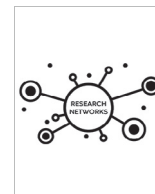


journal homepage: [www.elsevier.com/locate/csbj](http://www.elsevier.com/locate/csbj)

# Computational solutions for spatial transcriptomics

Iivari Kleino<sup>a,\*</sup>, Paulina Frolovaite<sup>a</sup>, Tomi Suomi<sup>a</sup>, Laura L. Elo<sup>a,b,\*</sup><sup>a</sup> Turku Bioscience Centre, University of Turku and Åbo Akademi University Turku, Turku, Finland<sup>b</sup> Institute of Biomedicine, University of Turku, Turku, Finland

## ARTICLE INFO

### Article history:

Received 22 April 2022

Received in revised form 18 August 2022

Accepted 18 August 2022

Available online 1 September 2022

### Keywords:

Spatial transcriptomics

Computational biology

Spatial data analysis frameworks

Single-cell analysis

Spatial deconvolution

## ABSTRACT

Transcriptome level expression data connected to the spatial organization of the cells and molecules would allow a comprehensive understanding of how gene expression is connected to the structure and function in the biological systems. The spatial transcriptomics platforms may soon provide such information. However, the current platforms still lack spatial resolution, capture only a fraction of the transcriptome heterogeneity, or lack the throughput for large scale studies. The strengths and weaknesses in current ST platforms and computational solutions need to be taken into account when planning spatial transcriptomics studies.

The basis of the computational ST analysis is the solutions developed for single-cell RNA-sequencing data, with advancements taking into account the spatial connectedness of the transcriptomes. The scRNA-seq tools are modified for spatial transcriptomics or new solutions like deep learning-based joint analysis of expression, spatial, and image data are developed to extract biological information in the spatially resolved transcriptomes. The computational ST analysis can reveal remarkable biological insights into spatial patterns of gene expression, cell signaling, and cell type variations in connection with cell type-specific signaling and organization in complex tissues.

This review covers the topics that help choosing the platform and computational solutions for spatial transcriptomics research. We focus on the currently available ST methods and platforms and their strengths and limitations. Of the computational solutions, we provide an overview of the analysis steps and tools used in the ST data analysis. The compatibility with the data types and the tools provided by the current ST analysis frameworks are summarized.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Contents

1. Introduction	4871
2. In situ spatial transcriptomics methods	4872
2.1. Strengths and limitations	4872

**Abbreviations:** AOI, area of illumination; Baysor, Bayesian Segmentation of Spatial Transcriptomics Data; BICCN, Brain Initiative Cell Census Network; BinSpect, Binary Spatial Extraction; BOLORAMIS, barcoded oligonucleotides ligated on RNA amplified for multiplexed and parallel *in situ* analyses; CCC, cell–cell communication; CCI, cell–cell interactions; CNV, copy-number variation; DbIT-Seq, Deterministic Barcoding in Tissue for spatial omics sequencing; DSP, digital spatial profiling; FA, factor analysis; FFPE, formalin-fixed, paraffin-embedded; FISH, fluorescence *in situ* hybridization; FISSEQ, fluorescence *in situ* sequencing of RNA; FOV, Field of view; GRNs, gene regulation networks; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; HDST, high definition spatial transcriptomics; HMRF, hidden Markov random field; ICG, interaction changed genes; ISH, *in situ* hybridization; ISS, *in situ* sequencing; JSTA, Joint cell segmentation and cell type annotation; KNN, k-nearest neighbor; LCM, Laser Capture Microdissection; LCM-seq, laser capture microdissection coupled with RNA sequencing; LOH, loss of heterozygosity analysis; MC, Molecular Cartography; MERFISH, multiplexed error-robust FISH; NMF (NNMF), Non-negative matrix factorization; PCA, Principal Component Analysis; PIXEL-seq, Polony (or DNA cluster)-indexed library-sequencing; PL-lig, padlock ligation; QC, quality control; RNAseq, RNA sequencing; ROI, region of interest; SCENIC, Single-Cell rEgulatory Network Inference and Clustering; scRNA-seq, single-cell RNA sequencing; scvi-tools, single-cell variational inference tools; seqFISH, sequential fluorescence *in situ* hybridization; sequ-smFISH, sequential single-molecule fluorescent *in situ* hybridization; SME, Spatial Morphological gene Expression normalization; smFISH, single molecule FISH; SPATA, SPATial Transcriptomic Analysis; ST, Spatial transcriptomics; ST Pipeline, Spatial Transcriptomics Pipeline; STARmap, spatially-resolved transcript amplicon readout mapping; t-SNE, t-distributed stochastic neighbor embedding; TIVA, Transcriptome in Vivo Analysis; TMA, tissue microarray; TME, tumor micro environment; UMAP, Uniform Manifold Approximation and Projection for Dimension Reduction; UMI, unique molecular identifier; ZipSeq, zipcoded sequencing.

\* Corresponding author.

E-mail addresses: [iivari.kleino@utu.fi](mailto:iivari.kleino@utu.fi) (I. Kleino), [laura.elo@utu.fi](mailto:laura.elo@utu.fi) (L.L. Elo).<https://doi.org/10.1016/j.csbj.2022.08.043>

2001-0370/© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

2.2.	Specific data processing steps with ST	4873
2.2.1.	Preprocessing and spot registration	4874
2.2.2.	Spatial segmentation	4874
3.	Spatial barcoding methods	4874
3.1.	Strengths and limitations	4874
3.2.	Specific data processing steps with spatial barcoding ST	4876
3.2.1.	Preprocessing and location matrix generation	4876
3.2.2.	Estimation of the spot-wise cell type compositions	4876
3.2.3.	Spatial segmentation	4876
4.	Regional selection spatial transcriptomics methods	4876
4.1.	Strengths and limitations	4876
4.2.	Specific data processing steps with regional selection ST	4877
5.	Advanced solutions in the analysis of spatial transcriptomics data	4877
5.1.	Available ST data analysis frameworks	4877
5.2.	Preprocessing and quality control steps	4878
5.3.	Finding patterns in spatial transcriptomics data	4878
5.3.1.	Finding cell type patterns from ST data	4878
5.3.2.	Detecting spatial gene expression patterns	4879
5.3.3.	Cellular neighborhoods, interactions, and communication	4879
5.3.4.	Interactive visualization	4880
5.4.	Summary and Outlook	4880
	Declaration of Competing Interest	4880
	Acknowledgements	4880
	References	4880

## 1. Introduction

To decipher the functions of the cellular systems and organs in multicellular organisms, one needs detailed information about the components and interactions between them at every scale level. In a homeostatic state, the cellular systems are in dynamic equilibrium. However, perturbations like immunological challenges (e.g. pathogens, autoimmune reactions), can flip the system into different states. While single-cell RNA-sequencing (scRNA-seq) provides detailed information about the gene expression profiles and their heterogeneity in dissociated cells, spatial transcriptomics (ST) links the transcriptomes to their cellular locations providing spatial context. This information can be used to lay out a map of the possible connections between cells, factors affecting the cells (e.g. signaling molecules, available nutritional resources, or pathogens), and connections to the other systems or organs at the organism level. These spatial relationships can reveal how the different cell types and genetic programs are interconnected with each other and with the surrounding environment. Thus, various ST methods integrated with the other spatial and non-spatial methods are currently paving the way to a deep understanding of the structure and function of the living organisms at the system level.

Spatial transcriptomics derives from tissue *in situ* hybridization techniques detecting single mRNA species with DNA-oligo probes. In current highly multiplexed *in situ* hybridization ST methods, thousands of mRNA species are detected simultaneously by using clever probing strategies. Currently, the widest probe sets detect expressions of over 18 000 known protein-coding genes, and the mRNA capture and sequencing-based methods can detect different mRNA species in an untargeted manner (including, but not limited to mRNAs, splice variants, lncRNAs, antisense RNAs, and structural RNAs). The transcriptomes can be analyzed from different sample types, including slices cut from freshly frozen and formalin-fixed, paraffin-embedded (FFPE) tissues. Each ST platform has its strengths and limitations, and therefore it is important to consider which platforms or their combinations are best for given research. The history of the spatial transcriptomics development is reviewed for example in Moses et al. 2022 [1] and Asp et al. [2].

The spatial transcriptomics methods can be categorized into three main approaches on the basis of how they capture and store

spatial information: *in situ* methods, spatial DNA-barcoding, and regional selection spatial transcriptomics methods (Table 1). *In situ* methods contain both the spatial information and the identity of the transcripts in the acquired images of the samples labeled with fluorescent probes [3,4] or displaying *in situ* sequencing (ISS) signals [5,6]. These methods provide a location for each individual detected transcript at the single-molecule level, and the expression profiles at the subcellular, cellular, or regional levels are decoded from the collected image data (Fig. 1A).

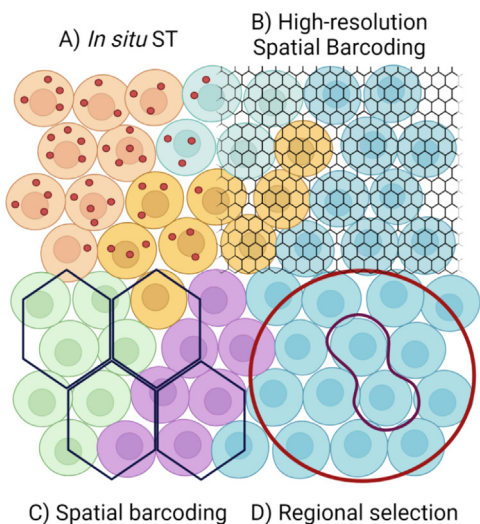
Spatial barcoding (DNA-barcoding) is a collection of methods that use mRNA capturing barcoded DNA-oligos to incorporate the positional information in a DNA format with the transcript sequences (Fig. 1B–C). The methods are based on DNA-oligos with known barcoded capture regions that form either a solid surface spot array or bead array to capture the mRNAs diffused from the tissue sample [7–10]. Alternatively, the barcoded capture-oligos are injected on the sample from arrayed inlets in a custom chip [10,11]. The positional information as DNA-barcodes is attached to the copies of the transcripts in an enzymatic reaction on the mRNA capture sites. The spatial origins of transcripts are resolved from the transcript-spatial-barcode DNA-libraries by sequencing and mapping the transcripts to the spot-level spatial coordinates with the DNA-barcodes. The size and shape of the identifiable capture region depend on the used methodology and the smallest spots in high-density arrays capture transcripts at subcellular resolution (Fig. 1B). However, these are not yet generally available, and the spots in current commonly used spatial barcode arrays cover the area of more than one cell (Fig. 1C). The transcripts and cell profiles at spots can be visualized on the image of the sample as a sample/spot image (tissue image). Computational methods are then used to resolve spot or region-level cell compositions and single or subcellular level transcriptomes from the data.

Alternative to spot spatial barcoding are regional illumination and collection methods (Fig. 1D). Region selection barcoding is accomplished by the use of photoactivatable markers and regional photo-activation with laser illumination. In different approaches, the transcripts or cells on the target regions are tagged with barcode-oligos (TIVA tag, Zip-Seq) [12,13], which requires sequential labeling of each targeted region and, therefore, strongly limits the feasible number of samples. On the other hand, these methods

**Table 1**  
Spatial transcriptomics methods.

Method	Principle	Spatial resolution	Data level	Coverage	Capture efficiency	Reference
<b>In situ ST methods</b>						
FISSEQ	RT,ISS	single molecule	subcellular	untargeted	200 UMI/cell	[5]
STARmap	PL-lig,ISS	single molecule	subcellular	1 k	2000 UMI/cell	[6]
BOLORAMIS	PL-lig,ISS/FISH	single molecule	subcellular	96	11%–35%	[18]
MERFISH	sequ-smFISH	single molecule	subcellular	10 k	60–99%	[19,20]
seqFISH+	sequ-smFISH	single molecule	subcellular	10 k	49 %	[4]
osmFISH	sequ-smFISH	single molecule	subcellular	33	NA	[21]
CosMx SMI	sequ-smFISH	single molecule	subcellular	1 k	96 %	[22]
<b>Spatial barcoding ST methods</b>						
Visium	spatial barcoding	100 μm	cell groups	untargeted	>6,9%	[23]
DBiT-seq	spatial barcoding	10–25 μm	cell level	untargeted	15.5 %	[10,11]
Slide-seq2	spatial barcoding	10 μm	cell level	untargeted	1/2 scRNA-seq	[8]
HDST	spatial barcoding	2 μm	subcellular	untargeted	1,30 %	[9]
Pixel-seq	spatial barcoding	1 μm	subcellular	untargeted	~scRNA-seq	[24]
Seq-Scope	spatial barcoding	0.6 μm	subcellular	untargeted	~scRNA-seq	[7]
Stereo-seq	spatial barcoding	0.6 μm	subcellular	untargeted	~scRNA-seq	[25]
<b>Regional selection ST methods</b>						
TIVA tag	photoactivatable tag	ROI	flexible	untargeted	na	[12]
GeoMx DSP	photo-release BC	ROI	flexible	18 k	na	[14]
LCM-seq	physical separation	ROI	flexible	untargeted	na	[16,17]
ZipSeq	photoactivatable cell-BC	ROI	single cell	untargeted	scRNA-seq	[13]
scRNA-seq	physical separation	NA	single cell	untargeted	10–40%	[26]

Abbreviations in Table 1: RT; Reverse transcription, ISS; *in situ* sequencing, lig; ligation, PL; padlock.



