POLARIS is its flexibility to optionally leverage histology images. To our knowledge, POLARIS is the first ST deconvolution method that can predict cell composition purely from a histological image. This functionality also empowers POLARIS to infer super-resolution cell composition based on images of areas without gene expression measurements (i.e., areas in between spots), as well as to predict cellular composition based purely on an original H&E stained image. The performance of POLARIS was evaluated on data from multiple tissues including the mouse cortex, developing human heart, and HER2+ breast cancer samples. POLARIS robustly

demonstrates high deconvolution accuracy across tissues compared to other state-of-the-art deconvolution methods, accurately predicts cell composition solely from images, and identifies LDE genes that are biologically relevant and meaningful. Our results showcase the advantages of POLARIS in the following three aspects: deconvolution accuracy, LDE gene identification and prediction with image.

Results

POLARIS method overview

POLARIS is a probabilistic-based inference method that assumes that gene expression counts in both scRNA-seq reference data and ST data follow a negative binomial distribution. As a first step, POLARIS maximizes likelihood to infer cell-type-specific gene expression profiles from scRNA-seq reference (**Fig. 1b**). The gene expression profile of each spot in ST data can then be viewed as a weighted sum of the negative binomial distribution derived from the scRNA-seq reference, where the weights are based on spot-level cell composition. As opposed to assuming that cell-type-specific gene expression profiles are invariant throughout a whole tissue slice, POLARIS assumes that only spots in similar biological or anatomical layers share the same gene expression profiles by introducing a layer-specific location parameter. Explicitly modeling layers is a unique feature of our POLARIS method. POLARIS accepts any user-specified layer annotations, e.g., derived manually (from pathologist annotation) or computationally (based on either morphological features or gene expression, e.g., using BayesSpace [22]). Note that the layer-specific parameters cannot be inferred from single-cell reference because there's no layer information by the nature of data generation. Using ST data with layer annotations, POLARIS enables layer-specific inference. By introducing the layer-specific shift parameters (**Online Methods**), we can obtain an updated location parameter for each layer in the ST data, allowing cell-type-specific gene expression profiles to vary across layers. By multiplying the updated location parameter with the cell composition parameter as well as the parameter to account for technical/batch effects, we simultaneously model the impact of cell composition and spatial location (as reflected by layers) on cell-type-specific gene expression, while controlling for potential batch effects. Parameters can be estimated using maximum a posteriori estimation (MAP) (**Online Methods**). So far, we have focused on inference with gene expression data only. When a co-registered histology image is available, POLARIS first employs MAE [32] to extract features from the image tile of each spot and the image tile of its neighborhood. These

two extracted features are then combined and used as inputs to build POLARIS's image network (**Fig. 1b**). The output of POLARIS's image network is cell composition for any input image (which can be from a completely independent histological image). The output of POLARIS includes inferred spot-level cell type composition and layer-specific gene expression profiles, as well as a trained POLARIS image network. The layer-specific gene expression profiles enable identification of LDE genes, and the pre-trained POLARIS image network allows resolution enhancement and cell composition inference from a new histology image.

POLARIS attains high deconvolution accuracy

The deconvolution accuracy of POLARIS was assessed both through simulation and in single-cell resolution ST data. Specifically, we simulated cells with gene expression counts from cell-type and layer-specific negative binomial distributions and randomly selected cells to create spot-level gene expression. For single-cell resolution real ST datasets, we clumped cells into spots according to their coordinates to mimic low-resolution spot-level ST data. We used data where we have cell type labels for the single cells such that we have the true cell type mixture in each clumped pseudo-spot. We quantified the performance using root mean square error (RMSE) where a smaller RMSE corresponds to better performance. We compared POLARIS with five state-of-art methods: CARD [26], DestVI [29], RCTD [28], stereoscope [27] and SPOTlight [30]. These methods were selected according to their specific methodological features and/or their high performance in previous benchmarking studies (**Online Methods**)[6, 31].

We began with a simulated scenario where all spots and layers share a similar composition of cells, but with layers differing in terms of their gene expression profiles. Under this scenario, gene expression variations are solely the result of variations in gene expression profiles across layers. Specifically, we first simulated a dataset with two "biological" layers, with cells from six cell types and expression values for 100 genes generated. We first simulated the layer of each cell and then the gene expression values for the cell were drawn from negative binomial distributions according to its layer and cell type. We then constructed pseudo spots by randomly selecting 10-16 cells from each layer. Specifically, we generated 200 spots with 50 spots in layer 1 and 150 spots in layer 2 (**Online Methods**). Under this simulation framework, genes could be classified into three categories: up-regulated in layer1 (e.g., *Gene36*), up-regulated in layer2 (e.g., *Gene100*) and no significant difference between layers (e.g., *Gene95*) (**Fig. 2a-b,**

Extended Data Fig. 1). The largest log2 fold change is seen in gene 36 and gene 100, and most genes display similar or non-differential expression patterns across layers. Applied to the simulated data, POLARIS outperforms all other methods as manifested by its lowest RMSE (**Fig. 2c**).

We further assessed POLARIS's deconvolution performance in a real single-cell resolution ST data from the mouse VISp region, a well-structured region in mouse cortex that has been extensively studied [2, 33]. Anatomical structure, major cell types, and layer-specific gene markers provide information about the layered and segmented structure of the mouse VISp (**Fig. 2d-f**) [33, 34]. We used single-cell resolution ST data from the STARmap platform [2], which consists of 1,020 genes measured in 973 cells. We divided the cells into 356 pseudo spots each of 400x400 square pixels (**Fig. 2d**). In order to perform deconvolution, we utilized the internal reference (that is, the STARmap single cell data itself as the reference). In this way, any systematic differences between the reference and the target ST data are eliminated as potential factors that may impair performance. This internal reference evaluation provides a baseline (or upper bound) for measuring the performance of deconvolution methods [6]. Since layer annotation is required when using POLARIS, we employed BayesSpace [22] to cluster the constructed pseudo spots, resulting in five distinct clusters , reflecting the expected layer structure of mouse VISp (**Fig. 2e**). In this mouse VISp dataset, POLARIS still achieves amongst the best performance in terms of RMSE (**Fig. 2g**). Moreover, POLARIS, based on its inferred cellular composition, successfully recovers the layer structure of mouse VISp (from top to bottom: Smc, eL2, eL3, eL4, eL5, eL6-1, eL6-2, Oligo, HPC, **Fig. 2h**).

Additionally, we tested POLARIS on the developing human heart tissue generated from the in situ sequencing (ISS) platform (**Fig. 2i**) [35]. The heart ISS data is also a single-cell resolution ST data, consisting of 24,371 cells and with only 65 genes measured in each cell. We gridded the cells into pseudo spots each of dimension 454 x 424 square pixels (**Fig. 2j**). The main purpose of this assessment is to evaluate POLARIS's performance with a limited number of genes. Instead of using the internal reference (ISS data itself), we used a scRNA-seq reference obtained from a similar biological sample [36]. Consequently, we can also evaluate the deconvolution performance when the reference and ST data are not perfectly matched. The heart ISS data provides us with a DAPI stained histology image, allowing us to measure the performance of POLARIS by including the histology image as an additional input (**Fig. 2i, Methods**). We clustered the spots into four layers using BayesSpace (**Fig. 2k**). The

BayesSpace inferred layers correspond reasonably well to the anatomy of the heart (red: epicardium, green: ventricles, light blue: atria, dark blue: outflow tract). POLARIS has maintained its best performer position. Specifically, POLARIS achieves the lowest/best mean of MSE (**Fig.2l**). In spite of the fact that the DAPI staining only contains one color channel, POLARIS with image input is able to effectively infer the type of cell, achieving accuracy close to the best performers.

