Deep Learning in Spatial Transcriptomics: Learning From the Next Next-Generation Sequencing

A. Ali Heydari[1, 2] and Suzanne S. Sindi[1, 2, 3]
[1)] *Department of Applied Mathematics, University of California, Merced*
[2)] *Health Sciences Research Institute, University of California, Merced*
[3)] *Corresponding author. Email: ssindi@ucmerced.edu*

(Dated: 28 February 2022)

Spatial transcriptomics (ST) technologies are rapidly becoming the extension of single-cell RNA sequencing (scRNAseq), holding the potential of profiling gene expression at a single-cell resolution while maintaining cellular compositions within a tissue. Having both expression profiles and tissue organization enables researchers to better understand cellular interactions and heterogeneity, providing insight into complex biological processes that would not be possible with traditional sequencing technologies. The data generated by ST technologies are inherently noisy, high-dimensional, sparse, and multi-modal (including histological images, count matrices, etc.), thus requiring specialized computational tools for accurate and robust analysis. However, many ST studies currently utilize traditional scRNAseq tools, which are inadequate for analyzing complex ST datasets. On the other hand, many of the existing ST-specific methods are built upon traditional statistical or machine learning frameworks, which have shown to be sub-optimal in many applications due to the scale, multi-modality, and limitations of spatially-resolved data (such as spatial resolution, sensitivity and gene coverage). Given these intricacies, researchers have developed deep learning (DL)-based models to alleviate ST-specific challenges. These methods include new state-of-the-art models in alignment, spatial reconstruction, and spatial clustering among others. However, deep-learning models for ST analysis are nascent and remain largely underexplored. In this review, we provide an overview of existing state-of-the-art tools for analyzing spatially-resolved transcriptomics, while delving deeper into the DL-based approaches. We discuss the new frontiers and the open questions in this field and highlight the domains in which we anticipate transformational DL applications.

I. INTRODUCTION

Although multicellular organisms contain a common genome within their cells, the morphology and gene expression patterns of cells are largely distinct and dynamic. These differences arise from internal gene regulatory systems and external environmental signals. Cells proliferate, differentiate and function in tissues while sending and receiving signals from their surroundings. These environmental factors cause cell fate to be highly dependent on the environment in which it exists. Therefore, monitoring a cell's behavior in the residing tissue is crucial to understanding cell function, as well as its past and future fate[1].

Advancements in single-cell sequencing have transformed the genomics and bioinformatics fields. The advent of single-cell RNA sequencing (scRNAseq) has enabled researchers to profile gene expression levels of various tissues and organs, allowing them to create comprehensive atlases in different species[2–6]. Moreover, scRNAseq enables the detection of distinct subpopulations present within a tissue; which has been paramount in discovering new biological processes, the inner workings of diseases, and effectiveness of treatments[7–14]. However, high-throughput sequencing of solid tissues requires tissue dissociation, resulting in the loss of spatial information[15,16]. To fully understand cellular interactions, data on tissue morphology and spatial information is needed, which scRNAseq alone can not provide. The placement of cells within a tissue are crucial from the developmental stages (e.g. asymmetric cell fate of mother and daughter cells[17]) and beyond cell differentiation (such as cellular functions, response to stimuli and tissue homeostasis[18]). These limitations would be alleviated by technologies that could preserve spatial information while measuring gene expression at the single-cell level.

Spatial Transcriptomics (ST) provide an *unbiased* view of tissue organization crucial in understanding cell fate, delineating heterogeneity, and other applications[19]. However, many current ST technologies suffer from lower sensitivities as compared to scRNAseq, while lacking the single-cell resolution that scRNAseq provides[20]. Targeted *in situ* technologies have tried to solve the issue of resolution and sensitivity, but are limited in gene throughput and often require *a priori* knowledge of target genes[20]. More specifically, *in situ* technologies (such as *in situ* sequencing[21], single-molecule fluorescence *in situ* hybridization (smFISH)[22–24], targeted expansion sequencing[25], cyclic-ouroboros smFISH (osmFISH)[26], multiplexed error-robust fluorescence *in situ* hybridization (MERFISH)[27], sequential FISH (seqFISH+)[28], and spatially resolved transcript amplicon readout mapping (STARmap)[29]), are typically limited to pre-selected genes that are on the order of hundreds, with the accuracy potentially dropping as more probes are added[29]. We will refer to these methods as *image-based* techniques.

On the other hand, Next Generation Sequencing (NGS)-based technologies (such as 10x Genomics' Visium and its predecessor[30,31], Slide-Seq[32], HDST[33]) barcode entire transcriptomes but have limited capture rates, and resolutions that are larger than a single cell[34] (50 μm - 100 μm for Visium and 10 μm for Slide-Seq). Moreover, unlike image-based technologies, NGS-based methods allow for unbiased profiling of large tissue sections without necessitating a set of target genes[35,36]. However, NGS-based technologies do not have single-cell resolution, requiring cellular features to be inferred or related to the histological scale using computational ap-

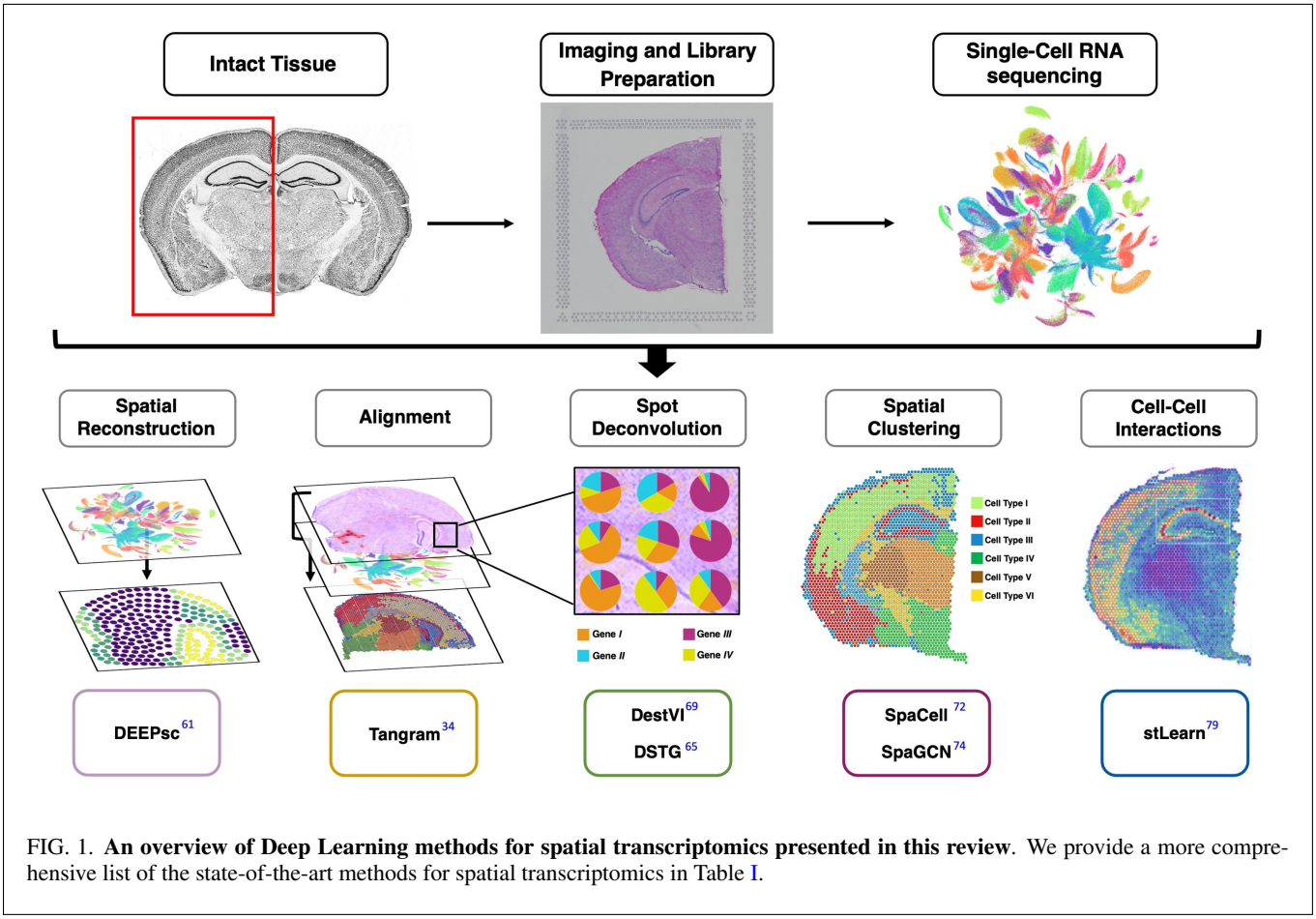

FIG. 1. **An overview of Deep Learning methods for spatial transcriptomics presented in this review**. We provide a more comprehensive list of the state-of-the-art methods for spatial transcriptomics in Table I.

proaches. Many current algorithms use traditional statistical or medical image processing frameworks that require human supervision[34,37,38], which is not ideal for large-scale analyses. Additionally, many algorithms are not generalizable across different sequencing platforms, which limit their utility and restrict multiomics integration efforts.

Deep Learning (DL) methods can use raw data to extract useful representations (or information) needed for performing a task, such as classification or detection[39]. This quality makes this class of Machine Learning (ML) algorithms ideal for applications where the available data is large, higher-dimensional, and noisy, such as single-cell omics. DL models have been extensively used in scRNAseq studies (e.g. preprocessing[40,41], clustering[42,43], cell-type identification[44,45] and data augmentation[46,47]), and have shown to significantly improve upon traditional methods[10], suggesting the potential of such methods in ST analysis. Moreover, DL models can leverage multiple data sources, such as images and text data, to learn a set of tasks[48]. Given that spatially-resolved transcriptomics are inherently multimodal (i.e. they consist of images and gene expression count data) and that downstream analysis consist of multiple tasks (e.g. clustering and cell-type detection), researchers have sought to develop ST-specific DL algorithms.

Spatially-resolved transcriptomics have been utilized to unravel complex biological processes in many diseases (e.g. COVID-19[49,50], arthritis[51,52] ,cancer[31,33,53–55], Alzheimer's[56], diabetes[57,58], etc.). Continuous improvements and commercialization of ST technologies (such as 10x's Visium) are resulting in wider use across individual labs. Therefore, scalable and platform-agnostic computational approaches are needed for accurate and robust analysis of ST data. So far, DL methods have shown promising results in handling the scale and multi-modality of spatially-resolved transcriptomics; however, DL-based models in this space remain nascent. Similar to scRNAseq analysis, we anticipate a suite of DL models to be developed in the near future to address many of the pressing challenges in spatial omics field. This review aims to provide an overview of the current state-of-the-art (SOTA) DL models developed for ST analysis. Due to the potentials and accessibility of NGS-based ST technologies, we primarily focus on methods and techniques developed for these technologies.

The remainder of this manuscript is organized as follows: We provide an overview of common scRNAseq and ST technologies in Section II, followed by a general description of common DL architectures used for ST analysis in Section III. Section IV is dedicated to the current DL methods developed for analyzing spatially-resolved transcriptomics. We conclude, in Section V, by discussing our outlook on the cur-

rent challenges and future research directions in ST domain. Table I provides the reader with a list of current SOTA methods for ST analysis. Given the pace of advancements in this field, the authors have compiled an online list of current DL methods for ST analysis on a dedicated repository(https://github.com/SindiLab/Deep-Learning-in-Spatial-Transcriptomics-Analysis), which will be maintained and continuously updated.

II. BIOLOGICAL BACKGROUND

A. Single-Cell RNA Sequencing (scRNAseq)

RNA sequencing (RNA-seq) provides comprehensive insights on cellular processes (such as identifying genes that are upregulated or downregulated, etc.). However, traditional bulk RNA-seq is limited to revealing the average expression from a collection of cells, and not disambiguation single-cell behavior. Thus, it is difficult to delineate cellular heterogeneity with traditional RNA-seq, which is a disadvantage since cellular heterogeneity has been shown to play a crucial role in understanding many diseases[82]. Therefore, researchers have turned to single-cell RNA-seq (scRNAseq) in order to identify cellular heterogeneity within tissues. ScRNAseq technologies have been instrumental in the study of key biological processes in many diseases, such as cancer[83], Alzheimer's[84], cardiovascular diseases[85], etcetera (see[82] for more details). RNA sequencing of cells at a single-cell resolution, scRNAseq, generally consists of four stages:

(i) **Isolation of Single-Cells and Lysing**: Cells are selected through laser microdirection, fluorescence-activated cell sorting (FACS), microfluidic/microplate Technology (MT) or a combination of these methods[86], with MT being highly complementary to NGS-based technologies[87]. MT encapsulates each single-cell into an independent microdroplet containing unique molecular identifiers (UMI), lysis buffer for cell lysis (to increase the capturing of as many RNA molecules as possible), oligonucleotide primers, and seoxynucleotide triphosphates (dNTPs) in addition to the cells themselves. Due to MT's higher isolation capacity, thousands of cells can be simultaneously tagged and analyzed, which is beneficial for large-scale scRNAseq studies.

(ii) **Reverse Transcription**: One challenge in RNA sequencing is that RNA can not be directly sequenced from cells, and thus RNA must first be converted to complementary DNA (cDNA)[88]. Although dist technologies employ different techniques, the reverse transcription phase generally involves capturing mRNA using poly[T] sequence primers that bind to mRNA ploy[A] tail prior to cDNA conversion. Based on the sequencing platform, other nucleotide sequences are added to the reverse-transcription; for example in NGS protocols, UMIs are added to tag unique mRNA molecules so that it could be trace back their originating cells, enabling the combination of different cells for sequencing.

(iii) **cDNA Amplification**: Given that RNA can not be directly sequenced from cells, single-stranded RNAs must first be reverse-transcribed to cDNA. However, due to the small amount of mRNA in cells, limited cDNA is produced which is not optimal for sequencing. Therefore, the limited quantity of cDNA must be amplified prior to library preparation and sequencing[89]. The amplification is often done by either PCR (exponential amplification process with its efficiency being sequence dependent) or IVT (a linear amplification method which requires an additional round of reverse transcription of the amplified RNA) before sequencing[88,90]. The final cDNA library consists of adaptor-ligated sequencing library attached to each end.

(iv) **Sequencing Library Construction**: Finally, every cell's tagged and amplified cDNA is combined for library preparation and sequencing similar to bulk RNA sequencing methods, followed by computational pipelines for processing and analysis.[91].

Fig. 2(A) illustrates an example of the workflow for scRNAseq. For more details of each stage and various scRNAseq workflows, we refer the reader to references[90,92–94].

B. Spatial Transcriptomics Technologies

More recently, technologies that profile gene expression while retaining spatial information have emerged. These technologies are collectively known as *spatial transcriptomics* (ST). The various ST technologies provide different advantages and are chosen based on experimental factors such as size of tissue to be assayed, the number of genes to be probed, *a priori* knowledge of taget genes, cost, etcetera. In general, ST technologies can be divided into two broad categories: imaging-based and next generation sequencing (NSG)-based technologies. In this section, we provide an overview of popular techniques, with more emphasis on NGS-based approaches. For a more comprehensive and technical reviews of ST technologies, we refer the reader to Asp *et al.*[95] and Rao *et al.*[20].

1. *Imaging-Based Technologies*

Imaging-based technologies are be broadly subdivided into *in situ* hybridization (ISH)-based, *in situ* sequencing (ISS)-based methods, or methods that borrow elements from both of these approaches. Unlike RNAseq methods described above, ISH and most ISS-based techniques require labeled probes. This means that the target genes must be known in advance and, moreover, the number of genes that can be measured is limited[20].

***In Situ* Hybridization (ISH)-based Approaches**: ISH-based methods aim to detect the absence or presence of target RNA (or DNA) sequences while localizing the information of the desired sequences to specific cells or chromosomal sites[96,97].

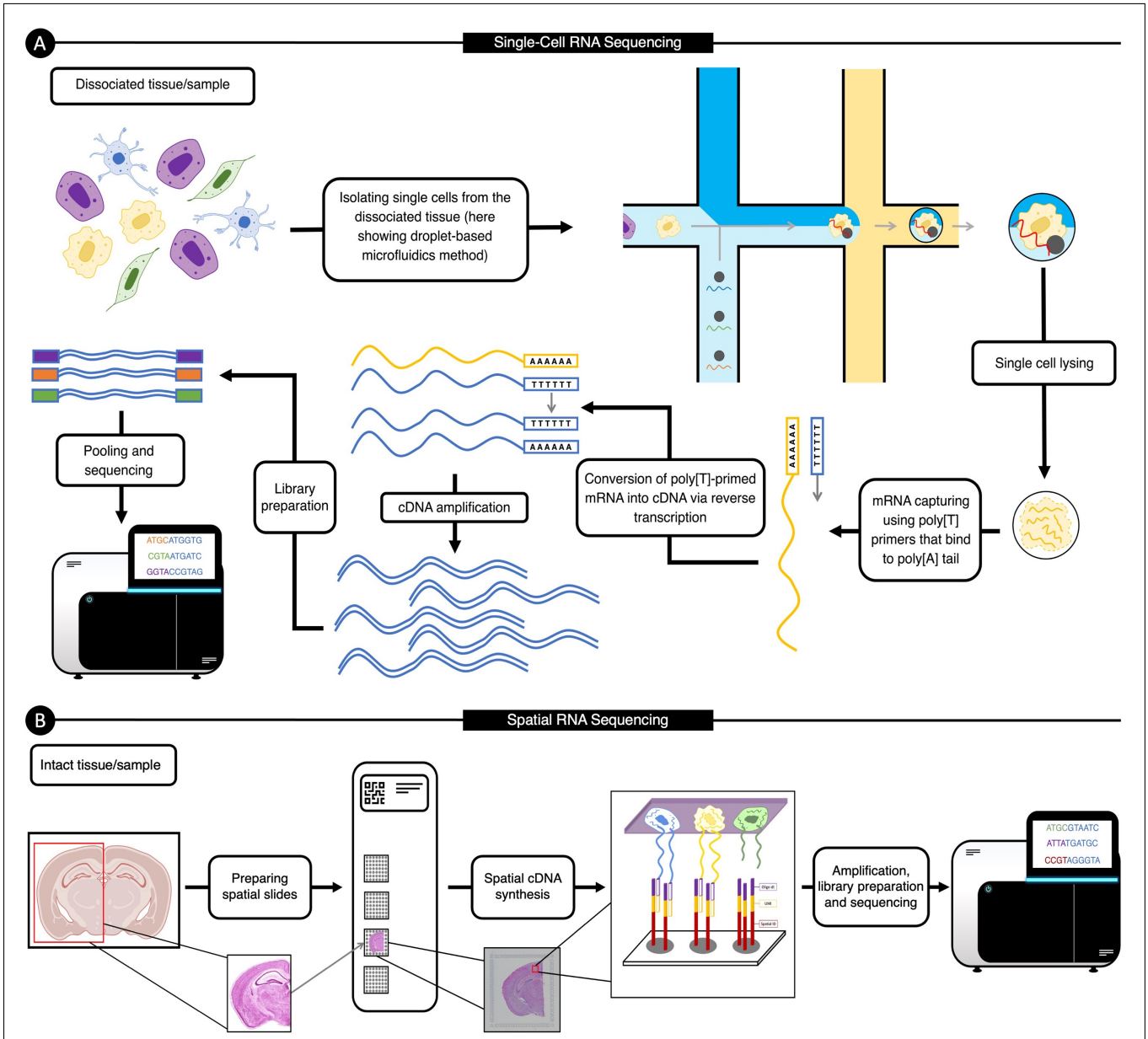

FIG. 2. **Example Single-Cell RNA Sequencing and Spatial Transcriptomics Workflows. (A)** A general overview of the single-cell RNA sequencing workflow (which we describe in section II A). **(B)** A visualization of the steps for next-generation sequencing based spatial transcriptomics (described in Section II B).

ISH-based techniques use labeled probes (usually made with DNA or RNA) which bind to desired sequences in fixed cells or tissue, therefore detecting the desired sequence through the hybridization of a complementary probe. The hybridized probes are then visualized through isotopic and nonisotopic (fluorescent and nonfluorescent) approaches[97]. The ISH-based techniques have been limited by the number of distinguishable transcripts, however, recent innovations have resulted in ample multiplexing capabilities[20].

***In Situ* Sequencing (ISS)-based Approaches**: ISS-based approaches aim to sequence the RNA content of a cell *in situ* using DNA balls that amplify the RNA signals: RNA is first reverse transcribed to cDNA, followed by circular amplification (to increase the number of transcripts) and sequencing[98]. Although the transcript can be localized at subcellular resolution, micrometer- or nanometer-sized DNA balls are often used to amplify the signals to reach sufficient signal for imaging[95]. Initially, the first ISS-based method[99] used targeted padlock probes (a single-stranded DNA molecule containing regions complementary to the target cDNA) followed by sequence-

TABLE I: **A list of relevant methods for the analysis of spatial transcriptomics data.** The italicized boldfaced methods are the ones which utilize deep learning (or elements closely aligned). We review these methods in depth in this paper.

Category	Method	Year	Framework	Language	Software Availability
Spatial Reconstruction	Seurat[59]	2015	Statistical	R	https://github.com/satijalab/seurat
	novoSpaRc[60]	2019	Optimization/Statistical	Python	https://github.com/rajewsky-lab/novosparc
	DEEPsc[61]	2021	Machine Learning	MATLAB	https://github.com/fmaseda/DEEPsc
Alignment and Integration	Spatial Backmapping[62]	2015	Scoring Scheme	R	https://github.com/jbogp/nbtspatialbackmapping
	Tangram[34]	2021	Machine Learning	Python	https://github.com/broadinstitute/Tangram
	GLUER[63]	2021	Machine Learning	Python	https://github.com/software-github/GLUER
Spot Deconvolution	Stereoscope[64]	2020	Statistical	Python	https://github.com/almaan/stereoscope
	DSTG[65]	2021	Machine Learning	Python, R	https://github.com/edward130603/BayesSpace
	SPOTlight[66]	2021	Machine Learning	R	https://github.com/MarcElosua/SPOTlight
	RTCD[67]	2021	Statistical	R	https://github.com/dmcable/RCTD
	SpatialDWLS[68]	2021	Optimization/Statistical	R	https://github.com/RubD/Giotto
	DestVI[69]	2021	Machine Learning	Python	https://github.com/YosefLab/scvi-tools
	Cell2location[70]	2022	Statistical	Python	https://github.com/BayraktarLab/cell2location
Spatial Clustering	HMRF[71]	2018	Statistical	R , Python , C	https://bitbucket.org/qzhudfci/smfishhmrf-py/
	SpaCell[72]	2019	Machine Learning	Python	https://github.com/BiomedicalMachineLearning/SpaCell
	BayesSpace[73]	2021	Statistical	R , C++	https://github.com/edward130603/BayesSpace
	SpaGCN[74]	2021	Machine Learning	Python	https://github.com/jianhuupenn/SpaGCN
Spatially Variable Genes Identification	Trendsceek[75]	2018	Statistial	R	https://github.com/edsgard/trendsceek
	SpatialDE[76]	2018	Statistical	Python	https://github.com/Teichlab/SpatialDE
	Spark[77]	2020	Statistical	R , C++	https://github.com/xzhoulab/SPARK
Cell-Cell Communication	SpaOTsc[78]	2020	Machine Learning	Python	https://github.com/zcang/SpaOTsc
	StLearn[79]	2020	Machine Learning	Python	https://github.com/BiomedicalMachineLearning/stLearn
	MISTy[80]	2020	Machine Learning	R	https://github.com/saezlab/mistyR
	Giotto[81]	2021	Statistical	R	https://github.com/RubD/Giotto

by-ligation[21] to detect desired genes. This method provided a subcellular resolution and an ability to detect single-nucleotide variants (SNVs). This ISS protocol is targeted and yields a detection efficiency of approximately 30%[100]. Several ISS protocols have built upon this approach to mitigate the number of cells that can be discriminated simultaneously, as well as to improve certain experimental aspect of the protocol. For example, a recently developed method, *barcode in-situ targeted sequencing* (BaristaSeq)[101], uses sequencing-by-synthesis and has led to increased read lengths, enabled higher throughput and cellular barcoding with improved detection efficiency compared to the initial ISS approach[101]. Another ISS-based technique is Spatially Resolved Transcript Amplicon Readout Mapping (STARmap)[29] that reduces noise and avoids the cDNA conversion complications by utilizing improved padlock-probe and primer design; STARmap adds a second primer to target the site next to the padlock probe in order to circumvent the reverse transcription step. STARmap also uses advanced hydrogel chemistry and takes advantage of an error-robust sequencing-by-ligation method, resulting in detection efficiency that is comparable to scRNAseq methods (around 40%)[29,95]. Although most ISS approaches (including the ones mentioned here) are targeted, ISS-based methods could also be untargeted[25,102] but this typically leads to much lower sensitivity (around 0.005%) and molecular crowding, affecting the rolling-circle amplification bias[102,103].