**Fig. 1.** Spatial transcriptomics methods produce data at different spatial resolutions. **A)** *In situ* ST methods detect selected targets at single-molecule resolution in their original location. Spatial information at molecular complex level localization is available. **B)** High-resolution spatial barcode arrays capture transcripts at subcellular resolution allowing cell organelle level localization. **C)** Lower resolution barcode arrays cover the area of more than one cell. Cell type analysis requires spot deconvolution methods. **D)** Regional illumination and collection methods offer flexibility in the target selection. The selected area can be any shape based on marker thresholding or the use of regular shapes (red or violet outline).

can be used even with living cells and can potentially be used to label cellular states differing in time in addition to location. Another method that uses light activation is the commercially available GeoMx system, which uses RNA-hybridization probe sets to detect the target RNAs. In GeoMx, the light is used to release the photocleavable indexing oligonucleotides from the probed samples in a cell-, marker-, or region-specific manner [14]. The target region can be of any shape and even discontinuous (Fig. 1D). The released probe indexing oligonucleotides are collected after each illumination and enzymatically associated with DNA-barcodes to indicate the collection set and through that the original region of

the transcripts in the sample. The laser capture microdissection (LCM) has also been used to capture the individual cells or regions from the samples for LCM coupled RNA-sequencing (LCM-seq) [15,16]. Similar to GeoMx, transcripts from each collected cell or cell group are barcoded before sequencing or they can be processed as independent sequencing libraries [17].

Already now, the insights gleaned with the developing ST platforms have shown the value of spatial transcriptomics in learning detailed information about the heterogeneity of the cells and interconnectedness of the transcriptomes in multicellular systems. However, the availability of a myriad of platforms to produce large and complex omics datasets raise the requirement of advanced computational tools and skills needed to analyze the data. Careful experimental design and data analysis planning can streamline the production of high-quality findings. The recent reviews on spatial transcriptomics and multiomics [1,27–32] cover the history, technology, and advances in the analysis in more detail; this review focuses on covering ways to help in choosing the suitable platform and analysis framework for spatial transcriptomics studies. We focus on the currently widely (and commercially) available platforms, their limitations, the available data analysis tools, and their suitability for the collection of different types of data. The state-of-the-art systems, not commercially available, are mentioned when appropriate and when expected to improve the methodology in the near future. The current computational ST analysis steps and frameworks are summarized, supplemented with optional standalone analysis options.

2. In situ spatial transcriptomics methods

2.1. Strengths and limitations

The sequential single-molecule fluorescent *in situ* hybridization (sequ-smFISH) and the *in situ* sequencing (ISS) offer the highest spatial resolution among the ST methods by detecting individual transcript molecules in the samples at the optical resolution of the microscope system (Table 1, Fig. 1A). Of the ST methods, the sequ-smFISH methods have also the highest sensitivity, as they can detect transcripts with over 80% efficiency even when thousands of RNAs are targeted [3,6,33–35]. However, the *in situ* ST

methods are not quite yet detecting different mRNAs at transcriptome coverage level like the spatial barcoding ST methods based on polyA targeting mRNA capture.

A particular strength of *in situ* ST methods is their exceptional resolution and capture efficiency. The high capture efficiency allows reliable detection of even very lowly expressed transcripts. A targeted design of probe sets focusing on, for example, signaling pathways or transcription and genomic imprinting regulation has proven a great strategy with *in situ* ST. A prime example of this is a study of the early T cell progenitor development [36], in which Zhou et al. designed a probe set to analyze the expression of otherwise elusive transcription factors and master regulators expressed in progenitor T cells. In combination with a scRNA-seq, the study was able to provide insight into the synchronized and asynchronized developmental patterns in gene expressions during the early T cell development. The study also elegantly demonstrated the feasibility of sequential use of seqFISH, smFISH, and immunostaining on the same sample. However, it should be noted that the signals from transcripts degrade each FISH cycle [21] setting limits to the number of successive probing rounds and sets the preferential probing order to start from the weakest expressing transcripts. Indicating the power of integration of *in situ* ST and scRNA-seq methods, seqFISH was used to identify the exact location of scRNA-seq characterized cell types in mouse organogenesis [37]. The Brain Initiative Cell Census Network (BICCN), a collaborative effort to produce reference brain transcriptome atlases is showing *in situ* ST use in a very large-scale cell mapping project [38].

MERFISH and seqFISH have a single-molecule level resolution which allows analysis of subcellular transcript distributions at the single-cell level adding on top of single-cell transcriptome profiling. The spatial distribution can be used to detect single-cell level features like mRNA targeting, cell polarization, and localization of specific molecular complexes [4]. For functional single-cell transcriptome analysis, the nuclear to cytoplasmic ratio of immature and mature transcripts can be used to infer the regulatory states of genes at the single-cell level through the so-called RNA velocity analysis [33]. RNA velocity data can be used, for example, to construct detailed transcript regulatory profiles and to organize cells to pseudotime trajectories to analyze, for example, cell cycle progression [33]. Showing the use of sequential smFISH at the molecular complex level, Takei et al. resolved the functional architecture of cell nucleus in detail by adapting seqFISH+ method to co-detect localization of 3660 chromosomal loci with 17 functional chromosomal markers, and 70 selected mRNA species [39]. The study indicated a good general correlation between gene activity, localization to different functional nuclear zones, and association of functional chromosomal markers. However, at the single-cell level discordant localization of individual genes with active nuclear zones and the actual transcriptional activity of the genes suggested slower temporal dynamics for a chromosomal spatial organization than for the gene activity regulation.

A major limitation with the *in situ* ST methods is that the optical crowding poses a problem with very high-plexed probe sets as the detection efficiency degrades along with the number of the target mRNAs. This technological limitation is particularly noticeable with ISS methods as the molecular amplification used for signal enhancement increases the optical size of the targets [5,6,18]. Detection of transcriptome level number of targets would require resolving up to hundreds of thousands of transcripts per cell. The current state-of-the-art seqFISH+ with 60 pseudocolor probe set combined with a clever transcript encoding scheme, optimized sequential probing, and adoption of confocal microscope platform for detection allowed detection of over 30 000 individual transcripts with ~10 000 genes in cultured fibroblasts [4].

The number of simultaneously detected RNA targets in *in situ* ST is not limited only by the physical limitations set by the used

microscopes. The sensitivity of smFISH is based on the number of *in situ* hybridization probes (ISH) targeting a given gene. For example, seqFISH+ needs the target RNA length to be at least 1 kb to accommodate the *in situ* hybridization probes for sensitive detection and encoding. This leaves out for example short mRNAs and many other short RNA species including miRNAs from the target repertoire. In MERFISH probes, the use of branched-DNA increased the detection sensitivity of MERFISH technology, and it can now detect even short 100–200 nucleotide long RNA targets with up to 60% efficiency and long RNAs with almost perfect efficiency [20]. This also allows the detection of a dramatically wider variety of RNA species including mRNA alternative splice variants. Remarkably, the increased sensitivity did not increase the optical size of the fluorescent pixel and with adoption of expansion microscope with more sensitive probes Xia et al. were able to measure the expression of around ~10 000 different RNAs simultaneously in cultured cells [33].

The ISS-based ST methods can detect very short RNA species, and the recently published new padlock ligation (PL-lig) based smFISH/ISS ST method BOLORAMIS can detect even miRNAs and single point mutations in cellular RNAs [18]. The sensitivity and specificity of BOLORAMIS and STARmap allow the detection of even different alternative splice forms of transcripts. However, the use of probes to cover transcripts in an untargeted manner is not feasible due to the cost of probes and without a technological leap in microscopy or probe design, the methodology is still limited by the optical crowding. The reverse transcription capture-based ISS methods [40,41] can detect RNA molecules in an untargeted manner but due to very low capture efficiency, they detect only a low number of transcripts per cell compared to other ST methods (Table 1).

The *in situ* ST methods are limited also by the speed of the imaging. In both ISS and sequential smFISH, the transcripts are identified by sequentially imaging fluorescence signals using a microscope. Each cycle involves FISH probing or re-probing and imaging steps. The speed of imaging of the multiple FISH cycles with a high-resolution microscope is thus a bottleneck limiting the sample and detection area throughput with the *in situ* ST methods. The use of 60 pseudocolours for encoding transcripts in seqFISH+ lowered the needed FISH cycles to four (with the included error correction cycle) for theoretical detection of the expression of up to 24 000 genes [4]. This methodological advancement shortened the sample processing time to 1/8th (imaging and hybridization) from the original in seqFISH [34]. Commercial platforms with automated liquid handling and imaging and with the use of optimized probe sets for *in situ* ST are expected to improve the throughput and increase the availability of *in situ* ST for researchers.

## 2.2. Specific data processing steps with ST

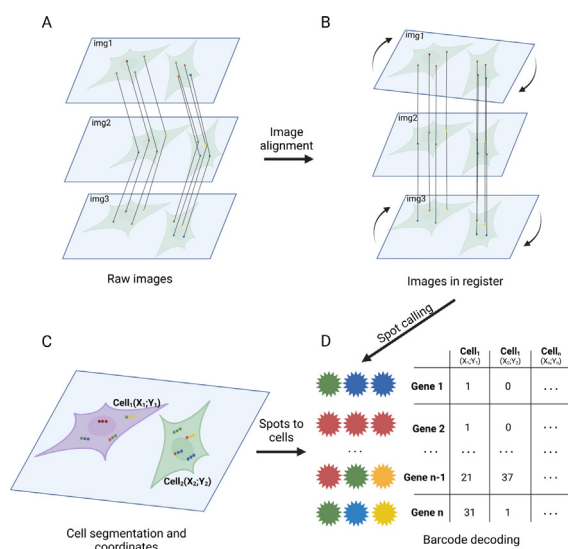
Deriving cell-by-gene and cell-by-location matrices for ST analysis from the large set of raw *in situ* microscopy image data takes several computational steps. For example, in a 69-bit 10 000 gene MERFISH experiment, for each of 256 tiled fields of views (FOV), 23 rounds of three color fluorescent images were taken at six different focal z-planes with an additional single image at fiducial bead z-plane [33]. This is 111 872 single-channel microscope images to process into a form that can be used to extract 69-bit transcript codes and locations to construct cellular transcriptome profiles. This section covers the *in situ* ST-specific data analysis steps from the raw image data to the spatial cell type pattern analysis. The more advanced analysis steps commonly used with many ST dataset types are covered in Section 5.



### 2.2.1. Preprocessing and spot registration

The probed transcripts show in raw images as signal spots that, due to limitations in automated microscopy imaging and due to chromatic aberrations, are not necessarily in the same position (in register) in the sequentially imaged FOVs (Fig. 2A). Therefore, the images in each FOV are aligned using cross-correlation of the fiducial marker peak signals or nuclear stains (Fig. 2B) [4,6,42]. Usually, the signal aligning and enhancing steps also include chromatic aberration and illumination correction with microscope setup specific control images, image deconvolution, and background subtraction [43]. For a composite representation of the whole imaged sample, the single or multi-channel FOVs can be stitched together with a process guided by the overlaps in the FOV tiles [28,44].

In aligned images, each positive spot/pixel/voxel is a potential transcript, with a location in the pixel coordinate system and transcript identifier encoded by the levels of the sequential image channels. The transcript spots are called by finding the local maxima of the images and by selecting the values that are above a certain pixel threshold [45]. The barcodes are then extracted as per location signal strings for decoding (Fig. 2D) [3,42]. Details of the above steps and the barcode decoding with error correction and spot quality control (QC) heavily depend on the used microscope setup and *in situ* ST method (MERFISH, seqFISH or ISS). Readers interested in these details are referred to the procedures explained in the original research referenced in Table 1. The decoded transcripts are arranged into a gene-by-location matrix containing the identifiers and 2D or 3D location of every identified transcript in the used coordinate system.