Polaris identifies layer-specific gene expression pattern

A major feature of POLARIS is its ability to model layer-specific parameters. The layer/structure of a tissue can be reflected in multiple dimensions, such as morphology, gene expression, and other omics levels. POLARIS focuses on leveraging the rich gene expression information provided by ST data. As detailed above, cell density, cellular composition, and the "real" spatially differentially expressed genes can all contribute to the observed gene expression variation. By incorporating layer-specific parameters into the cell type deconvolution process, POLARIS is able to identify such LDE genes while taking into account differential cell composition. POLARIS quantifies statistical significance for LDE genes using permutation tests, and magnitude of effect using log2 fold change in mean gene expression, based on the inferred layer-specific mean parameters (**Methods**). Through the elimination of potential confounding effects of cell composition, POLARIS ensures that the LDE genes identified are differentially expressed genes truly due to spatial factors.

As a starting point for assessing POLARIS' ability to infer LDE genes, we performed simulations where we know the truth. Following the same simulation framework used above to evaluate deconvolution efficiency, we evaluated the layer-specific location parameters. Again, since cellular composition is simulated from the same distribution across spots regardless of layer status, observed gene expression variation can only be attributed to truly differential expression patterns across layers (**Fig. 2a-b**). Consequently, genes could be classified into three categories: layer1-enriching genes, layer2-enriching genes, and genes with similar expression levels across layers (**Fig. 2b, 3a**). We applied POLARIS to perform deconvolution and simultaneously perform the permutation test and calculate the log2 fold change in mean expression, layer1 over layer2. POLARIS successfully identified genes that have different expression profiles across layers (**Fig. 3b**). The predicted log2 fold change well captures the true log2 fold change (**Fig. 3c**) when all the genes have layer-specific gene expression profiles.

In this particular simulation, we generated the genes such that all of them have layer-specific expression profiles, although the expression difference between layers of some genes could be small (detailed in **Online Methods**). To investigate the impact of the proportion of LDE genes on POLARIS performance, we further conducted simulations using the same setting but varying the proportion of genes with layer-specific expression profiles (**Extended Data Fig. 2**). As the proportion of LDE genes increases from 0 to 1, UMAP representations of cells become more separated by layers. It is difficult, however, to identify the sub-types of cells based on the UMAP representation, even with prior knowledge of the layer information when the proportion of LDE genes is < 0.5, thus failing to capture heterogeneity across layers (**Extended Data Fig. 2a**). POLARIS well controlled the type-I error and accurately estimated the true log2 fold change regardless of the proportion of LDE genes (**Extended Data Fig. 2b**). POLARIS also shows satisfactory performance on a sparse simulation setting (**Extended Data Fig. 3, Online Methods**). Note that layer-specific parameters are inferred only from the ST data because we do not have layer information from the scRNA-seq reference.

Encouraged by POLARIS's performance in simulated data, we proceeded to further test POLARIS's capability to detect LDE genes using real datasets. We began with the developing human heart single-cell resolution ISS ST data. We created pseudo-spots from the dataset and applied POLARIS to the pseudo-spots, feeding as input spot-level gene expression information along with layer information inferred from BayesSpace. LDE genes detected by POLARIS indeed have differential patterns of gene expression in different layers (**Fig. 3d-e)**. Although cell type affects gene expression, layer status also plays a significant role in shaping the spot-level expression profile. Furthermore, such layer-specific changes are shared by many cell types. For example, the expression of *EBF2* in most cell types is highest in layer 3 compared to the other three layers (**Fig. 3e**). POLARIS, by making accurate inference of gene expression profiles in different cell types, has successfully captured the LDE gene and has recovered the layer-specific variation present in most cell types. We further explore the links between LDE genes and biological functions of each layer. For example, among the LDE genes, *TCF21* has the lowest mean gene expression in layer 3 (**Fig. 3d-e**) where fibroblast-like cells are enriched (**Fig. 2j-k**). The transcription factor 21 (TCF21) encoded by *TCF21* plays a crucial role in regulating cell differentiation and cell fate determination through epithelial-mesenchymal transformations during cardiac development. More specifically, TCF21 has been reported to be capable of promoting the development of cardiac fibroblasts and inhibiting differentiation of epicardial cells into vascular smooth muscle cells [37], consistent with our observed down-regulation in layer 3.

We further applied POLARIS on a more complex tissue: breast cancer samples. There are several subtypes of breast cancer, among which the HER2-positive subtype is characterized by the increased expression of *ERBB2* (AKA *HER2*, human epidermal growth factor receptor 2) in tumor cells [38]. We obtained ST data of HER2-positive tumors from eight individuals (patient A-H) generated through the Spatial Transcriptomics platform [3, 38]. Each of the eight patients provided multiple slides, but only one slide from each patient was pathologist annotated. Annotations mark areas with one of the following five labels: in situ cancer (noninvasive ductal carcinoma in situ, DCIS), invasive breast cancer (IBC), adipose tissue, immune infiltrate, or connective tissue [38]. Here we highlight the results of two slides with both IBC and DCIS regions (results from the A1 slide in **Fig. 3f-k, Extended Data Fig. 5**, results from the G2 slide in **Extended Data Fig. 4,5**). We applied POLARIS on the pathologist annotated ST data using an external scRNA-seq reference [39]. POLARIS made reasonable inference regarding cell composition on slide A1 (**Fig. 3f**). For example, cancer epithelial cells are inferred to be enriched in the DCIS and IBC areas. POLARIS also identified several LDE genes in the DCIS area (**Fig. 3i-j**). These LDE genes, including *S100A14*, *MUC1*, *PITX1*, and *ERBB2* are mainly expressed in cancer epithelial cells (**Extended Data Fig. 6**). In general, genes that are primarily expressed in cancer epithelial cells are enriched either in the DCIS or the IBC region (**Fig. 3h-k**), which is expected since these two regions have similarly high proportions of cancer epithelial cells. The LDE genes identified by POLARIS successfully capture cancer epithelial cell specific genes, but differentially expressed in the two regions (**Fig. 3h-k**). For example, in the DCIS region, all POLARIS-identified LDE genes except *ERBB2* have a positive log2 fold change estimate (**Fig. 3j**), consistent with expression patterns shown in **Fig. 3k**. The observed down-regulation of *ERBB2* in the DCIS area most likely reflects an up-regulation of *ERBB2* elsewhere. As shown in **Fig. 3f**, the majority of slide A1 is the invasive cancer area. Cancer epithelial cells in the invasive cancer area presumably invade other areas resulting in an increased *ERBB2* expression in all other pathologist-identified areas except the DCIS and these invaded areas also exhibit an increase in the proportion of cancer epithelial cells (**Fig. 3h-i, Extended Data Fig. 4f**). ERBB2, the protein encoded by *ERBB2*, plays an important role in breast cancer. The over-expression of *ERBB2* disrupts normal cell-control mechanisms and gives rise of aggressive tumor cells and leads to increased breast cancer metastasis [40-45]. Interestingly, when applied to slide G2, *ERBB2* shows to be up-regulated in the DCIS area (**Extended Data Fig. 4e,g**). Together, these observations reflect the heterogeneity of the samples and are consistent with the literature that *ERBB2* is over-expressed in 30-35% of DCIS, while *ERBB2* is

only expressed in 15-25% of IBC [46-49]. POLARIS reveals such heterogeneity and complexity by showing differential gene expression profiles between DCIS and IBC regions on the same slide, and by revealing differential patterns across slides from the same patient as well as across patients. In addition to *ERBB2*, other LDE genes identified by POLARIS and the proteins encoded by those genes also play important roles in breast cancer. For example, copy number amplification of *S100A14*, significantly correlated with the increased *S100A14* mRNA expression, is present in 5.4%-20.7% of primary breast cancer patients and in approximately 26.1% of metastatic breast cancer patients [50]. For another example, *CXXC5* over-expression has been observed to be associated with a poor prognosis for estrogen receptor positive (ER+) breast cancer [51].

