Spatial components of molecular tissue biology

Giovanni Palla[1,2*], David S. Fischer[1,2*], Aviv Regev[3,4+], Fabian J. Theis[1,5+]

[1]Institute of Computational Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg 85764, Germany
[2]TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany
[3]Howard Hughes Medical Institute, Koch Institute for Integrative Cancer Research, Department of Biology, MIT, Cambridge, MA 02139, USA
[4]Current address: Genentech, 1 DNA Way, South San Francisco, CA
[5]Department of Mathematics, Technical University of Munich, Germany

*equal contribution
[+]correspondence to aviv.regev.sc@gmail.com, fabian.theis@helmholtz-muenchen.de

Methods for profiling RNA and protein expression in a spatially-resolved manner are rapidly evolving, making it possible to comprehensively characterize cells and tissues in health and disease. To maximize the biological insights obtained using these techniques, it is critical to both clearly articulate the key biological questions in spatial analysis of tissues and to develop the requisite computational tools to address them. Developers of analytical tools need to decide on the intrinsic molecular features of each cell that need to be considered and how cell shape and morphological features are incorporated into the analysis. Also, optimal ways to compare different tissue samples at various length scales are still being sought. Here we propose to group these biological problems and related computational algorithms into classes across length scales, thus characterizing common issues that need to be addressed to facilitate further progress in spatial transcriptomics and proteomics.

Function in multicellular organisms depends on the balance of interactions between cells in tissues, which are complex, structured and dynamic cellular ecosystems. In health, tissues maintain homeostasis through the joint action of multiple cells, with a dynamic division of labor between different parenchymal and accessory cell types[1]. In disease, malfunction often spans cells of multiple types and is accompanied by changes in the composition, structure and organization of tissues. Deciphering the relation between structure and function in tissues is the cornerstone of tissue biology and pathology, because the way in which cells and molecules are organized in histological patterns at different scales often reflects their joint functionality.

In recent years, there has been a dramatic growth in methods for spatial molecular profiling, which vary in resolution, scale and molecular multiplexing. Methods capture diverse information across different length scales: from single-molecule resolution in techniques such as MERFISH[2] or SeqFISH[3] to averages across cells spanning dozens of micrometers in spot-based protocols such as Spatial Transcriptomics[4]. Methods also vary in the number of molecular features acquired: from tens in fluorescent *in situ* hybridization (FISH), CyCIF[5] and imaging mass cytometry (IMC)[6] to hundreds or thousands in specialized probe-based spatial transcriptomics methods (MERFISH[2] or SeqFISH[3]) or Imaging Mass Spectrometry[7], and tens of thousands in spot-based spatial transcriptomics, such as Slide-Seq[8,9], Visium[10], DBiT-seq[11] and high-definition spatial transcriptomics (HDST)[12]. Such examples highlight a key feature of current spatial technologies: they are diverse in terms of resolution, throughput and multiplexing and should, accordingly, be employed to address different classes of questions.

Computational methods are key to extracting patterns from such data and will be especially powerful if they are designed to take into account the specific biological questions at hand as well as the distinct features and limitations of different measurement methods. Here, we review the computational methods for spatial molecular analysis organized by the biological questions they address and the spatial methods capable of measuring relevant parameters. We will focus specifically on the challenges that different length scales pose for experimental methods, and highlight the types of analysis methods that can be deployed in such studies. We define 'length scales' as the spatial context in which a biological process occurs: short-range length scales include direct cell-cell interactions, whereas long-range length scales include global gradients, such as in oxygen or metabolites (**Fig. 1).** In addition, we emphasize conceptual overlaps with and distinctions from computational methods currently used for analyzing single cell, dissociation-based methods to show how studies based on single-cell profiling approaches can be complemented by spatial approaches, and vice versa. We hope this conceptual and methodological roadmap will help drive the development of new computational methods for key biological questions in tissue biology, provide guidance to biologists seeking to apply methods, and help in sharpening concepts in cell and tissue biology.

Modeling variation at length scales for cell and tissue biology

A long-standing goal in biology is to understand how tissue organization ('structure') relates to tissue physiology ('function'). In a cell-centric model, tissue organization can be described by the different properties (or variables) that distinguish cells from each other **(Fig. 1a)**. Some of these components can be seen as dependent variables, with which we are able to measure variation between cells, whereas others can be seen as independent variables, with which we are able to explain the observed variation. In dissociated single cell profiling data, cellular phenotypes can be described based on dependent variables, such as overall gene expression, protein expression or chromatin accessibility. These dependent variables are a function of independent variables such as sample covariates, assay technology or sequencing batch. Spatial models of cellular variation extend this paradigm by incorporating additional dependent variables, such as subcellular spatial distribution of molecules and cell morphology, and independent variables, such as spatial context of the cell, opening directions to novel formulations of causal models and representation learning. Notably, the complementary view, with cells as independent variables "generating" the tissue is just as biologically compelling, but it is not our focus here.

We distinguish two main influences on cell states in the spatial context ("dependency classes"): direct cell-cell communication and tissue context **(Fig. 1)**. These dependency classes appear as high- and low-frequency patterns in a variance decomposition across a tissue, where the frequency of the pattern mirrors the length-scale of the biochemical phenomenon underlying the statistical dependency: High-frequency patterns can occur multiple times within the same tissue but are localized to the direct neighbourhood of a cell, and often correspond to cell-cell communication. Low-frequency patterns can represent global processes, such as developmental gradients in tissue ontogeny or oxygen availability gradients and other physiological characteristics. In a pathological setting, such as a tumor, a low frequency component could be associated with signals related to metabolic gradients, while a high frequency component could be associated with T cell cytotoxic activity **(Fig. 1b)**. As is often the case in biology, we expect nesting as well within scale. For example, within a cell, we expect both local patterns (molecular complexes and liquid-liquid phase separated "membraneless" organization), intermediate ones (organelles) and global ones (gradients in epithelia). Below, we stratify spatial effects and models by cell biology and tissue biology phenomena.

Cell biology reflected in spatial variation at cellular and subcellular resolution

In-situ cell biology has largely been investigated with microscopy-based techniques, usually by visualizing a limited number of molecular proteins or RNAs[13], whereas single-cell profiling has

characterized the molecular features of cell states and types and their underlying circuitry at unprecedented scale and molecular detail in dissociated samples[14–16]. Spatial molecular profiling offers the possibility of unifying these two worlds of cell biology and molecular biology, thus advancing our understanding in two core areas: cell state variation and extrinsic effectors of cell states.