**Fig. 2.** Preprocessing of raw *in situ* image data. **A)** Image alignment is required since the corresponding signal spots in raw images from sequential probing and imaging cycles (img1, img2, img3) are not in the register in shared Euclidean space. To align the spots, the images are moved and rotated in relation to each other. **B)** In the aligned sequential data, the corresponding signal spots are in the register in the whole image stack, and they form the sequ-FISH barcode. **C)** Cell segmentation assigns every location in the image to defined cells, nuclei, or background. Transcripts are assigned to cells based on their spatial coordinates in relation to cell mask coordinates. Cells are also assigned with spatial coordinates (X, Y) in the same Euclidean space. **D)** Connected strings of signal spots, which are called from the image stack in panel B, are the barcodes to identify the transcript/gene at that particular coordinate location (left). The gene identities are decoded from the barcodes and counted into the cell with an overlapping coordinate location in the cell to gene matrix (right).

### 2.2.2. Spatial segmentation

To construct single-cell or other spatial unit transcript profiles in the sample, segmentation and counting steps are performed to assign the detected transcripts into desired functional spatial units. Segmentation masks defining coordinates of the spatial units are done by tracing the shapes of the targeted features from the acquired images or algorithmically with the spatial distribution of the transcripts (Fig. 2C). Reflecting the challenges in the segmentation of different biological samples, several different computational methods and strategies have been developed for segmentation. These start from simple manual segmentation by marker thresholding and end with various automated machine learning utilizing strategies [43]. In a standard cell segmentation, the specific markers are used to divide the FOV area into nuclear, cytoplasmic, and empty regions. The individual cells are delineated by selecting the nuclei of each cell as a center and then propagating the cytoplasmic area around the nuclei. Different computational methods are employed to estimate the extent and shape of cell boundaries mathematically or by using marker signals or both. The various machine learning-based segmentation tools, like U-net [46], DeepCell [47], Mesmer [48], and CellPose [49], use ST images and pre- or post-stain marker data to predict spatial segmentation [50]. In cases where only transcript signals are present, the cell segmentation or segmentation-free cell transcriptome identification relies on algorithms that use spatial transcript distributions [51] or annotated transcript profiles are used in combination with the ST data to jointly segment and annotate the cell types, as in the recent methods SSAM [51], Baysor [52], and JSTA [53]. Finally, the identified transcripts within the spatial unit regions are counted based on the segmentation masks to construct the cell-by-gene and cell-by-location matrices for ST data analysis (Fig. 2D).

The image processing, segmentation, decoding, and counting steps can be performed by scripting with Python, R, or MatLab image processing packages or modules. The Starfish [54] and its fork SMART-Q with multiomics capabilities collect python tools and scripts for building pipelines to process images and get the cell-by-gene and cell-by-location matrices from raw microscope images of different *in situ* ST methods. Other options offering useful functionalities for image processing, spot calling, and cell segmentation include PySpots [53,55], Cellpose [49], and FISH-quant [56] in python, and EBIImage [57] and imager [58] in R. Of the complete ST data analysis frameworks Squidpy [59] has cell and nuclei segmentation capabilities available (Table 2).

## 3. Spatial barcoding methods

### 3.1. Strengths and limitations

Spatial barcoding ST methods are based on the limited diffusion of the target RNA molecules and their capture by the position-indicating barcoded DNA-oligos (Fig. 1B–C). Compared to *in situ* ST methods, spatial barcoding ST excels in availability, simplicity, and untargeted mRNA capture<sup>1</sup>. Nonetheless, it is behind in resolution and capture efficiency compared to many *in situ* ST methods (Table 1). Both the resolution and capture efficiency in spatial barcoding ST methods have increased gradually since early spot arrays [23] and custom microarrays [63], in which the center-to-center distance was in the 200 to 100-micron range. In these first-generation arrays, each “spot” captured transcripts from multiple cells, which complicated the cell and spatial data analysis.

<sup>1</sup> New strategies are developed to widen the functional RNA species being captured and sequenced in scRNA-seq analysis. [60–62].

**Table 2**  
Spatial data analysis frameworks.

Package	Giotto	Seurat	STUtility	SPATA2	Squidpy	scvi-tools	stLearn	GeoMx tools
Platforms	R/Python	R	R	R	Python	Python	Python	R
Input data	ig,mt,lc	ig,mt,lc	ig,mt,lc	ig,mt,lc	ig,mt,lc	mt,lc	ig,mt,lc	mt,lc
Datacontainer	Giotto	Seurat	Seurat	SPATA	Adt,img	Adt	Adt	S4
ST data types	is,sb	is,sb	sb	is,sb	is,sb	is,sb	is,sb	gmx
Spatial segmentation					••			
Nuclei count					•			
QC and preprocessing	•	•	•	•	•	•	•	•
Descriptive statistics	•	•	•	•	•	•	•	•
Dimensionality reduction	•••	••••••	••••••	••	•••	••••	••••	••
Cell/spot clustering	••••	••••	••••	••	••••	••••	••	•
Data visualizations	•	•	•	•	•	•	•	•
Factor analysis			•			•		
Differential expression	•	•	•	•	•	•	•	•
Cell type annotation						•	•	
Deconvolution	••	•	•		•	•••		•
Reverse deconvolution								•
Cell type signature inference	•							•
Spatial representations	•••	•	•	•	••	•	•	
Genes with spatial patterns	••••	•	•	•	•			
Spatial domains	••		•		•			
Cell neighborhood analysis	•				•			
Neighbor dependent genes	•				•			
Cell-cell interaction	•				•		•	
Ligand-receptor analysis	•				•		•	
Intergroup gene expression	•			•				
GSEA and GSVA				•				
CNV estimation				•				
Spatial visualization	•	•	••	•	•	•	•	
Interactive visualization	•	•	•	•	•		•	
Interactive annotation			•	•			•	
Image analysis	•		•		•			
Features extraction images					•			
Deep learning analysis					•	•	•	
Cell trajectory analysis				•			•	

Note: Each dot represents single method for the task. Squidpy and scvi are build on top of Scanpy. Table 2 abbreviations see 2.

Recently, the spatial resolution in spatial barcoding arrays has increased to a level of single-cell size and beyond. In Slide-seq, this was obtained by micro ball-arrays [8,64] and, in DbiT-Seq, by chips with microchannels injecting the spatial barcode oligos [10]. The spatial barcoding reached the subcellular resolution with high-definition spatial transcriptomics (HDST) oligo arrays [9]. The highest resolution and the best capture efficiencies have so far been achieved with Seq-Scope, Stereo-Seq, and Pixel-seq [7,24,25]. These recent high-resolution and high-density DNA-oligo spatial barcoding arrays not only increased the resolution to the sub-micron level, but the mRNA capture efficiency has also improved to a comparable level with the droplet-based scRNA-seq methods (Table 1). This allows robust untargeted detection of medium expression level genes at the single-cell level and even between the cellular compartments.

The HDST and Seq-Scope studies demonstrated that the high-resolution arrays can locate even rare cell types and resolve the gene expression differences at subcellular resolution, which makes for example the nuclear-to-cytoplasmic type of RNA-velocity of analysis feasible for spatial barcode ST methods [9,25]. The preprints showcase the capabilities and limitations of Stereo-Seq with tumor leading-edge samples [65] and high-resolution spatial transcriptome atlases produced from regenerating axolotl brains and the developmental stages of the mouse, zebrafish, and fruit fly embryos [7,66–68]. Already these detailed transcriptome atlases, after being released into the public domain, could provide a rich resource of transcriptomic data to analyze the developmental process and brain regeneration of multicellular organisms. It will also be a great resource for testing and developing bioinformatics, data storage, and data handling methods for the analysis of such large spatial transcriptomics datasets.

The high-density spatial barcoding arrays are a big developmental step towards a universal method in ST, even though the Seq-Scope, Stereo-seq, Pixel-seq, HDST, and DbiT-Seq are still behind in resolution and capture efficiency compared to the *in situ* ST methods. However, wide adoption of high-density spatial barcoding arrays may be bottlenecked by the array production and array indexing as these require special techniques like customized use of Illumina sequencer [7], custom microchip production [10], or highly optimized arraying of oligos on to a polyacrylamide gel matrix [24]. Even though the state-of-the-art spatial barcoding methods are not yet widely available, they have shown that the fast pace in the development of ST is still going on strong and that by optimizing spatial barcoding, it is possible to acquire untargeted high gene coverage subcellular resolution data.

The latest high-density spatial barcode arrays are not yet generally available for researchers, and hence only pioneering studies have been published. Nonetheless, the spatial barcoding methods with lower resolution arrays have been available for some time and several studies have indicated their power for biological discoveries. The lower resolution spatial barcoding methods have been successfully used to dissect regional gene expression profiles in homeostatic tissues and developmental settings. For example, Slide-seq was used to produce an ST atlas of the mouse testis to decipher the complex organization of mammalian spermatogenesis at an unprecedented level [69]. The spatial transcriptomics with different pattern detection algorithms has allowed a detailed characterization of the normal gut functions and the functions in inflammatory disease [70,71]. The spatial transcriptomics has also revealed many novel features in various tumor samples [63,72–75]. These studies have, for example, revealed regional enrichment patterns for cancer cell subtypes and co-enrichments of cancer

cells with different subtypes of non-malignant cells [63]. These regional gene expression patterns in turn help to identify candidate signaling pathways and mechanistic interactions between cancer, stromal, and immune cells and to identify prognostic markers [76,77].

### 3.2. Specific data processing steps with spatial barcoding ST

In spatial barcoding, the transcript and the location data are in a DNA format. The sequencing step converts this into a digital format that can be processed computationally. This section covers the spatial barcoding-specific data analysis steps from the raw sequence data to the spatial cell type pattern analysis. The more advanced analysis steps commonly used with many ST dataset types are covered in Section 5.

#### 3.2.1. Preprocessing and location matrix generation

The preprocessing steps of the spatial barcoding raw sequence data are relatively straightforward and similar to scRNA-seq data. In spatial transcriptomics, instead of a gene to droplet barcode matrix, a gene to spot matrix is generated. The information to build the gene expression (transcript sequences) to spot position (spatial barcodes) matrix is in the paired-end sequencing reads. To identify the expressed genes, the sequences are aligned against an annotated reference genome using, for instance, STAR aligner [78] or Kallisto pseudo-aligner [79]. A second alignment step against a decoy genome can be used to filter out unwanted contaminating sequences [80]. Each sequence pair also contains the spatial barcode sequence and a unique molecular identifier (UMI), which is used to remove the PCR copies of the captured transcripts arising during sequencing library preparation. The expression levels of the genes are counted from the deduplicated, aligned, and barcode-associated reads, and a spatial barcode-by-gene matrix is generated (analogous to Fig. 2D). STARsolo [81], bustools [79], ST Pipeline [80], Spaceranger count, and Slide-seq/drop-seq tools [69,82] are commonly used solutions to produce the spatial barcode-by-gene matrix.

The spatial barcode-by-location data connects each transcript to a location coordinate in the sample. The coordinates are used to construct spatial relationship graphs and grids for the spot or cell interconnection analyses, assign transcripts to cells after segmentation with subcellular resolution data, and for visualizations and joint analyses of the spots and different features on the associated tissue images. For instance, in the Visium platform, the barcode sequences and their positions in the spatial grid are fixed and the Spaceranger count creates a barcode-by-location matrix for spatial analyses. In spatial barcode methods with stochastic barcode spot locations, the spot barcode-by-location matrix is constructed in method-specific array sequencing and indexing step [7–9,24,25]. To visualize or jointly analyze transcripts with tissue images, the barcode coordinate system needs to be aligned with the one used with the tissue images. The Spaceranger count detects the spot positions (fiducial detection) from the bright-field image and creates data for the spot-image alignment. The spatial barcode-spot alignment with the images can also be done interactively [7,9] or by using custom scripts [8].