In imaging-based approaches, the generated image is segmented and processed to produce a cell-level gene-expression matrix. The gene-expression matrix is generated through processing the generated image(s), which can be done manually or automatically. However, given the biased and laborious nature of manual segmentation, there has been a shift towards designing general and automated techniques[104]. The accurate and general automation of this process still remains a challenge, therefore motivating the application of recent machine learning and computer vision approaches to this field[104–106], which have shown improvements compared to the traditional methods[107]. Although this manuscript focuses on methods for NGS-based technologies, many of those techniques (including ones in Section IV C and IV F) can be extended to image-based technologies as well.

C. Next Generation Sequencing (NGS)-Based Technologies

Due to the unbiased capture of mRNA, NGS-based technologies can shed light on the known and unknown morphological features using only the molecular characterization of tissues[1]. This unbiased and untargeted nature of NGS technologies makes them ideal for studying and exploring new systems[20], a major advantage compared to most image-based technologies which require target genes *a priori*. While NGS-based approaches differ in the specifics of the protocols, they all build on the idea of adding spatial barcodes before library preparation, which are then used to map transcripts back to the appropriate positions (known as spots or voxels). An example workflow of NGS-based spatial sequencing is depicted in Fig. 2(B). In the following subsections, we provide a general overview of the four most common spatial transcriptomics technologies. For a more complete review of these technologies, we refer the readers to references[20,95].

Ståhl *et al.*[31] were the first to successfully demonstrate the feasibility of using NGS for spatial transcriptomics (this initial approach is often referred to *Spatial Transcriptomics*). Their

innovation was to add spatial barcodes prior to library preparation, enabling the mapping of expressions to appropriate spatial spots. More specifically, Ståhl *et al.* positioned oligo(dT) probes and unique spatial barcodes as microarrays of spots on the surface of slides. Next, fresh frozen tissue slices were placed on the microarray and processed to release mRNA (using enzymatic permeabilization), which then hybridized with the probes on the surface of the slides. This approach consists of (i) collecting histological imaging (using standard fixation and staining techniques, including hematoxylin and eosin (HE) staining) for investigating morphological characteristics and (ii) sequencing spatially barcoded cDNA to profile gene expressions. In the initial experiments, each slide consisted of approximately 1000 spots, each of diameter 100 μm with 200 μm center-to-center distance[1]. This approach provides researchers with an unbiased technique for analyzing large tissue areas without the need for selecting target genes in advance[20,35,108].

After the initial success of *Spatial Transcriptomics*, 10x Genomics subsequently improved the resolution (shrinking the spot diameters to 55 μm with 100 μm centre-to-centre distance) and sensitivity (capturing more than 10^4 transcripts per spot) of the approach, and eventually commercializing it as Visium[30,109]). The development and commercialization of the spatial transcriptomics resulted in relatively rapid adoption across fields, such as cancer biology[110,111], developmental biology[112,113], neuroscience[114,115]. The histological imaging and gene expression profiling of Visium are similar to the initial approach: the staining and imaging of the tissues are through traditional staining techniques, including HE staining for visualizing tissue sections using a brightfield microscope and immunofluorescence staining to visualize protein detection in tissue sections through a fluorescent microscope. Visium protocol allows for both fresh frozen (FF) or Formalin-Fixed Paraffin-Embedded (FFPE) tissues. For FF tissues, similar to Ståhl *et al.*[31], the tissue is permeabilized, allowing the release of mRNA, which hybridizes to the spatially barcoded oligonucleotides present on the spots. The captured mRNA then goes through a reverse transcription process that results in cDNA, which are then barcoded and pooled for generating a library[116]. For FFPE tissues, tissue is permeabilized to release ligated probe pairs from the cells that bind to the spatial barcodes on slide, and the barcoded molecules are pooled for downstream processing to library generation[116].

Building on the *Spatial Transcriptomics*, Vickovic *et al.*[33] proposed High-Definition Spatial Transcriptomics (HDST) which improved the resolution to about 2 μm. Similar to the other approaches, HDST also employs specific barcodes ligated to beads that are coupled to a spot (prior to lysis), so that expressions are mapped to the tissue image. However, the innovation of HDST include the use of 2 μm beads places in hexagonal wells, enabling accurate compartmentalization and grouping of the biological materials in the experiment[33]. Simultaneously, Rodriques *et al.*[32] introduced SlideSeq which utilizes slides with randomly barcoded beads to capture mRNA, also increasing the resolution (to 10 μm) and sensitivity (500 transcripts per 54 bead) spatial-resolved sequencing compared to Ståhl *et al.*[31]. However, SlideSeq placed the barcoded beads in rubber and onto glass slides, as opposed to HDST's hexogonal beads, and determines the position of each random barcode by *in situ* -indexing[20,32].

Despite the differences, all NGS-based technologies use spatial barcodes to tag released RNAs, which then go through conventional processes for sequencing similar to scRNAseq. After sequencing, the data is processed to construct the spatial location of each read (using the spatial barcode) and to construct a gene-expression matrix (mapping the reads to the genome to identify the transcript of origin). Given that most technologies have resolutions larger than a single-cell (commonly having expression for 3 to 30 cells in each spot), the data processing and analysis procedures are relatively similar.

III. MACHINE LEARNING AND DEEP LEARNING BACKGROUND

With the technologies now defined, we next describe common *Machine Learning* (ML) methods used to analyze ST data. in this section, we first provide a discussion of the algorithmic development of ML and *Deep Learning* (DL) models, and then discuss common architectures used for spatially-resolved transcriptomics (and scRNAseq data).

ML refers to a computer algorithm's ability to acquire knowledge by extracting patterns and features from raw data[117]. All ML algorithms depend on data, which must be available before the methods can be used, and a defined mathematical objective. ML models' lifecycle consists of two phases, namely *training* and *evaluation*. During training, ML algorithms analyze the data to extract patterns and adjust their internal parameters based on optimizing their objectives (known as *loss function*). In the evaluation (or inference) stage, the trained model makes predictions (or performs the task it was trained to do) on *unseen* data.

There are two main types of ML algorithms: *supervised* and *unsupervised*. An ML algorithm is considered to be *unsupervised* if it utilizes raw inputs without any labels to optimize its objective function (an example would be the K-Means clustering algorithm[118]). Conversely, if an algorithm uses both raw data and the associated labels (or targets) in training, then it is a *supervised* learning algorithm. Supervised learning is the most common form of ML[39]. An example of supervised learning in scRNAseq analysis would be classifying cell subpopulations using prior annotations: this requires a labeled set of cell-types for training (the available annotations), an objective function for calculating learning statistics ("teaching" the model), and testing data for measuring how well the model can predict the cell-type (label) on data it has not seen before (*i.e.* generalizibility of the model). Another common example of supervised learning is regression, where a model predicts continuous values as opposed to outputting labels or categorical values in classification. For supervised tasks, a model is trained on the majority of the data (known as *training set*) and then evaluated on held-out data (*test set*). Depending on the size of our dataset, there can also be a third data split known as a *validation set*, which is used to measure the performance of

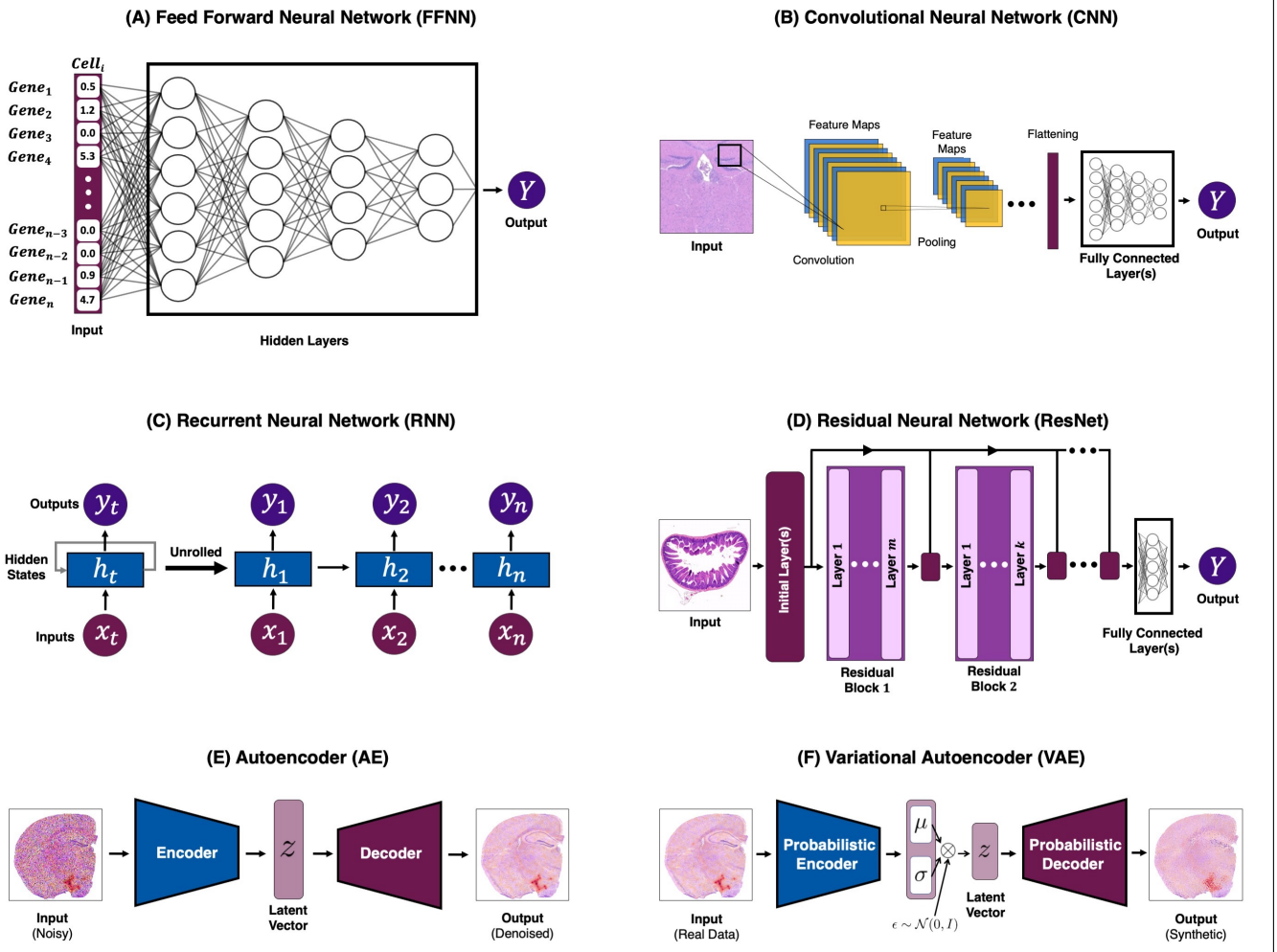

FIG. 3. **Examples of Deep Learning Architectures**. Models depicted in **(A), (B), (C), (D)** are examples of supervised learning, and networks shown in **(E), (F)** are unsupervised. **(A)** An example of an FFNN architecture with gene expression count as its input. **(B)** An example of CNN architecture, where the model passes the inputs through the three stages of a CNN (with non-linear activation not depicted) to extract features. Then, outputs are flattened and fed into a fully connected layer (or layers). **(C)** The general training flow of an RNN, with the unrolled version showing the timestep-dependent inputs, hidden state, and outputs. The inputs to RNNs need to have a sequential structure (*e.g.* time-series data). **(D)** An illustration of a ResNet. In traditional ResNets, there are identity mappings (or skip connections) that pass the input of a residual block to its output (often through addition). **(E)** Here we show the general architecture of a *trained* denoising AE in inference stage, with a noisy histology slide as its input, yielding a denoised version of the input image. **(6)** A depiction of a traditional VAE in inference stage. VAE's aim to generate synthetic data that closely resemble the original input. This is done through regularizing the latent space of an AE with the use of a probabilistic encoder and decoder.

the model throughout training to determine *early stopping*[119]: Early stopping is when we decide to stop the training of a model because its overfitting (or over optimization) on the training set. Overfitting on training data worsens the generalizability of the model on unseen data, which early stopping aims to avoid[119]. In addition to supervised and unsupervised algorithms, there are also *semi-supervised* learning, where a model uses a mix of both supervised and unsupervised tasks, and *self-supervised*, where the computer algorithm generates new or additional labels to improve its training, or to learn a new task.

Raw experimental data typically contains noise or other unwanted features, which present many challenges for ML algorithms. Therefore, it is often necessary to carefully preprocess data or to rely on domain-specific expertise in order to transform raw data into some internal representation from which ML models can learn[39]. Deep Learning (DL) algorithms, however, aim to use only raw data to automatically

extract and construct useful representations required for learning the tasks at hand. In a broad sense, DL models are able to learn from observations through constructing a hierarchy of concepts, where each concept is defined by its relation to simpler concepts. A graph representation of the hierarchy of concepts (and learning) will consist of many layers, with many nodes and edges connecting the vertices, somewhat resembling humans' neural network. This graph is referred to as an Artificial Neural Network (ANN). ANNs are composed of interconnected nodes ("artificial neurons") that resemble and mimic our brains' neuronal functions. An ANN is considered to be a DL model if it consists of many layers–often more than three, hence being called *deep*.

Many tasks that humans perform can be viewed as mappings between sets of inputs and outputs. For example, humans can take a snapshot image of their surroundings (input) and detect the relevant objects (the outputs). DL, and more generally Artificial Intelligence, aims to learn such mappings in order to model human-level intelligence. Mathematically, ANNs are universal function approximators, meaning that, theoretically, they can approximate any (continous) function[120–122]. Cybenko[120] proved this result for a one-layer neural network with arbitrary number of neurons (nodes) and a sigmoid activation function by showing that such architecture is dense within the space of continuous functions (this result has now been extended to ANNs with multiple layers[121]). While constructing arbitrarily-long single-layer ANNs is not possible, it has been shown that ANNs with many many layers (deeper) generally learn faster and more reliably than ANNs with few wide (many neurons) layers[123]. This has allowed researchers to employ deep networks for learning very complex functions through constructing simple non-linear layers which can transform the representation of each module (starting with the raw input) into a representation at a higher, slightly more abstract level[39].

DL models' ability to approximate highly non-linear functions has revolutionized many domains of science, including Computer Vision[124–126], Natural Language Processing[127–129] and Bioinformatics[130–133]. DL is becoming increasingly incorporated in many computational pipelines and studies, specially in genomics and bioinformatics, including scRNAseq and spatial transcriptomics analysis. In the following sections, we provide a brief overview of essential deep learning architectures that have been used in spatial transcriptomics and scRNAseq analysis. In Fig. 3, we present illustrations of the architectures discussed in the following sections. Note that for simplicity, we have categorized all Graph Convolution Networks (GCN)[134] as DL models; this is because (i) GCNs can easily be extended to include more layers (deeper networks), and (ii) lack of other existing methods which incorporate some elements of DL. A more comprehensive description of each architecture can be found in the seminal textbook by Goodfellow *et al.*[117].

A. Feed Forward Neural Network (FFNN)

FFNNs, the quintessential example of Artificial Neural Networks (ANNs), aim to approximate a function mapping a set of inputs to their corresponding targets (see Fig. 3(A)). More specifically, given an input $\mathbf{x} \in \mathbb{R}^n$ and a target $\mathbf{y} \in \mathbb{R}^m$, where $n, m \in \mathbb{R}$, FFNNs aim to *learn the optimal parameters* $\boldsymbol{\theta}$ *such that* $\mathbf{y} = f(\mathbf{x}; \boldsymbol{\theta})$. FFNNs are the building blocks of many more advanced architectures (*e.g.* convolutional neural networks), and therefore, of paramount importance in the field of ML[117]. As mentioned previously, ANNs are universal function approximators, and they represent a directed acyclic graph of function compositions hierarchy within the network. Each layer of a FFNN, $f^{(i)}(\mathbf{x}; \boldsymbol{\theta})$ ($i \in \mathbb{N}$ being the i-th layer), is often a simple linear function: For example, we can have a linear function for outputting $y \in \mathbb{R}$ of the form Eq. (1), with weight parameters $\mathbf{w} \in \mathbb{R}^n$ and a bias $b \in \mathbb{R}$:

$$y = f^{(1)}(\mathbf{x}; \boldsymbol{\theta}) = f^{(1)}(\mathbf{x}; \mathbf{w}, b) = \mathbf{x}^T \mathbf{w} + b. \tag{1}$$

However, a model composed of *only* linear functions can *only* approximate linear mappings. As such, we must consider *non-linear activation* functions to increase model capacity, enabling the approximation of complex non-linear functions. In the simplest case, Neural networks (NNs) use an affine transform (controlled by learned parameters) followed by a non-linear activation function, which, theoretically, enables them to approximate any non-linear function[135]. Moreover, we could compose many of such non-linear transformations to avoid infinitely wide-neural networks when approximating complex function. However, in this context, finding a set of optimal functions $f^{(i)} : \mathbb{R}^{q_i} \to \mathbb{R}^{d_i}$ ($q_i, d_i \in \mathbb{R}$) is a practically impossible task. As such, we restrict the class of function that we use for $f^{(i)}$ to the following form in Eq. (2):

$$f^{(i)}(\mathbf{x}^{(i-1)}; \boldsymbol{\theta}^{(i)}) = \sigma^{(i)}(W^{(i)}\mathbf{x}^{(i-1)} + \mathbf{b}^{(i)}), \tag{2}$$

where superscript i enumerates the layers, $\sigma(\cdot)$ is a non-linear activation function (usually a Rectified Linear Unit[136]), $\mathbf{x}^{(i-1)} \in \mathbb{R}^{q_i}$ denotes the output of the layer $(i-1)$ (with $\mathbf{x}^{(0)}$ indicating the input data), weights $W \in \mathbb{R}^{d_i \times q_i}$ and biases $\mathbf{b}^{(i)} \in \mathbb{R}^{d_i}$. Note that because of the dimensionality of the mapping, $W^{(i)}X^{(i-1)} \in \mathbb{R}^{d_i}$ and we must have a vector of biases $\mathbf{b}^{(i)} \in \mathbb{R}^{d_i}$). FFNNs are composed of such functions in chains; to illustrate, consider a three-layer neural network:

$$\mathbf{y} = f(\mathbf{x}; \boldsymbol{\theta}) \tag{3a}$$

$$= f^{(3)}(f^{(2)}(f^{(1)}(\mathbf{x}; \boldsymbol{\theta}^{(1)}); \boldsymbol{\theta}^{(2)}); \boldsymbol{\theta}^{(3)}) \tag{3b}$$

$$= f^{(3)}\left(h^{(2)}\left(h^{(1)}\left(\mathbf{x}; \mathbf{w}^{(1)}, \mathbf{b}^{(1)}\right); \mathbf{w}^{(2)}, \mathbf{b}^{(2)}\right); \mathbf{w}^{(3)}, \mathbf{b}^{(3)}\right). \tag{3c}$$

with h representing the *hidden states or hidden layers*.

FFNNs find the optimal contribution of each parameter (*i.e.* weights and biases) by minimizing a desired objective. The goal is to generalize the task to data the model has never seen before (testing data). Although the non-linearity increases the capacity of FFNNs, it causes most objective functions to become non-convex. In contrast to convex optimization, non-convex loss functions do not have global convergence guarantees, and are sensitive to initial starting point (parameters of

the network)[137]. Therefore, such optimization is often done through stochastic gradient descent (or some variant of it). Moreover, given the sensitivity to initial values, weights are typically chosen to be small random values, with biases initialized to zero or small positive values[117,138,139].

Almost all neural networks use iterative gradient descent (GD)-based optimizers to train. GD has three main variants, which differ in the amount of data utilized to calculate the gradients for updating the parameters. The classic GD variant, referred to as *batch* GD, uses all data points to make the updates to the parameters in *one iteration*. However, this approach is generally not feasible, since the amount of data required for training DL models almost never fits in memory[140]. The second variant of GD is *stochastic gradient descent* (SGD) where the parameters are updated for every training datum. Computationally, it has been shown that the noise in SGD accelerates its convergence compared to batch GD, but SGD also has the possibility of overshooting, specially for highly non-convex optimization functions[140,141]. The third variant, and the most frequently used one for deep learning, is mini-batch GD which updates the parameters for every batch of training data–if batch size is one then this variant is just SGD, and if batch size is the entire dataset then it is equivalent to batch GD. Conventionally, optimization of NNs is done through gradient descent performed backwards in the network, which itself consists of two components: a numerical algorithm for efficient computation of the chain rule for derivatives (backpropagation[142]) and a GD-based optimizer (*e.g.* , Adam[143] or AdaGrad[144]). The optimizer is an algorithm that performs gradient descent, while backpropagation is an algorithm for computing the expression for gradients during the backward pass of the model.

B. Convolutional Neural Network (CNN)

Learning from images, such as detecting edges and identifying objects, has been of interest for some time in computer science[145]. Images contain a lot of information, however, only a small amount of that information is often relevant to the task at hand. For example, an image of a stained tissue contains both important information, namely the tissue itself, and irrelevant pixels, such as the background. Prior to DL, researchers would hand-design a feature extractor to learn relevant information from the input. Much of the work had focused on the appropriate feature extractors for desired tasks (*e.g.* see the seminal work by Marr and Hildreth[146]). However, a main goal in ML is to extract features from raw inputs without hand-tuned kernels for feature extraction. CNNs[147,148] are a specialized subset of ANNs that use the convolution operation (in at least one of their layers) to learn appropriate kernels for extracting important feature beneficial to the task at hand. Mathematically, convolution between two functions f and w is defined as a commutative operation shown in Eq. (4)

$$(f*w)(x) \triangleq \int_{-\infty}^{\infty} f(s)w(s-x)ds. \tag{4}$$

Using our notation, we intuitively view convolution as the area under $f(s)$ weighted by $w(-s)$ and shifted by x. In most applications, discrete functions are used. As an example, assume we have a 2D kernel K that can detect edges in a 2D image I with dimension $m \times n$. Since I is discrete, we can use the discrete form of Eq. (4) for convolution of I and K over all pixels:

$$E(i,j) = (I*K)(i,j) = \sum_m \sum_n I(m,n)K(i-m,j-n). \tag{5}$$

However, since there is less variation in the valid range of m,n (the dimensions of the image) and the operation is commutative, most algorithms implement Eq. (5) equivalently:

$$E(i,j) = (K*I)(i,j) = \sum_m \sum_n I(i-m,j-n)K(m,n). \tag{6}$$

Typical CNNs consist of a sequence of layers (usually three) which include a layer performing convolution, hence called a *convolutional layer* (affine transform), a detector stage (non-linear transformation), and a pooling layer. The learning unit of a convolutional layer is called a *filter or kernel*. Each convolutional filter is a matrix, typically of small dimensions (*e.g.* 3x3 pixels), composed of a set of weights that acts as an object detector, with the weights being continuously calibrated during the learning process. CNNs' objective is to learn an optimal set of filters (weights) which can detected the needed features for specific tasks (*e.g.* image classification). The result of convolution between the input data and the filter's weights is often referred to as a *feature map* (as shown in Fig. 3(B)). Once a feature map is available, each value of this map is passed through a non-linearity (*e.g.* ReLU). The output of a convolutional layer consists of as many stacked feature maps as the number of filters present within the layer.