Molecular information collected with subcellular resolution characterizes cell states both by molecular abundance and the molecular distribution across the cell: Subcellular features can also extend the cell state space and can disentangle cell states that cannot be distinguished based on cell-level molecular features in unsupervised analyses **(Fig. 1a)**. Among the phenomena that can only be disentangled by subcellular-resolved molecular features are the effect of epidermal growth factor (EGF) signaling on the cell cycle[17] and organization of the transcriptome in the nucleus[18] and in the central nervous system[3,19]. In practice, if pixel-level molecular profiles are available, cellular phenotypes can be defined based on both organelle properties (e.g., nucleus size) and the distribution characteristics of molecules across the cell, including their covariance structure (to define complexes, membraneless compartments, etc). Subcellular features of cells can yield hypotheses on their functional states, where the distribution of proteins and mRNA in a cell hints to organelle or pathway activity.

Spatial assays are also used to profile the effects of cell-cell communication occurring at short length scales (μm) in cellular neighborhoods[20,21] via physical contact, compound exchange via tight junctions and synapses, and paracrine or autocrine signaling. Although single cell profiling methods have been adapted to capture some direct cell-cell interactions[22–24], many communication modes in tissues can currently only be directly observed by spatial protocols. These interactions between neighboring cells have the potential to explain the variation observed within cell types in dissociation-based protocols. For example, activated immune cell states can be understood based on the cells' neighborhoods, and correspond to variation observed in single-cell RNA-seq data[25,26]. It is important to carefully consider the experimental techniques used to probe cell-cell communication, as this inference depends both on the molecular features measured and whether true single-cell resolution is achieved (**Fig. 1b,c**, see box description).

Inference of cell-cell communication (a functional feature) from molecular profiling data has been addressed by statistical association methods that typically relate cell types to each other via prior knowledge annotation[27–30], or by co-variation patterns across samples[26,31]. A spatial neighborhood provides strong evidence of co-occurring molecular and structural phenotypes, and should therefore be encoded as a strong prior in any model of tissue biology[29,30,32,33]. Information on cell-cell communication also allows the development of causal inference frameworks, where directional models for cell state dependencies resulting from cell-cell communication in spatial neighborhoods may be used to yield causal and mechanistic hypotheses. Causal inference frameworks can e.g. elucidate the effect that a T cell has on a tumor cell[34,35], or how the metabolic states of the tumor microenvironment reshape protein expression of immune cells[36]. Importantly, a causally interpretable model of these extrinsic effects in the microenvironment requires dependencies between cells and violates the independence assumption commonly used in cell-centric (intrinsic) models of gene expression in many analyses of scRNA-seq[37,38,39]. An encoding of spatial proximity of cells as a spatial graph[40,41] is one promising prior for such cell-cell dependency models.

Finally, recent advances in pooled perturbation screens with high content readouts, either optically[42,43] or by single cell profiling[16,44], open the way to inform and test such causal models, by observing both cell biology and molecular features under a large number of perturbations in cell cultures or in a tissue. An orthogonal approach to large-scale perturbation modelling with spatial information would be to selectively perturb cells *in situ* in spatially-constrained regions of the system (for example by light activation in animal models), to observe the perturbation effect on cellular communication and tissue architecture[45].

Tissue biology reflected in spatial variation on super-cellular length scales

Deciphering the emergent properties of tissue, the modular, multicellular units of tissue function at different scales, and their dysregulation in disease will require spatial technologies with larger fields of view, together with computational methods that can identify functional units and structures.

These 'tissue modules' are constituted by recurrent cellular communities that exert specific functions in tissue, are spatially organized to execute this function, and may occur at different sites in varying compositions. Such structures emerge at different length scales, and therefore can be very different in terms of size and cell composition. For instance, both germinal centers in lymph nodes and glomeruli in the kidney can be viewed as tissue modules, although they are very different structures in terms of size and function.

Tissue modules are defined by phenotypes that characterize areas of a tissue, as opposed to cell- or sample-centric phenotypes that are often the focus of models of dissociation-based data. Clearly, models for identifying tissue-level modules require molecular profiles as well as spatial resolution[46,47]. However, spatial length scales are difficult to account for in such models. First, tissue modules are increasingly difficult to pinpoint as the length scale and the number of putative structures increase, because the relevant biological processes are likely to be convoluted with each other or with other co-occurring signatures. Second, the resolution of lab methods often decreases as the field of view increases in size (e.g. imaging time becomes limiting when imaging large volumes at high resolution) so that the potential for discovery of large and small structures is often anticorrelated. Third, the relation between length scales and functional structures or cellular communities will be different across tissues: the stereotypical layered organization of brains is different from the globular structure of the pancreas or the tree-like structure of lungs. Cross-length scale models that work in one organ might not be useful for another organ.

Metabolic and morphogenetic gradients[48–50] convey information about and arise from developmental processes as well as environmental differences, such as oxygen or nutrient availability, hormone concentration, metabolite concentration or physical stress[49]. They can be assayed with most of the current spatial techniques, as they do not strictly require single-cell resolution. Additionally, biophysical forces, such as the stiffness of the extracellular matrix and its interaction with cell membrane, or cell densities in specific tissues[51], could be inferred with spatial molecular techniques in combination with morphological features. Biophysical tissue features are relevant in many biological systems, from development to cancer biology[51,52], and they occur at multiple length scales. Coupling those phenotypic observations with molecular changes assayed by spatial omics has the potential to elucidate molecular determinants of mechanobiology.

Tissue-wide phenomena can be modeled via spatial variance decompositions and other unsupervised techniques[53–58]. Clustering approaches that account for spatial proximity[46] or morphological similarity[59] could also be used to discover tissue models that go beyond molecular similarity. Unsupervised representation learning has been tremendously successful and widely used for scRNA-seq analyses[37,60]. Advanced representation learning approaches can improve unsupervised models and offer promising tools to relate spatial patterns directly to tissue properties. These include established approaches in image-based deep learning that can deal with the absence of sufficiently informative labels, such as unsupervised, self-supervised and multi-task learning[61]. These techniques will be increasingly used for jointly modeling extrinsic features of cells from images, such as neighborhood density and morphology, as well as intrinsic features, such as expression profiles, thus learning a joint latent representation that encodes information from both modalities[62]. Biology-centered representation learning can be directly performed on image-structured data, which is often the raw version of spatial molecular profiling assays. However, this neglects the strong prior knowledge that cells are discrete

functional units in tissues. To leverage this information, we can represent tissue as a spatial graph of cells (nodes) and their proximity (edges). Representation learning on images and graphs relies on very different model classes: graph-based models will often have interpretability advantages as they are defined on an interpretable input space of cells rather than pixels, analogous to capturing 'social networks' of cells with cell–cell communication and larger cliques. However, information aggregation across graphs is not straightforward[63,64], posing limitations for representing large-scale effects and could miss some of the continuous nature of the tissue, as reflected in images.