#### 3.2.2. Estimation of the spot-wise cell type compositions

The spatial barcoding will locate the transcripts at multi- or subcellular spatial resolution. In multicellular resolution spatial barcoding, each spot can contain transcripts from multiple cells. The cell compositions and regional enrichment of different cell types and cell stages in the spots can be resolved with computational deconvolution, mapping, enrichment, and data-integration-based methods [83,84]. For instance, SPOTlight [85] uses non-negative matrix factorization (NNMF) and SpatialDecon [86]

log-normal regression for deconvolution of the transcriptomics data, whereas Cell2Location is based on a Bayesian model [87] and Tangram is a deep learning framework to resolve cell types. The ST-framework Giotto (discussed in section 5.1.) uses enrichment-based options for cell type composition analysis [88] and Seurat anchor-based integration [89]. The spot compositions are reported as cell type proportions or probabilities of occurrence in the spot. The latest deconvolution methods can also report the estimated transcriptomes of the identified cell types. Most deconvolution methods use annotated reference transcriptome profiles derived from scRNA-seq or bulk RNA-seq datasets, making the accuracy and resolution of the cell type identification highly dependent on the compatibility of the used reference profiles with the cells in the target sample. However, the recently published conditional autoregressive-based deconvolution (CARD) [90] and the latent Dirichlet allocation based STdeconvolve [91] offer reference-free deconvolution methods, which is useful when optimal reference scRNA-seq profiles are not available.

#### 3.2.3. Spatial segmentation

The objective of segmentation is to construct masks to assign the detected transcripts to individual cells, nuclei, or larger objects. The segmentation with spatial barcoding data is essentially the same process as with *in situ* based ST methods and is based on the stained image data of the target tissue in combination with the actual gene signals. Usually, ST data have associated image data with hematoxylin and eosin (H&E) staining, indicating the location of nuclei and selected structural features. In some cases, fluorescent staining for different cellular markers is available to guide the segmentation. Despite intense development, the segmentation of tissues with densely packed cells is still one of the most challenging steps in the analysis of spatial data. The decisions for segmentation strategy and methods depend heavily on the sample type, the platform used, and the available data to guide the segmentation. Several methods varying from manual assignment to statistical, supervised, semi-supervised, and unsupervised machine learning applications have been developed for the task with other data types (see *in situ* ST data segmentation). So far, the high-resolution spatial barcoding transcriptomics has not been widely available, and a simple grid segmentation has been used in some of the studies [7,65]. We anticipate the development of deep learning models for feature segmentation and cell type identification with the large spatial barcoding ST datasets.

## 4. Regional selection spatial transcriptomics methods

Laser capture microdissection coupled with RNA-sequencing (LCM-seq) and digital spatial profiling (DSP) are methods that use flexible regions of interest (ROI) binning in spatial transcriptomics [14–17]. Each selection bin contains transcripts from one ROI and they are barcoded for backtracking of the location in the sample. After sequencing and processing, the bins are used similarly to the spots in spatial barcoding arrays, except for that the regions can vary in size, shape, and can be discontinuous (Fig. 1D). Depending on the researchers' choices, each bin can contain transcripts from one or more cells and the ROIs can be selected, for example, to be homogenous in size and shape, cell type, cell marker, or consist of a functional region in the sample [14,92].

### 4.1. Strengths and limitations

Recently the DSP, under the commercial name GeoMx, has gained popularity as a nearly whole transcriptome level ST method with flexible ROI selection that works also with FFPE samples. In

GeoMx, the samples are hybridized with analyte probes detecting RNA or proteins, which upon laser illumination release their oligonucleotide tags with the identifiers of the target transcripts or proteins. The oligonucleotide tags are collected after each round of illumination and enzymatically barcoded for the collection bin identification. In LCM-seq the transcripts collected to each bin are barcoded by using barcoded DNA-oligos in reverse transcription reaction. In both methods, a DNA library containing the pooled bins is sequenced and the transcript count-to-bin matrix is resolved.

In GeoMx up to four fluorescent markers can be used to guide the sampling or ROIs can be selected based on the physical properties of the sample like distance from the feature of interest and/or by following the shape of the biological unit of interest. In LCM-seq the choice of selection markers depends on the used LCM-seq instrument. The strategies for ROI selection/sampling vary and are decided by the researcher based on prior knowledge and suitability for the research setting [14,92]. The flexible sampling allows the development of complex sampling strategies including for example selection of cell type-specific ROIs for profiling and then use of these in deconvolution of the mixed cell ROIs in the same sample. The available data analysis options are limited by the sampling strategy, hence the markers and the cell pooling to bins at the sampling phase should be carefully thought out before the experimental phase [92].

The spatial resolution of the GeoMx and LCM-seq data is determined by the sampling plan. The ROIs can be as small as single cells; however, the throughput in GeoMx and LCM-seq does not allow single-cell analysis on a scale comparable to *in situ* or spatial barcoding ST methods. GeoMx is also limited by the reliable quantification of transcripts, which in DSP requires 20 to 300 cells per ROI [14] and is thus below the current other ST methods. The transcript detection efficiency in LCM-seq depends on the sequencing library generation method and can be at the highest on the level with tube-based scRNA-seq. GeoMx does not offer untargeted mRNA capture like LCM-seq and spatial barcoding ST. The largest currently available pre-designed probe sets for GeoMx can detect over 18,000 protein-coding genes with a possibility for customization and co-detection of a number of protein targets. An advantage of the GeoMx probing protocol is that it is minimally destructive, and the tissue slides may be used for other applications like H&E staining or immunohistochemistry for retrieving additional information [14]. The GeoMx can shine in the analysis of sample cohorts in which only a limited number of biological compartments or cell types need to be sampled to answer the research questions. For example, GeoMx DSP results have revealed spatial heterogeneity in gene profiles in the host response to SARS-CoV-2 infection in lung samples [93]. DSP has also been used to profile diabetic foot ulcers [94], detect alterations in diabetic kidneys [95], and analyze intra-tumor and inter-tumor heterogeneity in various tumor types [96–99]. With the aid of deconvolution, GeoMx data was used to estimate cell level heterogeneity of tumor-infiltrating lymphocytes in different tumor microenvironments [86]. LCM-seq showed its usability in a study where neurons were collected from different regions of the brainstem and spinal cord for the identification of genes protecting neurons from spinal muscular atrophy associated cell death [100].

#### 4.2. Specific data processing steps with regional selection ST

The cells or regions for analysis are usually pooled at the ROI selection phase in GeoMx and LCM-seq either due to detection efficiency or throughput limitations. Unless carefully sampled, the ROIs will have cells in different stages or consist of heterogeneous cell types. Hence, the analysis depends heavily on the choices taken at ROI selection. In its simplest form, the transcriptomic pro-

files of GeoMx and LCM-seq ROIs are compared to each other to identify differentially expressed genes. GeoMx proprietary analysis software offers tools for basic analysis. For custom analysis, the GeoMx data can be converted from proprietary file formats to the more accessible spatial data format with the GeoMx tools package in R, which offers the capability for filtering ROIs and probes with quality control parameters associated with sequencing, alignment, and negative control probes included in the pre-designed probe sets. The GeoMx tools package also includes tools for normalization, dimensionality reduction, clustering, differential gene expression analysis, and options for visualization, such as UMAP, t-SNE, and volcano plots. Like spatial barcoding ST data, the GeoMx and LCM-seq multicell ROIs can be analyzed for cell compositions with deconvolution tools, such as the SpatialDecon R package for constrained log-normal regression-based deconvolution of GeoMx and other ST data [86]. For cell type prediction SpatialDecon uses provided cell type signatures or custom ones that can be inferred from RNAseq and scRNA-seq datasets. It can also form new cell profiles “on the fly” from pure cell ROIs and use tumor cell-type-specific ROIs to identify confusing genes in the target cell profiles to aid deconvolution of non-tumor cell types. To enhance estimates, SpatialDecon uses background counts of decoy probes and can also use nuclei counts to estimate total cell counts. For neighborhood analysis, the SpatialDecon includes tools to model the gene expression profiles and up and downregulation of genes in ROIs by using the estimated cell counts with a so-called reverse deconvolution method. Additional computational analyses can be done with custom scripts or other available ST-tools after the conversion of data to a suitable format. However, advanced analysis using spatial relationships (see Section 5) can be performed with the GeoMx and LCM-seq spatial datasets only if the spatial coordinates are resolved for the ROIs.

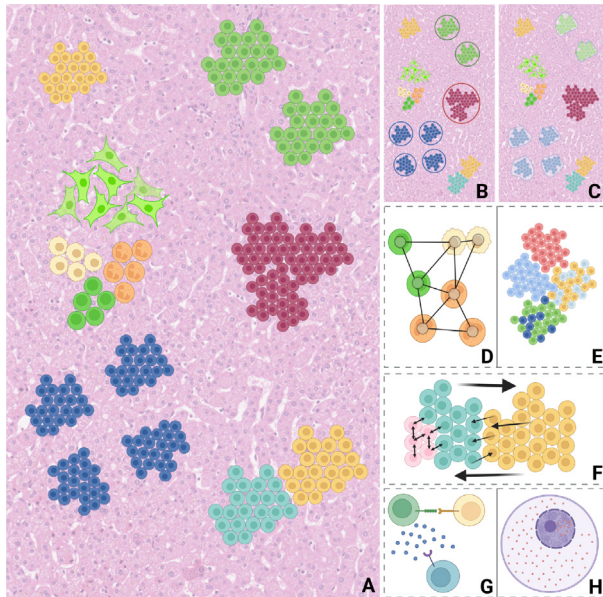
### 5. Advanced solutions in the analysis of spatial transcriptomics data

The common objectives in spatial transcriptomics data analysis include identification, quantitation, and annotation of the spatial patterns (domains) at multicellular, cellular, and molecular levels, and the statistical analysis of these quantitated features at different length scales (Fig. 3). Many processing and analysis steps with the spatially resolved single-cell transcriptomes are identical to common scRNA-seq analysis steps and the same tools can be used taking into account the method-specific confounding factors and limitations in the ST data. The following sections cover the commonly used data analysis steps and some advanced computational solutions to analyze biological phenomena in spatial data. We focus on methods specialized for ST data analysis. While a detailed description of all the methods mentioned here is beyond the scope of this review, more detailed descriptions of the ST data analysis methods can be found in the original research articles or the recent reviews [27–29,83,101,102].

#### 5.1. Available ST data analysis frameworks

Common scripting languages, such as R and Python, and various open-source data analysis and visualization packages can be used to build ST data analysis and visualization pipelines. Recently, some consolidated analysis solutions have been released which enable efficient and reproducible ST data analysis without extensive programming and scripting work. Different computational frameworks combine variable sets of commonly used and new scRNA-seq and ST data analysis tools chosen to work together. For a quick reference, the general capabilities in the analysis frameworks Giotto [88], Seurat [89,103], STUtility [104], SPATA2 [105],





**Fig. 3.** Overview of the information that may be extracted from spatial transcriptomics data. **A)** Overall cell type distribution in tissue space shows specialized cell type patterns that reflect communal functions and regulation. **B)** Cell type-based spatial patterns and **C)** domains can be detected based on transcriptome clustering. **D)** Cellular relationships are often represented with spatial graphs of cells. The gray circles (nodes) represent the cells, while the black lines (edges) correspond to the distances between the cells. **E)** Cell neighborhood analysis identifies spatially connected cell type pairs in spatial domains. **F)** Cell–cell communication and interactions happen at many length scales, illustrated by the black arrows. **G)** Cell–cell communication and interactions between receptors and ligands are often detected using curated ligand–receptor lists. The green-colored cell is releasing ligands (blue dots), while the green cell is communicating with the yellow-colored cell directly by receptors. **H)** Subcellular distribution of transcripts (orange dots) can be used in ST analysis, to understand the cellular scope.

Squidpy [59], scvi-tools [106], stLearn [107], and GeoMx tools [108] are summarized in Table 2.

At the time of writing, Giotto, SPATA2, Squidpy, scvi-tools, and stLearn can import many different *in situ* and spatial barcoding ST datasets for analysis. STUtility is limited to spatial barcoding type of data and Seurat can import Vizgen MERFISH, GeoMx, and spatial barcoding types of data at the moment. GeoMx tools is specialized for GeoMx DSP type of data analysis and covered in the previous section. There is some overlap in the frameworks as Squidpy and scvi-tools encompass the Scanpy package, whereas STUtility is built on Seurat extending Seurat's ST data analysis capabilities.