There are two key ideas behind the design of CNNs: First, in data with grid-like topology, local neighbors have highly correlated information. Second, equivariance to translation can be obtained if units at different locations *share weights*. In other words, sharing parameters in CNNs enabled the detection of features regardless of the locations that they appear in. An example of this would be detecting a car. In a dataset, a car could appear at any position in a 2D image, but the network should be able to detect it regardless of the specific coordinates[145]. These design choices provide CNNs with three main benefits compared to other ANNs: (i) sparse interactions, (ii) shared weights, and (iii) equivariant representations[147].

Another way of achieving equivariance to translation is to utilize pooling layers. Pooling decreases the dimension of learned representations, and makes the model insensitive to small shifts and distortions[39]. In the pooling layers, we use the outputs of the detector stage (at certain locations) to calculate a summary statistic for a rectangular window of values (*e.g.* calculating the mean of a 3x3 patch). There are many pooling operations, with common choices being max-pooling (taking the maximum value of a rectangular neighborhood), mean-pooling (taking the average), and L2 norm (taking the norm). In all cases, rectangular patches from one or several feature maps are inputted to the pooling layer, where semantically similar features are merged into one. CNNs typically have an ensemble of stacked convolution layers, non-linearity,

and pooling layers, followed by fully connected layers that produce the final output of the network. The backpropagation of gradients through CNNs is analogous to FFNNs, enabling the model to learn an optimal set of filters for the task(s) at hand. CNNs have been effectively used in many applications in computer vision and time-series analysis, and are being increasing utilized for analysis of ST data, since spatial-omics are multi-modal, with one of the modalities being images (as we discuss in Section IV).

C. Recurrent Neural Network (RNN)

Just as CNNs are specialized to process data with a grid-like toplogy, RNNs'[149] special characteristics make them ideal for processing sequential data $X = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \cdots, \mathbf{x}^{(n)}\}$, where $\mathbf{x}^{(i)}$ denotes the i-th element in the ordered sequence X. Examples of such sequence-like structure include times series and natural language. RNNs process sequential inputs one at a time and implicitly maintain a history of previous elements of the input sequence. We present an illustration of the conventional RNN architecture in Fig. 3(C). Similar to FFNNs or CNNs, RNNs can be composed of many layers, with each layer depending on the previous hidden state, $h^{(t-1)}$, and a shared set of parameters, $\boldsymbol{\theta}$. A deep RNN with n hidden states can be expressed as follows:

$$h^{(n)} = f(\mathbf{x}^{(n)}, h^{(n-1)}; \boldsymbol{\theta}); \boldsymbol{\theta}) \tag{7a}$$
$$= f(\mathbf{x}^{(n)}, f(\mathbf{x}^{(n-1)}, h^{(n-2)}; \boldsymbol{\theta}); \boldsymbol{\theta}) \tag{7b}$$
$$= f(f(\cdots f(\mathbf{x}^{(2)}, h^{(1)}(\mathbf{x}^{(1)}; \boldsymbol{\theta}); \boldsymbol{\theta}) \cdots; \boldsymbol{\theta}); \boldsymbol{\theta}). \tag{7c}$$

The idea behind sharing $\boldsymbol{\theta}$ in RNN states is similar to CNNs: parameter sharing across different time points allows RNNs to generalize the model to sequences of variable lengths, and share statistical strengths at different positions in time[117,150]. Similar to FFNNs, RNNs learn by propagating the gradients of each hidden state's inputs at discrete times. This process becomes more intuitive if we consider the outputs of hidden units at various time iterations as if they were the outputs of different neurons in a deep multi-layer network. However, due to the sequential nature of RNNs, the backpropagation of gradients shrinks or grows at each time step, causing the gradients to potentially vanish or blow up. This fact, and the inability to parallelize training at different hidden states (due to the sequential nature of RNNs) makes RNNs notoriously hard to train, specially for longer sequences[127,151]. However, when these issues are averted (via gradient clipping or other techniques), RNNs are powerful models and gain state-of-the-art capabilities in many domains, such natural language processing. The training challenges combined with the nature of scRNAseq data have resulted in fewer developments of RNNs for single-cell analysis. However, recently some studies have used RNNs and Long Short-Term Memory[152] (a variant of RNNs) used for predicting cell types and cell motility (*e.g.* see Kimmel et al.[153]).

D. Residual Neural Network (RestNet)

As mentioned above, deep RNNs may suffer from vanishing or exploding gradients. Such issues can also arise in other deep neural networks as well, where gradient information could diminish as the depth increases (though approaches such as Batch Normalization[154] aim to help with gradient issues). One way to alleviate vanishing gradients in very deep networks is to allow gradient information from successive layers to pass through, helping with maintaining information propagation even as networks become deeper. ResNets[155] achieve this by skip (or residual) connections that add the input to a block (a collection of sequential layers) to its output. For a FFNN, consider function f in Eq. (2). Using the same notation as in Eq. (2), ResNet's inner layers take the form shown in Eq. (8):

$$f^{(i)}(\mathbf{x}^{(i-1)}) = \mathbf{x}^{(i-1)} + \boldsymbol{\sigma}^{(i)}(W^{(i)}\mathbf{x}^{(i-1)} + \mathbf{b}). \tag{8}$$

The addition of $\mathbf{x}^{(i-1)}$, the input of the current layer (or the output of $(i-1)$-th layer), to the current i-th layer output is the *skip or residual* connection helps flow the information from the input deeper in the network, thus stabilizing training and avoiding vanishing gradient in many cases[155,156]. Indeed, this approach can be contextualized within the traditional time integration framework for dynamical system. For example, consider Eq. (9):

$$\dot{x}(t) = \frac{dx}{dt} = \mathscr{F}(t, x(t)), \quad x(t_0) = x_0. \tag{9}$$

In the simplest case, this system can be discretized and advanced using $x(t_n)$ and some scaled value of $\mathscr{F}(t_n, x(t_n))$, or a combination of scaled values of $\dot{x}(t_n)$. Forward Euler, perhaps the simplest time integrator, advances the solution as shown through the scheme in Eq. (10)

$$x_{n+1} = x_n + h\mathscr{F}(t_n, y_n), \tag{10}$$

where h is a *sufficiently small* real positive value. ResNets use this idea to propose a different way of calculating the transformations in each layer, as shown in Eq. (8).

ResNets consist of residual blocks (also called modules), each of which containing a series of layers. For visual tasks, these blocks often consist of convolutional layers, followed by activation functions, with the skip connection adding the input information to the output of the residual blocks (as opposed to the individual layers inside). ResNets have different depths and architectures, with a number usually describing the depth of the model (*e.g.* ResNet50 means there are 50 layers [there are 48 convolution layers, one MaxPool and one AveragePool layers]).

ResNets have transformed DL by enabling the training of *very* deep neural networks, setting the state-of-the-art performance in many areas, particularly in computer vision[155]. The pre-trained ResNets on ImageNet dataset[157] are widely used for transfer learning, where the network is either used as is or further fine-tuned on the specific dataset. Pre-trained ResNet models have also been used in spatial transcriptomics analysis, as we discuss later in this manuscript.

E. Autoencoder (AE)

AEs[158,159] are neural networks that aim to reconstruct (or copy) the original input via a *non-trivial mapping*. Conventional AEs have an "hour-glass" architecture (see Fig. 3(E)) consisting of two networks: (i) an encoder network, $Enc(\cdot)$, which maps an input $\mathbf{x} \in \mathbb{R}^n$ to a latent vector $\mathbf{z} \in \mathbb{R}^d$ where, ideally, $\mathbf{z}$ contains the most important information from $\mathbf{x}$ in a reduced space (*i.e.* $d \ll n$), (ii) the decoder network, $Dec(\cdot)$, which takes $\mathbf{z}$ as input and maps it back to $\mathbb{R}^n$, ideally, reconstructing $\mathbf{x}$ exactly; *i.e.* $\mathbf{x} = AE(\mathbf{x}) = Dec(Enc(\mathbf{x}))$. AEs were traditionally used for dimensionality reduction and denoising, trained by minimizing a mean squared error (MSE) objective between the input data and the reconstructed samples (outputs of the decoder).

Over time, the AE framework has been generalized to stochastic mappings, *i.e.* probabilistic encoder-decoder mappings, $p_{Enc}(\mathbf{z}|\mathbf{x})$ and $p_{Dec}(\mathbf{x}|\mathbf{z})$. A well-known example of such generalization is Variational Autoencoders (VAEs)[160], where by using the same hour-glass architecture, one can use probabilistic encoders and decoders to generate new samples drawn from an approximated posterior. Both traditional AEs and VAEs have practical applications in many biological fields, and have been used extensively in scRNAseq (see reference[10] for an overview of these models), and are becoming more frequently employed in spatial transcriptomics analysis, which we overview later in this work.

F. Variational Autoencoder (VAE)

One can describe VAEs[160] as AEs that regularize the encoding distribution, enabling the model to generate new synthetic data. The general idea behind VAEs is to encode the inputs as a *distribution over the latent space*, as opposed to a single point (which is done by AEs). More specifically, VAEs draw samples $\mathbf{z}$ from an encoding distribution, $p_{model}(\mathbf{z})$, and subsequently feed the sample through a differentiable generator network, obtaining $Gen(\mathbf{z})$. Then, $\mathbf{x}$ is sampled from a distribution $p_{model}(\mathbf{x}; Gen(\mathbf{z})) = p_{model}(\mathbf{x}|\mathbf{z})$. Moreover, VAEs utilize an approximate inference network $q(\mathbf{z}|\mathbf{x})$ (*i.e.* the encoder) to obtain $\mathbf{z}$. With this approach, $p_{model}(\mathbf{x}|\mathbf{z})$ now is considered a decoder network, decoding $\mathbf{z}$ that comes from $q(\mathbf{z}|\mathbf{x})$. VAEs can take advantage of gradient-based optimization for training through *maximizing the variational lower bound*, $\mathscr{L}$, associated with $\mathbf{x}$. Fig.3(F) depicts the architecture of traditional VAEs.

Mathematically, we can express the objective function as in Eq. (11):

$$\mathscr{L}(q) = \mathbb{E}_{\mathbf{z}\sim q(\mathbf{z}|\mathbf{x})} \log p_{model}(\mathbf{z},\mathbf{x}) + \mathscr{H}(q(\mathbf{z}|\mathbf{x})) \tag{11a}$$

$$= \mathbb{E}_{\mathbf{z}\sim q(\mathbf{z}|\mathbf{x})} \log p_{model}(\mathbf{x}|\mathbf{z}) - \mathbb{KL}(q(\mathbf{z}|\mathbf{x})||p_{model}(\mathbf{z})) \tag{11b}$$

$$\leq \log p_{model}(\mathbf{x}) \tag{11c}$$

where $\mathscr{H}(\cdot)$ denotes entropy and $\mathbb{KL}$ is the Kullback-Leibler divergence. The first term in Eq.(11c) is the joint log-likelihood of the hidden and visible variables under the approximate posteriors over the latent variables. The second term of Eq. (11c) is the entropy of the approximate posterior. This entropy term encorages the variational posterior to increase the probability mass on a range of $\mathbf{z}$ which could have produced $\mathbf{x}$, as opposed to mapping to a one point estimate of the most likely value[117].

Compared to other generative models (*e.g.* Generative Adversarial Networks (GANs)[161]), VAEs have desirable mathematical properties and training stability[117]. However, they suffer from two major weaknesses: (i) classic VAEs create "blurry" samples (those that adhere to an average of the data points), rather than the sharp samples that GANs generate due to GANs' adversarial training. (ii) The other major issue with VAEs is posterior collapse: when the variational posterior and actual posterior are nearly identical to the prior (or collapse to the prior), which results in poor data generation quality[162]. To alleviate these issues, different algorithms have been developed, which have been shown to significantly improve the quality of data generation[163–168]. VAEs are used extensively for the analysis of single-cell RNA sequencing (see Erfanian et al.[10]), and we anticipate them to be applied to a wide range of spatial transcriptomics analysis as well.

IV. DEEP LEARNING MODELS FOR SPATIALLY-RESOLVED TRANSCRIPTOMICS ANALYSIS

In the following sections, we describe the use of ML and DL to problems emerging from spatial transcriptomics.

A. Spatial Reconstruction

Prior to the advancement of spatial transcriptomics, several studies aimed to reconstruct spatial information using gene expression data, with most of these works using a statistical framework. As perhaps one of the most influential models in this space, Satija *et al.*[59] introduced Seurat: a tool which utilized spatial reference maps constructed from a few landmark *in situ* patterns to infer the spatial location of cells from corresponding gene expression profiles (*i.e.* scRNAseq data). This approach showed promising results: Satija *et al.* tested seurat's capabilities and performance on developing zebrafish embryo dataset (containing 851 cells) and a reference atlas constructed from colorigenic *in situ* data for 47 genes[59], confirming Seurat's accuracy with several experimental assays. Additionally, they showed that Seurat can accurately identify and localize rare cell populations. Satija *et al.* also demonstrated that Seurat was a feasible computational solution for handling stochastic noise in omics data, and finding a correspondence between ST and scRNAseq data.

Although Seurat proved to be successful in some applications, it had the limitation of requiring spatial patterns of marker genes expression[60]. To alleviate Seurat's limitations, newer methods that did not require spatial reference atlases were developed. A more recent and an influential model in this space is *novoSpaRc*[60], with the ability to infer spatial mappings of single cells *de novo*. For novoSpaRc, Nitzan *et al.*[60] assume that cells which are closer to one-another physi-

cally have similar gene expressions as well, therefore searching for spatial arrangement possibilities which place cells with similar expressions closer in space. Nitzan *et al.* formulate this search through a generalized optimal-transport problem for probabilistic embedding.

NovoSpaRc shows very promising results when it is applied to spatially reconstruct mammalian liver and intestinal epithelium, and embryos from fly and zebrafish from gene expression data[60]. However, novoSpaRc (and similar models) use a generic framework and can not be easily adapted to specific biological systems, which may be required given the vast diversity of biological processes and organisms. For this reason, many have utilized ML algorithms to specifically adapt to the biological system by learning from the data, as opposed to using pre-defined algorithms that remain unchanged. Indeed, we anticipate that DL models will soon play a salient role in spatial reconstruction of scRNAseq, given their ability to extract features from raw data while remaining flexible across different applications. In this section, we review DEEPsc[61], a system-adaptive ML model which aims to impute spatial information onto non-spatial scRNAseq data.

DEEPsc is a spatial reconstruction method which requires a reference atlas (see Fig. 4). This reference map can be expressed as a matrix $M_{spatial} \in \mathbb{R}^{n_{positions} \times n_{genes}}$ where $n_{positions} \in \mathbb{N}$ is number of spatial locations and $n_{genes} \in \mathbb{N}$ the number of genes. Maseda *et al.* start by selecting common genes between $M_{spatial}$ and the gene expression matrix, $M_{expression} \in \mathbb{R}^{m_{cells} \times m_{genes}}$ (where $m_{cells} \in \mathbb{N}$ is the number of cells and $m_{genes} \in \mathbb{N}$ the number of genes), resulting in a spatial matrix $S \in \mathbb{R}^{n_{positions} \times g}$ and an expression matrix $E \in \mathbb{R}^{m_{cells} \times g}$, where $g \in \mathbb{N}$ is the common genes between the two matrices. Next, S is projected into a lower dimension using principal component analysis (PCA), and the same PCA coefficients are used to project E into these principal components. In the last step of processing, both matrices are normalized by their largest elements, resulting all elements of the matrices E and S to be in [0, 1]. Let us denote the normalized and PCA-reduced spatial and gene expression matrices as $\tilde{S}$ and $\tilde{E}$, respectively.

DEEPsc requires known spatial expression to learn the correct spatial positions, given the gene expression. More specifically, Maseda *et al.* construct training vectors of size $Inp_{ij} = [pos_i; pos_j] \in \mathbb{R}^{2N}$ (with N being the number of features preserved in the reduced matrix $\tilde{S}$). The first N elements of Inp_{ij} correspond to the spatial expression at the i-th position , and the last N elements correspond to some position j in the reference atlas, including the position $j = i$. During training, DEEPsc's goal is produce the highest likelihood when $j = i$ (meaning that $Inp_{ij} = [pos_i; pos_i]$), and assign low likelihood when $j \neq i$. DEEPsc also adds Gaussian noise to pos_i (the first N elements of Inp_{ij}), which aims to preserve robustness and avoid overfitting. The addition of noise can lead to DEEPsc learning a complex nonlinear mapping between the spatial positions in the reference atlas rather than a simple step-like function which activates when an exact match is inputted. During inference stage (*i.e.* after DEEPsc is trained), pos_i is replaces with the gene expression feature vector, which are the elements of $\tilde{E}$, and the goal is to predict the likelihood of the expression vector being originated from all possible positions j.

DEEPsc's network is a FFNN with two hidden layers, with each $h^{(1),(2)} \in \mathbb{R}^N$, mapping to an $y \in [0,1]$, where y can be viewed as a likelihood that the input cell originated from the input spatial position[169]. Given that for each training data Inp_{ij} there will be $n_{positions} - 1$ non-matches (labels of zero) and only one match for when $j = i$, the training labels will have many more zeros than ones. Therefore, Maseda *et al.* propose a non-traditional objective function which accounts for the imbalance between zeros and ones in the training data. This objective function is shown in Eq. (12)

$$\mathscr{L}(Y^{true}, Y^{pred}) = \sum_{i=1}^{p} \frac{(y_i^{true} - y_i^{pred})}{1.001 - y_i^{true}}, \tag{12}$$

where y_i^{pred} is the networks predicted outputs and y_i^{true} is the true target ($y_i^{true} = 1$ if it exactly matches, and $y_i^{true} = 0$ otherwise). This allows DEEPsc to avoid producing trivially zero outputs, which is important given the sparsity of the data. Maseda *et al.* also employ strategies in data splitting which helps to account for the inherent sparsity in the targets[61]. It is important to note that Maseda *et al.* also formulate a novel system-adaptive scoring scheme to evaluate the performance of DEEPsc using the spatial reference. However, the scoring scheme does not fall within the scope of this manuscript.

Maseda *et al.* apply DEEPsc for spatial imputation of four different biological systems (Zebrafish[59], Drosophila[170], Cortex[171] and Follicle[172]), achieving accuracy comparable to several existing models while having higher precision and robustness[61]. DEEPsc also shows better consistency across the different biological systems tested, which can be attributed to its system-adaptive design. In addition, the authors attribute the performance and generalizibility of DEEPsc to the use of FFNN (which have been noted before in other biological applications as well[45,47]) and the various strategies for robustness used during the training of DEEPsc. On the other hand, a weakness of DEEPsc is its training time, which depends nonlinearly on the number of locations available. However, this issue can be potentially mitigated by considering a small subset of possible locations, or a more optimized design when training the model.

B. Alignment

Alignment in ST analysis refers to the process of mapping scRNAseq data to a physical domain while aiming to match the geometry with the available spatial data. As previously stated, NGS-based technologies suffer from limited capture rates and significant dropout[173] (specially at higher resolutions). Before the use of DL in ST analysis, many computational approaches aimed to spatially reconstruct key marker genes scRNAseq data by assuming continuity in the gene space[60], or by leveraging local alignment information[59]. Moreover, most techniques for alignment or deconvolution of spatial data either learned a program dictionary[32] or estimated a probabilistic distribution of the data[64] for the cell-types at each spot. However, such approaches are not generalizable to

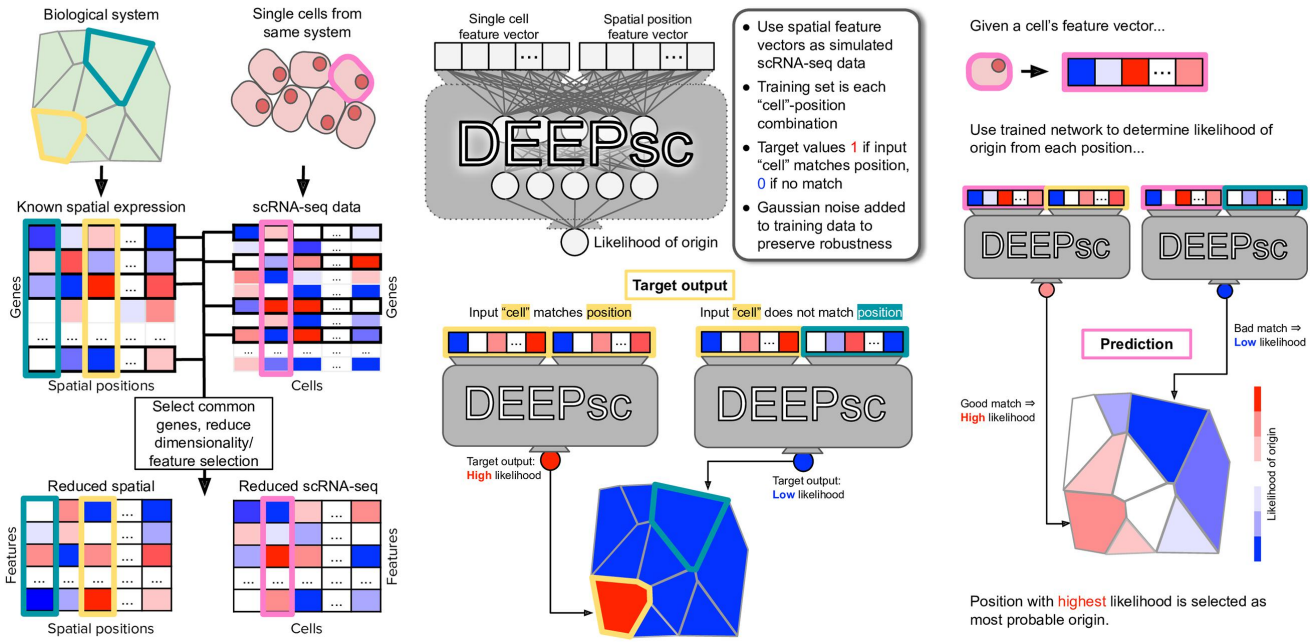

FIG. 4. **An overview of DEEPsc training and inference**. **(A)** Maseda *et al.* find the common genes in both spatial and scRNAseq data, and perform dimensionality reduction on each data modality (with the final matrices having the same number of features). **(B)** During the training, DEEPsc uses spatial expression to "simulate" single-cell gene expression vectors. More specifically, every feature vector from the spatial expression is concatenated with all other vectors (labeled as "non-match") and also itself (labeled as "match") to form the input data to the neural network. **(C)** During inference, the scRNAseq feature vectors are concatenated with all spatial feature vectors, where the model should place a high probability for locations where the gene expression could have originated from. This figure was obtained from Maseda *et al.*[61].

all experimental settings, since finding the mapping of sparse or sporadically distributed genes to the spots is difficult, and is error-prone due to dropouts[34].