Whether such tissue modules are identified by unsupervised approaches or expert annotation, methods are necessary to harmonize them and coherently join them across samples and individuals to validate pattern discoveries. This validation requires a common coordinate framework for molecular variation across organs and individuals in which tissue modules can be registered across samples[65].

Tissue-level models of cell fate decision making

A key question addressed with single-cell genomic is inference of cell fate cell lineage maps: arranging cells from one or more snapshots into continuous transitions across the molecular state space that explain cell fate decision making[66], as well as defining which cells are related by shared lineage through cell division events. Current spatial profiling techniques face the same issue of observing only a snapshot of biological processes, but they capture substantial auxiliary information that impacts cell-fate decisions, cell lineage and cell differentiation: given the limited movement of cells in tissue in many contexts, physical proximity often reflects relatedness (*e.g.*, between stem cells in an intestinal crypt and their progeny in the villus), as well as the dependence of cells' differentiation on external signals provided by neighboring cells and morphogenetic gradients (*e.g.*, between stromal cells and stem cells in the crypt).

Analysis of single cell profiles has enhanced our understanding of cell-fate decisions and cell differentiation. Methods developed for dissociation-based data can be directly applied to spatial data, thus providing trajectory visualization in a spatial context[2,67,68]. Moreover, spatial data sets allow cell fate-inference methods to interpolate in not only the molecular cell state space, but also the spatial coordinates, with the potential to resolve subtle differences in cellular fates that were previously hidden (**Fig. 2**). From a causal modeling perspective, branching events might be explained by the spatial context; in particular, it may be possible to infer the paracrine signals affecting developmental processes[69,70], readily moving from a cell-intrinsic view of differentiation to a more comprehensive tissue view. Developmental processes where morphogenic factor localization is tightly regulated on large length scales, such as in *Drosophila* embryos[71], could be modeled by additionally accounting for cell-cell interactions and other spatial effects on short length scales. Overall, the spatial context of cell fate decision making could help to disentangle intrinsic and extrinsic effects of cell differentiation and developmental processes[72–75].

To determine cell lineages, efforts to integrate spatial molecular information with lineage-tracing techniques in engineered models[76–78] and native human tissue[79] can also spur the development of new modeling approaches to disentangle spatial contribution to lineage formation. For instance, a recent study[80] has used lineage barcoding experiments in cerebral organoids to show that cellular clones have a strong regionalization during development and accumulate in specific brain regions. Another study[81] has developed a lineage barcoding system suitable for spatial imaging techniques, such as multiplexed FISH, and applied it to the *Drosophila* brain, observing that gene expression similarity between different clones was not related to spatial proximity, but that cells from the same lineage (i.e., members of the same clone) had higher expression similarity when proximal in space than cells from the same lineage that were farther apart. With such data, lineage tracing can be used

not only to validate proposed lineage relationships but also to constrain inference models and infer intrinsic and extrinsic factors of cell identity.

A key innovation in the field of trajectory inference from single-cell profiles has been the discovery of molecular states that relate to velocities in the state space (RNA velocity[67,82]). Similarly, RNA velocity could be directly measured *in situ*. For example, MERFISH could determine RNA velocity by distinguishing intra-nuclear from cytoplasmic concentrations for specific transcripts[2]. These state-space velocities could again be coupled to biological axes of variation. In a spatio-molecular lineage, one could extend molecular state space velocities with spatial velocities of migrating cells. Taken together, these examples highlight the opportunities to directly translate, extend or re-formulate concepts developed for single cell genomics to spatial molecular profiling.

Spatial signatures of disease

Most disease processes involve the complex ecosystem of the entire tissue, and histology is accordingly a cornerstone of pathological characterization and definition of disease. Advances in single cell profiling have highlighted the aberrations in cell states, compositions, and interactions that are associated with disease onset, progression, or response to therapy[83,84] and have begun to delineate the coordinated changes across them[34,35]. Spatial information can complement these signatures, explain the changes observed in dissociation-based protocols, disentangle additional alterations of tissue organization, and relate them to the rich legacy of histopathology, which underlies both biomedical research and patient care. In addition, disease and treatment labels provide strong indicators of emergent properties of a tissue, such as tissue modules, and help in their discovery.

Spatial dimensions of disease state entail both tissue-scale deranged cellular communities and short-range perturbed cell-cell communications, therefore requiring models tailored to capturing effects at the relevant length scales. Several recent studies follow this approach in conditions as diverse as cancer[34,35,85–89], myocardial infarction[90] and neurodegenerative diseases[91]. For example, a recent study reported immune cell distributional signatures of T cells in colorectal cancer that do not manifest on a compositional level, at least for broad cell categories[35]. Specifically, T cell aggregation correlated with core tumor phenotypes, whereas T cell number did not. Another study applied spatial transcriptomics to biopsies from prostate cancer tumors and identified cancer foci as multicellular sub-tissue compartments, often not identified by typical histopathological analysis[87]. An extensive multicenter study of breast cancer tissue showed how cellular communities identified by IMC explain inter- and intra-sample variation, as well as define cancer subtypes with strong associations to patient survival[92]. These communities, newly discovered tissue modules, have characteristic cell type enrichments and putative cell–cell interactions. An integrative analysis of scRNA-seq and spatial transcriptomics of the same pancreatic cancer tumors, showed co-localization of stress-response cancer cells and inflammatory fibroblasts[86] forming a cancer specific cellular neighborhood. Such studies are also relevant in non-oncological pathologies. For example, spatial transcriptomics and FISH-based measurements of brain sections from a mouse model of Alzheimer's disease showed a local, multicellular interplay in cellular response to amyloid plaques[91]. These examples show how spatial information can both elucidate novel disease features, and highlight the potential causal mechanism that connects a deranged cellular state to diseased tissue.

Analytical challenges and current approaches

So far, we have described how spatial components of variation can be mapped to cell and tissue biology questions, and highlighted some of the modeling approaches to disentangle spatial components of molecular variation. Next, we describe methods and analysis strategies that have

been already proposed, and the practical problems of working with spatial molecular data, in terms of processing, representation, integration and analysis. We highlight available solutions and the open challenges that still need to be addressed.