While Giotto's main functions are run in R, it interacts with Python modules through the reticulate interface and has a wide set of tools available to statistically detect and analyze different spatial patterns. Seurat, in addition to being a versatile scRNA-seq data analysis framework, has a variety of functions for ST analysis and visualization and connectivity with other ST analysis packages in R. In addition to general ST analysis, STUtility specializes in analysis and visualization of sequential spatial sample layers in 3D. As modular Python frameworks<sup>2</sup>, Squidpy and scvi-tools, in addition to statistical analysis, can leverage the powerful Python deep learning environments by offering a standardized interface to higher-level machine learning packages. stLearn is another ST analysis framework using Python with a focus on integrative deep learning-based analysis and spatial trajectory and pseudotime anal-

yses. SPATA2 is focused on trajectory and pseudotime analyses and integrates tools for gene set enrichment and variance analysis (GSEA and GSVA) and to estimate genomic copy number variation (CNV) based on transcriptome profiling along the genomic locations [109].

## 5.2. Preprocessing and quality control steps

The data imported into the ST data analysis frameworks (Giotto, Squidpy, Seurat, STUtility, SPATA2, scvi-tools, stLearn, and GeoMx) at a minimum consists of cell/spot-by-gene matrix and cell/spot-by-location matrix, and optional associated image data of the sample tissue (Table 2). All the ST frameworks include the common preprocessing tools for normalization, filtering, and dimensional reduction used to ensure the quality and consistency of the ST data [83,110]. There is no single solution for all datasets as the data may contain different levels of variation from the confounding sources. The usual standard preprocessing steps are the same as used for scRNA-seq data, including normalization across data points, scaling, filtering of bad quality cells or spots, and low abundance genes based on the total number of detected molecules [59,104,106,111]. The scRNA-seq specific methods like SCTransformation [112], SCnorm [113], or SCRAN [114] can also be used to remove unwanted variations. A recent preview of the stLearn package showcased a deep learning-aided normalization method specialized for spatial data, which takes into account image data and the neighboring spatial data to adjust the gene expression values [107]. Usually, data is treated also with different dimensionality reducing algorithms for easing computationally heavy analysis tasks like clustering and for visualization of the variation of the data in low dimensional space.

## 5.3. Finding patterns in spatial transcriptomics data

Cells and molecules are spatially organized to optimize the structure and function of the biological units they are part of (Fig. 3A). The main goal of ST analysis is to identify spatial patterns and domains at different scales in the ST data. In many tissues and tumors, the specific cell types or cell stages form distinct spatial homo- or heterotypic cell and gene expression patterns, which can be used to identify the functional units or cell niches. Strategies starting from cell types or from gene expression patterns are used to detect spatial patterns and identify spatially distinct domains and niches.

### 5.3.1. Finding cell type patterns from ST data

A simple strategy to get an overview of the cellular patterns is to cluster and annotate the cells or spots based on their transcriptomes and then visualize the spatial variation on tissue overlay plots (spatial plots). For transcriptome-based cell type profiling, the available ST data analysis frameworks offer dimensionality reduction and clustering methods and options for visualizations as dimensionally reduced data plots and spatial plots (Fig. 3B). In addition to the widely used clustering alternatives for scRNA-seq data, such as K-means, hierarchical, Louvain, and Seurat has a specific sample clustering algorithm for ST data based on modularity tuning in the KNN (k-nearest neighbor) distance graph and STUtility has implemented NMF dimensionality reduction that is based on the recent NMF implementation [115]. Additionally, standalone methods tuned for scRNA-seq data clustering can be used, including BayesSpace [116], SC3 [117], lloReg [118], and SIMLR [119]. Of these, BayesSpace offers capabilities to infer higher resolution delineation of the spatial domains by leveraging spatial information in clustering and lloReg optimizes genes used for clustering with a probabilistic feature extraction step before clustering. Intuitive and clear descriptions of many clustering algorithms useful in ST data analysis can be found in [83]. With multicell resolu-

<sup>2</sup> Recently a scverse consortium started maintaining the Python ST-frameworks Scanpy, Squidpy, and scvi-tools [scverse.org/](https://scverse.org/).

tion spatial barcoding data, the deconvolution tools discussed in Section 3.2.3. provide probabilistic cell type compositions of the spots that can be used to divide the sample into cellular regions and determine cell neighborhood composition. (Fig. 3B).

For cell type annotation of single-cell resolution ST data, the same strategies and tools can be used as with the scRNA-seq data. Typically, the available ST data analysis frameworks have one or more differential gene expression analysis tools for marker gene profiling and manual marker-based cell type annotation. For instance, scvi-tools has two automated cell type annotation methods, and the option to use the scArches [120] transfer learning models, in addition to dataset integration, to annotate new datasets even with different modality reference data. Standalone reference-based annotation tools like SingleR [121], CHETAH [122], and scGate [123] in R, and Cello [124] in Python can be included in the ST analysis solutions for cell type annotation. Additionally, methods are available that use both the tissue image and ST data for the annotation. For example, the deep learning-based solution SpaCell [125] uses tissue image and ST data jointly for cell type and disease stage annotation. The identified cell types or unannotated cell clusters can be visually and statistically analyzed for their spatial organization by plotting on the sample image and then comparing to, for example, available tissue atlases or a pathologist's tissue annotation of the corresponding H&E stained tissues image.

Regulon-based analysis could provide further cell stage clustering by identifying coregulated genes and transcription factor networks (regulons) underlying the functional cell states. Currently, none of the ST data analysis frameworks incorporate such gene regulatory network analysis by default. For scRNA-seq data, for instance, SCENIC (Single-Cell rEgulatory Network Inference and Clustering) and pySCENIC are tools to detect regulons [126,127]. SCENIC uses the cellular regulon activity patterns to cluster the cells.

### 5.3.2. Detecting spatial gene expression patterns

At the molecular level, genes may show spatial expression patterns which reflect their cell type-specific intrinsic programs and extrinsic effects on the local cell community or tissue microdomain (Fig. 3C). Cell molecular and organellar subcellular distribution data can be used jointly with spatial transcriptomics data to provide more detailed information about gene-phenotype associations in cells (Fig. 3H) [28,29]. RNA velocity analysis based on exon-intron ratios in transcripts [128] or inferred from nuclear to cytoplasmic gene expression ratios can be used to order cells to pseudotime trajectories for cell fate analysis [33].

Many ST data analysis frameworks implement methods developed for the identification of genes showing spatial expression patterns with the aid of the spatial representations of the ST data (Table 2) [29,88]. A simple form is a spatial grid in which the average expression of genes in each grid box area is measured or calculated. Many spatial barcode methods naturally produce data in such a form as the spots are spatially organized in a grid with a fixed size, relative locations, and the number of neighboring spots, and each spot integrates gene expressions from the covered areas. For high-resolution ST data, the grid box dimensions can be tuned to computationally integrate expressions from subcellular to multicellular-sized regions. A common approach is to construct a spatial network with nodes/vertices representing the cells and links/edges representing the connections between the cells, where the edges can have associated weights derived from their physical distance or other connection metrics (Fig. 3D).

For detecting genes with spatial trends in their expression, Giotto implements their Binary Spatial Extraction (BinSpect) method, which finds genes with spatially coherent expression patterns [88], as well as the previously published tools SpatialDE

[129], SPARK [130], and trendsceek [131]. SpatialDE<sup>3</sup> uses Gaussian process regression to identify genes with spatial expression patterns, SPARK identifies trending expressions using generalized spatial linear models, whereas trendsceek uses nonparametric marked point processes to detect spatial trends in gene expression. Seurat has its own implementation similar to trendsceek, SPATA2 implements SPARK, Squidpy and STUtility use autocorrelation-based detection of genes showing patterned expression, whereas stLearn uses factor analysis based on tissue microenvironment detection to detect genes correlated with the identified factors. Additionally, standalone solutions exist, such as MERINGUE [133], which uses an autocorrelation-based analysis method, and Hotspot [134], which uses a graph-based approach to identify the most informative genes.

In addition to detecting genes showing spatially distinct expression patterns, another typical goal of ST data analysis is to detect spatial domains with coherent gene expression. These may differ from the patterns detected based on the cell type analysis [135]. In the available ST data analysis frameworks, various computational solutions are used to specify the spatial domains by summarizing the detected genes with distinct coherent gene expression patterns into metagenes. For instance, SPARK and SpatialDE use clustering to form metagenes from the detected spatial genes. Additionally, Giotto implements a hidden Markov random field (HMRF) method [135] to detect spatial domains, which uses the relationship information in the spatial graphs to compare the gene expression of each cell to its neighborhood (domain) to search for coherent gene expression patterns and to assign cells into a predefined number of spatial domains. The standalone solution MERINGUE uses cross-correlations for grouping the identified expression patterns into spatial domains and for spatially informed cell clustering [133].

The statistically detected spatial domains of coherent gene expression can then be corroborated by comparing them to domains identified from the corresponding image data, for instance, by manual annotation or deep learning-based feature detection [136]. The joint analysis of molecular and image data is leveraged in recently developed methods, including stLearn [107], SpaGCN [137], and SpaCell [125]. stLearn uses a two-step clustering method in which the clustered expression data is further clustered with the spatial information to find finer spatial domains [107]. SpaGCN integrates the image RGB pixel data along with the spatial coordinates to tune spatial expression graph weights and then uses graph convolution and iterative clustering to find spatial domains and spatial genes and metagenes with differential expression [137]. SpaCell uses deep learning-based analysis of the images of H&E stained samples jointly with the matched measured spatial molecular data to classify the cell types and disease stages on the ST samples.

### 5.3.3. Cellular neighborhoods, interactions, and communication

Understanding the compositions of cell neighborhoods and molecular level cell-cell interactions (CCI) and cell-cell communication (CCC) in tissues and organs will open the path for building comprehensive models of biological systems. Also, detection of critical molecular components of cellular connectedness would allow identification of crucial molecular pathways causing pathologies and potential ways to treat these [138]. The molecule level cell connectedness analysis has progressed with giant steps recently with the availability of dissociated scRNA-seq datasets and a number of computational data analysis tools [138,139]. Now, with spatially aware transcriptome data, modeling and inference of molecular interconnectedness of cells at different length scales is becoming reality.

<sup>3</sup> GPU enhanced SpatialDE2 has been launched [132].

Spatial cell neighborhood analysis detects preferred cell–cell adjacencies and cell community compositions that can suggest homo- and heterotypic attractive or repulsive connections between different cell types (Fig. 3E). The analysis is based on statistical enrichment of cell–cell interactions. Of the available ST data analysis frameworks, only Giotto and Squidpy offer cell neighborhood analysis by default. Statistical analysis in both is based on the enrichment method introduced in HistoCat [140], where the occurrences of cell–cell interactions are compared to the randomized baseline in the data.

Spatial information also allows the prediction of cell adhesion and cell signaling ligand–receptor pairs mediating cell connections in tissues at auto-, juxta-, para-, and endocrine length scales (Fig. 3F–G). The basis for statistical analysis is a cell–cell relationship representation and a list of known ligand–receptor and cell–adhesion ligand–receptor pairs, including also molecular compositions, as the receptors and ligands may be homo- or hetero-multimeric in nature. The ligand–receptor list can be constructed from protein interaction datasets. However, standalone tools like CellphoneDB [141,142], NicheNet [143], and CellChat [144] usually provide manually curated ligand–receptor lists for the analysis, and collections like FANTOM5 [145] and Omnipath [146] have larger ligand–receptor datasets available. Typically, the cell–communication scores are calculated for each ligand–receptor pair in all relevant pairs of cell types. The highest-scoring pairs are the most likely ones mediating connections between the particular cell types.

For instance, Giotto implements a spatially-aware method that calculates cell–communication scores associated with the cell type neighborhoods, whereas stLearn and Squidpy have their own implementations of the enrichment test of CellphoneDB [141,142], which is a method commonly used to identify important ligand–receptor pairs in scRNA-seq data analysis. Two recent standalone methods, tensor composition-based Tensor-cell2cell [147], and random forest-based MISTy [148], aim to identify significant ligand–receptor pairs simultaneously at multiple length scales (Fig. 3F). Also, spatially-aware versions of the CellphoneDB [149] and NicheNet [143], and several additional standalone spatially aware ligand–receptor analysis tools based on different statistical and machine learning methods have been released and are described in recent reviews [138,139].