DL frameworks have the potential of providing robust models that can adapt to the specific species or technologies, while being generalizable to other datasets and platforms. The potential application of DL in alignment of spatially-resolved transcriptomics came to fruition recently through the work of Biancalani *et al.* , called **Tangram**[34]. Tangram is a framework that, among many of its capabilities, can align scRNAseq or single-nucleus(sn) RNAseq profiles to spatial data; for the sake of simplicity, we refer to both data types as scRNAseq, although there are differences between the two methods (see reference[174] for a systematic comparison of scRNAseq and snRNAseq approaches). Tangram aims to: (i) learn the transcriptome-wide spatial gene expression map at a single-resolution, and (ii) relate the spatial information back to histological and anatomical data obtained from the same samples. Tangram's general workflow is to learn a mapping between the data modalities, and then to construct specific models for the downsteam tasks (such as deconvolution, correcting low-quality data, etcetera). We first summarize Tangram's alignment algorithm, and then provide the applications in which DL models are utilized.

Tangram's general objective is to learn a spatial matrix $S \in \mathbb{R}^{n_{cells} \times n_{genes}}$ describing the spatial alignments for the cells, with n_{cells}, n_{genes} denoting the number of single-cells and number of genes, respectively. Let the expression of gene k in cell i be denoted by $S_{ik} \in \mathbb{R}_{[0,\infty)}$, a non-negative value. Next, Tangram partitions ("voxelizes") the spatial volume at the finest possible resolution (depending on the spatial technology) as a one-dimensional array. This allows Tangram to construct (1) a matrix $G \in \mathbb{R}^{n_{voxels} \times n_{genes}}_{[0,\infty)}$ where G_{jk} is a non-negative value denoting the expression of gene k in voxel j, and (2) a cell-density vector $\mathbf{v} = \{v_1, v_2, \cdots, v_{n_{voxels}}\}$, where $0 \leq v_j \leq 1$ is the cell density in voxel j (with the total density for each voxel summing to 1).

The learning of transcriptome-wide spatial gene expression map at a single-resolution happens through learning a mapping operator $M \in \mathbb{R}^{n_{cells} \times n_{voxels}}_{[0,1]}$ where M_{ij} denotes the probability of cell i being in voxel j. Moreover, given any matrix $\tilde{M} \in \mathbb{R}^{n_{cells} \times n_{voxels}}$, each element of the operator M is assigned according Eq. (13)

$$M_{ij} = \frac{e^{\tilde{M}_{i,j}}}{\sum_{q=0}^{n_{voxels}} e^{\tilde{M}_{q,j}}}, \tag{13}$$

ensuring that $\sum_{j=1}^{n_{voxels}} M_{ij} = 1$, *i.e.* assigning a probability distribution along the voxels using the well-known $softmax(\cdot)$

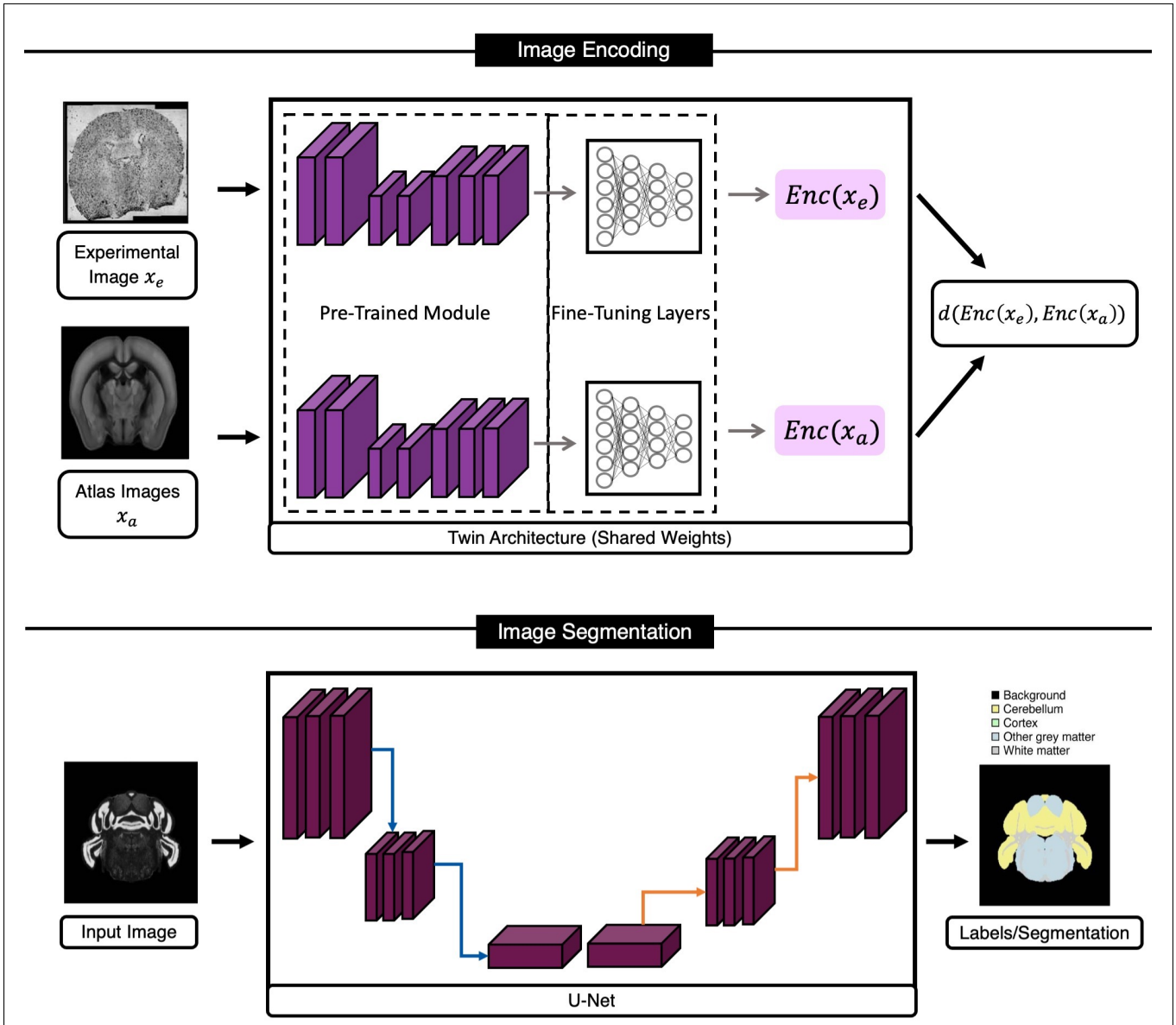

FIG. 5. **Tangram's DL Framework for Aligning/Integrating Histology and Anatomical Data with Molecular Data**. Tangram's model for this task is a combination of encoding (using a twin network) and segmentation modules. The twin network learns a similarity metric for brain sections based on anatomical features in images, while the U-Net model is trained to segment five different classes on mouse brain images. This figure was recreated for this manuscript using images from Biancalani *et al.*[34].

function. Biancalani *et al.* define an additional quantity, $M^T S$, which denotes the spatial gene expression as predicted by the operator M, and a vector $\mathbf{m} = \{m_1, \cdots, m_{n_{cells}}\}$ where $m_j = \sum_i^{n_{voxels}} \frac{M_{ij}}{n_{cells}}$ is the *predicted* cell density for each voxel j.

Given the preliminary quantities, we can now write Tangram's generic objective function as shown in Eq. (14)

$$\mathscr{L}(S,M) = \sum_{k=1}^{n_{genes}} \cos_{sim}\left((M^T S)_{[:],k}, G_{[:],k}\right), \tag{14}$$

where "[:]" denotes the matrix slicing and $\cos_{sim}$ is the cosine similarity, defined as Eq. (15)

$$\cos_{sim}(\mathbf{a},\mathbf{b}) \triangleq \frac{a \cdot b}{\|a\|\|b\|}. \tag{15}$$

The objective function, Eq. (14), learns a proportional mapping of the genes to the voxels. Additionally, this loss function can be further modified to incorporate prior knowledge. Indeed, Biancalani *et al.* modify this to regularize for the learned density distributions and the cells contained within each voxel, as shown in Eq. (16)

$$\mathscr{L}(S,M) = \mathbb{KL}(\mathbf{m},\mathbf{v}) - \sum_{k=1}^{n_{genes}} \cos_{sim}\left((M^T S)_{[:],k}, G_{[:],k}\right) - \sum_{j=1}^{n_{voxels}} \cos_{sim}\left((M^T S)_{j,[:]}, G_{j,[:]}\right), \tag{16}$$

where minimizing the divergence (first term) enforces that the learned density distribution and the expected distribution are similar, and the additional loss over the voxels (third term) penalize the model if predicted gene expression is not proportional to the expected gene expression. Biancalani *et al.* minimize the objective shown in Eq. (16) through gradient-based optimizers implemented in PyTorch[175,176]. After optimizing Eq. (16), Tangram is able to map all scRNAseq profiles onto the physical space, thus performing alignment. It worthy to note that although Tangram learns a linear operator M, this mapping could be replaced with a deep neural network as well.

Tangram utilizes DL to integrate anatomical and molecular features, specifically for mouse brain images. To do so, the authors use an image segmentation network (U-Net[177]) in combination with a "Twin" network[178] to produce segmentation masks of anatomical images, with both networks being CNNs. We present a general overview of these architectures in Fig. 5. The twin network uses DenseNet[179]: a deep NNs which concatenate the outputs at each layer to propagate salient information to deeper layers in the network (refer to section III D for the motivation behind such approaches). More specifically, Tangram uses a pre-trained DenseNet encoder (trained on ImageNet) to encode images and remove technical noise and artifacts. Biancalani *et al.* also add two additional layers to the pre-trained encoder, which map the outputs to a smaller latent space. The encoder of the twin network is fine-tuned on learning the prediction of spatial depth difference between two images: Two random images are inputted to the twin network with their spatial difference depth being the desired target output, d^{true}. The network then tries to predict the depth, d^{pred}, for all N inputs, ultimately comparing them against the corresponding true depth differences, d^{true}(as shown in Eq. (17))

$$MSE(d^{pred}, d^{true}) = \frac{1}{N}\sum_{i=1}^{N}\left(d_i^{pred} - d_i^{true}\right)^2. \tag{17}$$

The segmentation model of Tangram generates five custom segmentation masks (background, cortex, cerebellum, white matter, and other gray matter) which are compatible with existing Allen ontology atlas. The segmentation model is a U-Net, which uses a pre-trained ResNet50[155] as its core. Finally for each pixel in input images, the model's last layer (a softmax function) assign a probability of belonging to one of the five segmentation classes. Tangram's segmentation model aims to optimize the superposition of the cross entropy and Jaccard index, as presented in Eq. (18)

$$\mathscr{L}(g,p) = -g\cdot\log(p) - \frac{p\cap g}{p\cup g}, \tag{18}$$

with p denoting the model prediction and g referring to the ground truth image.

Biancalani *et al.* demonstrate that Tangram learns an accurate mapping between the spatial data and scRNAseq gene expression when applied to fine or coarse grained spatial atlases. The authors show that their approach works well across different technologies (namely ISH, smFISH, Visium , STARmap and MERFISH) at different resolutions and gene coverage, and is able to learn a robust and accurate alignment mapping for the isocortex of the adult healthy mouse brain[34]. While Tangram can offer different advantages based on the spatial technology, it can produce consistent spatial mappings and overcoming limitations in resolution or throughput, which is beneficial in many ST experiments and studies.

C. Spot Deconvolution

One downside of using NGS-based technologies remains to be their resolution: Despite the recent technological advancements, most ST platforms (*e.g. Spatial Transcriptomics*, Visium, DBiT-seq[180], Nanostring GeoMx[181] and SlideSeq) do not have a single-cell resolution. The number of cells captured in each spot still varies based on the tissues (about 1-10[182]) and the technology used. On the other hand, we can not assume that all cells within a spot are the same, due to the heterogeneity of the cells. Therefore, it is necessary to use computational approaches for inferring the cell types in each spot or voxel. Such estimations would be possible if there were a complementary scRNAseq dataset. The process of inferring the cellular composition of each spot is known as *cell-type deconvolution*. Deconvolution has been at the forefront of computational efforts and it is important in building oragan atlases[20,183,184]. In fact, cell-type deconvolution is an existing procedure for inferring cell-type composition in RNAseq data using scRNAseq. However, methods developed for bulk RNAseq do not account for the spatial components of ST datasets, and are therefore generally inadequate. Given that deconvolution is an existing practice in RNAseq studies, we will refer to spatial deconvolution problem as *spot deconvolution* to distinguish between the traditional methods and the ones developed for ST analysis.

We divide spot deconvolution methods into three categories: (i) Statistical methods, (ii) Machine Learning and (iii) Deep Learning, with many of the current models falling into the first two categories. We now dive deeper into the two existing models which use DL for performing spot deconvolution.

D. DestVI

DestVI (DEconvolution of Spatial Transcriptomics profiles using Variation Inference) is a Baysian model for spot deconvolution. DestVI employs a conditional deep generative model (similar to scVI[185], a popular model for scRNAseq analysis) to learn cell-type profiles and continuous sub-cell-type variations, aiming to recover the cell-type frequency and the average transcriptions state of cell-types at each spot. To

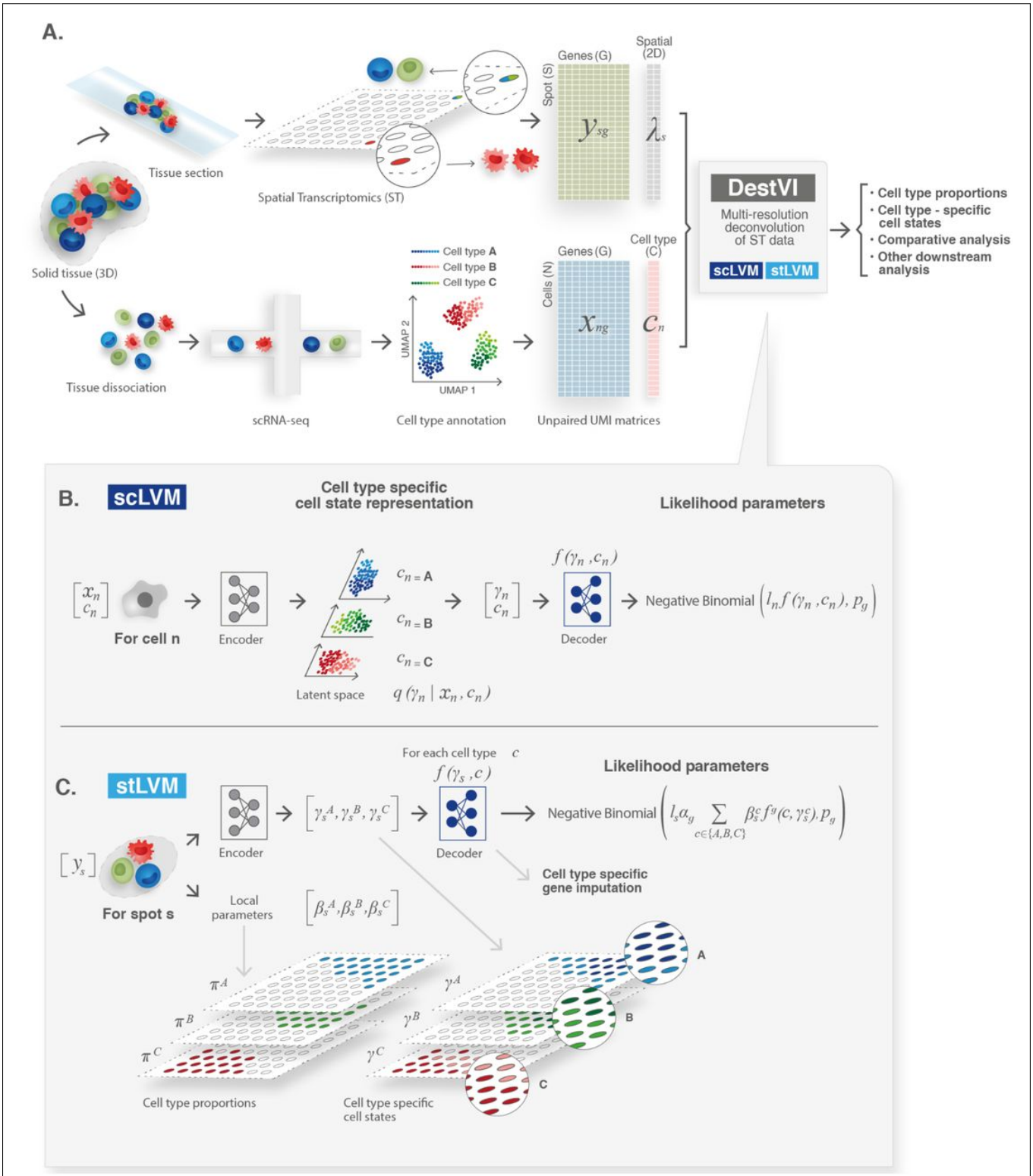

FIG. 6. **Visualization of DestVI's Computation Workflow for Spot Deconvolution.** DestVI uses information from both data modalities of ST data (shown in **A**). DestVI defines two latent variable models (LVMs) for each data modality: an LVM for modeling scRNAseq data (**B**) and one that aims to model the ST data (**C**). We describe each one in section IV C. This image is re-used from Lopez *et al.*[69]

do so, DestVI takes a pair of transcriptomics datasets as inputs: (i) a reference scRNA-seq data and (ii) a query spatial transcriptomics data (from the same samples). DestVI then outputs the expected proportion of cell types for every spot and a continuous estimation of cell-state for the cell types present in each spots, which can be viewed as the average state the cell-types in each spot, which Lopez *et al.* suggest as useful for downstream analysis and formulation of biological hypotheses[69].

DestVI uses two different latent variable models (LVMs) for distinguishing cell-type proportions and delineating cell-type-specific sub-states (shown in Fig. 6). The first LVM is for single-cell data (therefore named *scLVM*) which assumes the counts follow a negative binomial (NB) distribution, which has shown to model RNAseq count data well[185–187]. Specifically, Lopez *et al.* assume that for each gene g and cell n, the count of observed transcripts, x_{ng}, follows a NB distribution paramaterized with (r_{ng}, p_g): $r_{ng} = l_n \cdot f(\gamma_n, c_n; \boldsymbol{\theta})$ is a parameter which depends on the type assigned to the cell c_n, the total number of detected molecules l_n, and a low-dimensional latent vector γ_n (which lopez *et al.* set $\gamma_n = 5$) that describes the variability of cell-type assignment to cell c_n, and a neural network f parameters $\boldsymbol{\theta}$ (in this case, a two layer NN). The second parameter of the NB, p_g, is optimized using variational Bayesian inference. We can summarize the assumptions for scLVM as shown in Eq. (19):

$$x_{ng} \sim NB(l_n f(c_n, \gamma_n), p_g), \tag{19}$$

with the latent variable $\gamma_n \sim \mathcal{N}(0, I)$. Each c_n (the annotations) are represented by a one-hot encoded vector, which is concatenated with γ to serve as the input of the NN f. Lopez *et al.* use a VAE to optimize for the marginal conditional likelihood $\log p_\theta(x_n | l_n, c_n)$.

Finally for scLVM, a mean-field Gaussian distribution $q_\phi(\gamma | c_n, x_n)$, parametrized by another two-layer NN g, is inferred for each cell which quantifies the cell state and the associated uncertainty. The NN g takes a concatenation of (i) the gene expression vector x_n and (ii) the one-hot encoded cell annotations as its inputs. The network g outputs the mean and variance of the variational distribution for γ_n, obtained through optimizing Eq. (20)

$$\mathbb{E}_{q_\phi(\gamma_n | x_n, c_n)} \log p_\theta(x_n, \gamma_n | l_n, c_n) - \mathbb{KL}(q_\phi(\gamma_n | x_n, c_n) || p_\theta(\gamma_n)) \leq \log p_\theta(x_n | l_n, c_n),$$

where $p_\theta(\gamma_n)$ is the prior likelihood for γ_n. Similar to training any other VAE, the observations are split in mini-batches and sampling from the variational distribution is done using the reparameterization trick described in Kingma *et al.*[160]. The computational workflow for scLVM is visualized in Fig. 6(B).

The second LVM aims to model the spatial transcriptomics data (hence called *stLVM*) with the assumption that the number of observed transcripts x_{sg} at each spot s for each gene g follow a NB distribution. Additionally, Lopez *et al.* also assume that each spot has $C(s)$ cells, with each cell n in spot s being generated from the latent variables (c_{ns}, γ_{ns}). For stLVM's NB distribution, the rate parameter $r_{sg} = \alpha_g l_s f_g(c_{ns}, \gamma_{ns}; \theta_g)$, where α_g is a correction factor for the gene-specific bias between spatial and scRNAseq data, l_g is the overall number of molecules observed in each spot, and f_g is a NN network with parameters θ_g. These assumptions and quantities allow Lopez *et al.* formulate the total gene expression x_{sg} as shown in Eq. (21)

$$x_{st} \sim NB(l_s \alpha_g f_g(c_{ns}, \gamma_{ns}), p_g). \tag{21}$$

Moreover, using a parameter to designate the abundance of every cell type in every spot, β_{sc}, and NB's rate-shape parameterization property (see Aragón et al.[188]), Eq. (21) can be rewritten as in Eq. (22)

$$x_{st} \sim NB(l_s \alpha_g \sum_{n=1}^{C(s)} \beta_{sc} f_g(c_{ns}, \gamma_{ns}), p_g), \tag{22}$$

supposing that cells from a given cell type c in a spot s must come from the same covariate γ_s^c.

The covariate γ_s^c in DestVI allows for the model to account for ST technology discrepancies by assuming various empirical priors (refer to Fig. 6). Lopez *et al.* simplify the problem of identifying every cell type in each spot to determining the density cell-types, *i.e.* assuming that there cannot be significantly different cell states of the same cell types within a spot. Lopez *et al.* use a penalized likelihood method to infer point estimates for γ^c, α, and β. With the addition two strategies to stablize the training of DestVI, the final objective function for stLVM consists of (i) the negative binomial likelihood (ii) the likelihood of the empirical prior and (iii) the variance penalization for α.

Lopez *et al.* use simulations to present DestVI's ability to provide higher resolution compared to the existing methods and estimate gene expression by every cell type in all spots. Furthermore, they show that DestVI is able to accurately deconvolute spatial organization when applied mouse tumor model. In the cases tested, Lopez *et al.* demonstrate that DestVI is capable of identifying important cell-type-specific changes in gene expression between different tissue regions or between conditions, and that it can provide a high resolution and accurate spatial characterization of the cellular organization of tissues.

E. DSTG

Deconvoluting spatial transcriptomics data through graph-based convolutional network (DSTG)[65] is a recent semi-supervised model which employs graph convolutional networks (GCN)[134] for spot deconvolution. DSTG utilizes scRNAseq to construct a pseudo-ST data, and then building a link graph which represents the similarity between all spots in both real and pseudo ST data. The pseudo-ST is generated by combining scRNAseq transcriptomics of multiple cells to mimic the expression profiles at each spot; while the real ST data is unlabeled, the pseudo-ST has labels. To construct the link graph, DSTG first reduces the dimensionality of both real and pseudo data using canonical correlation analysis[189], and then identifies mutual nearest neighbors[190]. Next, a GCN is used on the link graph to propagate the real and pseudo ST data

into a latent space that is turned into a probability distribution of the cell compositions for each spot.