Image processing, segmentation, coordinate registration and data structure

In contrast to dissociation-based protocols, most spatial methods require a computational mapping of the measured molecules to their cellular origin, either by segmentation or deconvolution[9394–104] . It is likely that segmentation-based approaches will need to be fine-tuned for the data at hand, due to the high diversity in spatial technologies (both in terms of resolution, multiplexing and modality) as well as variation across tissues. Semantic segmentation methods that assign each pixel of an image to a label of interest, such as blood vessels, stroma or connective tissue, offer an orthogonal approach to annotate tissue regions by morphology features. Their representation could be used for downstream tasks, such as integration with molecular data[105]. Benchmarks and evaluations of segmentation methods, as well as integrating them in a processing pipeline[106–108] could help decision making on nuclear versus cytoplasmic segmentation, different segmentation algorithms[109,110], and experimental design consideration on segmentation-auxiliary membrane stains[74]. For imaging-based spatial technologies, alternative segmentation-free approaches have emerged. Such methods are often based on maximum likelihood estimation and graph representation learning, and aim at finding pseudo-cells based on spatial aggregates of molecular probes[111–114]. They present a viable alternative when experimental settings hinder robust segmentation-based approaches.

In addition to the molecular profiles associated with segmented cells, images convey invaluable statistics on cell area, morphology and population density that can help uncover principles of tissue organization and higher-order structure across tissue samples. Image analysis techniques are also suitable for registering multiple tissue samples to a common coordinate framework[65]. For example, methods that leverage convolutional neural networks for feature extraction and sample mapping have already been proposed in the context of spatial molecular data[115–118]. Furthermore, in order to build 3D molecular maps of complex tissues and organs, techniques such as image registration and stitching will be essential[115–117,119–121].

A key computational question concerns data representation: spatial dimensions can be maintained as vector coordinates or transformed in a graph, while image information, such as extracted features and segmentation masking, should be easily accessible and directly linked to observations. Most of the existing software for (spatial) single cell analysis should accommodate such data types associated with segmented cells in its existing framework[59,108,122–128]. Efficient and flexible integration of image-structured data is often still difficult in cell-centric workflows as both data handling infrastructure and algorithmic approaches are missing (**Fig. 3a, Table 1**).

Integrating modalities

Some spatial methods capture rich molecular information at the cost of resolution, but computational analysis can be used for deconvolution and mapping by integrating the spatial data with microscopy data or single cell profiles that have been determined using dissociation-based methods. For example, spot-based protocols such as HDST[12], Slide-seq[8,9] and Spatial Transcriptomics/Visium[4] capture RNA or proteins from areas spanning more (or less) than one cell, without consideration of cell boundaries. When spots are smaller than average cells, computational analysis can reveal cell identities by clustering, as has been demonstrated in HDST[12] and is possible through grouping of spots by segmentation masks defined in bright-field microscopy (FISH-based protocols and IMC[3,6,129]). If spots are larger than one cell (*e.g.*, Visium), deconvolution methods can relate transcriptomes measured in a spot to the cell profiles present in tissue (**Fig. 3b, Table 1**)[130–139]. In contrast to

deconvolution approaches, that explicitly model the spatial observation as an aggregate of cells, label projection methods have been developed to map the cellular states identified in scRNA-seq experiments to lower resolution spatial data[86,140,141]. These methods transfer cell type labels from scRNA-seq to spot transcriptomics by manifold alignment techniques[142]. Novel methods should include the additional information provided by spatial omics data, such as morphology features derived from imaging data and spatial neighbor graph computed from spatial coordinates. Such additional information could include smoothing constraints on cell-type proportions in space or morphological similarity derived from imaging data of cellular niches. The histology stain collected by Spatial Transcriptomics methods is invaluable for such mappings as it can be used to derive nuclei densities under each spot array, as well as other constraints in terms of morphology-based cell variability. For example, a recent method accounts for nuclei densities derived from tissue and adopts a probabilistic framework to assign cell type proportions[116,143].

A second challenge lies in the feature space mapping: probe-based protocols like FISH and IMC yield many fewer features and, in the case of protein measurements, also a different modality than in single cell RNA-seq, and several methods have been proposed for mapping unobserved gene expression profiles in sparse spatial data by learning a correlation-based representation from single cell references[74,140,141,144–146]. Beyond feature imputation, multiple tools are focused on mapping single cell transcriptomes to spatial coordinate systems, using the spatial data as a reference map[71,147–149].

Recently, new methods emerged that reconstruct the spatial arrangement of cells mainly based on scRNA-seq data with little or no spatially resolved experimental data[29,150]. They rely only on a few markers to perform the mapping, based on the assumption that similarity in gene expression space correlates with spatial distances, and that such a trend is monotonic in a defined neighborhood. While reference-free methods have been applied for biological systems where there is a strong stereotypical organization, for example due to global gradients (e.g., *Drosophila* embryo), it's unclear how they perform in tissues that display disordered spatial heterogeneity, salt-and-pepper patterns, and perturbation states, such as human organs and diseases. Learning such patterns from some spatial data, including the nature of cell-cell interactions, may help us find ways to introduce additional constraints for reference-free mapping, for example through receptor-ligand interactions. Mapping cell states to spatial coordinates is a challenging problem that we expect to remain an active area of research for other spatial molecular data.

Finally, there is a rich set of opportunities around integration of molecular and cellular information with the non-molecular features of the image, to better define both cells and their tissue relationships (**Fig. 3c, Table 1**). Image analysis techniques and deep learning approaches can gather additional information from imaging data, including explicit or latent representation of cell area, morphology and population density[151]. These can help improve performances on different tasks such as cancer cell classification[152,153] and mapping molecular counts at higher resolution[154]. Multimodal representation learning that captures both imaging and molecular features can enhance our definition of cells (and neighborhoods), and allow training models that generate one modality from the other. On paired data, such models could be learnt jointly yet should encode sources of variation separately across modalities, enhancing interpretability and biological discovery. When models are learnt from uncoupled, separate data modalities[62] (e.g., a collection of microscopy images and of single cell RNA-seq profiles), joint measurements provides validation data with which to test predictions[155]. Disjoint data sources are very likely to outmatch joint measurements in terms of datasets sizes (e.g., H&E tissue slides for imaging, and single cell atlases for molecular data). For these very large datasets of different data types measured on different samples, (self-)supervised approaches such as contrastive learning could be used for pre-training and transfer learning[156,157]. Learnt representations could then be integrated ex-post with interpretable linear methods[158,159]. Such models could also be integrated with high-resolution imaging efforts, such as recently proposed high resolution cellular

maps[160,161]. We forecast transfer learning efforts to be broadly useful for enriching legacy data (measured only in one modality) and facilitating future studies.