Polaris enables prediction purely from histology image

After demonstrating that even a single-color histology image is able to generate high-accuracy deconvolution inference in the developing human heart tissue, we continued to apply POLARIS to other spot-level ST data with H&E staining images. We utilized the mouse primary somatosensory cortex area (SSp) data generated from the 10x Visium platform [52]. Similar to the mouse VISp region, mouse SSp is also an area in the mouse cortex with well-defined anatomical and functional structure. Specifically, the glutamatergic neuron types exhibit clear layered patterns [6, 33, 53]. Using an independent scRNA-seq from similar SSp regions [33] as the external reference, POLARIS trained an image network using four SSp ST slides. Each slide was clustered into six groups using BayesSpace (**Extended Data Fig. 7**). POLARIS successfully captures expected patterns of glutamatergic composition and reveals layer structure consistent with data from the Allen Brain Atlas [34].

Spot-level ST technologies cannot measure every part of a tissue slide. As shown in **Fig. 4a**, gene expression levels are not available for any region outside the measured spots (**Fig. 4a**). POLARIS, through its trained image network, can determine cell type composition using the image of unmeasured areas. Specifically, POLARIS-trained image network can be applied to cropped images of the same size as the original grid in training. Sliding across the entire histological image of the SSp slide and applying the POLARIS-trained image network to the image of each sliding window, one can obtain super-resolution inference, encompassing areas not initially covered by spatial spots. Such super-resolution inference empowers us to gain finer details of the layered structure (**Fig. 4b**). Following the training of POLARIS's image network,

we applied it to a new SSp slide to test whether the trained image network is able to make reasonable inference on images from independent slides (**Fig. 4c**). We compared cell compositions inferred using the image network trained by POLARIS from other slides, with those inferred using the new slide's own gene expression and histology image on measured spots (**Fig. 4d-e, Extended Data Fig. 8a-b**). Although the two sets of inference show differences, results obtained solely from POLARIS's pre-trained image network still show strong correlation with the inference results using its own image and gene expression, especially in glutamatergic neurons (For example, the Pearson's correlation for L6b CTX is 0.8). POLARIS-trained image network successfully recovers the layered structure in most cell types, suggesting that the approach could be applied to new histology slides as long as we have a network pre-trained by POLARIS using ST data from similar regions.

POLARIS-trained image network was further examined using breast cancer data. We again used pathologist annotated ST data and the external scRNA-seq from Wu et al. as reference to perform deconvolution on slide H1 (**Fig. 4f**) [38, 39]. Afterwards, we employed the image network trained by POLARIS from slide H1 to enhance resolution and obtain super-resolution cell composition inference on the H1 slide itself (**Fig. 4g**). The inferred cell compositions are consistent with those in other slides where cancer epithelial cells are enriched in the DCIS and IBC regions. It appears that cancer epithelial cell proportion is able to accurately capture the cancerous areas. For example, spots with >0.4 proportion of cancer epithelial cells are enriched within regions labeled as DCIS and IBC by pathologists (**Extended Data Fig. 8e**). We then further applied the POLARIS image network trained on H1 to two other slides of the same patients (H2 and H3), which have no pathologist's annotation (**Fig. 4h-i, Extended Data Fig. 8c-g**). By closely examining the histology images, it is evident that regions characterized by high proportions of cancer epithelial cells are indeed primarily cancerous areas. Our results therefore suggest that POLARIS enables a new method of registering histology images to different anatomical/functional regions, for example cancerous areas in this analysis. By examining cell composition in each spot, we are able to group the spots into layers. In summary, with a trained POLARIS image network, we could obtain super-resolution cell composition inference which reveals finer layer structure of a tissue.

POLARIS allows cell composition inference on new histology images without gene expression and consequently, is able to identify anatomical and functional regions. Compared to existing supervised classification methods for registering histology image tiles to different regions or

layers which require annotated layer information, POLARIS is unsupervised in terms of histology information and does not require pathologist annotation. Instead, POLARIS relies on pre-knowledge about the relationship between the targeted layer and the cell composition.

Discussion

ST technologies are rapidly evolving. In the near future, we expect to be able to measure gene expression levels at single cell resolution and of all genes in the transcriptome. Currently, spot-level resolution ST technologies such as the Visium and Spatial Transcriptomes still have their advantages in terms of throughput (both in terms of number of genes measured and number of spatial spots examined) and the ability to obtain high-resolution H&E staining images. Because of their advantages, researchers are generating a deluge of these data. It is, however, imperative to perform cell type deconvolution at each spot in order to mitigate or eliminate potential confounding caused by differential cell composition across spots. Despite numerous methods developed for ST deconvolution, two pieces of information have been under-utilized. First, often there is a layer structure, or at least areas reflecting different anatomical or functional regions in an ST slide. Second, histological images, carrying information complementary to spot-level gene expression profiles, have not been fully explored in their value for cell type decomposition. In this work, we present POLARIS, a unified framework that leverages layer structure information and/or histological images, for cell type deconvolution both at spots with expression measurement and in regions with only image information, as well as for the revelation of LDE genes.

We demonstrate the performance of POLARIS on simulation and real datasets including developing human heart, mouse cortex VISp and SSp region, and human HER2+ breast cancer samples. POLARIS robustly achieves best or close to best deconvolution performance compared to other state-of-the-art methods. POLARIS's inference on spot-level ST data reveals layered structures that are consistent with gene expression profiles, histological images, and known/established anatomical/functional layers/regions in the corresponding tissue samples.

Equally if not more importantly, POLARIS accurately infers layer-specific expression profiles across different cell types which leads to the identification of LDE genes. We demonstrate POLARIS's power to identify LDE genes using simulation data as well as the single-cell resolution ST data from the developing human heart, where we have knowledge regarding the

true LDE genes. The LDE genes identified by POLARIS in developing human heart data indeed exhibit different gene expression profiles across layers, beyond what can be attributed to differential cell type compositions. Finally, applying POLARIS to ST data from breast cancer patients, we found that POLARIS identified LDE genes reveal complex heterogeneous across-layer/region differential expression across samples and/or patients. The detected LDE genes are consistent with established knowledge regarding breast cancer pathology, including metastasis and prognosis, but offer more granular sample-level and patient-level information that can potentially empower personalized diagnosis and treatment. POLARIS assumes that the layer-specific shift parameter is shared across cell types. This assumption is based on observations in single-cell resolution ST data (e.g. heart ISS data, **Fig. 3e**). In addition, due to the model identification problem, we chose to go with a parsimonious model and therefore did not implement a layer shift specific to each cell type. Considering that we are inferring both the cell type proportion and the layer-specific shift at the same time, modeling a layer shift specific to each cell type will very likely cause the model to confuse the proportion and shift, which will result in poor and unstable estimation of both parameters. We consider this as a limitation of our model which will likely to be resolved in future efforts.

Another key feature of POLARIS is its ability to leverage image data. In the ST deconvolution field, gene expression itself has proved its ability to infer cell composition. Imaging information, however, has been under-utilized. Recent work [18, 19] showed the potential of histology images accompanying ST data. We believe that histology images can be further leveraged for ST inference. For example, histology images alone are widely used to segment cells with deep learning models [13, 14]. POLARIS can take an accompanying image as input to train an image network, and employ a pre-trained image network on a completely new image. Our pre-trained POLARIS image network offers a novel method for tissue registration, which extracts and reveals tissue anatomical or functional structures either from the histological image alone or jointly with gene expression. The major barriers that prevent the full potential of integrating histological images with ST data include quality of the co-registered image and, most importantly, the absence of pathologist annotations. In order to accomplish the task with the currently available data, POLARIS incorporates training of the image network into the inference of cell composition. Instead of training a model with inferred cell composition as the goal and using MSE as the loss function, our intention is to use histology images to help with the estimation of cell composition, under the rationale that spots with similar histological images and similar neighborhoods tend to share similar cell composition. Nevertheless, we fail to

demonstrate that the image improves deconvolution performance due to the limitation of the current data: single-cell level resolution ST data only provides DAPI stained images which only comprise one color panel, while spot-level ST data has no gold standard truth. Despite these limitations to quantify the performance, POLARIS with image input still achieved high accuracy among the state-of-art methods in single-cell resolution ST data, and in spot-level ST data, cell type composition inferred by POLARIS agreed with single-cell level data and the expected biological layers (e.g. glutamatergic neurons in the mouse cortex and cancer epithelial cells in the breast cancer slides). POLARIS introduces a novel approach for inferring cell composition purely from histological images that has not previously been explored by ST deconvolution. We believe that the versatile ability of POLARIS to incorporate histological images to elucidate layer-specific gene expression patterns will empower novel discoveries in spatial biology.