Song *et al.* form the link graph by taking the number of spots as the number of vertices, resulting in a graph $G = (V, E)$ where $|V|$ denotes the number of spots and E represents the edges between them. DSTG takes two inputs: (i) the adjacency matrix of graph G, represented by A, and (ii) a combination of both real and pseudo datasets $X = [x_{pseudo}; x_{real}] \in \mathbb{R}^{m \times N}$ where m is the number of variable genes, and $N = S_{pseudo} + S_{real}$ (the total number of spots in both datasets) with S_{pseudo} and S_{real} indicating the number of spots in the pseudo and real ST datasets, respectively. Next, Song *et al.* normalize the adjacency matrix (for efficient training of DSTG) using the diagonal degree of A, denoted by D, as shown in Eq. (23)

$$\hat{A} = D^{-\frac{1}{2}} A_I D^{-\frac{1}{2}}, \tag{23}$$

where $A_I = A + I$ (with I denoting the identity matrix). Given the two inputs, DSTG's graph convolution layers take the following form:

$$h^{(i+1)} = \begin{cases} \sigma(\hat{A} X^T W^{(i)}) & , i = 0 \\ \sigma(\hat{A} h^{(i)} W^{(i)}) & , i > 0 \end{cases} \tag{24}$$

with $h^{(i)}$ denoting the i-th hidden layer of DSTG, and $\sigma(\cdot)$ being a non-linear activation function (in this case $\sigma(\cdot) = ReLU(\cdot)$). The output of DSTG is denoted by $y_{s,t}$, the proportion of cell type $t = \{0, \cdots, T\}$ at each each spot $s = \{0, \cdots, N\}$. Song *et al.* design DSTG's architecture as shown in (25):

$$Y_{pred} = softmax(\hat{A}\sigma(\cdots(\hat{A}\sigma(\hat{A} X^T W^{(0)}) W^{(1)}) \cdots W^{(k)}), \tag{25}$$

where k is the last layer, and $Y^{pred} = [Y^{pred}_{pseudo}; Y^{pred}_{real}] \in \mathbb{R}^{N \times T}$ is the predicted proportions at each spot in the pseudo and real data, denoted by Y_{pseudo} and Y_{real}, respectively. It is important to note that Song *et al.* chose a GCN with three layers after performing an ablation study on the number of layers. Finally, DSTG is trained by optimizing the cross entropy loss:

$$\mathscr{L}(Y^{pred}_{pseudo}, Y^{true}_{pseudo}) = -\sum_{s=0}^{S_{pseudo}} \sum_{t=1}^{T} y^{true}_{s,t} \log(y^{pred}_{s,t}), \tag{26}$$

with $y^{\{true,pred\}}_{s,t}$ denoting the label for true/predicted cell type t at spot s. Note that this constitutes a semi-supervised training for DSTG, since only labels for the pseudo ST are used in training, but the model will also learn to predict labels for the real dataset as well (refer to Eq. (25)).

Song *et al.* note that, compared to traditional approaches, DSTG provides three key advantages: (i) Given that DSTG uses variable genes and a non-linear GCN, it allows for learning complex deconvolution mappings from ST data. (ii) The weights assigned to the different cell types in the pseudo-ST and the semi-supervised scheme allow DSTG identify key features which allow the model to learn the cellular composition in real data. (iii) DSTG's scalability and adaptability will be beneficial in ST analysis, given that the sequence depth of ST data is expected to increase. Song *et al.* show that DSTG consistently outperforms the benchmarked state-of-the-art model (SPOTlight) on both synthetic data and real data. More specifically, DSTG is evaluated on simulated data generated from PBMC where it shows high accuracy between the predicted cell compositions and the true proportions. Song *et al.* also find that DSTG's deconvolution of ST data from complex tissues including mouse cortex, hippocampus, and human pancreatic tumor slices is consistent with the underlying cellular mixtures[65].

F. Spatial Clustering

Clustering allows the aggregation of data into subpopulations based on some shared metric of distance or "closeness". In RNA sequencing studies, clustering is the first step of identifying cell clusters, often followed by laborious manual annotation (*e.g.* through identifying differentially expressed genes) or some automated workflows[191]. Clustering has been a crucial step in many scRNAseq studies, often performed using graph-based community detection algorithms (such as Louvain[192] or Leiden[193]) or more traditional methods (such as K-Means[118]). Although the scRNAseq techniques can be used in some ST studies (*e.g.* for multiplexed FISH data where single-cell resolution is available), the result may be discontiguous or erroneous since the spatial coordinates have not been taken into account[74]. Therefore, there is a need for ST-specific methods that can utilize both gene expression and histology data to produce clusters that are coherent, both in gene expression and physical space.

Recently, new frameworks for spatial clustering of ST data have emerged which utilize both spatial and expression information available. Zhu *et al.*[71] introduced a Hidden-Markov Random Field (HMRF)-based method to model the spatial dependency of gene expression using both the sequencing and imaging-based single-cell transcriptomic profiling technologies. HMRF is a graph-based model used to model the spatial distribution of signals. Using the ST data, Zhu *et al.* create a grid where neighboring nodes are connected to each other. However, the spatial pattern can not be observed directly (since it is "hidden"), and it must be inferred through observations that depend on the hidden states probabilisticly. Similar to Zhu *et al.* , BayesSpace[73] employs a Bayesian formulation of HMRF, and uses the Markov chain Monte Carlo (MCMC) algorithm to estimate the model parameters. Despite the ability of these methods to cluster voxels (or cells) into distinct subpopulations, these approaches suffer from the lack of versatility required to handle different modalities present in ST data[74].

With the emergence of newer technologies, the scale and variability within datasets are increasing, requiring more general and flexible models for accurate and robust analysis of these studies. A few ML-based approaches have been proposed to combat some of these mentioned challenges. Below, we review the ML-based methods which offer scalability and are generally more applicable to various experimental settings.

G. SpaCell

SpaCell[72] is a double-stream DL framework which utilizes both histology images and the associated spot gene counts. For the histology data, Tan *et al.* first preprocess the images (removing low-quality images, stain normalization, normalizing the pixels using a z-transform and removing background noise). Next, they split each histology image into tiles that contain one spot each (sub-images of 299 × 299 pixels). For the preprocessing of the count matrix (which contains the reads at each spot), Tan *et al.* follow traditional scRNAseq preprocessing workflows, including count normalization, removal of outlier genes and cells with too few genes. After the preprocessing stage, each tile (containing the image of a spot) corresponds to a column in the count matrix (reads from the same spot). At this point, each image $X_i \in \mathbb{R}^{299\times299}$ and the count matrix M will be in a $\mathbb{R}^{n_{spots}\times n_{genes}}$ space. However, Tan *et al.* reduce the count matrix to only contain 2048 most variable genes at each spot, therefore resulting in a new count matrix $\hat{M} \in \mathbb{R}^{n_{spots}\times 2048}$. Let us denote the i^{th} spot of $\hat{M}$ as $\hat{m}_i \in \mathbb{R}^{2048}$, which has a corresponding image x_i.

In order to spatially cluster cells of the same type, both image and count data must be used. The first step in SpaCell is to pass on the spot images, x_i, to a pre-trained ResNet50 (trained on ImageNet data) in order to output feature vectors, $\hat{x}_i \in \mathbb{R}^{2048}$ (each having the same dimensionality as columns of $\hat{M}$). Next, to extract features from both modalities, SpaCell uses two separate AEs for the image feature vectors, $\hat{X} \in \mathbb{R}^{n_{spots}\times 2048}$, and the most-variable-genes counts, $\hat{M}$, with both AEs having the same latent dimension (we discuss the reason behind this later). Let us denote the AE for images as $AE_I(\cdot) = Dec_I(Enc_I(\cdot))$, and the gene counts AE as $AE_G(\cdot) = Dec_G(Enc_G(\cdot))$, with $Enc_{\{I,G\}}(\cdot)$ and $Dec_{\{I,G\}}(\cdot)$ indicating the encoder and decoders, respectively.

Given N spots, each AE in spaCell aims to minimize three objective functions for their respective inputs [*i.e.* $\hat{x}_i$ for $AE_I(\cdot)$ and $\hat{m}_i$ for $AE_G(\cdot)$]: (i) the mean squared error (MSE) between the input and output (shown in Eq. (27)), (ii) the KL divergence between the probability distributions for input and constructed output of all N spots (denoted by p and q in Eq. (28) respectively) and (iii) Binary Cross Entropy (BCE), shown in Eq. (29):

$$MSE_{\{I,G\}}(v_i,\tilde{v}_i) = \frac{1}{N}\sum_{i=1}^{N}(v_i - AE_{\{I,G\}}(v_i))^2 \tag{27}$$

$$\mathbb{KL}(p||q) = \sum_{i=1}^{N} p(v_i)\frac{\log p(v_i)}{\log q(v_i)} \tag{28}$$

$$BCE(q) = -\frac{1}{N}\sum_{i=1}^{N}\left[v_i\log(p) + (1-v_i)\log(1-p)\right]. \tag{29}$$

Once training has concluded, spaCell encodes both images and gene counts, *i.e.* $Enc_I(\hat{x}_i)$ and $Enc_G(\hat{m}_i)$, to be used for clustering. More specifically, clustering is performed on a matrix that is the concatenation of the latent vectors produced by each AE, $C = [Enc_I(\hat{x}_i); Enc_G(\hat{m}_i)]$. This is why the latent spaces of $AE_{I,G}(\cdot)$ have the same dimension. After obtaining the concatenated matrix, the downstream clustering is performed using K-Means (which can be substituted for other algorithms as well). Through this procedure, spaCell uses both data modalities and can produce clusters that are highly accurate when compared to the true clusters (annotated by pathologists).

H. SpaGCN

SpaGCN[74] is a graph convolution network (GCN) that integrates both spatial information and histology images to perform spatial clustering. Using each spot as vertices, Hu *et al.* create a weighted undirected graph, $G = (V,E)$, where $|V|$ is the total number of spots and E is the set of edges with prescribed weights representing the similarity between the nodes. The weight of each of these edges is determined by (i) the distance between the two spots (nodes) that the edge connects, and (ii) the associated histology information (in this case, pixel intensity). This means that two spots are deemed similar if they are physically close to one another *and* they seem similar in the histology image.

In order to attribute the pixel information to each spot, Hu *et al.* use mean RGB pixel intensity of each spot within a window of size 50 × 50 pixels. That is, given a spot s with physical coordinates (x_s, y_s) and pixel coordinates (x_{ps}, y_{ps}), SpaGCN calculates the mean and variance of all the pixels present within a 50 × 50 pixels centered at (x_{ps}, y_{ps}). Let ps_r, ps_g, ps_b denote the means, and $var_r(ps)$, $var_g(ps)$, $var_b(ps)$ refer to the variance for the red, green and blue channels, respectively. SpaGAN then summarizes the pixel mean and variance information as a unified value, as shown in Eq. (30)

$$z_s = \frac{(ps_r \cdot var_r(ps)) + (ps_g \cdot var_g(ps)) + (ps_b \cdot var_b(ps))}{var_r(ps) + var_g(ps) + var_b(ps)}. \tag{30}$$

Furthermore, z_s is rescaled using the mean and standard deviation of each coordinate (including the newly-created z axis), with an additional scaling factor which can put more emphasis on histology data when needed. Let μ_z denote the mean of z_s, and let $\sigma_{x,y,z}$ be the standard deviation of x_s, y_s, z_s with $s \in V$, then we can formulate the rescaling as the following:

$$\tilde{z}_s = \alpha\frac{(z_s - \mu_z)(\max\{\sigma_x, \sigma_y\})}{\sigma_z}, \tag{31}$$

where α denotes the scaling factor described previously ($\alpha = 1$ by default).

Using the rescaled value in Eq. (31), the weight of each edge between two vertices s and k is calculated as shown in Eq. (32)

$$w(s,k) = e^{-d(s,k)^2/(2l^2)} \tag{32}$$

where l denotes the characteristic length scale and $d(s,k)$ is the traditional Euclidean distance, as shown in (33)

$$d(s,k) = \left((x_s - x_k)^2 + (y_s - y_k)^2 + (\tilde{z}_s - \tilde{z}_k)^2\right)^{\frac{1}{2}}. \tag{33}$$

SpaGCN's network construction (and backpropogation) is similar to other GCNs, inspired by Kipf *et al.*[134] (for an overview of GCNs, refer to section IV E). The network intakes the adjacency matrix A to represent the graph G, and a reduced-dimension representation of the gene expression matrix, which Hu *et al.* achieve using PCA with 50 principal components. The outputs of the GCN network is matrix which includes combined information on histology, gene expression, spatial position. SpaGCN then uses the output of the GCN to perform unsupervised clustering of the spatial data.

SpaGCN uses the Louvain algorithm (an iterative unsupervised clustering algorith) on the output of GCN to initialize cluster centroids, with the number of clusters (controlled by Louvain's *resolution* parameter) being optimized on maximizing the Silhouette score[194]). The iterative updates are based on optimizing a metric that defines the distance between each spot and all cluster centroids using the t-distribution as a kernel. For a centroid c_j, a total of N clusters, and the embedded point h_i for spot i, this metric can be defined as the probability of assigning cell i to cluster j, as shown in Eq. (34)

$$q_{ij} = \frac{(1+h_i-\mu_j^2)^{-1}}{\sum_{c=1}^{N}(1+h_i-\mu_c^2)^{-1}}. \tag{34}$$

Hu *et al.* further refine the clusters using an auxiliary target distribution (shown in Eq. (35)) which prefers spots assignments with the highest confidence, and normalizes the centroid contribution to the overall loss function as the following:

$$p_{ij} = \frac{q_{ij}^2}{\sum_i^S q_{ij}} \cdot \left(\sum_{c=1}^{N} \left(\frac{q_{ic}^2}{\sum_i^S q_{ic}} \right) \right)^{-1}. \tag{35}$$

Lastly, spaGCN is trained by optimizing the $\mathbb{KL}$ divergence between the p and q distributions, as shown in Eq. (36)

$$\mathscr{L} = \mathbb{KL}(P||Q) = \sum_{i=1}^{S}\sum_{j=0}^{N} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \tag{36}$$

Hu *et al.* demonstrate that SpaGCN can accurately identify spatial clusters that are consistent with manual annotations, since SpaGCN utilizes information from both gene expression and histology. The authors perform spatial clustering with SpaGCN on human dorsolateral prefrontal cortex, and human primary pancreatic cancer and multiple mouse tissue data, showing that SpaGCN performs consistently well, outperforming other state-of-the-art models (stLearn, BayesSpace, and Louvain). These results show the feasibility and potential of SpaGCN for clustering spatial-resolved transcriptomics.

I. Cell-Cell Interactions

Multicellular organisms depend on intricate cell–cell interactions (CCIs) which dictate cellular development, homeostasis, and single-cell functions[195]. Unravelling such interaction within tissues can present unique insights on complex biological processes and disease pathogenesis[195–198]. CCI has been investigated using both scRNAseq and RNAseq, wherein most approaches test for enrichment in ligand-receptor profiles in the expression data[199–201]. However, ST data can offer a more comprehensive view of CCI, since the distance traveled by ligand signal is crucial in determining the type of cell–cell signaling[182]. Given the importance of CCI and the advantages that ST data provides, several computational approaches for inferring cellular interactions using ST data have been developed, such as SpaOTsc[78], Giotto[81], MISTy[80].

SpaOTsc is a model that can be used in integrating scRNAseq data with spatial measurements, and in inferring cellular interactions in spatial-resolved transcriptomics data. SpaOTsc aims to estimate cellular interactions by analyzing the relationships between ligand-receptor pairs and their downstream genes. SpaOTsc formulates a spatial metric using the optimal transport algorithm, returning a mapping that contains the probability distribution of each scRNA-seq cell over a spatial region. SpaOTsc also utilizes a random forest in order to infer the spatial range of ligand-receptor signaling and subsequently removing the long-distance connections. Another approach is Giotto[81]: Giotto is an extended and comprehensive toolbox designed for ST analysis and visualization, which includes a CCI model which calculates an enrichment score (the weighted mean expression of a ligand and the corresponding receptor in the two neighboring cells). Giotto then constructs an empirical null distribution by moving the locations for each cell-type, subsequently calculating corresponding statistical significance (P-value), and ordering the ligand-receptors pair-wise for all neighboring cells.

Although the mentioned models have shown to discover simple cellular interactions, such approaches often fail to detect complex gene-gene interactions, which is essential in understanding many diseases. DL models learn such complicated interactions from raw data, further utilizing ST data in studying CCI. For this purpose, **StLearn**[79] is a recent DL model that, among many of its capabilities, can learn CCI from spatially-resolved transcriptomics. StLearn's DL components lies within its Spatial Morphological gene Expression (SME) normalization. The SME normalization aims to combine critical information from Hematoxylin and Eosin stained (H&E) tissue images and transcriptome-wide gene expression profile to then take advantage of in downstream analysis, such as clustering, spatial trajectory inference, and CCI.

The SME normalization procedure includes *(i) spatial location:* In order to use the spatial positions for selecting neighboring spot pairs, Pham *et al.* consider two spots s_i and s_j as neighbors if the center-to-center euclidean distance between two spots,$|C(s_i)-C(s_j)|$, is less than a specified distance r, *i.e.* $|C(s_i)-C(s_j)|< r$. Pham *et al.* include all paired spots s_i and s_j as input to adjust for the gene expression of the center spot s_i. The next step in SME normalization is *(ii) Morphological similarity:* stLearn calculates the morphological similarity between spots using feature vectors produces by an ImageNet-pre-trained ResNet50. More specifically, all H&E images corresponding to each spot s_i is inputted to the ResNet50 model, which then produces a feature vector $x_i \in \mathbb{R}^{2048}$. Subsequently, stLearn performs PCA on each feature vector x_i, resulting in reduced-dimension feature vectors

$\hat{x}_i \in \mathbb{R}^{50}$. To calculate the morphological distance (MD) between two neighboring spots s_i and s_j (according to criterion defined in *(i)*), Pham *et al.* measure the cosine similarity between two reduced feature vectors (refer to Eq. (15) for definition of Cosine Similarity); this MD is shown in Eq. (37),

$$MD(s_i, s_j) \triangleq \cos_{sim}(\hat{x}_i, \hat{x}_j), \tag{37}$$

As a last step in SME normalization, the gene expression at each spot s_i is normalized using the MD distance, as shown in Eq.(38)

$$\hat{GE}_i = GE_i + \frac{\sum_{j=1}^{n} (GE_j \cdot MD(s_i, s_j))}{n}, \tag{38}$$

where GE_i denotes raw gene expression counts at spot s_i, and n is the total number of neighbors identified for spot s_i.

After SME normalization, stLearn can perform multiple downstream tasks, including the identification of tissue regions with high CCI activities[79]. StLearn's CCI algorithm finds ligand–receptor (L-R) co-expression between neighboring spots, and tests for the enrichment of L-R pairs between two cell types, which are compared to a random null distribution using CellPhoneDB[200]. After this initial testing, significant L-R pairs are selected to calculate cell-cell interactions. These interactions are measured using the nearest neighbors for a spot, and are queried to the cKDTree algorithm[202] to validate that the neighboring cells express ligand or receptor genes that are above a pre-defined threshold. Next, Pham *et al.* form a matrix where significant L-R pairs represent the features (columns) for each spot coordinates (the rows). Using this matrix, stLearn can cluster the spatial regions with the most similar L-R co-expression values, which combined with the CCI measures, can identify tissue regions that have high L-R co-expression, indicating areas that have a high likelihood of active CCI. This approach consitutes stLearn as one of the the first methods which combines both spatial cell populations (identified through clustering) and L-R interactions to detect tissue region with a high likelihood of CCI. Pham *et al.* apply stLearn's CCI method to breast cancer tissue and identify spatial regions and L-R pairs in cancer-immune cell interactions, indicating a great potentials for shedding light on CCI using ST data.

V. CONCLUSIONS AND OUTLOOK

The ST field is rapidly growing, with new datasets and analysis pipelines released weekly. The innovations in biological methods will continue to spur the creativity in algorithm development, with an emphasis on ML-based frameworks. Although the space of DL models for ST analysis is currently small, we anticipate the field to experience a paradigm shift towards deep-learned models. In this review, our goal was to provide readers with the necessary biological, mathematical, and computational background for understanding the existing approaches, and expanding upon the current models to address the challenges posed by the ST domain.

In this manuscript, we provided an overview of current DL-based techniques for alignment and integration of ST data, spatial clustering, spot deconvolution, inferring cell-cell communication, and approaches for reconstructing spatial coordinates using scRNAseq data (with limited or no spatial reference atlas). The DL methods we presented, in comparison to their conventional counterparts, offer accuracy and scalability advantages. However, DL methods are not always the preferred choice as they are computationally expensive and may lack biological interpretability. As more methods for ST analysis are developed, we believe that standard datasets for benchmarking new models as well as comprehensive accuracy and efficiency analysis of existing techniques will be of significant value to the field. Though the existing methods set the new state-of-the-art in their respective categories, room for improvements remains large. Among the ST downstream analyses, applications of DL algorithms for studying cell-cell communication and identification of spatially-variable genes remain mostly underexplored. Given DL models' ability to extract sophisticated patterns from raw data, we anticipate that DL approaches will prove useful in unraveling complex biological processes, aiding the efforts in identifying cellular interactions and highly variable genes in a spatial context.

Recent technological advancements have enabled researchers to utilize various single-cell omics sources to construct multi-omics datasets, providing comprehensive view of many diseases (e.g. COVID19[203,204] and cancer[205]), and developmental processes[206,207]. As the single-cell analysis enters the multi-omics age, the need for integrating ST data with other single-cell sources will increase. Therefore, we expect an increase in ML-based frameworks for data integration and alignment, spearheaded by DL-based approaches. Additionally, due to the noise and multi-modality of ST data, there exists an unmet need for methods that account for batch effects in spatial and gene expression data. Given the success of DL techniques for batch effect removal in scRNAseq, we foresee DL models being widely used for batch effect correction of spatially-resolved transcriptomics data.

Despite the recentness of ST technologies, researchers have successfully used these technologies to generate spatially-resolved cell atlases, providing new insights on a wide range of biological processes and organs[208–212]. Such studies show the tremendous potential that ST technologies hold, but also highlight the need for scalable and efficient analyses tools. The application of DL to ST analysis remains a rapidly evolving nascent domain, demonstrating promising great prospects in advancing the field of ST, and the integration of ST datasets with other omics data.

ACKNOWLEDGMENTS

We wish to acknowledge Oscar Davalos for offering fruitful discussions and providing useful insights on earlier versions of this work. We also would like to thank Tommaso Buvoli, Maia Powell, and the anonymous reviewers for providing valuable feedback on earlier drafts of this manuscript. The authors received support from the National Institutes of Health (R15-HL146779 and R01-GM126548) and the National Science Foundation (DMS-1840265).

GRAPHICS ACKNOWLEDGMENTS

The Visium slide and its visualizations in Fig. 1 and 3 were accessed from 10x Genomics[213]. The mouse brain histology image in Fig. 1 was taken from reference[214]. The illustration of mouse brain in Fig. 2 was obtained from BioRender[215].

AUTHOR CONTRIBUTIONS

A.A.H. and S.S.S. wrote the manuscript, edited the drafts and conceptualized the figures, which A.A.H then created.

REFERENCES

[1] L. Larsson, J. Frisén, and J. Lundeberg, "Spatially resolved transcriptomics adds a new dimension to genomics," Nature Methods **18**, 15–18 (2021).