#### 5.3.4. Interactive visualization

All the ST data analysis frameworks provide multiple options for production of graphs and spatial feature visualizations of the performed data analysis. Most of the frameworks offer also some level of interactive visualization for exploratory data browsing and visual inspection of the results along with the analysis. Giotto enables exploratory viewing and selection of subsets of data for reanalysis and SPATA2 has interactive tool for manual spatial trajectory drawing and annotation. Squidpy has its own image data container type and connects to Napari, a Python-based GPU accelerated image analysis software, for advanced data visualizations and image-based analysis. Squidpy allows the use of machine learning packages for feature extraction from the image data (H&E and fluorescent staining), including cell and nuclei segmentation used in subcellular spatial domain computations. Seurat and stLearn have interactive viewing options with “click and select” types of gene selection and visualization adjustments. STUtility provides tools for image trimming and for automatic and manual image alignment of consecutive sample sections of spatial barcoding data for the construction of a rotating 3D stack of the spot images. At the moment scvi-tools does not offer interactive spatial visualization integrated into the framework, however the use of anndata data container offers workaround to import of analysis

results to other software for interactive visualization. Various spatial transcriptomics visualization options are reviewed in [83].

#### 5.4. Summary and Outlook

Both the methodological and computational solutions for spatial transcriptomics are evolving at a fast pace, providing coverage to larger sample areas at higher resolution thus producing spatial datasets with increasing complexity and size. New advanced commercial spatial transcriptomics platforms allowing wide adoption of spatial transcriptomics in biological and medical research fields have been promised to be launched by the end of the year 2022. The newest ST methodologies allow the collection of nearly complete transcriptomic datasets with subcellular resolution at the organ and systems level. The collected datasets will contain a wealth of biological information about the spatial relationships of the cells that can be used to supplement, interpret, and find causal relationships for connecting the genetic programs in multicellular systems.

Methods to mine these vast datasets by combining statistical modeling, supervised and unsupervised machine learning methods, and data integration techniques with the current knowledge of signaling, metabolic and biochemical pathways, and gene regulation networks holds promise to identify complete sets of components taking part in biological processes and to modeling dynamic cellular functions in interconnected multicellular systems.

Computationally, already the size of tens of terabytes of raw data per sample holding spatial information of hundreds of millions individual transcripts in the forthcoming *in situ* datasets [22] will pose a challenge for use of many current computational analysis methods with current computational resources. On the other hand, these new datasets open an unprecedented opportunity for the development of new computational methods for mining and interpreting the vast amounts of biological and biomedical data.

The exploratory visualization and sharing of the data and analysis results in interactive formats with collaborators and in the public domain are also becoming an important part of the computational analysis of complex biological datasets. The option to share the spatial data as standalone datasets with the possibility for interactive visualization with dedicated viewers is currently not easily attainable with the available ST data analysis platforms. To provide maximum benefit for all communities, scientific and other, the vast information in spatial datasets could be provided also in interactive visualization formats.

#### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgements

Prof. Elo reports grants from the European Research Council ERC (677943), European Union's Horizon 2020 research and innovation programme (955321), Academy of Finland (310561, 314443, 329278, 335434, 335611 and 341342), and Sigrid Juselius Foundation during the conduct of the study. Our research is also supported by Biocenter Finland, and ELIXIR Finland. The figures were created with BioRender.com.

#### References

- [1] Moses L, Pachter L. Museum of spatial transcriptomics. *Nature Methods* 2022;19:534–46. <https://doi.org/10.1038/s41592-022-01409-2>.



- [2] Asp M, Bergenstråhle J, Lundeberg J. Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *BioEssays* 2020;42:1–6. <https://doi.org/10.1002/bies.201900221>.
- [3] Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015;348:1360–3. <https://doi.org/10.1126/science.aaa6090>.
- [4] Eng CHL, Lawson M, Zhu Q, Dries R, Kouloula N, Takei Y, et al. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 2019;568:235–9. <https://doi.org/10.1038/s41586-019-1049-y>.
- [5] Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Ferrante TC, Terry R, et al. Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nature Protocols* 2015;10:442–58. <https://doi.org/10.1038/nprot.2014.191>.
- [6] Wang X, Allen WE, Wright MA, Sylvestrak EL, Samusik N, Vesuna S, et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 2018;361. doi:10.1126/science.aat5691.
- [7] Cho CS, Xi J, Si Y, Park SR, Hsu JE, Kim M, et al. Microscopic examination of spatial transcriptome using Seq-Seq. *Cell* 2021;184:3559–3572.e22. <https://doi.org/10.1016/j.cell.2021.05.010>.
- [8] Stickels RR, Murray E, Kumar P, Li J, Marshall JL, Di Bella DJ, et al. Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nature Biotechnology* 2021;39:313–9. <https://doi.org/10.1038/s41587-020-0739-1>.
- [9] Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, Schapiro D, et al. High-definition spatial transcriptomics for in situ tissue profiling. *Nature Methods* 2019;16:987–90. <https://doi.org/10.1038/s41592-019-0548-y>.
- [10] Su G, Qin X, Enninfu A, Bai Z, Deng Y, Liu Y, et al. Spatial multi-omics sequencing for fixed tissue via DBIT-seq. *STAR Protocols* 2021;2. <https://doi.org/10.1016/j.xpro.2021.100532>.
- [11] Liu Y, Yang M, Deng Y, Su G, Enninfu A, Guo CC, et al. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* 2020;183:1665–1681.e18. <https://doi.org/10.1016/j.cell.2020.10.026>.
- [12] Lovatt D, Ruble BK, Lee J, Dueck H, Kim TK, Fisher S, et al. Transcriptome in vivo analysis (TIVA) of spatially defined single cells in live tissue. *Nature Methods* 2014;11:190–6. <https://doi.org/10.1038/nmeth.2804>.
- [13] Hu KH, Eichorst JP, McGinnis CS, Patterson DM, Chow ED, Kersten K, et al. ZipSeq: barcoding for real-time mapping of single cell transcriptomes. *Nature Methods* 2020;17:833–43. <https://doi.org/10.1038/s41592-020-0880-2>.
- [14] Merritt CR, Ong GT, Church SE, Barker K, Danaher P, Geiss G, et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nature Biotechnology* 2020;38:586–99. <https://doi.org/10.1038/s41587-020-0472-9>.
- [15] Chen J, Suo S, Tam PP, Han JDJ, Peng G, Jing N. Spatial transcriptomic analysis of cryosectioned tissue samples with Geo-seq. *Nature Protocols* 2017;12:566–80. <https://doi.org/10.1038/nprot.2017.003>.
- [16] Brosch M, Kattler K, Herrmann A, von Schönfels W, Nordström K, Seehofer D, et al. Epigenomic map of human liver reveals principles of zoned morphogenic and metabolic control. *Nature Communications* 2018;9:1. <https://doi.org/10.1038/s41467-018-06611-5>.
- [17] Nichterwitz S, Chen G, Aguila Benítez J, Yilmaz M, Storvall H, Cao M, et al. Laser capture microscopy coupled with Smart-seq2 for precise spatial transcriptomic profiling. *Nature Communications* 2016;7:1. <https://doi.org/10.1038/ncomms12139>.
- [18] Liu S, Puthambaker S, Iyer EPR, Ferrante T, Goodwin D, Fürth D, et al. Barcoded oligonucleotides ligated on RNA amplified for multiplexed and parallel in situ analyses. *Nucleic Acids Research* 2021;49. doi:10.1093/nar/gkab120.
- [19] Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 2015;348:1360–3. <https://doi.org/10.1126/science.aaa6090>.
- [20] Xia C, Babcock HP, Moffitt JR, Zhuang X. Multiplexed detection of RNA using MERFISH and branched DNA amplification. *Scientific Reports* 2019;9:1–3. <https://doi.org/10.1038/s41598-019-43943-8>.
- [21] Codeluppi S, Borm LE, Zeisel A, La Manno G, van Lunten JA, Svensson CI, et al. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature Methods* 2018;15:932–5. <https://doi.org/10.1038/s41592-018-0175-z>.
- [22] He S, Bhatt R, Brown C, Brown EA, Buhr DL, Chantranuvattana K, et al. High-plex Multicomic Analysis in FFPE at Subcellular Level by Spatial Molecular Imaging. *bioRxiv* 2022;2021.11.03.467020. doi:10.1101/2021.11.03.467020.
- [23] Stahl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 2016;353:78–82. <https://doi.org/10.1126/science.aaa6403>.
- [24] Fu X, Sun L, Chen JY, Dong R, Lin Y, Palmiter RD, et al. Continuous Polony Gels for Tissue Mapping with High Resolution and RNA Capture Efficiency. *bioRxiv* 2021;2021.03.17.435795. doi:10.1101/2021.03.17.435795.
- [25] Chen A, Liao S, Cheng M, Ma K, Wu L, Lai Y, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* 2022;185:1777–1792.e21. <https://doi.org/10.1016/j.cell.2022.04.003>.
- [26] Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine* 2018;50:1–4. <https://doi.org/10.1038/s12276-018-0071-8>.
- [27] Longo SK, Guo MG, Ji AL, Khavari PA. Integrating single-cell and spatial transcriptomics to elucidate intercellular tissue dynamics. *Nature Reviews Genetics* 2021;22:627–44. <https://doi.org/10.1038/s41576-021-00370-8>.
- [28] Dries R, Chen J, Del Rossi N, Khan MM, Sisti A, Yuan GC. Advances in spatial transcriptomic data analysis. *Genome Research* 2021;31:1706–18. <https://doi.org/10.1101/gr.275224.121>.
- [29] Palla G, Fischer DS, Regev A, Theis FJ. Spatial components of molecular tissue biology. *Nature Biotechnology* 2022. <https://doi.org/10.1038/s41587-021-01182-1>.
- [30] Huo L, Jiao Li J, Chen L, Yu Z, Hutvagner G, Li J. Single-cell multi-omics sequencing: application trends, COVID-19, data analysis issues and prospects. *Briefings in Bioinformatics* 2021;22:1–9. <https://doi.org/10.1093/bib/bbab229>.
- [31] Strell C, Hilscher MM, Laxman N, Svedlund J, Wu C, Yokota C, et al. Placing RNA in context and space – methods for spatially resolved transcriptomics. *FEBS Journal* 2019;286:1468–81. <https://doi.org/10.1111/febs.14435>.
- [32] Liao J, Lu X, Shao X, Zhu L, Fan X. Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends in Biotechnology* 2021;39:43–58. <https://doi.org/10.1016/j.tibtech.2020.05.006>.
- [33] Xia C, Fan J, Emanuel G, Hao J, Zhuang X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 2019;116:19490–9. <https://doi.org/10.1073/pnas.1912459116>.
- [34] Shah S, Lubeck E, Zhou W, Cai L. In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* 2016;92:342–57. <https://doi.org/10.1016/j.neuron.2016.10.001>.
- [35] Shah S, Lubeck E, Zhou W, Cai L. seqFISH Accurately Detects Transcripts in Single Cells and Reveals Robust Spatial Organization in the Hippocampus. *Neuron* 2017;94:752–758.e1. <https://doi.org/10.1016/j.neuron.2017.05.008>.
- [36] Zhou W, Yui MA, Williams BA, Yun J, Wold BJ, Cai L, et al. Single-Cell Analysis Reveals Regulatory Gene Expression Dynamics Leading to Lineage Commitment in Early T Cell Development. *Cell Systems* 2019;9:321–337.e9. <https://doi.org/10.1016/j.cels.2019.09.008>.
- [37] Lohoff T, Ghazanfar S, Missarova A, Kouloula N, Pierson N, Griffiths JA, et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nature Biotechnology* 2022;40:74–85. <https://doi.org/10.1038/s41587-021-01006-2>.
- [38] Callaway EM, Dong HW, Ecker JR, Hawrylycz MJ, Huang ZJ, Lein ES, et al. A multimodal cell census and atlas of the mammalian primary motor cortex. *Nature* 2021;598:86–102. <https://doi.org/10.1038/s41586-021-03950-0>.
- [39] Takei Y, Yun J, Zheng S, Ollikainen N, Pierson N, White J, et al. Integrated spatial genomics reveals global architecture of single nuclei. *Nature* 2021;590:344–50. <https://doi.org/10.1038/s41586-020-03126-2>.
- [40] Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, Wählby C, et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nature Methods* 2013;10:857–60. <https://doi.org/10.1038/nmeth.2563>.
- [41] Shah S, Takei Y, Zhou W, Lubeck E, Yun J, Eng CHL, et al. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* 2018;174:363–376.e16. <https://doi.org/10.1016/j.cell.2018.05.035>.
- [42] Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X. High-throughput single-cell gene expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proceedings of the National Academy of Sciences of the United States of America* 2016;113:11046–51. <https://doi.org/10.1073/pnas.1612826113>.
- [43] Watson ER, Taherian Fard A, Mar JC. Computational Methods for Single-Cell Imaging and Omics Data Integration. *Frontiers in Molecular Biosciences* 2022;8:1–25. <https://doi.org/10.3389/fmolb.2021.768106>.
- [44] Eng CHL, Shah S, Thomassie J, Cai L. Profiling the transcriptome with RNA SPOTS. *Nature Methods* 2017;14:1153–5. <https://doi.org/10.1038/nmeth.4500>.
- [45] Zhu Q, Shah S, Dries R, Cai L, Yuan GC. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nature Biotechnology* 2018;36:1183–90. <https://doi.org/10.1038/nbt.4260>.
- [46] Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-Net: deep learning for cell counting, detection, and morphology. *Nature Methods* 2019;16:67–70. <https://doi.org/10.1038/s41592-018-0261-2>.
- [47] Van Valen DA, Kudo T, Lane KM, Macklin DN, Quach NT, DeFelice MM, et al. Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments. *PLoS Computational Biology* 2016;12. <https://doi.org/10.1371/journal.pcbi.1005177>.
- [48] Greenwald NF, Miller G, Moen E, Kong A, Kagel A, Dougherty T, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology* 2021. <https://doi.org/10.1038/s41592-021-01094-0>.
- [49] Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. *Nature Methods* 2021;18:100–6. <https://doi.org/10.1038/s41592-020-01018-x>.
- [50] Cameron WD, Bennett AM, Bui CV, Chang HH, Rocheleau JV. Leveraging multimodal microscopy to optimize deep learning models for cell segmentation. *APL Bioeng* 2021;5. <https://doi.org/10.1063/5.0027993016101>.