Material and methods

POLARIS inference algorithm

We use *g* for gene index, *c* for cell index, *z* for cell type index, and *L* for layer index. Z_c is the cell type of cell *c*, L_c is the layer of cell *c* and L_s is the layer of spot *s*. We assume that in scRNA-seq reference dataset, X_{cg}, expression count of gene *g* in cell *c* follows the following negative binomial distribution:

$$X_{cg} \sim NB\left(S_c\, Softplus\left(\theta_{gz_c} + T_{L_c g}\right), P_g\right)$$

$Softplus(x) = \log(1 + \exp(x)))$ is used to guarantee that $R_{gz_c} > 0$. $P_g = Sigmoid\left(o_g\right)$, where $Sigmoid(x) = \frac{1}{1+\exp(-x)}$ is the sigmoid function. S_c is the library size of cell *c* and Z_c is the cell type of cell *c*. NB is the negative binomial distribution. θ_{gz_c} is the mean location parameter shared across layers. To capture the gene expression variation between layers, we introduce a new parameter T_{Lg}, a mean shift parameter for layer *L* and gene *g*. We assume that for a gene, all cell types share the same shift parameter across layers. We further assume that $T_{Lg} \sim N(0,1)$. However, in real world, we do not have the layer information in the scRNA-seq reference data. We make the following assumption

$$X_{cg} \sim NB\left(S_c\, Softplus\left(\theta_{gz_c}\right), P_g\right)$$

when inferring the parameters in the reference data. Estimates for the parameters are then obtained by finding the MLE (maximum likelihood estimates), given the provided scRNA-seq reference data via the gradient-based optimization using the PyTorch library in Python.

In the ST data, we make similar distributional assumption. Specifically, we assume that X_{scg}, expression count of gene *g* in cell *c* in spot *s* follows the following negative binomial distribution:

$$X_{scg} \sim NB\left(\beta_g Softplus\left(\theta_{gz_c} + T_{L_c g}\right), P_g\right)$$

β_g is the parameter measuring technical bias between ST and scRNA-seq reference for gene *g*. Note that the bias parameter is gene-specific. Considering the additive property of negative binomial distribution and summing across all single cells within spot *s*, the resulting distribution of gene *g* in spot *t* also follows a negative binomial distribution:

$$X_{sg} \sim NB\left(\sum_{z=1}^{Z} \beta_g Softplus\left(\theta_{gz} + T_{L_s g}\right) n_{sz}, P_g\right)$$

Z is total number of cell types and n_{sz} is the number of cells of type *z* in spot *s*. All the parameter are inferred using maximum a posteriori estimation (MAP) with $\widehat{\theta_g}$, $\widehat{P_g}$ obtained from scRNA-seq reference. When image data is utilized in the model, we incorporate a POLARIS image network to infer the cell type composition. POLARIS image network takes MAE extracted features as input and the output is the cell composition. The image network is trained using MAP which means that the loss function is the negative posterior likelihood. At last, cell type proportion can be calculated as $v_{sz} = \frac{n_{sz}}{\sum n_{sz}}$.

Image feature extraction using masked autoencoder (MAE)

The MAE codes are obtained from https://github.com/facebookresearch/mae . We only use the pretrained encoder part to extract the image features [32]. We used mae_visualize_vit_large_ganloss.pth as the pretrained model. We used MAE to extract image features for the spot image and spot neighborhood image. The spot image is a $r * r$ square covering the spot. r is defined specific to the dataset used (details about the used *r* could be found in the source code). The neighborhood image is defined as a $3r * 3r$ square sharing the same center point as the spot image. The two MAE-extracted 1024-length vectors are combined as the image input of the spot. The POLARIS image network combines two levels of fully connected layers (2048 to 512, 512 to the number of cell types). The θ_{gz} and $T_{L_s g}$ are fixed during the training of image network.

Simulation

We begin by generating a single-cell reference. A total of 100 genes are simulated in six cell types. The number of cells in each cell type is simulated from $N(500, 100^2)$ and rounded to the

nearest integer. For each cell, we simulate the layer from $Binomial(0.3)$ where we treat a simulated value of 0 as layer 1 and value of 1 as layer 2. We then simulate the gene expression of gene *g* in cell *c* from $NB(Softplus(\theta_{gz_c} + T_{L_c g}), P_g)$, where NB denotes the negative binomial distribution. Z_c represents the cell type of cell *c* and L_c represents the layer of cell *c*. In the first simulation, both θ_{gz_c} and T_{Lg} are simulated from $N(0,1)$ and P_g is simulated from $Uniform(0.2,0.8)$. In the sparse simulation, θ_{gz_c} are simulated from $N(-2,1)$. Note T_{Lg} represents the layer-specific gene expression profile, thus allowing expression profiles to vary across layers, even for the same gene *g* and cell type. We then construct pseudo spots by randomly selecting cells in layer 1 and layer 2 with replacement from single cells simulated above. In each spot, the number of cells is determined by sampling a random number from Uniform(10,16). We simulated 50 spots in layer1 and 150 spots in layer2. For simulation with different proportion of genes with layer-specific expression profiles, we define the proportion as m. Then the T_{Lg} for $100 * (1 - m)$ % genes are set as 0 while the T_{Lg} for the rest $100 * m$ % genes are simulated from $N(0,1)$.

Calculate between-layer fold change

To identify LDE genes, we compare the gene expression profile across layers by comparing layer-specific location parameters:

$$Softplus(\theta_{gz} + T_{Lg})$$

Specifically, the fold change of gene *g* in layer L compare to other layers in cell type *z* are calculated as

$$F_{Lg} = (\theta_{gz} + T_{Lg})/(\sum_{L' \neq L} (\theta_{gz} + T_{L'g})/(S - 1))$$

Where *S* is the total number of layers. Then we take the maximum fold change across cell types to quantify the across-layer fold change of gene *g*.

LDE gene identification

We perform permutation test to identify LDE genes. With a given layer annotation $Layer^{obs}$, we perform cell type deconvolution and infer T_{Lg}^{obs} and F_{Lg}^{obs}. Then we randomize the layer annotation for N times and generate pseudo layer annotation $Layer^{sim1}, \dots, Layer^{simN}$. Consequently, we are able to infer $T_{Lg}^{sim1}, \dots, T_{Lg}^{simN}$. Then the p value of gene *g* in layer L is defined as

$$p\ value = \frac{\#\ |T_{Lg}^{sim}| > |T_{Lg}^{obs}|}{N}$$

Genes with $p\ value < \frac{0.05}{Total\ number\ of\ genes}$ (Bonferroni correction) and $\left|\log_2 F_{Lg}^{obs}\right| > 1$ are considered as the LDE genes in layer L (marked as pink in the figures). We used $N = 10000$ (permutation times) in all the analysis. For the LDE analysis in the manuscript, we only used gene expression (i.e., histological images not used) in the deconvolution process in the permutation test.

Data preprocessing

For the STARmap data, we only keep genes that are present in at least 2% of spots.

In the breast cancer analysis, we only keep HER2+ subtype ST data for deconvolution evaluation. Similarly, we keep only the HER2+ patients' scRNA-seq data as reference. For scRNA-seq reference, genes expressed in at least 3 cells and cells expressing at least 200 genes are kept. Similarly for ST data, genes expressed in at least 3 spots and spots expressing at least 100 genes are kept. Only genes that exist in both scRNA-seq and ST data are employed in further analysis. Top 2000 highly variable genes (HVGs) are used. The gene subsettings are accomplished using the R package Seurat [54]. HVGs are selected using feature variance calculated by the FindVariableFeatures function with default settings. To reduce the effect of excessive zeros, we recommend removing genes with an expression level of less than 2%.

For the heart ISS DAPI-stained image, we reverse the color and enhance the contrast using ImageEnhance function in the Pillow package [55].