Spatially variable features and cellular neighborhoods

Although scRNA-seq has taught us how gene expression differs between individual cells, cell types and between different experimental conditions, spatial methods add another dimension to the analysis of gene expression regulation. Several methods have been proposed to identify spatially variable genes in the tissue of interest (**Fig. 3d, Table 1**)[162–167]. These methods vary greatly in terms of assumptions and model classes employed. They can be broadly categorized into Gaussian Process based frameworks[164,165,168–170], methods that employ hidden markov random fields[171–173] and graph-based methods[114,174]. The power and sensitivity of these methods to detect spatial variability depends on the experimental technology and sample characteristics and thus has to be chosen carefully for the data at hand. Currently, the metrics that quantify spatially variability" are still very heterogeneous, because of the diversity in underlying algorithms, and this problem will need to be stratified in the future. Furthermore, established approaches such as differential expression analysis across conditions should be expanded to account for the variation of the spatial dependency over samples, as well as model latent spatial variance components of the data[53]. We expect that techniques borrowed from the field of spatial statistics will also be useful to model spatial components of molecular variation, in particular accounting for different views of tissue organization: from direct cell-cell interaction to higher scale interactions such as paracrine communication to metabolic gradients.

Spatial techniques will also be useful to study cellular networks and cell-cell communications (**Fig. 3e**). Few methods have been proposed that aim at estimating cell type neighborhood enrichments in spatial context[59,106,175], and inferring cell interaction and signaling genes from spatial graph structure[33,41,58,59,106,172,176]. The modeled phenomena differ drastically between these methods, and will both need to be stratified in the future.

Spatial power analysis

Power analysis enables researchers to determine the sample sizes required to observe an effect of interest. In the context of spatial molecular data, such effects are subject not only to sample variation but also to its spatial coordinates in the samples. This is of particular importance in spatial systems since realistically attainable sample sizes (image size or image number) depend strongly on the chosen experimental protocol. For clarity, we propose to frame the problem in two sub-challenges: gene-centric and cell-centric power analysis (**Fig. 3.f**).

‘Gene-centric’ power analysis entails the power calculation to find significant spatially variable genes in tissue. In this context, each gene is considered separately and a power calculation can be performed based on an estimated effect size of gene variability in non-random regions of the tissue samples. A recent comprehensive power analysis, using synthetic datasets[165], evaluated several methods that rely on different test statistics[162,163164,165]. Importantly, the superposition of multiple confounding spatial expression trends, such as high- and low-frequency patterns **(Fig. 1b),** further complicates trend discovery. Notably, asking an algorithm to identify any dependence of expression on spatial coordinates is often too general a question, whereas a user might be interested in finding patterns of a certain shape, such as radial symmetries in tumors. Such pattern-based trend discovery has previously been attempted in spatial variable gene detection, but may be limited by the hypothesis that can be framed as patterns[164,165].

‘Cell centric’ power analysis concerns features of tissue organization, such as cellular neighborhoods, cellular communities and tissue composition. For instance, the optimal sample size (total number of cells) to observe a particular enrichment of cellular neighborhoods depends on the frequency of the participating cell types and the required significance threshold. Moreover, large scale tissue features often correspond directly to images so that the sample size is the number of images obtained, posing trade-offs between number and size of images. This problem has been investigated in the context of dissociation based assays[177,178] but is yet unexplored for spatial molecular data.

Conclusions

Spatial profiling carries the promise of connecting variation in high-dimensional single-cell omics data sets with interpretable biological phenomena in tissue. These spatial dependencies exist on drastically different length scales and correspond to cell biological effects with different implications at tissue level. Here, we considered effects that can be modeled across these length scales and coupled the existing spatial profiling methods that provide pertinent measurements. We formulated a roadmap for modeling dependencies in spatial data based on the length-scale of underlying molecular phenomena. We also classified current processing and analysis approaches to spatial molecular data. As highlighted by different application settings, such as spatial signatures of disease, models for spatial variation can resolve cellular states and tissue phenotypes that were previously hidden. Accordingly, model-based experimental design of spatial profiling studies is not only limited by resolution but is also a compromise among resolution, section size and sample size.

Data analysis currently still presents many bottlenecks, as opportunities inherent in integrating data from spatial to non-spatial and between spatial assays of different resolutions and for different analytes are not fully realized. Major challenges remain in the analysis of image-structured data in terms of cell segmentation, image processing and the relation of tissue structures observed in the image to the of the measured molecular profiles. Novel ways of thinking about processing imaging data in light of modelling particular length scales may alleviate data processing bottlenecks as spatial models move from pipelines to end-to-end models. On the other hand, mathematical models for length-scale specific and cross-length-scale spatial variation have only started to emerge. Ultimately, we will find solutions to the remaining analytical and experimental problems that will allow the creation of single-cell resolved and spatially aware tissue atlases.

Acknowledgements

We gratefully thank Anna Hupalowska for help with figure graphics and Jennifer Rood for useful comments and suggestions that greatly improved the manuscript. G.P. is supported by the Helmholtz Association under the joint research school "Munich School for Data Science-MUDS". F.J.T. acknowledges support by the BMBF (grant# 01IS18036B and grant# 01IS18053A) and by the Helmholtz Association’s Initiative and Networking Fund through Helmholtz AI [grant number: ZT-I-PF-5-01] and sparse2big [grant # ZT-I-0007]. A.R. was an Investigator of the Howard Hughes Medical Institute. D.S.F. acknowledges support from a German Research Foundation (DFG) fellowship through the Graduate School of Quantitative Biosciences Munich (QBM) [GSC 1006 to D.S.F.]. G.P. and D.S.F. acknowledge support by the Joachim Herz Stiftung.

Conflicts of interest

F.J.T. reports receiving consulting fees from Roche Diagnostics GmbH and Cellarity Inc., and ownership interest in Cellarity, Inc. and Dermagnostix. A.R. is a founder of and equity holder in

Celsius Therapeutics, an equity holder in Immunitas Therapeutics, and was a scientific advisory board member for ThermoFisher Scientific, Syros Pharmaceuticals and Neogene Therapeutics until August 1, 2020. From August 1, 2020, A.R. is an employee of Genentech. A.R. is a named inventor on several patents and patent applications filed by the Broad Institute in the area of single cell and spatial genomics.