- [51] Park J, Choi W, Tiesmeyer S, Long B, Borm LE, Garren E, et al. Cell segmentation-free inference of cell types from in situ transcriptomics data. *Nature Communications* 2021;12:1–3. <https://doi.org/10.1038/s41467-021-23807-4>.
- [52] Petukhov V, Xu RJ, Soldatov RA, Cadinu P, Khodosevich K, Moffitt JR, et al. Cell segmentation in imaging-based spatial transcriptomics. *Nature Biotechnology* 2022;40:345–54. <https://doi.org/10.1038/s41587-021-01044-w>.
- [53] Littman R, Hemminger Z, Foreman R, Arneson D, Zhang G, Gómez-Pinilla F, et al. Joint cell segmentation and cell type annotation for spatial transcriptomics. *Molecular Systems Biology* 2021;17. <https://doi.org/10.1525/msb.202010108>.
- [54] Axelrod S, Cai M, Carr A, Freeman J, Ganguli D, Kiggins J, et al. Starfish: Scalable Pipelines for Image-Based Transcriptomics. *Journal of Open Source Software* 2021;6:2440. <https://doi.org/10.21105/joss.02440>.
- [55] Janout H, Buchegger B, Haghofer A, Hoeglender D, Jacak J, Winkler S, et al. PySpot: A python based framework for the assessment of laser-modified 3D microstructures for Windows and Raspbian. *BIOIMAGING 2020 - 7th Int Conf Bioimaging, Proceedings; Part 13th Int Jt Conf Biomed Eng Syst Technol BIOSTEC 2020* 2020;2:135–42. <https://doi.org/10.5220/0008948001350142>.
- [56] Imbert A, Ouyang W, Safieddine A, Coleno E, Zimmer C, Bertrand E, et al. FISH-quant v2: a scalable and modular tool for smFISH image analysis. *RNA* 2022;28:786–95. <https://doi.org/10.1261/rna.079073.121>.
- [57] Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. EBImage-an R package for image processing with applications to cellular phenotypes. *Bioinformatics* 2010;26:979–81. <https://doi.org/10.1093/bioinformatics/btq046>.
- [58] Barthelmé S, Roediger S, Robotham A, Roth M, Wijffels J, Ooi H. imager, R package for image processing 2019.
- [59] Palla G, Spitzer H, Klein M, Fischer D, Schaar AC, Kuemmerle LB, et al. Squidpy: a scalable framework for spatial omics analysis. *Nature Methods* 2022;19:171–8. <https://doi.org/10.1038/s41592-021-01358-2>.
- [60] Isakova A, Neff N, Quake SR. Single-cell quantification of a broad RNA spectrum reveals unique noncoding patterns associated with cell types and states. *Proceedings of the National Academy of Sciences* 2021;118:1–9. <https://doi.org/10.1073/pnas.2113568118>.
- [61] Johnson BK, Rhodes M, Wegener M, Himadewi P, Foy K, Schipper JL, et al. STORM-seq reveals differentiation trajectories of primary human Fallopian tube epithelium. *bioRxiv* 2022:1–33. doi://doi.org/10.1101/2022.03.14.484332.
- [62] Salmen F, Jonghe JD, Kaminski TS, Alemany A, Parada G, Verity-legg J, et al. Droplet-based Single-cell Total RNA-seq Reveals Differential Non-Coding Expression and Splicing Patterns during Mouse Development In recent years, single-cell transcriptome sequencing has revolutionized biology, allowing for the unbiased characterization. *bioRxiv* 2021:1–67. doi:10.1101/2021.09.15.460240.
- [63] Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M, et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nature Biotechnology* 2020;38:333–42. <https://doi.org/10.1038/s41587-019-0392-8>.
- [64] Rodrigues SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, et al. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 2019;363:1463–7. <https://doi.org/10.1126/science.aaw1219>.
- [65] Wu L, Yan J, Bai Y, Chen F, Xu J, Zou X, et al. Spatially-resolved transcriptomics analyses of invasive fronts in solid tumors. *bioRxiv* 2021:2021.10.21.465135.
- [66] Wei X, Fu S, Li H, Liu Y, Wang S, Feng W, et al. Spatiotemporal transcriptome at single-cell resolution reveals key radial glial cell population in axolotl telencephalon development and regeneration. *bioRxiv* 2021:2021.10.23.465550.
- [67] Wang M, Hu Q, Lv T, Wang Y, Lan Q, Xiang R, et al. High-resolution 3D spatiotemporal transcriptomic maps of developing *Drosophila* embryos and larvae. *Dev Cell* 2022;57:1271–83. <https://doi.org/10.1016/j.devcel.2022.04.006>.
- [68] Liu C, Li R, Li Y, Lin X, Zhao K, Liu Q, et al. Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. *Dev Cell* 2022;57:1284–1298.e5. <https://doi.org/10.1016/j.devcel.2022.04.009>.
- [69] Chen H, Murray E, Sinha A, Laumas A, Li J, Lesman D, et al. Dissecting mammalian spermatogenesis using spatial transcriptomics. *Cell Reports* 2021;37:1. <https://doi.org/10.1016/j.celrep.2021.109915>.
- [70] Fawcner-Corbett D, Antanaviciute A, Parikh K, Jagielowicz M, Gerós AS, Gupta T, et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell* 2021;184:810–826.e23. <https://doi.org/10.1016/j.cell.2020.12.016>.
- [71] Elmentaite R, Kumasaka N, Roberts K, Fleming A, Dann E, King HW, et al. Cells of the human intestinal tract mapped across space and time. *Nature* 2021;597:250–5. <https://doi.org/10.1038/s41586-021-03852-1>.
- [72] Wu SZ, Al-Eryani G, Roden DL, Junankar S, Harvey K, Andersson A, et al. A single-cell and spatially resolved atlas of human breast cancers. *Nature Genetics* 2021;53:1334–47. <https://doi.org/10.1038/s41588-021-00911-1>.
- [73] Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J, et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nature Communications* 2018;9. doi:10.1038/s41467-018-04724-5.
- [74] Ji AL, Rubin AJ, Thrane K, Jiang S, Reynolds DL, Meyers RM, et al. Multimodal Analysis of Composition and Spatial Architecture in Human Squamous Cell Carcinoma. *Journal of Cleaner Production* 2020:497–514. <https://doi.org/10.1016/j.jclep.2020.05.039>.
- [75] Andersson A, Larsson L, Stenbeck L, Salmén F, Ehinger A, Wu S, et al. Spatial deconvolution of HER2-positive breast tumors reveals novel intercellular relationships. *bioRxiv* 2020:1–64. doi:10.1101/2020.07.14.200600.
- [76] Fredrik Salmén, Sanja Vickovic, Ludvig Larsson, Linnea Stenbeck, Johan Vallon-Christersson, Anna Ehinger, et al. Multidimensional transcriptomics provides detailed information about immune cell distribution and identity in HER2+ breast tumors. *bioRxiv* 2018. doi:10.1101/358937.
- [77] Zhao N, Zhang Y, Cheng R, Zhang D, Li F, Guo Y, et al. Spatial maps of hepatocellular carcinoma transcriptomes highlight an unexplored landscape of heterogeneity and a novel gene signature for survival. *Cancer Cell International* 2022;22:1–8. <https://doi.org/10.1186/s12935-021-02430-9>.
- [78] Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- [79] Melsted P, Boeshaghi AS, Liu L, Gao F, Lu L, Min KH(Joseph), et al. Modular, efficient and constant-memory single-cell RNA-seq preprocessing. *Nature Biotechnology* 2021;39:813–8. doi:10.1038/s41587-021-00870-2.
- [80] Navarro JF, Sjöstrand J, Salmén F, Lundberg J, Ståhl PL. ST Pipeline: An automated pipeline for spatial mapping of unique transcripts. *Bioinformatics* 2017;33:2591–3. <https://doi.org/10.1093/bioinformatics/btx211>.
- [81] Kaminow B, Yunusov D, Dobin A. STARsolo: accurate, fast and versatile mapping/ quantification of single-cell and single-nucleus RNA-seq data. *bioRxiv Preprint* 2021. doi:10.1101/2021.05.05.442755.
- [82] Petukhov V, Guo J, Baryawno N, Severe N, Scadden DT, Samsonova MG, et al. dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biology* 2018;19:78. <https://doi.org/10.1186/s13059-018-1449-6>.
- [83] Liu B, Li Y, Zhang L. Analysis and Visualization of Spatial Transcriptomic Data. *Frontiers in Genetics* 2022;12:1–5. <https://doi.org/10.3389/fgene.2021.785290>.
- [84] Noel T, Wang QS, Greka A, Marshall JL. Principles of Spatial Transcriptomics Analysis: A Practical Walk-Through in Kidney Tissue. *Frontiers in Physiology* 2022;12:1–7. <https://doi.org/10.3389/fphys.2021.809346>.
- [85] Elosua-Bayes M, Nieto P, Mereu E, Gut I, Heyn H. SPOTlight: Seeded NMF regression to deconvolute spatial transcriptomics spots with single-cell transcriptomes. *Nucleic Acids Research* 2021;49:E50. <https://doi.org/10.1093/nar/gkab043>.
- [86] Danaher P, Kim Y, Nelson B, Griswold M, Yang Z, Piazza E, et al. Advances in mixed cell deconvolution enable quantification of cell types in spatial transcriptomic data. *Nature Communications* 2022;13:385. <https://doi.org/10.1038/s41467-022-28020-5>.
- [87] Kleshchevnikov V, Shmatko A, Dann E, Aivazidis A, King HW, Li T, et al. Cell 2location maps fine-grained cell types in spatial transcriptomics. *Nature Biotechnology* 2022. <https://doi.org/10.1038/s41587-021-01139-4>.
- [88] Dries R, Chen J, Del Rossi N, Khan MM, Sistig A, Yuan GC. Advances in spatial transcriptomic data analysis. *Genome Research* 2021;31:1706–18. <https://doi.org/10.1101/gr.275224.121>.
- [89] Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* 2018;36:411–20. <https://doi.org/10.1038/nbt.4096>.
- [90] Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. *Nature Biotechnology* 2022. <https://doi.org/10.1038/s41587-022-01273-7>.
- [91] Miller BF, Huang F, Atta L, Sahoo A, Fan J. Reference-free cell type deconvolution of multi-cellular pixel-resolution spatially resolved transcriptomics data. *Nature Communications* 2022;13:2339. <https://doi.org/10.1038/s41467-022-30033-z>.
- [92] Bergholtz H, Carter JM, Cesano A, Cheang MCU, Church SE, Divakar P, et al. Best practices for spatial profiling for breast cancer research with the GeoMx (R) digital spatial profiler. *Cancers* 2021;13. <https://doi.org/10.3390/cancers13174456>.
- [93] Desai N, Neyaz A, Szabolcs A, Shih AR, Chen JH, Thapar V, et al. Temporal and spatial heterogeneity of host response to SARS-CoV-2 pulmonary infection. *Nature Communications* 2020;11. doi:10.1038/s41467-020-20139-7.
- [94] Theodoridis G, Thomas BE, Sarkar D, Mumme HL, Pilcher WJR, Dwivedi B, et al. Single cell transcriptomic landscape of diabetic foot ulcers. *Nature Communications* 2022;13. doi:10.1038/s41467-021-27801-8.
- [95] Zimmerman SM, Propf R, Kulasekara BR, Griswold M, Appelbe O, Bahrami A, et al. Spatially resolved whole transcriptome profiling in human and mouse tissue using Digital Spatial Profiling. *bioRxiv* 2021:2021.09.29.462442. doi:10.1101/2021.09.29.462442.
- [96] Brady L, Kriner M, Coleman I, Morrissey C, Roudier M, True LD, et al. Inter- and intra-tumor heterogeneity of metastatic prostate cancer determined by digital spatial gene expression profiling. *Nature Communications* 2021;12:1–6. <https://doi.org/10.1038/s41467-021-21615-4>.
- [97] Omilian AR, Sheng H, Hong CC, Bandera EV, Khoury T, Ambrosone CB, et al. Multiplexed digital spatial profiling of invasive breast tumors from Black and White women. *Molecular Oncology* 2022;16:54–68. <https://doi.org/10.1002/1878-0261.13017>.