Comparing to other state-of-art methods

We compared the performance of POLARIS with several state-of-the-art deconvolution methods developed for ST data. We included RCTD and stereoscopes because of their top performance in the probabilistic-based model. Our inclusion of DestVI was based on the fact that it claimed to identify a continuous state that goes beyond discrete cell types, which we consider to be similar to our objective. CARD was included because it is capable of inferring the cell type composition in areas that are directly measured. Because SPOTlight is the top performer among the NMF+NNLS techniques, we also included it in our analysis.

We followed the instructions of each method on their corresponding website. Among all the methods, only RCTD has a built-in gene filtering method, where only genes with normalized

gene expression >= 0.0002 are included, and it selects cell type marker genes based on a log-fold-change threshold of 0.75 [28]. We used the default parameters of RCTD and ran RCTD in full mode. Only selected cell type marker genes were fed into RCTD. For all other methods, we used all genes without any further filtering from the preprocessed data described in the previous section.

Tissue detection

To use POLARIS-trained image network, we need to automatically detect the tissue section from a histological image. We utilized the filter_entropy function in https://github.com/CODAIT/deep-histopath and followed the instructions in https://developer.ibm.com/articles/an-automatic-method-to-identify-tissues-from-big-whole-slide-images-pt3/ . We used entropy, which measures tissue complexity, to detect the percentage of tissue in each spot. Areas such as the slide background are less complex than the tissue area. We used the default threshold of 5 in the analysis. Pixels with an entropy value greater than 5 are counted as tissue regions. Spots with greater than 5% tissue area is kept in the super-resolution composition inference.

Acknowledgements

Funding: The research is supported by National Institutes of Health (NIH) grants (R01HL163972 and U01HG011720). GPG is supported by grants from the NIH/NCI (R37CA227837) and the V Foundation (T2019-010).

Data and code availability: The processed developing human heart scRNA-seq, ISS ST data, mouse cortex SSp ST data, SSp scRNA-seq and VISp scRNA-seq reference could be obtained from the review paper of Chen et al. [6]. The STARmap data are obtained from https://www.starmapresources.org/data and the breast cancer data are obtained from https://doi.org/10.5281/zenodo.4751624 (ST) and https://singlecell.broadinstitute.org/single_cell/study/SCP1039 (scRNA-seq reference). The processed STARmap data and breast cancer data and the source code of POLARIS could be found at https://doi.org/10.5281/zenodo.7302022 . Figure 1 is created via bioRender.

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

Author contributions: J.C., Y.L. and M. J developed the method. J.C and T.L performed the computational experiment. J.L advised the heart study. Y.L. and G.G. both supervised the study. J.C. and Y.L. wrote the manuscript, and all authors commented and approved the final version. We thank Li lab members for providing advice on data preprocessing, method selection and feedback on the manuscript.

Reference

1. Xia, C., et al., *Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression.* Proceedings of the National Academy of Sciences, 2019. **116**(39): p. 19490.
2. Wang, X., et al., *Three-dimensional intact-tissue sequencing of single-cell transcriptional states.* Science, 2018. **361**(6400): p. eaat5691.
3. Ståhl Patrik, L., et al., *Visualization and analysis of gene expression in tissue sections by spatial transcriptomics.* Science, 2016. **353**(6294): p. 78-82.
4. Stickels, R.R., et al., *Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2.* Nature Biotechnology, 2021. **39**(3): p. 313-319.
5. Eng, C.-H.L., et al., *Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+.* Nature, 2019. **568**(7751): p. 235-239.
6. Chen, J., et al., *A comprehensive comparison on cell-type composition inference for spatial transcriptomics data.* Briefings in Bioinformatics, 2022: p. bbac245.
7. Moses, L. and L. Pachter, *Museum of spatial transcriptomics.* Nature Methods, 2022. **19**(5): p. 534-546.
8. Rao, A., et al., *Exploring tissue architecture using spatial transcriptomics.* Nature, 2021. **596**(7871): p. 211-220.
9. Bae, S., H. Choi, and D.S. Lee, *Discovery of molecular features underlying the morphological landscape by integrating spatial transcriptomic data with deep features of tissue images.* Nucleic Acids Research, 2021. **49**(10): p. e55-e55.
10. Zong, Y., et al., *conST: an interpretable multi-modal contrastive learning framework for spatial transcriptomics.* bioRxiv, 2022: p. 2022.01.14.476408.
11. Hu, J., et al., *SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network.* Nature Methods, 2021. **18**(11): p. 1342-1351.
12. He, B., et al., *Integrating spatial gene expression and breast tumour morphology via deep learning.* Nature Biomedical Engineering, 2020. **4**(8): p. 827-834.
13. Xu, J., et al., *Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images.* IEEE Transactions on Medical Imaging, 2016. **35**(1): p. 119-130.
14. Tofighi, M., et al. *Deep Networks with Shape Priors for Nucleus Detection*. in *2018 25th IEEE International Conference on Image Processing (ICIP)*. 2018.
15. Daly, A.C., K.J. Geras, and R. Bonneau, *A convolutional neural network for common coordinate registration of high-resolution histology images.* Bioinformatics, 2021. **37**(22): p. 4216-4226.
16. Coudray, N., et al., *Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning.* Nature Medicine, 2018. **24**(10): p. 1559-1567.
17. Liao, H., et al., *Deep learning-based classification and mutation prediction from histopathological images of hepatocellular carcinoma.* Clinical and Translational Medicine, 2020. **10**(2): p. e102.
18. Biancalani, T., et al., *Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram.* Nature Methods, 2021. **18**(11): p. 1352-1362.
19. Zubair, A., et al., *Cell type identification in spatial transcriptomics data can be improved by leveraging cell-type-informative paired tissue images using a Bayesian probabilistic model.* Nucleic Acids Research, 2022: p. gkac320.

20. Zhu, J., S. Sun, and X. Zhou, *SPARK-X: non-parametric modeling enables scalable and robust detection of spatial expression patterns for large spatial transcriptomic studies.* Genome Biology, 2021. **22**(1): p. 184.
21. Svensson, V., S.A. Teichmann, and O. Stegle, *SpatialDE: identification of spatially variable genes.* Nature Methods, 2018. **15**(5): p. 343-346.
22. Zhao, E., et al., *Spatial transcriptomics at subspot resolution with BayesSpace.* Nature Biotechnology, 2021. **39**(11): p. 1375-1384.
23. Zhu, Q., et al., *Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data.* Nature Biotechnology, 2018. **36**(12): p. 1183-1190.
24. Miller, B.F., et al., *Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data.* Nature Communications, 2022. **13**(1): p. 2339.
25. Bernstein, M.N., et al., *SpatialCorr: Identifying Gene Sets with Spatially Varying Correlation Structure.* bioRxiv, 2022: p. 2022.02.04.479191.
26. Ma, Y. and X. Zhou, *Spatially informed cell-type deconvolution for spatial transcriptomics.* Nature Biotechnology, 2022.
27. Andersson, A., et al., *Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography.* Communications Biology, 2020. **3**(1): p. 565.
28. Cable, D.M., et al., *Robust decomposition of cell type mixtures in spatial transcriptomics.* Nature Biotechnology, 2021.
29. Lopez, R., et al., *DestVI identifies continuums of cell types in spatial transcriptomics data.* Nature Biotechnology, 2022.
30. Elosua-Bayes, M., et al., *SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes.* Nucleic Acids Research, 2021. **49**(9): p. e50-e50.
31. Li, B., et al., *Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution.* Nature Methods, 2022. **19**(6): p. 662-670.
32. He, K., et al., *Masked Autoencoders Are Scalable Vision Learners*. 2021, arXiv.
33. Yao, Z., et al., *A taxonomy of transcriptomic cell types across the isocortex and hippocampal formation.* Cell, 2021. **184**(12): p. 3222-3241.e26.
34. *Allen Reference Atlas – Mouse Brain [brain atlas]. Available from atlas.brain-map.org.*
35. Asp, M., et al., *A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart.* Cell, 2019. **179**(7): p. 1647-1660.e19.
36. Eralp, I., et al., *Epicardium-derived cells are important for correct development of the Purkinje fibers in the avian heart.* The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology, 2006. **288A**(12): p. 1272-1280.
37. Hu, H., et al., *The Role of Transcription Factor 21 in Epicardial Cell Differentiation and the Development of Coronary Heart Disease.* Frontiers in Cell and Developmental Biology, 2020. **8**.
38. Andersson, A., et al., *Spatial deconvolution of HER2-positive breast cancer delineates tumor-associated cell type interactions.* Nature Communications, 2021. **12**(1): p. 6012.
39. Wu, S.Z., et al., *A single-cell and spatially resolved atlas of human breast cancers.* Nature Genetics, 2021. **53**(9): p. 1334-1347.
40. Hung, M.C., et al., *Molecular cloning of the neu gene: absence of gross structural alteration in oncogenic alleles.* Proceedings of the National Academy of Sciences, 1986. **83**(2): p. 261-264.
41. Tan, M., J. Yao, and D. Yu, *Overexpression of the c-erbB-2 gene enhanced intrinsic metastasis potential in human breast cancer cells without increasing their transformation abilities.* Cancer Res, 1997. **57**(6): p. 1199-205.