Figures

Figure 1: Components of variation in spatial profiling assays. a): Molecular profiles from spatial technologies entail dissociation-based measurements such as gene expression and sample-based covariates, as well as additional measurements such as subcellular variation, cell morphology, multi-modal measurements in spatial context and spatial coordinates. In the figure, we highlight how such systems entail both long-range and short-range effects that should be accounted for in modelling molecular variation in spatial dimensions. **b)** The study of several biological phenomena such as morphogenic gradients in development, T cell cytotoxic activity and the tumor microenvironment can greatly benefit from the use of spatial technologies. In this figure, we also discriminate between observed variables (in solid lines) and latent variables (biological processes, dashed lines) that have to be unveiled from the data. Finally, the resolution of spatial technologies is also different and should be taken into account in the context of interest.

Figure 2: Trajectory inference in new dimensions. Spatial data allows trajectories inferred from the transcriptome to be placed in spatial context. It also opens up analytical approaches to infer trajectories with additional spatial information.

Figure 3: Analytical challenges and opportunities of spatial molecular data. a) Image processing, segmentation, registration and data structure: spatial molecular data are diverse and require tailored solutions for processing and data infrastructure. **b) Deconvolution and data integration**: spot-based technologies require deconvolution analysis to approximate proportions of cell composition within spots. Integration methods can also be used to map known cell phenotypes to spatial data. **c) Multi modality:** integration of shapes and morphometric features, as well as additional molecular profiles improve cell state identification and tissue phenotyping. **d) Spatially variable genes:** regression frameworks that aim at finding spatially variable molecular features are key in understanding cell states and tissue organization. **e) Cellular neighborhoods:** tissue coordinates are key in understanding spatial communities and cellular communication, and graph abstraction is a suitable representation of the data. **f) Spatial power analysis:** understanding tissue-level effects across samples and individuals requires power analysis that accounts for spatial distribution. **Table 1** contains an overview of tasks and methods for spatial single cell analysis.

Box descriptions

Spatial molecular profiling technologies differ in terms of resolution and multiplexing. Therefore, technology capability should be accounted for by modelling approaches. Here, we show how techniques can be grouped together based on the molecular entity they are able to capture, as well as the resolution (and therefore length scale) that they are able to profile.

Transcriptome: spot-based spatially barcoded microarrays that capture transcriptome-wide gene expression profiles: Low resolution (spatial transcriptomics (ST), Visium, Slide-seqV1) and high resolution (HDST, Slide-seqV2). *In-situ* transcriptome profiles with barcoded oligos: High throughput (MERFISH, seqFISH, NanoString), low throughput (SCRINSHOT (CARTANA), Barista-seq, osmFISH), sequencing by synthesis in situ (STARMAP) and others.

Proteins: Imaging mass cytometry (MALDI-TOF, MiBI-TOF) and multiplexed IHC (4i, cyCIF, CODEX, Immuno-SABER).
Metabolites: Imaging mass spectrometry
Several reviews covered experimental techniques for spatial molecular profiling: [179–191]

Table 1: Challenges and current solutions for spatial single cell analysis*

Challenge	References
Image processing, segmentation, registration and data structures	
- Cell entity identification via segmentation - Cell entity identification via segmentation-free methods - Image processing - Image registration and alignment - Toolkit for spatial molecular data analysis	94–100,109 111–113,192–195 101–103,107,119,196–200 115,116,119,126,201 59,107,108,122,125–128,167,202–206
Deconvolution and multimodal integration	
- Cellular-interaction inference - Data integration and deconvolution - Gene-expression mapping and imputation - Spatial mapping of gene expression profiles - Multi-modal integration	27,28,32,33,41,53,207,208 130–134,136–139,141,143,209–213 104,144–146,154,214,215 29,147,148,150,216 54,62,152,153,155,159,217,218
Spatially variable features, spatial communities and spatial power analysis	
- Spatially variable genes - Spatial decomposition and clustering - Spatial power analysis	114,163–165,168–170,172,174,219–222 46,47,54,56–59,114,162,171–173,195,223–226 165,170,227–229

Tasks in the data analysis pipeline can be broadly divided into pre-processing, integration and spatially variable features identification. For each step in the analysis, we outline tasks and proposed solutions present in the literature.

References

1. Okabe, Y. & Medzhitov, R. Tissue biology perspective on macrophages. *Nat. Immunol.* **17**, 9–17 (2016).
2. 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. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19490–19499 (2019).
3. Eng, C.-H. L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
4. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
5. Lin, J.-R., Fallahi-Sichani, M., Chen, J.-Y. & Sorger, P. K. Cyclic Immunofluorescence (CycIF), A Highly Multiplexed Method for Single-cell Imaging. *Curr. Protoc. Chem. Biol.* **8**, 251–264 (2016).
6. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
7. Alexandrov, T. Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence. *Annu. Rev. Biomed. Data Sci.* **3**, 61–87 (2020).
8. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
9. Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with Slide-seqV2. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0739-1.
10. Salmén, F. *et al.* Barcoded solid-phase RNA capture for Spatial Transcriptomics profiling in mammalian tissue sections. *Nat. Protoc.* **13**, 2501–2534 (2018).
11. Liu, Y. *et al.* High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell* **0**, (2020).

12. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
13. Dey, P. Fluorescent In Situ Hybridization Techniques in Pathology: Principle, Technique and Applications. in *Basic and Advanced Laboratory Techniques in Histopathology and Cytology* (ed. Dey, P.) 213–220 (Springer Singapore, 2018).
14. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160 (2016).
15. Tabula Muris Consortium. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. *Nature* **583**, 590–595 (2020).
16. Dixit, A. *et al.* Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* **167**, 1853–1866.e17 (2016).
17. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, (2018).
18. Fazal, F. M. *et al.* Atlas of Subcellular RNA Localization Revealed by APEX-Seq. *Cell* **178**, 473–490.e26 (2019).
19. Wang, G. *et al.* Spatial organization of the transcriptome in individual neurons. *Cold Spring Harbor Laboratory* 2020.12.07.414060 (2020) doi:10.1101/2020.12.07.414060.
20. Ramilowski, J. A. *et al.* A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat. Commun.* **6**, 7866 (2015).
21. Balkwill, F. R., Capasso, M. & Hagemann, T. The tumor microenvironment at a glance. *J. Cell Sci.* **125**, 5591–5596 (2012).
22. Boisset, J.-C. *et al.* Mapping the physical network of cellular interactions. *Nat. Methods* **15**, 547–553 (2018).
23. Medaglia, C. *et al.* Spatial reconstruction of immune niches by combining photoactivatable reporters and scRNA-seq. *Science* **358**, 1622–1626 (2017).
24. Halpern, K. B. *et al.* Paired-cell sequencing enables spatial gene expression mapping of

liver endothelial cells. *Nat. Biotechnol.* **36**, 962–970 (2018).