[2] J. S. Packer, Q. Zhu, C. Huynh, P. Sivaramakrishnan, E. Preston, H. Dueck, D. Stefanik, K. Tan, C. Trapnell, J. Kim, R. H. Waterston, and J. I. Murray, "A lineage-resolved molecular atlas of c. elegans embryogenesis at single-cell resolution," Science **365**, eaax1971 (2019), https://www.science.org/doi/pdf/10.1126/science.aax1971.

[3] X. Han, R. Wang, Y. Zhou, L. Fei, H. Sun, S. Lai, A. Saadatpour, Z. Zhou, H. Chen, F. Ye, D. Huang, Y. Xu, W. Huang, M. Jiang, X. Jiang, J. Mao, Y. Chen, C. Lu, J. Xie, Q. Fang, Y. Wang, R. Yue, T. Li, H. Huang, S. H. Orkin, G.-C. Yuan, M. Chen, and G. Guo, "Mapping the mouse cell atlas by microwell-seq," Cell **172**, 1091–1107.e17 (2018).

[4] N. Schaum, J. Karkanias, N. F. Neff, A. P. May, S. R. Quake, T. Wyss-Coray, S. Darmanis, J. Batson, O. Botvinnik, M. B. Chen, S. Chen, F. Green, R. C. Jones, A. Maynard, L. Penland, A. O. Pisco, R. V. Sit, G. M. Stanley, J. T. Webber, F. Zanini, A. S. Baghel, I. Bakerman, I. Bansal, D. Berdnik, B. Bilen, D. Brownfield, C. Cain, M. B. Chen, M. Cho, G. Cirolia, S. D. Conley, A. Demers, K. Demir, A. de Morree, T. Divita, H. du Bois, L. B. T. Dulgeroff, H. Ebadi, F. H. Espinoza, M. Fish, Q. Gan, B. M. George, A. Gillich, G. Genetiano, X. Gu, G. S. Gulati, Y. Hang, S. Hosseinzadeh, A. Huang, T. Iram, T. Isobe, F. Ives, R. C. Jones, K. S. Kao, G. Karnam, A. M. Kershner, B. M. Kiss, W. Kong, M. E. Kumar, J. Y. Lam, D. P. Lee, S. E. Lee, G. Li, Q. Li, L. Liu, A. Lo, W.-J. Lu, A. Manjunath, A. P. May, K. L. May, O. L. May, M. McKay, R. J. Metzger, M. Mignardi, D. Min, A. N. Nabhan, N. F. Neff, K. M. Ng, J. Noh, R. Patkar, W. C. Peng, R. Puccinelli, E. J. Rulifson, S. S. Sikandar, R. Sinha, R. V. Sit, K. Szade, W. Tan, C. Tato, K. Tellez, K. J. Travaglini, C. Tropini, L. Waldburger, L. J. van Weele, M. N. Wosczyna, J. Xiang, S. Xue, J. Youngyunpipatkul, M. E. Zardeneta, F. Zhang, L. Zhou, A. P. May, N. F. Neff, R. V. Sit, P. Castro, D. Croote, J. L. DeRisi, G. M. Stanley, J. T. Webber, A. S. Baghel, M. B. Chen, F. H. Espinoza, B. M. George, G. S. Gulati, A. M. Kershner, B. M. Kiss, C. S. Kuo, J. Y. Lam, B. Lehallier, A. N. Nabhan, K. M. Ng, P. K. Nguyen, E. J. Rulifson, S. S. Sikandar, S. Y. Tan, K. J. Travaglini, L. J. van Weele, B. M. Wang, M. N. Wosczyna, H. Yousef, A. P. May, S. R. Quake, G. M. Stanley, J. T. Webber, P. A. Beachy, C. K. F. Chan, B. M. George, G. S. Gulati, K. C. Huang, A. M. Kershner, B. M. Kiss, A. N. Nabhan, K. M. Ng, P. K. Nguyen, E. J. Rulifson, S. S. Sikandar, K. J. Travaglini, B. M. Wang, K. Weinberg, M. N. Wosczyna, S. M. Wu, B. A. Barres, P. A. Beachy, C. K. F. Chan, M. F. Clarke, S. K. Kim, M. A. Krasnow, M. E. Kumar, C. S. Kuo, A. P. May, R. J. Metzger, N. F. Neff, R. Nusse, P. K. Nguyen, T. A. Rando, J. Sonnenburg, B. M. Wang, I. L. Weissman, S. M. Wu, S. R. Quake, T. T. M. Consortium, O. coordination, L. coordination, O. collection, processing, L. preparation, sequencing, C. data analysis, C. type annotation, W. group, S. text writing group, and P. investigators, "Single-cell transcriptomics of 20 mouse organs creates a tabula muris," Nature **562**, 367–372 (2018).

[5] A. Regev, S. A. Teichmann, E. S. Lander, I. Amit, C. Benoist, E. Birney, B. Bodenmiller, P. Campbell, P. Carninci, M. Clatworthy, H. Clevers, B. Deplancke, I. Dunham, J. Eberwine, R. Eils, W. Enard, A. Farmer, L. Fugger, B. Göttgens, N. Hacohen, M. Haniffa, M. Hemberg, S. Kim, P. Klenerman, A. Kriegstein, E. Lein, S. Linnarsson, E. Lundberg, J. Lundeberg, P. Majumder, J. C. Marioni, M. Merad, M. Mhlanga, M. Nawijn, M. Netea, G. Nolan, D. Pe'er, A. Phillipakis, C. P. Ponting, S. Quake, W. Reik, O. Rozenblatt-Rosen, J. Sanes, R. Satija, T. N. Schumacher, A. Shalek, E. Shapiro, P. Sharma, J. W. Shin, O. Stegle, M. Stratton, M. J. T. Stubbington, F. J. Theis, M. Uhlen, A. van Oudenaarden, A. Wagner, F. Watt, J. Weissman, B. Wold, R. Xavier, and N. Yosef, eng"The human cell atlas." Elife **6** (2017), 10.7554/eLife.27041.

[6] K. Davie, J. Janssens, D. Koldere, M. De Waegeneer, U. Pech, Ł. Kreft, S. Aibar, S. Makhzami, V. Christiaens, C. Bravo González-Blas, S. Poovathingal, G. Hulselmans, K. I. Spanier, T. Moerman, B. Vanspauwen, S. Geurs, T. Voet, J. Lammertyn, B. Thienpont, S. Liu, N. Konstantinides, M. Fiers, P. Verstreken, and S. Aerts, eng"A single-cell transcriptome atlas of the aging drosophila brain." Cell **174**, 982–998 (2018).

[7] Y. Zhang, D. Wang, M. Peng, L. Tang, J. Ouyang, F. Xiong, C. Guo, Y. Tang, Y. Zhou, Q. Liao, X. Wu, H. Wang, J. Yu, Y. Li, X. Li, G. Li, Z. Zeng, Y. Tan, and W. Xiong, "Single-cell RNA sequencing in cancer research," Journal of Experimental & Clinical Cancer Research **40**, 81 (2021).

[8] A. Derakhshani, Z. Asadzadeh, H. Safarpour, P. Leone, M. A. Shadbad, A. Heydari, B. Baradaran, and V. Racanelli, "Regulation of ctla-4 and pd-l1 expression in relapsing-remitting multiple sclerosis patients after treatment with fingolimod, ifn-1, glatiramer acetate, and dimethyl fumarate drugs," Journal of Personalized Medicine **11** (2021), 10.3390/jpm11080721.

[9] K. Anthony, R.-S. Ciro, F. Vicente, C. S. J., M. B. J., S. Hayley, P. Emanuela, S. Grégory, A. Ferhat, V. Pandurangan, and O. C. H., "Severely ill patients with covid-19 display impaired exhaustion features in sars-cov-2–reactive cd8+ t cells," Science Immunology **6**, eabe4782 (2021).

[10] N. Erfanian, A. A. Heydari, P. Iañez, A. Derakhshani, M. Ghasemigol, M. Farahpour, S. Nasseri, H. Safarpour, and A. Sahebkar, "Deep learning applications in single-cell omics data analysis," bioRxiv (2022), 10.1101/2021.11.26.470166, https://www.biorxiv.org/content/early/2022/01/20/2021.11.26.470166.full.pdf.

[11] S. R. Park, S. Namkoong, L. Friesen, C.-S. Cho, Z. Z. Zhang, Y.-C. Chen, E. Yoon, C. H. Kim, H. Kwak, H. M. Kang, and J. H. Lee, "Single-cell transcriptome analysis of colon cancer cell response to 5-fluorouracil-induced dna damage," Cell Reports **32**, 108077 (2020).

[12] Y. Su, D. Chen, D. Yuan, C. Lausted, J. Choi, C. L. Dai, V. Voillet, V. R. Duvvuri, K. Scherler, P. Troisch, P. Baloni, G. Qin, B. Smith, S. A. Kornilov, C. Rostomily, A. Xu, J. Li, S. Dong, A. Rothchild, J. Zhou, K. Murray, R. Edmark, S. Hong, J. E. Heath, J. Earls, R. Zhang, J. Xie, S. Li, R. Roper, L. Jones, Y. Zhou, L. Rowen, R. Liu, S. Mackay, D. S. O'Mahony, C. R. Dale, J. A. Wallick, H. A. Algren, M. A. Zager, W. Wei, N. D. Price, S. Huang, N. Subramanian, K. Wang, A. T. Magis, J. J. Hadlock, L. Hood, A. Aderem, J. A. Bluestone, L. L. Lanier, P. D. Greenberg, R. Gottardo, M. M. Davis, J. D. Goldman, and J. R. Heath, "Multi-omics resolves a sharp disease-state shift between mild and moderate covid-19," Cell **183**, 1479–1495.e20 (2020).

[13] F. Iqbal, A. Lupieri, M. Aikawa, and E. Aikawa, "Harnessing single-cell RNA sequencing to better understand how diseased cells behave the way they do in cardiovascular disease," Arteriosclerosis, Thrombosis, and Vascular Biology **41**, 585–600 (2021).

[14] M. Heming, X. Li, S. Räuber, A. K. Mausberg, A.-L. Börsch, M. Hartlehnert, A. Singhal, I.-N. Lu, M. Fleischer, F. Szepanowski, O. Witzke, T. Brenner, U. Dittmer, N. Yosef, C. Kleinschnitz, H. Wiendl, M. Stettner, and G. Meyer Zu Hörste, "Neurological manifestations of covid-19 feature t cell exhaustion and dedifferentiated monocytes in cerebrospinal fluid," Immunity **54**, 164–175.e6 (2021).

[15] A. A. Pollen, T. J. Nowakowski, J. Shuga, X. Wang, A. A. Leyrat, J. H. Lui, N. Li, L. Szpankowski, B. Fowler, P. Chen, N. Ramalingam, G. Sun, M. Thu, M. Norris, R. Lebofsky, D. Toppani, D. W. Kemp, M. Wong, B. Clerkson, B. N. Jones, S. Wu, L. Knutsson, B. Alvarado, J. Wang, L. S. Weaver, A. P. May, R. C. Jones, M. A. Unger, A. R. Kriegstein, and J. A. A. West, "Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex," Nature Biotechnology **32**, 1053–1058 (2014).

[16]B. Treutlein, D. G. Brownfield, A. R. Wu, N. F. Neff, G. L. Mantalas, F. H. Espinoza, T. J. Desai, M. A. Krasnow, and S. R. Quake, "Reconstructing lineage hierarchies of the distal lung epithelium using single-cell rna-seq," Nature **509**, 371–375 (2014).

[17]M. J. Barresi and S. F. Gilbert, *Developmental Biology* (Oxford University Press, 2019).

[18]R. Dries, J. Chen, N. Del Rossi, M. M. Khan, A. Sistig, and G.-C. Yuan, "Advances in spatial transcriptomic data analysis," Genome Research **31**, 1706–1718 (2021).

[19]T. Noel, Q. S. Wang, A. Greka, and J. L. Marshall, "Principles of spatial transcriptomics analysis: A practical walk-through in kidney tissue," Frontiers in Physiology **12** (2022), 10.3389/fphys.2021.809346.

[20]A. Rao, D. Barkley, G. S. França, and I. Yanai, "Exploring tissue architecture using spatial transcriptomics," Nature **596**, 211–220 (2021).

[21]R. Ke, M. Mignardi, A. Pacureanu, J. Svedlund, J. Botling, C. Wählby, and M. Nilsson, "In situ sequencing for RNA analysis in preserved tissue and cells," Nature Methods **10**, 857–860 (2013).

[22]S. Codeluppi, L. E. Borm, A. Zeisel, G. L. Manno, J. A. van Lunteren, C. I. Svensson, and S. Linnarsson, "Spatial organization of the somatosensory cortex revealed by cyclic smfish," bioRxiv (2018), 10.1101/276097, https://www.biorxiv.org/content/early/2018/03/04/276097.full.pdf.

[23]A. Raj, P. van den Bogaard, S. A. Rifkin, A. van Oudenaarden, and S. Tyagi, "Imaging individual mRNA molecules using multiple singly labeled probes," Nature Methods **5**, 877–879 (2008).

[24]A. M. Femino, F. S. Fay, K. Fogarty, and R. H. Singer, "Visualization of single RNA transcripts in situ," Science **280**, 585–590 (1998), https://www.science.org/doi/pdf/10.1126/science.280.5363.585.

[25]S. Alon, D. R. Goodwin, A. Sinha, A. T. Wassie, F. Chen, E. R. Daugharthy, Y. Bando, A. Kajita, A. G. Xue, K. Marrett, R. Prior, Y. Cui, A. C. Payne, C.-C. Yao, H.-J. Suk, R. Wang, C.-C. J. Yu, P. Tillberg, P. Reginato, N. Pak, S. Liu, S. Punthambaker, E. P. R. Iyer, R. E. Kohman, J. A. Miller, E. S. Lein, A. Lako, N. Cullen, S. Rodig, K. Helvie, D. L. Abravanel, N. Wagle, B. E. Johnson, J. Klughammer, M. Slyper, J. Waldman, J. Jané-Valbuena, O. Rozenblatt-Rosen, A. Regev, null null, G. M. Church, A. H. Marblestone, E. S. Boyden, H. R. Ali, M. A. Sa'd, S. Alon, S. Aparicio, G. Battistoni, S. Balasubramanian, R. Becker, B. Bodenmiller, E. S. Boyden, D. Bressan, A. Bruna, M. Burger, C. Caldas, M. Callari, I. G. Cannell, H. Casbolt, N. Chornay, Y. Cui, A. Dariush, K. Dinh, A. Emenari, Y. Eyal-Lubling, J. Fan, A. Fatemi, E. Fisher, E. A. González-Solares, C. González-Fernández, D. Goodwin, W. Greenwood, F. Grimaldi, G. J. Hannon, O. Harris, S. Harris, C. Jauset, J. A. Joyce, E. D. Karagiannis, T. Kovačević, L. Kuett, R. Kunes, A. K. Yoldaş, D. Lai, E. Laks, H. Lee, M. Lee, G. Lerda, Y. Li, A. McPherson, N. Millar, C. M. Mulvey, F. Nugent, C. H. O'Flanagan, M. Paez-Ribes, I. Pearsall, F. Qosaj, A. J. Roth, O. M. Rueda, T. Ruiz, K. Sawicka, L. A. Sepúlveda, S. P. Shah, A. Shea, A. Sinha, A. Smith, S. Tavaré, S. Tietscher, I. Vázquez-García, S. L. Vogl, N. A. Walton, A. T. Wassie, S. S. Watson, J. Weselak, S. A. Wild, E. Williams, J. Windhager, T. Whitmarsh, C. Xia, P. Zheng, and X. Zhuang, "Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems," Science **371**, eaax2656 (2021), https://www.science.org/doi/pdf/10.1126/science.aax2656.

[26]S. Codeluppi, L. E. Borm, A. Zeisel, G. La Manno, J. A. van Lunteren, C. I. Svensson, and S. Linnarsson, "Spatial organization of the somatosensory cortex revealed by osmfish," Nature Methods **15**, 932–935 (2018).

[27]K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang, "Spatially resolved, highly multiplexed RNA profiling in single cells," Science **348**, aaa6090 (2015), https://www.science.org/doi/pdf/10.1126/science.aaa6090.

[28]C.-H. L. Eng, M. Lawson, Q. Zhu, R. Dries, N. Koulena, Y. Takei, J. Yun, C. Cronin, C. Karp, G.-C. Yuan, and L. Cai, "Transcriptome-scale super-resolved imaging in tissues by RNA seqfish+," Nature **568**, 235–239 (2019).

[29]X. Wang, W. E. Allen, M. A. Wright, E. L. Sylwestrak, N. Samusik, S. Vesuna, K. Evans, C. Liu, C. Ramakrishnan, J. Liu, G. P. Nolan, F.-A. Bava, and K. Deisseroth, "Three-dimensional intact-tissue sequencing of single-cell transcriptional states," Science **361**, eaat5691 (2018), https://www.science.org/doi/pdf/10.1126/science.aat5691.

[30]10x Genomics, "Spatial transcriptomics," `https://www.10xgenomics.com/spatial-transcriptomics` (2021).

[31]P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, A. Mollbrink, S. Linnarsson, S. Codeluppi, Å. Borg, F. Pontén, P. I. Costea, P. Sahlén, J. Mulder, O. Bergmann, J. Lundeberg, and J. Frisén, "Visualization and analysis of gene expression in tissue sections by spatial transcriptomics," Science **353**, 78–82 (2016), https://www.science.org/doi/pdf/10.1126/science.aaf2403.

[32]S. G. Rodriques, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, and E. Z. Macosko, "Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution," Science **363**, 1463–1467 (2019), https://www.science.org/doi/pdf/10.1126/science.aaw1219.

[33]S. Vickovic, G. Eraslan, F. Salmén, J. Klughammer, L. Stenbeck, D. Schapiro, T. Äijö, R. Bonneau, L. Bergenstråhle, J. Navarro, J. Gould, G. K. Griffin, Å. Borg, M. Ronaghi, J. Frisén, J. Lundeberg, A. Regev, and P. L. Ståhl, "High-definition spatial transcriptomics for in situ tissue profiling," Nature Methods **16**, 987–990 (2019).

[34]T. Biancalani, G. Scalia, L. Buffoni, R. Avasthi, Z. Lu, A. Sanger, N. Tokcan, C. R. Vanderburg, Å. Segerstolpe, M. Zhang, I. Avraham-Davidi, S. Vickovic, M. Nitzan, S. Ma, A. Subramanian, M. Lipinski, J. Buenrostro, N. B. Brown, D. Fanelli, X. Zhuang, E. Z. Macosko, and A. Regev, "Deep learning and alignment of spatially resolved single-cell transcriptomes with tangram," Nature Methods **18**, 1352–1362 (2021).

[35]F. Salmén, P. L. Ståhl, A. Mollbrink, J. Navarro, S. Vickovic, J. Frisén, and J. Lundeberg, "Barcoded solid-phase RNA capture for spatial transcriptomics profiling in mammalian tissue sections," Nature Protocols **13**, 2501–2534 (2018).

[36]A. Jemt, F. Salmén, A. Lundmark, A. Mollbrink, J. Fernández Navarro, P. L. Ståhl, T. Yucel-Lindberg, and J. Lundeberg, "An automated approach to prepare tissue-derived spatially barcoded rna-sequencing libraries," Scientific Reports **6**, 37137 (2016).

[37]N. J. Tustison, P. A. Cook, A. Klein, G. Song, S. R. Das, J. T. Duda, B. M. Kandel, N. van Strien, J. R. Stone, J. C. Gee, and B. B. Avants, "Large-scale evaluation of ants and freesurfer cortical thickness measurements," NeuroImage **99**, 166–179 (2014).

[38]G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "Voxelmorph: A learning framework for deformable medical image registration," IEEE Transactions on Medical Imaging **38**, 1788–1800 (2019).

[39]Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature **521**, 436–444 (2015).

[40]G. Eraslan, L. M. Simon, M. Mircea, N. S. Mueller, and F. J. Theis, "Single-cell rna-seq denoising using a deep count autoencoder," Nature Communications **10**, 390 (2019).

[41]M. B. Badsha, R. Li, B. Liu, Y. I. Li, M. Xian, N. E. Banovich, and A. Q. Fu, "Imputation of single-cell gene expression with an autoencoder neural network," Quantitative Biology **8**, 78–94 (2020).

[42]X. Li, K. Wang, Y. Lyu, H. Pan, J. Zhang, D. Stambolian, K. Susztak, M. P. Reilly, G. Hu, and M. Li, "Deep learning enables accurate clustering with batch effect removal in single-cell rna-seq analysis," Nature Communications **11**, 2338 (2020).

[43]J. Hu, X. Li, G. Hu, Y. Lyu, K. Susztak, and M. Li, "Iterative transfer learning with neural network for clustering and cell type classification in single-cell rna-seq analysis," Nature Machine Intelligence **2**, 607–618 (2020).

[44]X. Shao, H. Yang, X. Zhuang, J. Liao, P. Yang, J. Cheng, X. Lu, H. Chen, and X. Fan, "scdeepsort: a pre-trained cell-type annotation method for single-cell transcriptomics using deep learning with a weighted graph neural network," Nucleic Acids Research **49**, e122–e122 (2021).

[45]F. Ma and M. Pellegrini, "Actinn: automated identification of cell types in single cell RNA sequencing," Bioinformatics **36**, 533–538 (2020).

[46]A. A. Heydari, O. A. Davalos, L. Zhao, K. K. Hoyer, and S. S. Sindi, "ACTIVA: realistic single-cell RNA-seq generation with automatic cell-type identification using introspective variational autoencoders," Bioinformatics (2022), 10.1093/bioinformatics/btac095, btac095, https://academic.oup.com/bioinformatics/advance-article-pdf/doi/10.1093/bioinformatics/btac095/42628775/btac095.pdf.

[47]M. Marouf, P. Machart, V. Bansal, C. Kilian, D. S. Magruder, C. F. Krebs, and S. Bonn, "Realistic in silico generation and augmentation of single-cell rna-seq data using generative adversarial networks," Nature Communications **11**, 166 (2020).

[48]K. Bayoudh, R. Knani, F. Hamdaoui, and A. Mtibaa, "A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets," The Visual Computer (2021), 10.1007/s00371-021-02166-7.

[49]D. Butler, C. Mozsary, C. Meydan, J. Foox, J. Rosiene, A. Shaiber, D. Danko, E. Afshinnekoo, M. MacKay, F. J. Sedlazeck, N. A. Ivanov, M. Sierra, D. Pohle, M. Zietz, U. Gisladottir, V. Ramlall, E. T. Sholle, E. J. Schenck, C. D. Westover, C. Hassan, K. Ryon, B. Young, C. Bhattacharya, D. L. Ng, A. C. Granados, Y. A. Santos, V. Servellita, S. Federman, P. Ruggiero, A. Fungtammasan, C.-S. Chin, N. M. Pearson, B. W. Langhorst, N. A. Tanner, Y. Kim, J. W. Reeves, T. D. Hether, S. E. Warren, M. Bailey, J. Gawrys, D. Meleshko, D. Xu, M. Couto-Rodriguez, D. Nagy-Szakal, J. Barrows, H. Wells, N. B. O'Hara, J. A. Rosenfeld, Y. Chen, P. A. D. Steel, A. J. Shemesh, J. Xiang, J. Thierry-Mieg, D. Thierry-Mieg, A. Iftner, D. Bezdan, E. Sanchez, T. R. Campion, J. Sipley, L. Cong, A. Craney, P. Velu, A. M. Melnick, S. Shapira, I. Hajirasouliha, A. Borczuk, T. Iftner, M. Salvatore, M. Loda, L. F. Westblade, M. Cushing, S. Wu, S. Levy, C. Chiu, R. E. Schwartz, N. Tatonetti, H. Rennert, M. Imielinski, and C. E. Mason, "Shotgun transcriptome, spatial omics, and isothermal profiling of sars-cov-2 infection reveals unique host responses, viral diversification, and drug interactions," Nature Communications **12**, 1660 (2021).