42. Moody, S.E., et al., *Conditional activation of Neu in the mammary epithelium of transgenic mice results in reversible pulmonary metastasis.* Cancer Cell, 2002. **2**(6): p. 451-61.
43. Holbro, T., G. Civenni, and N.E. Hynes, *The ErbB receptors and their role in cancer progression.* Exp Cell Res, 2003. **284**(1): p. 99-110.
44. Reese, D.M. and D.J. Slamon, *HER-2/neu signal transduction in human breast and ovarian cancer.* Stem Cells, 1997. **15**(1): p. 1-8.
45. Tan M, Y.D., *Molecular Mechanisms of ErbB2-Mediated Breast Cancer Chemoresistance. In: Madame Curie Bioscience Database [Internet]. Austin (TX): Landes Bioscience; 2000-2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK6194/*.
46. Curigliano, G., et al., *Risk of subsequent in situ and invasive breast cancer in human epidermal growth factor receptor 2-positive ductal carcinoma in situ.* Ann Oncol, 2015. **26**(4): p. 682-687.
47. Thorat, M.A., et al., *Prognostic and Predictive Value of HER2 Expression in Ductal Carcinoma In Situ: Results from the UK/ANZ DCIS Randomized Trial.* Clin Cancer Res, 2021. **27**(19): p. 5317-5324.
48. Cooke, T., et al., *HER2 as a prognostic and predictive marker for breast cancer.* Ann Oncol, 2001. **12 Suppl 1**: p. S23-8.
49. Gajria, D. and S. Chandarlapaty, *HER2-amplified breast cancer: mechanisms of trastuzumab resistance and novel targeted therapies.* Expert Rev Anticancer Ther, 2011. **11**(2): p. 263-75.
50. Li, X., et al., *A S100A14-CCL2/CXCL5 signaling axis drives breast cancer metastasis.* Theranostics, 2020. **10**(13): p. 5687-5703.
51. Fang, L., et al., *Overexpression of CXXC5 is a strong poor prognostic factor in ER+ breast cancer.* Oncol Lett, 2018. **16**(1): p. 395-401.
52. Kleshchevnikov, V., et al., *Cell2location maps fine-grained cell types in spatial transcriptomics.* Nature Biotechnology, 2022.
53. Codeluppi, S., et al., *Spatial organization of the somatosensory cortex revealed by osmFISH.* Nature Methods, 2018. **15**(11): p. 932-935.
54. Hao, Y., et al., *Integrated analysis of multimodal single-cell data.* Cell, 2021. **184**(13): p. 3573-3587.e29.
55. Clark, A., *Pillow (PIL Fork) Documentation. readthedocs. Retrieved from* https://pillow.readthedocs.io/en/stable/. 2022.
56. Zhao, P., et al., *Modeling zero inflation is not necessary for spatial transcriptomics.* Genome Biology, 2022. **23**(1): p. 118.

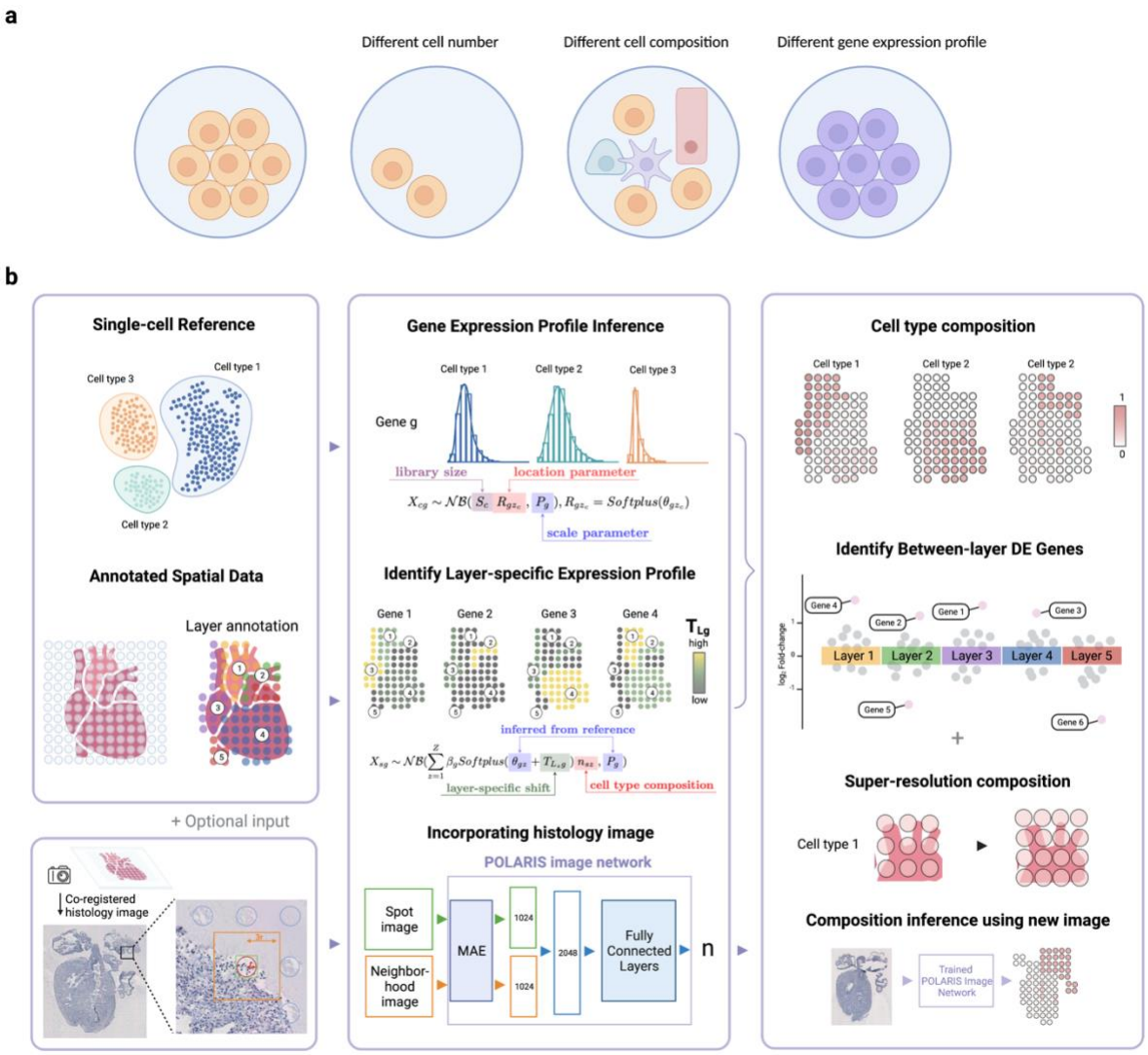

Fig. 1 | POLARIS overview. (a) Three reasons that can explain gene expression variation across spots. Each circle represents a spatial spot. Compared to the leftmost spot, the three spots on the right differ primarily in cell number, cell composition and gene expression profile. (b) POLARIS workflow. POLARIS takes as input single-cell reference and annotated spatial data, infers cell-type-specific gene expression profiles and identifies layer-specific expression profiles. When a co-registered histology image (as an optional input) is provided, POLARIS will additionally train an image network. The output of POLARIS includes inferred spot-level cell type composition, identified layer-wise differentially expressed (DE) genes and a pre-trained POLARIS image network that can be applied to independent images.