25. Codeluppi, S. *et al.* Spatial organization of the somatosensory cortex revealed by osmFISH. *Nature Methods* vol. 15 932–935 (2018).
26. Jerby-Arnon, L. & Regev, A. Mapping multicellular programs from single-cell profiles. *Cold Spring Harbor Laboratory* 2020.08.11.245472 (2020) doi:10.1101/2020.08.11.245472.
27. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
28. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
29. Cang, Z. & Nie, Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. *Nat. Commun.* **11**, 2084 (2020).
30. Almet, A. A., Cang, Z., Jin, S. & Nie, Q. The landscape of cell-cell communication through single-cell transcriptomics. *Current Opinion in Systems Biology* (2021) doi:10.1016/j.coisb.2021.03.007.
31. Pelka, K. *et al.* Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* **184**, 4734–4752.e20 (2021).
32. Armingol, E. *et al.* Inferring the spatial code of cell-cell interactions and communication across a whole animal body. *Cold Spring Harbor Laboratory* 2020.11.22.392217 (2020) doi:10.1101/2020.11.22.392217.
33. Tanevski, J., Flores, R. O. R., Gabor, A., Schapiro, D. & Saez-Rodriguez, J. Explainable multi-view framework for dissecting inter-cellular signaling from highly multiplexed spatial data. *bioRxiv* 2020.05.08.084145 (2020) doi:10.1101/2020.05.08.084145.
34. Phillips, D. *et al.* Immune cell topography predicts response to PD-1 blockade in cutaneous T cell lymphoma. *Oncology* (2020) doi:10.1101/2020.12.06.20244913.

35. Schürch, C. M. *et al.* Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *bioRxiv* 743989 (2019) doi:10.1101/743989.
36. Hartmann, F. J. *et al.* Single-cell metabolic profiling of human cytotoxic T cells. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0651-8.
37. Lopez, R., Gayoso, A. & Yosef, N. Enhancing scientific discoveries in molecular biology with deep generative models. *Mol. Syst. Biol.* **16**, e9198 (2020).
38. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
39. Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S. & Theis, F. J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* **10**, 390 (2019).
40. Borgwardt, K., Ghisu, E., Llinares-López, F., O'Bray, L. & Rieck, B. Graph Kernels: State-of-the-Art and Future Challenges. *arXiv [cs.LG]* (2020).
41. Yuan, Y. & Bar-Joseph, Z. GCNG: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol.* **21**, 300 (2020).
42. Feldman, D. *et al.* Optical Pooled Screens in Human Cells. *Cell* **179**, 787–799.e17 (2019).
43. Wang, C., Lu, T., Emanuel, G., Babcock, H. P. & Zhuang, X. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 10842–10851 (2019).
44. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
45. Legnini, I. *et al.* Optogenetic perturbations of RNA expression in tissue space. *bioRxiv* 2021.09.26.461850 (2021) doi:10.1101/2021.09.26.461850.
46. Zhao, E. *et al.* BayesSpace enables the robust characterization of spatial gene expression architecture in tissue sections at increased resolution. 2020.09.04.283812

(2020) doi:10.1101/2020.09.04.283812.

47. Chen, Z., Soifer, I., Hilton, H., Keren, L. & Jojic, V. Modeling Multiplexed Images with Spatial-LDA Reveals Novel Tissue Microenvironments. *J. Comput. Biol.* (2020) doi:10.1089/cmb.2019.0340.
48. Kietzmann, T. Metabolic zonation of the liver: The oxygen gradient revisited. *Redox Biol* **11**, 622–630 (2017).
49. Rogers, K. W. & Schier, A. F. Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* **27**, 377–407 (2011).
50. Saviano, A., Henderson, N. C. & Baumert, T. F. Single-cell genomics and spatial transcriptomics: Discovery of novel cell states and cellular interactions in liver physiology and disease biology. *J. Hepatol.* (2020) doi:10.1016/j.jhep.2020.06.004.
51. Ladoux, B. & Mège, R.-M. Mechanobiology of collective cell behaviours. *Nat. Rev. Mol. Cell Biol.* **18**, 743–757 (2017).
52. Vining, K. H. & Mooney, D. J. Mechanical forces direct stem cell behaviour in development and regeneration. *Nat. Rev. Mol. Cell Biol.* **18**, 728–742 (2017).
53. Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J. & Stegle, O. Modeling Cell-Cell Interactions from Spatial Molecular Data with Spatial Variance Component Analysis. *Cell Rep.* **29**, 202–211.e6 (2019).
54. Velten, B., Braunger, J. M., Arnol, D., Argelaguet, R. & Stegle, O. Identifying temporal and spatial patterns of variation from multi-modal data using MEFISTO. *Cold Spring Harbor Laboratory* 2020.11.03.366674 (2020) doi:10.1101/2020.11.03.366674.
55. Argelaguet, R. *et al.* MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* **21**, 111 (2020).
56. Walter, F. C., Stegle, O. & Velten, B. FISHFactor: A Probabilistic Factor Model for Spatial Transcriptomics Data with Subcellular Resolution. *bioRxiv* 2021.11.04.467354 (2021) doi:10.1101/2021.11.04.467354.

57. William Townes, F. & Engelhardt, B. E. Nonnegative spatial factorization. *arXiv [stat.ME]* (2021).

58. Fischer, D. S., Schaar, A. C. & Theis, F. J. Learning cell communication from spatial graphs of cells. *bioRxiv* 2021.07.11.451750 (2021) doi:10.1101/2021.07.11.451750.

59. Pham, D. *et al.* stLearn: integrating spatial location, tissue morphology and gene expression to find cell types, cell-cell interactions and spatial trajectories within undissociated tissues. *Bioarxiv* (2020) doi:10.1101/2020.05.31.125658.

60. Eraslan, G., Avsec, Ž., Gagneur, J. & Theis, F. J. Deep learning: new computational modelling techniques for genomics. *Nat. Rev. Genet.* **20**, 389–403 (2019).

61. Cheplygina, V., de Bruijne, M. & Pluim, J. P. W. Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med. Image Anal.* **54**, 280–296 (2019).