[50]T. M. Delorey, C. G. K. Ziegler, G. Heimberg, R. Normand, Y. Yang, Å. Segerstolpe, D. Abbondanza, S. J. Fleming, A. Subramanian, D. T. Montoro, K. A. Jagadeesh, K. K. Dey, P. Sen, M. Slyper, Y. H. Pita-Juárez, D. Phillips, J. Biermann, Z. Bloom-Ackermann, N. Barkas, A. Ganna, J. Gomez, J. C. Melms, I. Katsyv, E. Normandin, P. Naderi, Y. V. Popov, S. S. Raju, S. Niezen, L. T. Y. Tsai, K. J. Siddle, M. Sud, V. M. Tran, S. K. Vellarikkal, Y. Wang, L. Amir-Zilberstein, D. S. Atri, J. Beechem, O. R. Brook, J. Chen, P. Divakar, P. Dorceus, J. M. Engreitz, A. Essene, D. M. Fitzgerald, R. Fropf, S. Gazal, J. Gould, J. Grzyb, T. Harvey, J. Hecht, T. Hether, J. Jané-Valbuena, M. Leney-Greene, H. Ma, C. McCabe, D. E. McLoughlin, E. M. Miller, C. Muus, M. Niemi, R. Padera, L. Pan, D. Pant, C. Pe'er, J. Pfiffner-Borges, C. J. Pinto, J. Plaisted, J. Reeves, M. Ross, M. Rudy, E. H. Rueckert, M. Siciliano, A. Sturm, E. Todres, A. Waghray, S. Warren, S. Zhang, D. R. Zollinger, L. Cosimi, R. M. Gupta, N. Hacohen, H. Hibshoosh, W. Hide, A. L. Price, J. Rajagopal, P. R. Tata, S. Riedel, G. Szabo, T. L. Tickle, P. T. Ellinor, D. Hung, P. C. Sabeti, R. Novak, R. Rogers, D. E. Ingber, Z. G. Jiang, D. Juric, M. Babadi, S. L. Farhi, B. Izar, J. R. Stone, I. S. Vlachos, I. H. Solomon, O. Ashenberg, C. B. M. Porter, B. Li, A. K. Shalek, A.-C. Villani, O. Rozenblatt-Rosen, and A. Regev, "Covid-19 tissue atlases reveal sars-cov-2 pathology and cellular targets," Nature **595**, 107–113 (2021).

[51]S. Vickovic, D. Schapiro, K. Carlberg, B. Lötstedt, L. Larsson, M. Korotkova, A. H. Hensvold, A. I. Catrina, P. K. Sorger, V. Malmström, A. Regev, and P. L. Ståhl, "Three-dimensional spatial transcriptomics uncovers cell type dynamics in the rheumatoid arthritis synovium," bioRxiv , 2020.12.10.420463 (2020).

[52]K. Carlberg, M. Korotkova, L. Larsson, A. I. Catrina, P. L. Ståhl, and V. Malmström, "Exploring inflammatory signatures in arthritic joint biopsies with spatial transcriptomics," Scientific Reports **9**, 18975 (2019).

[53]E. Berglund, J. Maaskola, N. Schultz, S. Friedrich, M. Marklund, J. Bergenstråhle, F. Tarish, A. Tanoglidi, S. Vickovic, L. Larsson, F. Salmén, C. Ogris, K. Wallenborg, J. Lagergren, P. Ståhl, E. Sonnhammer, T. Helleday, and J. Lundeberg, "Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity," Nature Communications **9**, 2419 (2018).

[54]R. Moncada, D. Barkley, F. Wagner, M. Chiodin, J. C. Devlin, M. Baron, C. H. Hajdu, D. M. Simeone, and I. Yanai, "Integrating microarray-based spatial transcriptomics and single-cell rna-seq reveals tissue architecture in pancreatic ductal adenocarcinomas," Nature Biotechnology **38**, 333–342 (2020).

[55]A. L. Ji, A. J. Rubin, K. Thrane, S. Jiang, D. L. Reynolds, R. M. Meyers, M. G. Guo, B. M. George, A. Mollbrink, J. Bergenstråhle, L. Larsson, Y. Bai, B. Zhu, A. Bhaduri, J. M. Meyers, X. Rovira-Clavé, S. T. Hollmig, S. Z. Aasi, G. P. Nolan, J. Lundeberg, and P. A. Khavari, "Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma," Cell **182**, 497–514.e22 (2020).

[56]W.-T. Chen, A. Lu, K. Craessaerts, B. Pavie, C. Sala Frigerio, N. Corthout, X. Qian, J. Laláková, M. Kühnemund, I. Voytyuk, L. Wolfs, R. Mancuso, E. Salta, S. Balusu, A. Snellinx, S. Munck, A. Jurek, J. Fernandez Navarro, T. C. Saido, I. Huitinga, J. Lundeberg, M. Fiers, and B. De Strooper, "Spatial transcriptomics and in situ sequencing to study alzheimer's disease," Cell **182**, 976–991.e19 (2020).

[57]J. Bäckdahl, L. Franzén, L. Massier, Q. Li, J. Jalkanen, H. Gao, A. Andersson, N. Bhalla, A. Thorell, M. Rydén, P. L. Ståhl, and N. Mejhert, "Spatial mapping reveals human adipocyte subpopulations with distinct sensitivities to insulin," Cell Metabolism **33**, 1869–1882.e6 (2021).

[58]G. Theocharidis, B. E. Thomas, D. Sarkar, H. L. Mumme, W. J. R. Pilcher, B. Dwivedi, T. Sandoval-Schaefer, R. F. Sîrbulescu, A. Kafanas, I. Mezghani, P. Wang, A. Lobao, I. S. Vlachos, B. Dash, H. C. Hsia, V. Horsley, S. S. Bhasin, A. Veves, and M. Bhasin, "Single cell transcriptomic landscape of diabetic foot ulcers," Nature Communications **13**, 181 (2022).

[59]R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, and A. Regev, "Spatial reconstruction of single-cell gene expression data," Nature Biotechnology **33**, 495–502 (2015).

[60]M. Nitzan, N. Karaiskos, N. Friedman, and N. Rajewsky, "Gene expression cartography," Nature **576**, 132–137 (2019).

[61]F. Maseda, Z. Cang, and Q. Nie, "Deepsc: A deep learning-based map connecting single-cell transcriptomics and spatial imaging data," Frontiers in Genetics **12**, 348 (2021).

[62]K. Achim, J.-B. Pettit, L. R. Saraiva, D. Gavriouchkina, T. Larsson, D. Arendt, and J. C. Marioni, "High-throughput spatial mapping of single-cell rna-seq data to tissue of origin," Nature Biotechnology **33**, 503–509 (2015).

[63]T. Peng, G. M. Chen, and K. Tan, "Gluer: integrative analysis of single-cell omics and imaging data by deep neural network," bioRxiv (2021), 10.1101/2021.01.25.427845, https://www.biorxiv.org/content/early/2021/01/26/2021.01.25.427845.full.pdf.

[64]A. Andersson, J. Bergenstråhle, M. Asp, L. Bergenstråhle, A. Jurek, J. Fernández Navarro, and J. Lundeberg, "Single-cell and spatial transcriptomics enables probabilistic inference of cell type topography," Communications Biology **3**, 565 (2020).

[65]Q. Song and J. Su, "DSTG: deconvoluting spatial transcriptomics data through graph-based artificial intelligence," Briefings in Bioinformatics **22** (2021), 10.1093/bib/bbaa414, bbaa414, https://academic.oup.com/bib/article-pdf/22/5/bbaa414/40261325/bbaa414.pdf.

[66]M. Elosua-Bayes, P. Nieto, E. Mereu, I. Gut, and H. Heyn, "SPOTlight: seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes," Nucleic Acids Research **49**, e50–e50 (2021), https://academic.oup.com/nar/article-pdf/49/9/e50/37998836/gkab043.pdf.

[67]D. M. Cable, E. Murray, L. S. Zou, A. Goeva, E. Z. Macosko, F. Chen, and R. A. Irizarry, "Robust decomposition of cell type mixtures in spatial transcriptomics," Nature Biotechnology (2021), 10.1038/s41587-021-00830-w.

[68]R. Dong and G.-C. Yuan, "Spatialdwls: accurate deconvolution of spatial transcriptomic data," Genome Biology **22**, 145 (2021).

[69]R. Lopez, B. Li, H. Keren-Shaul, P. Boyeau, M. Kedmi, D. Pilzer, A. Jelinski, E. David, A. Wagner, Y. Addad, M. I. Jordan, I. Amit, and N. Yosef, "Multi-resolution deconvolution of spatial transcriptomics data reveals continuous patterns of inflammation," bioRxiv (2021), 10.1101/2021.05.10.443517, https://www.biorxiv.org/content/early/2021/05/11/2021.05.10.443517.full.pdf.

[70]V. Kleshchevnikov, A. Shmatko, E. Dann, A. Aivazidis, H. W. King, T. Li, R. Elmentaite, A. Lomakin, V. Kedlian, A. Gayoso, M. S. Jain, J. S. Park, L. Ramona, E. Tuck, A. Arutyunyan, R. Vento-Tormo, M. Gerstung, L. James, O. Stegle, and O. A. Bayraktar, "Cell2location maps fine-grained cell types in spatial transcriptomics," Nature Biotechnology (2022), 10.1038/s41587-021-01139-4.

[71]Q. Zhu, S. Shah, R. Dries, L. Cai, and G.-C. Yuan, "Identification of spatially associated subpopulations by combining scrnaseq and sequential fluorescence in situ hybridization data," Nature Biotechnology **36**, 1183–1190 (2018).

[72]X. Tan, A. Su, M. Tran, and Q. Nguyen, "SpaCell: integrating tissue morphology and spatial gene expression to predict disease cells," Bioinformatics **36**, 2293–2294 (2020), https://academic.oup.com/bioinformatics/article-pdf/36/7/2293/33027676/btz914.pdf.

[73]E. Zhao, M. R. Stone, X. Ren, J. Guenthoer, K. S. Smythe, T. Pulliam, S. R. Williams, C. R. Uytingco, S. E. B. Taylor, P. Nghiem, J. H. Bielas, and R. Gottardo, "Spatial transcriptomics at subspot resolution with bayesspace," Nature Biotechnology **39**, 1375–1384 (2021).

[74]J. Hu, X. Li, K. Coleman, A. Schroeder, N. Ma, D. J. Irwin, E. B. Lee, R. T. Shinohara, and M. Li, "Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network," Nature Methods **18**, 1342–1351 (2021).

[75]D. Edsgärd, P. Johnsson, and R. Sandberg, "Identification of spatial expression trends in single-cell gene expression data," Nature Methods **15**, 339–342 (2018).

[76]V. Svensson, S. A. Teichmann, and O. Stegle, "Spatialde: identification of spatially variable genes," Nature Methods **15**, 343–346 (2018).

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

[78]Z. Cang and Q. Nie, "Inferring spatial and signaling relationships between cells from single cell transcriptomic data," Nature Communications **11**, 2084 (2020).

[79]D. Pham, X. Tan, J. Xu, L. F. Grice, P. Y. Lam, A. Raghubar, J. Vukovic, M. J. Ruitenberg, and Q. Nguyen, "stlearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues," bioRxiv (2020), 10.1101/2020.05.31.125658, https://www.biorxiv.org/content/early/2020/05/31/2020.05.31.125658.full.pdf.

[80]J. Tanevski, R. O. R. Flores, A. Gabor, D. Schapiro, and J. Saez-Rodriguez, "Explainable multi-view framework for dissecting intercellular signaling from highly multiplexed spatial data," bioRxiv (2021), 10.1101/2020.05.08.084145, https://www.biorxiv.org/content/early/2021/07/13/2020.05.08.084145.full.pdf.

[81]R. Dries, Q. Zhu, R. Dong, C.-H. L. Eng, H. Li, K. Liu, Y. Fu, T. Zhao, A. Sarkar, F. Bao, R. E. George, N. Pierson, L. Cai, and G.-C. Yuan, "Giotto: a toolbox for integrative analysis and visualization of spatial expression data," Genome Biology **22**, 78 (2021).

[82]R. K. Gupta and J. Kuznicki, "Biological and medical importance of cellular heterogeneity deciphered by single-cell RNA sequencing," Cells **9**, 1751 (2020).

[83]P. van Galen, V. Hovestadt, M. H. Wadsworth II, T. K. Hughes, G. K. Griffin, S. Battaglia, J. A. Verga, J. Stephansky, T. J. Pastika, J. Lombardi Story, G. S. Pinkus, O. Pozdnyakova, I. Galinsky, R. M. Stone, T. A. Graubert, A. K. Shalek, J. C. Aster, A. A. Lane, and B. E. Bernstein, "Single-cell rna-seq reveals aml hierarchies relevant to disease progression and immunity," Cell **176**, 1265–1281.e24 (2019).

[84]T. Masuda, R. Sankowski, O. Staszewski, C. Böttcher, L. Amann, Sagar, C. Scheiwe, S. Nessler, P. Kunz, G. van Loo, V. A. Coenen, P. C. Reinacher, A. Michel, U. Sure, R. Gold, D. Grün, J. Priller, C. Stadelmann, and M. Prinz, "Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution," Nature **566**, 388–392 (2019).

[85]J. M. Churko, P. Garg, B. Treutlein, M. Venkatasubramanian, H. Wu, J. Lee, Q. N. Wessells, S.-Y. Chen, W.-Y. Chen, K. Chetal, G. Mantalas, N. Neff, E. Jabart, A. Sharma, G. P. Nolan, N. Salomonis, and J. C. Wu, "Defining human cardiac transcription factor hierarchies using integrated single-cell heterogeneity analysis," Nature Communications **9**, 4906 (2018).

[86]A. Gross, J. Schoendube, S. Zimmermann, M. Steeb, R. Zengerle, and P. Koltay, "Technologies for single-cell isolation," International Journal of Molecular Sciences **16**, 16897–16919 (2015).

[87]S. Ma, T. W. Murphy, and C. Lu, "Microfluidics for genome-wide studies involving next generation sequencing," Biomicrofluidics **11**, 021501–021501 (2017).

[88]B. Hwang, J. H. Lee, and D. Bang, "Single-cell RNA sequencing technologies and bioinformaticspipelines," Experimental & Molecular Medicine **50**, 1–14 (2018).

[89]A. Haque, J. Engel, S. A. Teichmann, and T. Lönnberg, "A practical guide to single-cell rna-sequencing for biomedical research and clinical applications," Genome Medicine **9**, 75 (2017).

[90]R. Stark, M. Grzelak, and J. Hadfield, "RNA sequencing: the teenage years," Nature Reviews Genetics **20**, 631–656 (2019).

[91]E. L. van Dijk, H. Auger, Y. Jaszczyszyn, and C. Thermes, "Ten years of next-generation sequencing technology," Trends in Genetics **30**, 418–426 (2014).

[92]X. Zhang, T. Li, F. Liu, Y. Chen, J. Yao, Z. Li, Y. Huang, and J. Wang, "Comparative analysis of droplet-based ultra-high-throughput single-cell rna-seq systems," Molecular Cell **73**, 130–142.e5 (2019).

[93]A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann, "The technology and biology of single-cell RNA sequencing," Molecular Cell **58**, 610–620 (2015).

[94]X.-t. Huang, X. Li, P.-z. Qin, Y. Zhu, S.-n. Xu, and J.-p. Chen, "Technical advances in single-cell RNAsequencing and applications in normal and malignant hematopoiesis," Frontiers in Oncology **8** (2018), 10.3389/fonc.2018.00582.

[95]M. Asp, J. Bergenstråhle, and J. Lundeberg, "Spatially resolved transcriptomes—next generation tools for tissue exploration," BioEssays **42**, 1900221 (2020), https://onlinelibrary.wiley.com/doi/pdf/10.1002/bies.201900221.

[96]T. Rautenstrauss, Bernd W; Liehr, *FISH Technology* (Springer: Verlag Berlin, 2002).

[97]E. Jensen, "Technical review: In situ hybridization," The Anatomical Record **297**, 1349–1353 (2014), https://anatomypubs.onlinelibrary.wiley.com/doi/pdf/10.1002/ar.22944.

[98]M. M. Hilscher, D. Gyllborg, C. Yokota, and M. Nilsson, "In situ sequencing: A high-throughput, multi-targeted gene expression profiling technique for cell typing in tissue sections," in *In Situ Hybridization Protocols*, edited by B. S. Nielsen and J. Jones (Springer US, New York, NY, 2020) pp. 313–329.

[99]J. A. Weinstein, A. Regev, and F. Zhang, "Dna microscopy: Optics-free spatio-genetic imaging by a stand-alone chemical reaction," Cell **178**, 229–241.e16 (2019).

[100]C. Larsson, I. Grundberg, O. Söderberg, and M. Nilsson, "In situ detection and genotyping of individual mRNA molecules," Nature Methods **7**, 395–397 (2010).

[101]X. Chen, Y.-C. Sun, G. M. Church, J. H. Lee, and A. M. Zador, "Efficient in situ barcode sequencing using padlock probe-based BaristaSeq," Nucleic Acids Research **46**, e22–e22 (2017), https://academic.oup.com/nar/article-pdf/46/4/e22/24214417/gkx1206.pdf.

[102]J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, T. C. Ferrante, R. Terry, B. M. Turczyk, J. L. Yang, H. S. Lee, J. Aach, K. Zhang, and G. M. Church, "Fluorescent in situ sequencing (fisseq) of RNA for gene expression profiling in intact cells and tissues," Nature Protocols **10**, 442–458 (2015).

[103]J. H. Lee, "Quantitative approaches for investigating the spatial context of gene expression," WIREs Systems Biology and Medicine **9**, e1369 (2017), https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wsbm.1369.

[104]P. R. Gudla, K. Nakayama, G. Pegoraro, and T. Misteli, "Spotlearn: Convolutional neural network for detection of fluorescence in situ hybridization (fish) signals in high-throughput imaging approaches." Cold Spring Harbor symposia on quantitative biology **82**, 57–70 (2017).

[105]E. Pardo, J. M. T. Morgado, and N. Malpica, "Semantic segmentation of mfish images using convolutional networks," Cytometry Part A **93**, 620–627 (2018), https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.a.23375.

[106]Z. Frankenstein, N. Uraoka, U. Aypar, R. Aryeequaye, M. Rao, M. Hameed, Y. Zhang, and Y. Yagi, "Automated 3d scoring of fluorescence in situ hybridization (fish) using a confocal whole slide imaging scanner," Applied Microscopy **51**, 4 (2021).

[107]R. Littman, Z. Hemminger, R. Foreman, D. Arneson, G. Zhang, F. Gómez-Pinilla, X. Yang, and R. Wollman, "Joint cell segmentation and cell type annotation for spatial transcriptomics," Molecular Systems Biology **17**, e10108 (2021), https://www.embopress.org/doi/pdf/10.15252/msb.202010108.

[108]R. Littman, Z. Hemminger, R. Foreman, D. Arneson, G. Zhang, F. Gómez-Pinilla, X. Yang, and R. Wollman, "Joint cell segmentation and cell type annotation for spatial transcriptomics," Molecular Systems Biology **17**, e10108 (2021), https://www.embopress.org/doi/pdf/10.15252/msb.202010108.

[109]x. Genomics, "Visium spatial gene expression reagent kits - user guide," Online, 6230 Stoneridge Mall Road, Pleasanton, CA 94588 USA (2021).

[110]M. V. Hunter, R. Moncada, J. M. Weiss, I. Yanai, and R. M. White, "Spatially resolved transcriptomics reveals the architecture of the tumor/microenvironment interface," bioRxiv (2021), 10.1101/2020.11.05.368753, https://www.biorxiv.org/content/early/2021/06/01/2020.11.05.368753.full.pdf.

[111]A. L. Ji, A. J. Rubin, K. Thrane, S. Jiang, D. L. Reynolds, R. M. Meyers, M. G. Guo, B. M. George, A. Mollbrink, J. Bergenstråhle, L. Larsson, Y. Bai, B. Zhu, A. Bhaduri, J. M. Meyers, X. Rovira-Clavé, S. T. Hollmig, S. Z. Aasi, G. P. Nolan, J. Lundeberg, and P. A. Khavari, "Multimodal analysis of composition and spatial architecture in human squamous cell carcinoma," Cell **182**, 497–514.e22 (2020).

[112]D. Fawkner-Corbett, A. Antanaviciute, K. Parikh, M. Jagielowicz, A. S. Gerós, T. Gupta, N. Ashley, D. Khamis, D. Fowler, E. Morrissey, C. Cunningham, P. R. Johnson, H. Koohy, and A. Simmons, "Spatiotemporal analysis of human intestinal development at single-cell resolution," Cell **184**, 810–826.e23 (2021).

[113]M. Asp, S. Giacomello, L. Larsson, C. Wu, D. Fürth, X. Qian, E. Wärdell, J. Custodio, J. Reimegård, F. Salmén, C. Österholm, P. L. Ståhl, E. Sundström, E. Åkesson, O. Bergmann, M. Bienko, A. Månsson-Broberg, M. Nilsson, C. Sylvén, and J. Lundeberg, "A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart," Cell **179**, 1647–1660.e19 (2019).

[114]M. Y. Batiuk, T. Tyler, S. Mei, R. Rydbirk, V. Petukhov, D. Sedmak, E. Frank, V. Feher, N. Habek, Q. Hu, A. Igolkina, L. Roszik, U. Pfisterer, Z. Petanjek, I. Adorjan, P. V. Kharchenko, and K. Khodosevich, "Selective vulnerability of supragranular layer neurons in schizophrenia," bioRxiv (2021), 10.1101/2020.11.17.386458, https://www.biorxiv.org/content/early/2021/01/17/2020.11.17.386458.full.pdf.

[115]C. Ortiz, J. F. Navarro, A. Jurek, A. Märtin, J. Lundeberg, and K. Meletis, "Molecular atlas of the adult mouse brain," Science Advances **6**, eabb3446 (2020), https://www.science.org/doi/pdf/10.1126/sciadv.abb3446.

[116]10x Genomics, "Spatial gene expression," `https://www.10xgenomics.com/products/spatial-gene-expression` (2022).

[117]I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (The MIT Press, 2016).

[118]S. Lloyd, "Least squares quantization in pcm," IEEE Transactions on Information Theory **28**, 129–137 (1982).

[119]L. Prechelt, "Early stopping-but when?" in *Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop* (Springer-Verlag, Berlin, Heidelberg, 1998) pp. 55–69.

[120]G. Cybenko, "Approximation by superpositions of a sigmoidal function," Mathematics of Control, Signals and Systems **2**, 303–314 (1989).

[121]F. Scarselli and A. Chung Tsoi, "Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results," Neural Networks **11**, 15–37 (1998).

[122]T. Chen and H. Chen, "Universal approximation to nonlinear operators by neural networks with arbitrary activation functions and its application to dynamical systems," IEEE Transactions on Neural Networks **6**, 911–917 (1995).

[123]M. Uzair and N. Jamil, "Effects of hidden layers on the efficiency of neural networks," in *2020 IEEE 23rd International Multitopic Conference (INMIC)* (2020) pp. 1–6.