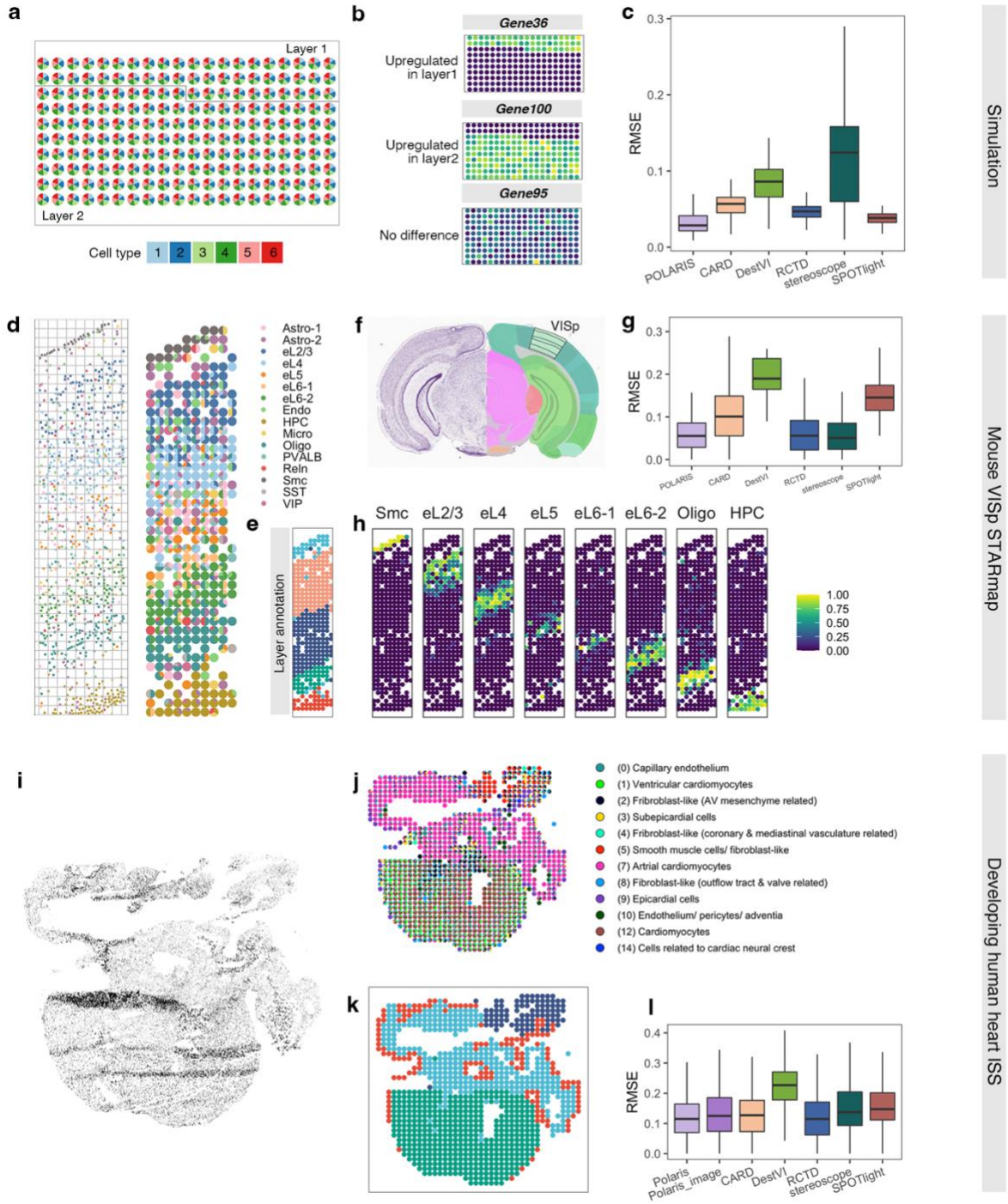

Fig. 2 | Deconvolution accuracy of POLARIS. Simulation: (a) cell type composition of the simulated ST data. (b) Three categories of gene expression pattern: up-regulated in layer1, up-regulated in layer2, no significant difference across layers. (c) RMSE of POLARIS along with other state-of-the-art methods on simulated data. **On mouse VISp STARmap data**: (d) (left)

cell map and (right) clumped pseudo spots along with their cell compositions visualized by pie charts (e) BayesSpace identified clusters, consistent with layer structure of mouse VISp (f) Nissl staining (left) and anatomical annotations (right) from the Allen Mouse Brain Atlas and Allen Reference Atlas - Mouse Brain. The black lined area indicates the layer structure of the VISp region. (g) RMSE of POLARIS along with other state-of-the-art methods on mouse VISp STARmap data. (h) POLARIS inferred spot-level composition of the eight major cell types, along cortex depth. **On developing human heart ISS data**: (i) Processed DAPI-stained histology image of the developing human heart ISS data. (j) clumped pseudo spots in the ISS data along with their cell compositions visualized by pie charts. (k) BayesSpace identified clusters in the ISS data. (l) RMSE of POLARIS along with RMSE of other state-of-the-art methods on the heart ISS data.