- [98] Rico K, Duan S, Pandey RL, Chen Y, Chakrabarti JT, Starr J, et al. Genome analysis identifies differences in the transcriptional targets of duodenal versus pancreatic neuroendocrine tumours. *BMJ Open Gastroenterology* 2021;8:–. <https://doi.org/10.1136/bmjgast-2021-000765> e000765.
- [99] Wang N, Wang R, Li X, Song Z, Xia L, Wang J, et al. Tumor Microenvironment Profiles Reveal Distinct Therapy-Oriented Proteogenomic Characteristics in Colorectal Cancer. *Frontiers in Bioengineering and Biotechnology* 2021;9:1–2. <https://doi.org/10.3389/fbioe.2021.757378>.
- [100] Nichterwitz S, Nijssen J, Storvall H, Schweingruber C, Comley LH, Allodi I, et al. LCM-seq reveals unique transcriptional adaptation mechanisms of resistant neurons and identifies protective pathways in spinal muscular atrophy. *Genome Research* 2020;30:1083–96. <https://doi.org/10.1101/gr.265017.120>.
- [101] Zeng Z, Li Y, Li Y, Luo Y. Statistical and machine learning methods for spatially resolved transcriptomics data analysis. *Genome Biology* 2022;23:83. <https://doi.org/10.1186/s13059-022-02653-7>.
- [102] Zhang M, Sheffield T, Zhan X, Li Q, Yang DM, Wang Y, et al. Spatial molecular profiling: Platforms, applications and analysis tools. *Briefings in Bioinformatics* 2021;22:1–9. <https://doi.org/10.1093/bib/bbaa145>.
- [103] Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell* 2021;184:3573–3587. e29. <https://doi.org/10.1016/j.cell.2021.04.048>.
- [104] Bergenstråhle J, Larsson L, Lundeberg J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* 2020;21:1–7. <https://doi.org/10.1186/s12864-020-06832-3>.
- [105] Kueckelhaus J, Ravi VM, Will P, Joseph K, Beck J, Ulrich G, et al. Abstract Spatial transcriptomic is a technology to provide deep transcriptomic profiling by preserving the spatial 2020;49:1–2. doi: 10.1101/2020.10.20.346544.
- [106] Gayoso A, Lopez R, Xing G, Boyeau P, Valiollah Pour Amiri V, Hong J, et al. A Python library for probabilistic analysis of single-cell omics data. *Nature Biotechnology* 2022;40:163–6. <https://doi.org/10.1038/s41587-021-01206-w>.
- [107] Pham DT, Tan X, Xu J, Grice LF, Lam PY, Raghubar A, et al. stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *bioRxiv* 2020;2020.05.31.125658. doi:10.1101/2020.05.31.125658.
- [108] Ortogero N, Yang Z, Ortogero N, Yang Z (2021). *GeomxTools: NanoString GeoMx Tools*. R package version 2.0.0. 2021. doi:10.18129/B9.bioc.GeoMxTools.
- [109] Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* 2016;352:189–96. <https://doi.org/10.1126/science.aad0501>.
- [110] Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology* 2019;15. <https://doi.org/10.15252/msb.20188746>.
- [111] Wolf FA, Angerer P, Theis FJ. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biology* 2018;19:1–5. <https://doi.org/10.1186/s13059-017-1382-0>.
- [112] Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology* 2019;20:1–5. <https://doi.org/10.1186/s13059-019-1874-1>.
- [113] Bacher R, Chu LF, Leng N, Gasch AP, Thomson JA, Stewart RM, et al. SCnorm: Robust normalization of single-cell RNA-seq data. *Nature Methods* 2017;14:584–6. <https://doi.org/10.1038/nmeth.4263>.
- [114] Lun ATL, Bach K, Marioni JC. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology* 2016;17:1–4. <https://doi.org/10.1186/s13059-016-0947-7>.
- [115] Lin X, Boutros PC. Optimization and expansion of non-negative matrix factorization. *BMC Bioinformatics* 2020;21:1. <https://doi.org/10.1186/s12859-019-3312-5>.
- [116] Zhao E, Stone MR, Ren X, Guenther J, Smythe KS, Pulliam T, et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nature Biotechnology* 2021;39:1375–84. <https://doi.org/10.1038/s41587-021-00935-2>.
- [117] Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, Chandra T, et al. SC3: Consensus clustering of single-cell RNA-seq data. *Nature Methods* 2017;14:483–6. <https://doi.org/10.1038/nmeth.4236>.
- [118] Smolander J, Junttila S, Venäläinen MS, Elo LL. ILoReg: A tool for high-resolution cell population identification from single-cell RNA-seq data. *Bioinformatics* 2021;37:1107–14. <https://doi.org/10.1093/bioinformatics/btaa919>.
- [119] Wang B, Zhu J, Pierson E, Ramazzotti D, Batzoglou S. Visualization and analysis of single-cell maseq data by kernel-based similarity learning. *Nature Methods* 2017;14:414–6. <https://doi.org/10.1038/nmeth.4207>.
- [120] Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi N, Büttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. *Nature Biotechnology* 2022;40:121–30. <https://doi.org/10.1038/s41587-021-01001-7>.
- [121] Aran D, Looney AP, Liu L, Wu E, Fong V, Hsu A, et al. Reference-based analysis of lung singlecell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology* 2019;20:163–72. <https://doi.org/10.1038/s41590-018-0276-y>.
- [122] de Kanter JK, Lijnzaad P, Candelli T, Margaritis T, Holstege FCP. CHETAH: a selective, hierarchical cell type identification method for single-cell RNA sequencing. *Nucleic Acids Research* 2019;47:–. <https://doi.org/10.1093/nar/gkz543e95>.
- [123] Andreatta M, Berenstein AJ, Carmona SJ. scGate: marker-based purification of cell types from heterogeneous single-cell RNA-seq datasets. *Bioinformatics* 2022;2021.11.08.467740. doi: 10.1093/bioinformatics/btac141.
- [124] Bernstein MN, Ma Z, Gleicher M, Dewey CN. Cello: comprehensive and hierarchical cell type classification of human cells with the Cell Ontology. *iScience* 2021;24:101913. doi:10.1016/j.isci.2020.101913.
- [125] Tan X, Su A, Tran M, Nguyen Q. SpaCell: Integrating tissue morphology and spatial gene expression to predict disease cells. *Bioinformatics* 2020;36:2293–4. <https://doi.org/10.1093/bioinformatics/btz914>.
- [126] Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselms G, et al. SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods* 2017;14:1083–6. <https://doi.org/10.1038/nmeth.4463>.
- [127] Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselms G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols* 2020;15:2247–76. <https://doi.org/10.1038/s41596-020-0336-2>.
- [128] Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity—current challenges and future perspectives. *Molecular Systems Biology* 2021;17:1–9. <https://doi.org/10.15252/msb.202110282>.
- [129] Svensson V, Teichmann SA, Stegle O. SpatialDE: Identification of spatially variable genes. *Nature Methods* 2018;15:343–6. <https://doi.org/10.1038/nmeth.4636>.
- [130] Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nature Methods* 2020;17:193–200. <https://doi.org/10.1038/s41592-019-0701-7>.
- [131] Edsgård D, Johnsson P, Sandberg R. Identification of spatial expression trends in single-cell gene expression data. *Nature Methods* 2018;15:339–42. <https://doi.org/10.1038/nmeth.4634>.
- [132] Kats I, Vento-Tormo R, Stegle O. SpatialDE2: Fast and localized variance component analysis of spatial transcriptomics. *bioRxiv* 2021;2021.10.27.466045. doi:10.1101/2021.10.27.466045.
- [133] Miller BF, Bambach-Mukku D, Dulac C, Zhuang X, Fan J. Characterizing spatial gene expression heterogeneity in spatially resolved single-cell transcriptomic data with nonuniform cellular densities. *Genome Research* 2021;31:1843–55. <https://doi.org/10.1101/gr.271288.120>.
- [134] DeTomaso D, Yosef N. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Systems* 2021;12:446–456.e9. <https://doi.org/10.1016/j.cels.2021.04.005>.
- [135] Zhu Q, Shah S, Dries R, Cai L, Yuan GC. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nature Biotechnology* 2018;36:1183–90. <https://doi.org/10.1038/nbt.4260>.
- [136] Lee K, Lockhart JH, Xie M, Chaudhary R, Slebos RJC, Flores ER, et al. Deep Learning of Histopathology Images at the Single Cell Level. *Frontiers in Artificial Intelligence* 2021;4:1–4. <https://doi.org/10.3389/frai.2021.754641>.
- [137] Hu J, Li X, Coleman K, Schroeder A, Ma N, Irwin DJ, et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature Methods* 2021;18:1342–51. <https://doi.org/10.1038/s41592-021-01255-8>.
- [138] Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics* 2021;22:71–88. <https://doi.org/10.1038/s41576-020-00292-x>.
- [139] Almet AA, Cang Z, Jin S, Nie Q. The landscape of cell–cell communication through single-cell transcriptomics. *Current Opinion in Systems Biology* 2021;26:12–23. <https://doi.org/10.1016/j.coisb.2021.03.007>.
- [140] Schapiro D, Jackson HW, Raghuraman S, Fischer JR, Zanotelli VRT, Schulz D, et al. HistoCAT: Analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature Methods* 2017;14:873–6. <https://doi.org/10.1038/nmeth.4391>.
- [141] Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, et al. Singlecell reconstruction of the early maternal–fetal interface in humans. *Nature* 2018;563:347–53. <https://doi.org/10.1038/s41586-018-0698-6>.
- [142] Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. Cell PhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature Protocols* 2020;15:1484–506. <https://doi.org/10.1038/s41596-020-0292-x>.
- [143] Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nature Methods* 2020;17:159–62. <https://doi.org/10.1038/s41592-019-0667-5>.
- [144] Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell–cell communication using Cell Chat. *Nature Communications* 2021;12:1–20. <https://doi.org/10.1038/s41467-021-21246-9>.
- [145] Ramilowski JA, Goldberg T, Harshbarger J, Kloppman E, Lizio M, Satagopam VP, et al. A draft network of ligand–receptor-mediated multicellular

- signalling in human. *Nature Communications* 2015;6:7866. <https://doi.org/10.1038/ncomms8866>.
- [146] Türei D, Valdeolivas A, Gul L, Palacio-Escat N, Klein M, Ivanova O, et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Molecular Systems Biology* 2021;17:1–6. <https://doi.org/10.15252/msb.20209923>.
- [147] Armingol E, Baghdassarian HM, Martino C, Perez-Lopez A, Knight R, Lewis NE. Context-aware deconvolution of cell-cell communication with Tensor-cell2cell. *bioRxiv* 2021:2021.09.20.461129.
- [148] Tanevski J, Flores ROR, Gabor A, Schapiro D, Saez-Rodriguez J. Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biol* 2022;23:97. <https://doi.org/10.1186/s13059-022-02663-5>.
- [149] Garcia-Alonso L, Handfield LF, Roberts K, Nikolakopoulou K, Fernando RC, Gardner L, et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nature Genetics* 2021;53:1698–711. <https://doi.org/10.1038/s41588-021-00972-2>.