[124]A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, Vol. 25, edited by F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Curran Associates, Inc., 2012).

[125]X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "Tph-yolov5: Improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios," (2021), arXiv:2108.11539 [cs.CV].

[126]M. Tan and Q. V. Le, "Efficientnetv2: Smaller models and faster training," (2021), arXiv:2104.00298 [cs.CV].

[127]A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, Vol. 30, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017).

[128]T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," (2020), arXiv:2005.14165 [cs.CL].

[129]J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, edited by J. Burstein, C. Doran, and T. Solorio (Association for Computational Linguistics, 2019) pp. 4171–4186.

[130]R. Poplin, P.-C. Chang, D. Alexander, S. Schwartz, T. Colthurst, A. Ku, D. Newburger, J. Dijamco, N. Nguyen, P. T. Afshar, S. S. Gross, L. Dorfman, C. Y. McLean, and M. A. DePristo, "A universal snp and small-indel variant caller using deep neural networks," Nature Biotechnology **36**, 983–987 (2018).

[131]H. Li, U. Shaham, K. P. Stanton, Y. Yao, R. R. Montgomery, and Y. Kluger, "Gating mass cytometry data by deep learning," Bioinformatics **33**, 3423–3430 (2017), https://academic.oup.com/bioinformatics/article-pdf/33/21/3423/25166108/btx448.pdf.

[132]H. Huang, Z. Li, R. He, Z. Sun, and T. Tan, ": Introspective variational autoencoders for photographic image synthesis," in *Advances in Neural Information Processing Systems 31*, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018) pp. 52–63.

[133]J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis, "Highly accurate protein structure prediction with alphafold," Nature **596**, 583–589 (2021).

[134]T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings* (OpenReview.net, 2017).

[135]K. Hornik, "Approximation capabilities of multilayer feedforward networks," Neural Networks **4**, 251–257 (1991).

[136]V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10 (Omnipress, Madison, WI, USA, 2010) p. 807–814.

[137]J. Nocedal and S. Wright, *Numerical optimization* (Springer Science & Business Media, 2006).

[138]X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, Vol. 9, edited by Y. W. Teh and M. Titterington (PMLR, Chia Laguna Resort, Sardinia, Italy, 2010) pp. 249–256.

[139]K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imageNet classification," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (2015).

[140]S. Ruder, "An overview of gradient descent optimization algorithms," CoRR **abs/1609.04747** (2016), 1609.04747.

[141]S. Smith, E. Elsen, and S. De, "On the generalization benefit of noise in stochastic gradient descent," in *Proceedings of the 37th International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 119, edited by H. D. III and A. Singh (PMLR, 2020) pp. 9058–9067.

[142]D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature **323**, 533–536 (1986).

[143]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, edited by Y. Bengio and Y. LeCun (2015).

[144]J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," Journal of Machine Learning Research **12**, 2121–2159 (2011).

[145]Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE Transactions on Neural Networks and Learning Systems **30**, 3212–3232 (2019).

[146]D. Marr and E. Hildreth, "Theory of edge detection," Proceedings of the Royal Society of London. Series B. Biological Sciences **207**, 187–217 (1980).

[147]Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Computation **1**, 541–551 (1989).

[148]Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," in *The Handbook of Brain Theory and Neural Networks* (MIT Press, Cambridge, MA, USA, 1998) pp. 255–258.

[149]D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature **323**, 533–536 (1986).

[150]Note that if the model chose a separate parameter for each $\mathbf{x}^{(i)}$, for $i = 1, \cdots, n$, then the model could not generalize to any inputs where $|X| > n$ (size of X is greater that n).

[151]J. F. Kolen and S. C. Kremer, "Gradient flow in recurrent nets: The difficulty of learning longterm dependencies," in *A Field Guide to Dynamical Recurrent Networks* (IEEE, 2001) pp. 237–243.

[152]S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput. **9**, 1735–1780 (1997).

[153]J. C. Kimmel, A. S. Brack, and W. F. Marshall, "Deep convolutional and recurrent neural networks for cell motility discrimination and prediction," IEEE/ACM Transactions on Computational Biology and Bioinformatics **18**, 562–574 (2021).

[154]S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on Machine Learning*, Proceedings of Machine Learning Research, Vol. 37, edited by F. Bach and D. Blei (PMLR, Lille, France, 2015) pp. 448–456.

[155]K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016) pp. 770–778.

[156]A. Shah, E. Kadam, H. Shah, S. Shinde, and S. Shingade, "Deep residual networks with exponential linear unit," in *Proceedings of the Third International Symposium on Computer Vision and the Internet* (2016) pp. 59–65.

[157]ImageNet[216] is the standard dataset for benchmarking performance of machine learning algorithms in classification and object recofnition. ImageNet contains more than 14 million hand-annotated images.

[158]D. H. Ballard, "Modular learning in neural networks." in *Aaai*, Vol. 647 (1987) pp. 279–284.

[159]G. E. Hinton, "20 - connectionist learning procedures11this chapter appeared in volume 40 of artificial intelligence in 1989, reprinted with permission of north-holland publishing. it is a revised version of technical report cmu-cs-87-115, which has the same title and was prepared in june 1987 while the author was at carnegie mellon university. the research was supported by contract n00014-86-k-00167 from the office of naval research and by grant ist-8520359 from the national science foundation." in *Machine Learning*, edited by Y. Kodratoff and R. S. Michalski (Morgan Kaufmann, San Francisco (CA), 1990) pp. 555–610.

[160]D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *International Conference on Learning Representations* (2014).

[161]I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, Vol. 27, edited by Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger (Curran Associates, Inc., 2014).

[162]J. He, D. Spokoyny, G. Neubig, and T. Berg-Kirkpatrick, "Lagging inference networks and posterior collapse in variational autoencoders," in *International Conference on Learning Representations* (2019).

[163]Z. Yang, Z. Hu, R. Salakhutdinov, and T. Berg-Kirkpatrick, "Improved variational autoencoders for text modeling using dilated convolutions," in *International conference on machine learning* (PMLR, 2017) pp. 3881–3890.

[164]A. A. Heydari and A. Mehmood, "SRVAE: super resolution using variational autoencoders," in *Proc.SPIE*, Vol. 11400 (2020).

[165]S. Semeniuta, A. Severyn, and E. Barth, "A hybrid convolutional variational autoencoder for text generation," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (Association for Computational Linguistics, Copenhagen, Denmark, 2017) pp. 627–637.

[166]A. A. Heydari, C. A. Thompson, and A. Mehmood, "Softadapt: Techniques for adaptive loss weighting of neural networks with multi-part loss functions," arXiv preprint arXiv:1912.12355 (2019).

[167]I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf, "Wasserstein auto-encoders," in *International Conference on Learning Representations* (2018).

[168]T. Daniel and A. Tamar, "SoftIntroVAE: Analyzing and improving the introspective variational autoencoder," (2021), arXiv:2012.13253 [cs.LG].

[169]The astute reader will note that although Maseda *et al.* refer to DEEPsc as DL model, the methods's two-layer FFNN is not considered deep model in most definitions.

[170]K. Nikos, W. Philipp, A. Jonathan, B. Anastasiya, A. Salah, K. Claudia, K. Christine, R. Nikolaus, and Z. R. P., "The drosophila embryo at single-cell transcriptome resolution," Science **358**, 194–199 (2017).

[171]B. Tasic, V. Menon, T. N. Nguyen, T. K. Kim, T. Jarsky, Z. Yao, B. Levi, L. T. Gray, S. A. Sorensen, T. Dolbeare, D. Bertagnolli, J. Goldy, N. Shapovalova, S. Parry, C. Lee, K. Smith, A. Bernard, L. Madisen, S. M. Sunkin, M. Hawrylycz, C. Koch, and H. Zeng, "Adult mouse cortical cell taxonomy revealed by single cell transcriptomics," Nature Neuroscience **19**, 335–346 (2016).

[172]S. Joost, A. Zeisel, T. Jacob, X. Sun, G. La Manno, P. Lönnerberg, S. Linnarsson, and M. Kasper, "Single-cell transcriptomics reveals that differentiation and spatial signatures shape epidermal and hair follicle heterogeneity," Cell Systems **3**, 221–237.e9 (2016).

[173]Dropout refers to the scenario when a gene is observed at a moderate or high expression level in a subset of cells, but not detected in other cells.

[174]J. Ding, X. Adiconis, S. K. Simmons, M. S. Kowalczyk, C. C. Hession, N. D. Marjanovic, T. K. Hughes, M. H. Wadsworth, T. Burks, L. T. Nguyen, J. Y. H. Kwon, B. Barak, W. Ge, A. J. Kedaigle, S. Carroll, S. Li, N. Hacohen, O. Rozenblatt-Rosen, A. K. Shalek, A.-C. Villani, A. Regev, and J. Z. Levin, "Systematic comparison of single-cell and single-nucleus rna-sequencing methods," Nature Biotechnology **38**, 737–746 (2020).

[175]PyTorch is one of the most popular DL library in Python. `https://pytorch.org/`.

[176]A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," CoRR **abs/1912.01703** (2019), 1912.01703.

[177]O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention –MICCAI 2015*, edited by N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Springer International Publishing, Cham, 2015) pp. 234–241.

[178]G. R. Koch, "Siamese neural networks for one-shot image recognition," (2015).

[179]G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017) pp. 2261–2269.

[180]Y. Liu, M. Yang, Y. Deng, G. Su, A. Enninful, C. C. Guo, T. Tebaldi, D. Zhang, D. Kim, Z. Bai, E. Norris, A. Pan, J. Li, Y. Xiao, S. Halene, and R. Fan, "High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue," Cell **183**, 1665–1681.e18 (2020).

[181]Https://www.nanostring.com/.

[182]J. Hu, A. Schroeder, K. Coleman, C. Chen, B. J. Auerbach, and M. Li, "Statistical and machine learning methods for spatially resolved transcriptomics with histology," Computational and Structural Biotechnology Journal **19**, 3829–3841 (2021).

[183]L. Garcia-Alonso, L.-F. Handfield, K. Roberts, K. Nikolakopoulou, R. C. Fernando, L. Gardner, B. Woodhams, A. Arutyunyan, K. Polanski, R. Hoo, C. Sancho-Serra, T. Li, K. Kwakwa, E. Tuck, V. Kleshchevnikov, A. Tarkowska, T. Porter, C. I. Mazzeo, S. van Dongen, M. Dabrowska, V. Vaskivskyi, K. T. Mahbubani, J.-e. Park, M. Jimenez-Linan, L. Campos, V. Kiselev, C. Lindskog, P. Ayuk, E. Prigmore, M. R. Stratton, K. Saeb-Parsy, A. Moffett, L. Moore, O. A. Bayraktar, S. A. Teichmann, M. Y. Turco, and R. Vento-Tormo, "Mapping

the temporal and spatial dynamics of the human endometrium in vivo and in vitro," bioRxiv (2021), 10.1101/2021.01.02.425073, https://www.biorxiv.org/content/early/2021/01/04/2021.01.02.425073.full.pdf.

[184] M. Asp, S. Giacomello, L. Larsson, C. Wu, D. Fürth, X. Qian, E. Wärdell, J. Custodio, J. Reimegård, F. Salmén, C. Österholm, P. L. Ståhl, E. Sundström, E. Åkesson, O. Bergmann, M. Bienko, A. Månsson-Broberg, M. Nilsson, C. Sylvén, and J. Lundeberg, "A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart," Cell **179**, 1647–1660.e19 (2019).

[185] R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, and N. Yosef, "Deep generative modeling for single-cell transcriptomics," Nature Methods **15**, 1053–1058 (2018).

[186] D. Grün, L. Kester, and A. van Oudenaarden, eng"Validation of noise models for single-cell transcriptomics." Nat Methods **11**, 637–640 (2014).

[187] C. H. Grønbech, M. F. Vording, P. N. Timshel, C. K. Sønderby, T. H. Pers, and O. Winther, "scVAE: variational auto-encoders for single-cell gene expression data," Bioinformatics **36**, 4415–4422 (2020).

[188] J. Aragón, D. Eberly, and S. Eberly, "Existence and uniqueness of the maximum likelihood estimator for the two-parameter negative binomial distribution," Statistics & Probability Letters **15**, 375–379 (1992).

[189] J. R. Kettenring, "Canonical analysis of several sets of variables," Biometrika **58**, 433–451 (1971).

[190] L. Haghverdi, A. T. L. Lun, M. D. Morgan, and J. C. Marioni, eng"Batch effects in single-cell rna-sequencing data are corrected by matching mutual nearest neighbors." Nat Biotechnol **36**, 421–427 (2018).

[191] G. Pasquini, J. E. Rojo Arias, P. Schäfer, and V. Busskamp, "Automated methods for cell type annotation on scrna-seq data," Computational and Structural Biotechnology Journal **19**, 961–969 (2021).

[192] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment **2008**, P10008 (2008).

[193] V. A. Traag, L. Waltman, and N. J. van Eck, "From louvain to leiden: guaranteeing well-connected communities," Scientific Reports **9**, 5233 (2019).

[194] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," Journal of Computational and Applied Mathematics **20**, 53–65 (1987).

[195] E. Armingol, A. Officer, O. Harismendy, and N. E. Lewis, "Deciphering cell–cell interactions and communication from gene expression," Nature Reviews Genetics **22**, 71–88 (2021).

[196] D. Arneson, G. Zhang, Z. Ying, Y. Zhuang, H. R. Byun, I. S. Ahn, F. Gomez-Pinilla, and X. Yang, "Single cell molecular alterations reveal target cells and pathways of concussive brain injury," Nature Communications **9**, 3894 (2018).

[197] D. A. Skelly, G. T. Squiers, M. A. McLellan, M. T. Bolisetty, P. Robson, N. A. Rosenthal, and A. R. Pinto, "Single-cell transcriptional profiling reveals cellular diversity and intercommunication in the mouse heart," Cell Reports **22**, 600–610 (2018).

[198] M. Cohen, A. Giladi, A.-D. Gorki, D. G. Solodkin, M. Zada, A. Hladik, A. Miklosi, T.-M. Salame, K. B. Halpern, E. David, S. Itzkovitz, T. Harkany, S. Knapp, and I. Amit, "Lung single-cell signaling interaction map reveals basophil role in macrophage imprinting," Cell **175**, 1031–1044.e18 (2018).

[199] B. He, L. Bergenstråhle, L. Stenbeck, A. Abid, A. Andersson, Å. Borg, J. Maaskola, J. Lundeberg, and J. Zou, "Integrating spatial gene expression and breast tumour morphology via deep learning," Nature Biomedical Engineering **4**, 827–834 (2020).

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

[201] S. Cabello-Aguilar, M. Alame, F. Kon-Sun-Tack, C. Fau, M. Lacroix, and J. Colinge, "Singlecellsignalr: inference of intercellular networks from single-cell transcriptomics," Nucleic Acids Research **48**, e55–e55 (2020).

[202] `https://docs.scipy.org/doc/scipy/reference/generated/scipy.spatial.cKDTree.html`.

[203] E. Stephenson, G. Reynolds, R. A. Botting, F. J. Calero-Nieto, M. D. Morgan, Z. K. Tuong, K. Bach, W. Sungnak, K. B. Worlock, M. Yoshida, N. Kumasaka, K. Kania, J. Engelbert, B. Olabi, J. S. Spegarova, N. K. Wilson, N. Mende, L. Jardine, L. C. S. Gardner, I. Goh, D. Horsfall, J. McGrath, S. Webb, M. W. Mather, R. G. H. Lindeboom, E. Dann, N. Huang, K. Polanski, E. Prigmore, F. Gothe, J. Scott, R. P. Payne, K. F. Baker, A. T. Hanrath, I. C. D. Schim van der Loeff, A. S. Barr, A. Sanchez-Gonzalez, L. Bergamaschi, F. Mescia, J. L. Barnes, E. Kilich, A. de Wilton, A. Saigal, A. Saleh, S. M. Janes, C. M. Smith, N. Gopee, C. Wilson, P. Coupland, J. M. Coxhead, V. Y. Kiselev, S. van Dongen, J. Bacardit, H. W. King, S. Baker, J. R. Bradley, G. Dougan, I. G. Goodfellow, R. K. Gupta, C. Hess, N. Kingston, P. J. Lehner, N. J. Matheson, W. H. Owehand, C. Saunders, K. G. C. Smith, C. Summers, J. E. D. Thaventhiran, M. Toshner, M. P. Weekes, A. Bucke, J. Calder, L. Canna, J. Domingo, A. Elmer, S. Fuller, J. Harris, S. Hewitt, J. Kennet, S. Jose, J. Kourampa, A. Meadows, C. O'Brien, J. Price, C. Publico, R. Rastall, C. Ribeiro, J. Rowlands, V. Ruffolo, H. Tordesillas, B. Bullman, B. J. Dunmore, S. Fawke, S. Gräf, J. Hodgson, C. Huang, K. Hunter, E. Jones, E. Legchenko, C. Matara, J. Martin, C. O'Donnell, L. Pointon, N. Pond, J. Shih, R. Sutcliffe, T. Tilly, C. Treacy, Z. Tong, J. Wood, M. Wylot, A. Betancourt, G. Bower, A. De Sa, M. Epping, O. Huhn, S. Jackson, I. Jarvis, J. Marsden, F. Nice, G. Okecha, O. Omarjee, M. Perera, N. Richoz, R. Sharma, L. Turner, E. M. D. D. De Bie, K. Bunclark, M. Josipovic, M. Mackay, A. Michael, S. Rossi, M. Selvan, S. Spencer, C. Yong, A. Ansaripour, L. Mwaura, C. Patterson, G. Polwarth, P. Polgarova, G. d. Stefano, J. Allison, H. Butcher, D. Caputo, D. Clapham-Riley, E. Dewhurst, A. Furlong, B. Graves, J. Gray, T. Ivers, M. Kasanicki, E. L. Gresley, R. Linger, S. Meloy, F. Muldoon, N. Ovington, S. Papadia, I. Phelan, H. Stark, K. E. Stirrups, P. Townsend, N. Walker, J. Webster, A. J. Rostron, A. J. Simpson, S. Hambleton, E. Laurenti, P. A. Lyons, K. B. Meyer, M. Z. Nikolić, C. J. A. Duncan, K. G. C. Smith, S. A. Teichmann, M. R. Clatworthy, J. C. Marioni, B. Göttgens, M. Haniffa, C. I. of Therapeutic Immunology, and I. D.-N. I. of Health Research (CITIID-NIHR) COVID-19 BioResource Collaboration, "Single-cell multi-omics analysis of the immune response in covid-19," Nature Medicine **27**, 904–916 (2021).

[204] J. P. Bernardes, N. Mishra, F. Tran, T. Bahmer, L. Best, J. I. Blase, D. Bordoni, J. Franzenburg, U. Geisen, J. Josephs-Spaulding, P. Köhler, A. Künstner, E. Rosati, A. C. Aschenbrenner, P. Bacher, N. Baran, T. Boysen, B. Brandt, N. Bruse, J. Dörr, A. Dräger, G. Elke, D. Ellinghaus, J. Fischer, M. Forster, A. Franke, S. Franzenburg, N. Frey, A. Friedrichs, J. Fuß, A. Glück, J. Hamm, F. Hinrichsen, M. P. Hoeppner, S. Imm, R. Junker, S. Kaiser, Y. H. Kan, R. Knoll, C. Lange, G. Laue, C. Lier, M. Lindner, G. Marinos, R. Markewitz, J. Nattermann, R. Noth, P. Pickkers, K. F. Rabe, A. Renz, C. Röcken, J. Rupp, A. Schaffarzyk, A. Scheffold, J. Schulte-Schrepping, D. Schunk, D. Skowasch, T. Ulas, K.-P. Wandinger, M. Wittig, J. Zimmermann, H. Busch, B. F. Hoyer, C. Kaleta, J. Heyckendorf, M. Kox, J. Rybniker, S. Schreiber, J. L. Schultze, and P. Rosenstiel, eng"Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe covid-19." Immunity **53**, 1296–1314 (2020).

[205] Z. Zhang, K. Huang, C. Gu, L. Zhao, N. Wang, X. Wang, D. Zhao, C. Zhang, Y. Lu, and Y. Meng, "Molecular subtyping of serous ovarian cancer based on multi-omics data," Scientific Reports **6**, 26001 (2016).

[206] J. Lee, D. Y. Hyeon, and D. Hwang, "Single-cell multiomics: technologies and data analysis methods," Experimental & Molecular Medicine **52**, 1428–1442 (2020).

[207] J. D. Welch, V. Kozareva, A. Ferreira, C. Vanderburg, C. Martin, and E. Z. Macosko, eng"Single-cell multi-omic integration compares and contrasts features of brain cell identity." Cell **177**, 1873–1887 (2019).

[208] S. Maniatis, T. Äijö, S. Vickovic, C. Braine, K. Kang, A. Mollbrink, D. Fagegaltier, Ž. Andrusivová, S. Saarenpää, G. Saiz-Castro, M. Cuevas, A. Watters, J. Lundeberg, R. Bonneau, and H. Phatnani, "Spatiotemporal dynamics of molecular pathology in amyotrophic lateral sclerosis," Science **364**, 89–93 (2019), https://www.science.org/doi/pdf/10.1126/science.aav9776.

[209] S. Z. Wu, G. Al-Eryani, D. L. Roden, S. Junankar, K. Harvey, A. Andersson, A. Thennavan, C. Wang, J. R. Torpy, N. Bartonicek, T. Wang, L. Larsson, D. Kaczorowski, N. I. Weisenfeld, C. R. Uytingco, J. G. Chew, Z. W. Bent, C.-L. Chan, V. Gnanasambandapillai, C.-A. Dutertre, L. Gluch, M. N. Hui, J. Beith, A. Parker, E. Robbins, D. Segara, C. Cooper, C. Mak, B. Chan, S. Warrier, F. Ginhoux, E. Millar, J. E. Powell, S. R. Williams, X. S. Liu, S. O'Toole, E. Lim, J. Lundeberg, C. M. Perou, and A. Swarbrick, "A single-cell and spatially resolved atlas of human breast cancers," Nature Genetics **53**, 1334–1347 (2021).

[210]R. C. V. Tyser, E. Mahammadov, S. Nakanoh, L. Vallier, A. Scialdone, and S. Srinivas, "Single-cell transcriptomic characterization of a gastrulating human embryo," Nature **600**, 285–289 (2021).

[211]M. Zhang, S. W. Eichhorn, B. Zingg, Z. Yao, K. Cotter, H. Zeng, H. Dong, and X. Zhuang, "Spatially resolved cell atlas of the mouse primary motor cortex by merfish," Nature **598**, 137–143 (2021).

[212]M. Asp, S. Giacomello, L. Larsson, C. Wu, D. Fürth, X. Qian, E. Wärdell, J. Custodio, J. Reimegård, F. Salmén, C. Österholm, P. L. Ståhl, E. Sundström, E. Åkesson, O. Bergmann, M. Bienko, A. Månsson-Broberg, M. Nilsson, C. Sylvén, and J. Lundeberg, eng"A spatiotemporal organ-wide gene expression and cell atlas of the developing human heart." Cell **179**, 1647–1660 (2019).

[213]Link: `https://www.10xgenomics.com/products/spatial-gene-expression`.

[214]Z. Mousavi, M. Kourosh-Arami, M. Mohsenzadegan, and A. Komaki, "An immunohistochemical study of the effects of orexin receptor blockade on phospholipase c-3 level in rat hippocampal dentate gyrus neurons," Biotechnic & Histochemistry **96**, 191–196 (2021).

[215]Link:`https://biorender.com/`.

[216]J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009) pp. 248–255.