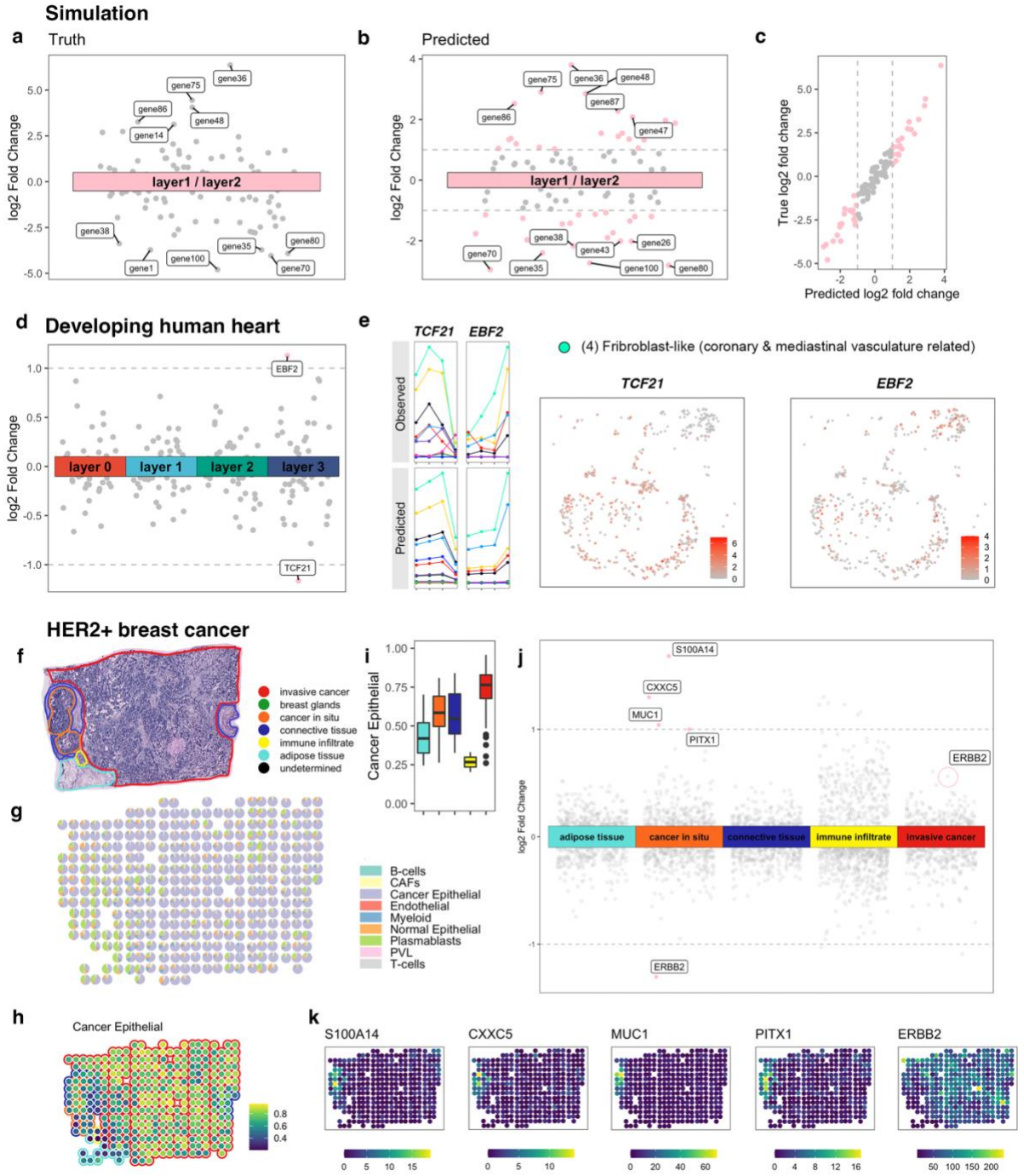

Fig. 3 | LDE genes identified by POLARIS on three datasets. On simulated data: (a) true log2 fold change of mean gene expression, layer1 over layer2. (b) POLARIS inferred log2 fold change of mean gene expression, layer1 over layer2. POLARIS identified LDE genes are colored as pink, otherwise, gray. (c) POLARIS inferred log2 fold change of gene expression across layers achieves high correlation (Pearson's correlation = 0.974) with the true log2 fold change. **On developing human heart data**: (d) POLARIS inferred log2 fold change of gene

expression across layers. POLARIS identified LDE genes are colored as pink, otherwise, gray. (e) (left) Observed mean gene expression in scRNA-seq data (top) and POLARIS inferred gene expression location parameter (bottom) of each cell type across layers. Lines are colored by cell types. X axis indicates layer status: from left to right is layer 0,1,2,3. (right) Gene expression of TCF21 and EBF2 in the Fibroblast-like (coronary & mediastinal vasculature related) cells. **On HER2+ breast cancer data**: (f) Pathologist annotation on slide A1. (g) POLARIS inferred cell composition (h) POLARIS inferred cancer epithelial cell proportion (i) Distribution of POLARIS inferred cancer epithelial cell proportions in each layer (color scheme is the same as in Fig. 3e) (j) POLARIS inferred log2 fold change of gene expression across layers. Points with absolute value greater than 1 are colored as pink, otherwise, gray. (k) Gene expression profiles of POLARIS identified LDE genes in the cancer in situ layer.

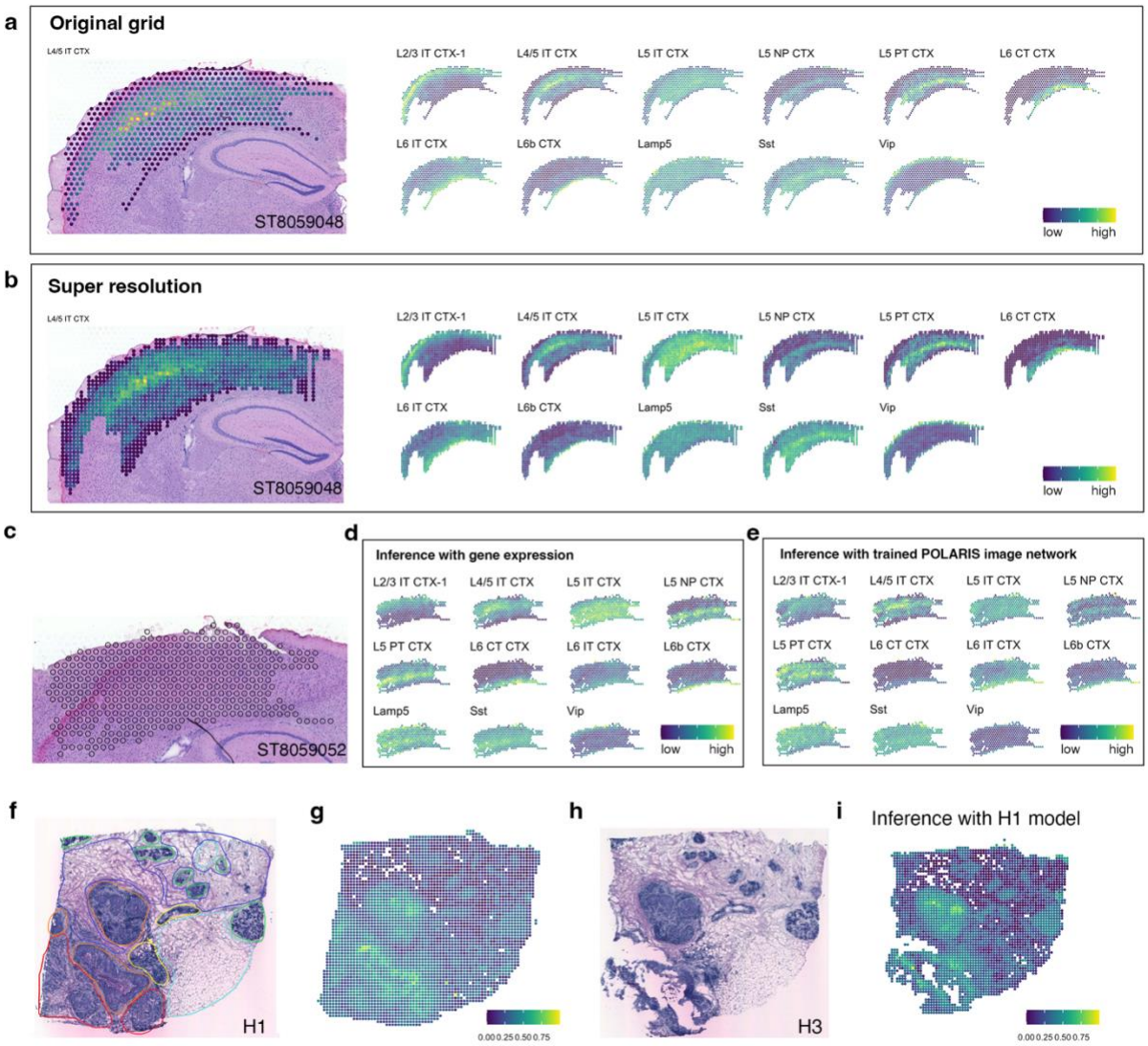

Fig. 4 | POLARIS achieves super-resolution cell composition inference when using histology image as input. On mouse SSp 10x Visium slide ST8059048: (a) The original grid and POLARIS inferred cell composition. (b) The super-resolution cell composition inferred using a POLARIS-trained image network. Points in (a) and (b) are colored by the corresponding cell proportion (from blue to yellow corresponds to low to high). **On slide ST8059052**: (c) The original grid of the slide (d) POLARIS inferred cell composition inference using spot-level gene expression and histology image. (e) Inferred cell composition using a POLARIS-trained image network (trained on the other 4 slides), based on histology image only. Similarly, points in (d) and (e) are colored by the corresponding cell proportion (from blue to yellow corresponds to low to high). **On Her2+ breast cancer**: (f) pathologist's annotation of slide H1. (g) Super-resolution

inference of H1 using POLARIS image network trained on H1. (h) Histology image of slide H3. (i) Super-resolution inference of H3 using POLARIS image network, again trained on H1. The points in (g) and (i) are colored by the cancer epithelial cell proportion. CTX: isocortex, IT: intratelencephalic, PT: pyramidal tract, NP: near-projecting.