62. Yang, K. D. *et al.* Multi-Domain Translation between Single-Cell Imaging and Sequencing Data using Autoencoders. *Cold Spring Harbor Laboratory* 2019.12.13.875922 (2019) doi:10.1101/2019.12.13.875922.

63. Cheung, M., Shi, J., Jiang, L., Wright, O. & Moura, J. M. F. Pooling in Graph Convolutional Neural Networks. *2019 53rd Asilomar Conference on Signals, Systems, and Computers* (2019) doi:10.1109/ieeeconf44664.2019.9048796.

64. Hetzel, L., Fischer, D. S., Günnemann, S. & Theis, F. J. Graph representation learning for single-cell biology. *Current Opinion in Systems Biology* (2021) doi:10.1016/j.coisb.2021.05.008.

65. Rood, J. E. *et al.* Toward a Common Coordinate Framework for the Human Body. *Cell* **179**, 1455–1467 (2019).

66. Forrow, A. & Schiebinger, G. A Unified Framework for Lineage Tracing and Trajectory Inference. 2020.07.31.231621 (2020) doi:10.1101/2020.07.31.231621.

67. Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to

transient cell states through dynamical modeling. *Nat. Biotechnol.* (2020) doi:10.1038/s41587-020-0591-3.

68. Lange, M. *et al.* CellRank for directed single-cell fate mapping. *Cold Spring Harbor Laboratory* 2020.10.19.345983 (2020) doi:10.1101/2020.10.19.345983.
69. Schiebinger, G. *et al.* Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176**, 928–943.e22 (2019).
70. Fischer, D. S. *et al.* Inferring population dynamics from single-cell RNA-sequencing time series data. *Nat. Biotechnol.* **37**, 461–468 (2019).
71. Karaiskos, N. *et al.* The Drosophila embryo at single-cell transcriptome resolution. *Science* **358**, 194–199 (2017).
72. van den Brink, S. C. *et al.* Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* **582**, 405–409 (2020).
73. Asp, M. *et al.* A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. *Cell* **179**, 1647–1660.e19 (2019).
74. Lohoff, T. *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.* (2021) doi:10.1038/s41587-021-01006-2.
75. Tyser, R. C. V. *et al.* A spatially resolved single cell atlas of human gastrulation. *bioRxiv* 2020.07.21.213512 (2020) doi:10.1101/2020.07.21.213512.
76. Alemany, A., Florescu, M., Baron, C. S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. *Nature* **556**, 108–112 (2018).
77. McKenna, A. *et al.* Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, (2016).
78. Spanjaard, B. *et al.* Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
79. Ludwig, L. S. *et al.* Lineage Tracing in Humans Enabled by Mitochondrial Mutations and

Single-Cell Genomics. *Cell* **176**, 1325–1339.e22 (2019).

80. He, Z. *et al.* Lineage recording reveals dynamics of cerebral organoid regionalization. *bioRxiv* 2020.06.19.162032 (2020) doi:10.1101/2020.06.19.162032.
81. Chow, K.-H. K. *et al.* Imaging cell lineage with a synthetic digital recording system. *Science* **372**, (2021).
82. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
83. Keren-Shaul, H. *et al.* A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276–1290.e17 (2017).
84. Smillie, C. S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714–730.e22 (2019).
85. Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373–1387.e19 (2018).
86. Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
87. Berglund, E. *et al.* Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* **9**, 2419 (2018).
88. Andersson, A. *et al.* Spatial Deconvolution of HER2-positive Breast Tumors Reveals Novel Intercellular Relationships. *Cold Spring Harbor Laboratory* 2020.07.14.200600 (2020) doi:10.1101/2020.07.14.200600.
89. Lomakin, A. *et al.* Spatial genomics maps the structure, character and evolution of cancer clones. *bioRxiv* 2021.04.16.439912 (2021) doi:10.1101/2021.04.16.439912.
90. Kuppe, C. *et al.* Spatial multi-omic map of human myocardial infarction. *Cold Spring Harbor Laboratory* 2020.12.08.411686 (2020) doi:10.1101/2020.12.08.411686.
91. Chen, W.-T. *et al.* Spatial Transcriptomics and In Situ Sequencing to Study Alzheimer's Disease. *Cell* **182**, 976–991.e19 (2020).

92. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).

93. Liu, Z. *et al.* Integrating single-cell RNA-seq and imaging with SCOPE-seq2. *Sci. Rep.* **10**, 19482 (2020).

94. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

95. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**, e2005970 (2018).

96. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).

97. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Cold Spring Harbor Laboratory* 2021.03.01.431313 (2021) doi:10.1101/2021.03.01.431313.

98. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell Detection with Star-convex Polygons. *arXiv [cs.CV]* (2018).

99. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

100. Mandal, S. & Uhlmann, V. SplineDist: Automated Cell Segmentation With Spline Curves. *Cold Spring Harbor Laboratory* 2020.10.27.357640 (2021) doi:10.1101/2020.10.27.357640.

101. Hörl, D. *et al.* BigStitcher: reconstructing high-resolution image datasets of cleared and expanded samples. *Nat. Methods* **16**, 870–874 (2019).

102. Andersson, A., Diego, F., Hamprecht, F. A. & Wählby, C. ISTDECO: In Situ Transcriptomics Decoding by Deconvolution. *Cold Spring Harbor Laboratory* 2021.03.01.433040 (2021) doi:10.1101/2021.03.01.433040.

103. Chen, S. *et al.* BARcode DEmixing through Non-negative Spatial Regression

(BarDensr). *PLoS Comput. Biol.* **17**, e1008256 (2021).

104. Cleary, B. *et al.* Compressed sensing for highly efficient imaging transcriptomics. *Nat. Biotechnol.* 1–7 (2021).

105. Asgari Taghanaki, S., Abhishek, K., Cohen, J. P., Cohen-Adad, J. & Hamarneh, G. Deep semantic segmentation of natural and medical images: a review. *Artificial Intelligence Review* **54**, 137–178 (2021).

106. Dries, R. *et al.* Giotto, a pipeline for integrative analysis and visualization of single-cell spatial transcriptomic data. *bioRxiv* 701680 (2019) doi:10.1101/701680.

107. Axelrod, S. *et al.* Starfish: Open Source Image Based Transcriptomics and Proteomics Tools. (2018).

108. Palla, G. *et al.* Squidpy: a scalable framework for spatial single cell analysis. *bioRxiv* 2021.02.19.431994 (2021) doi:10.1101/2021.02.19.431994.

109. von Chamier, L. *et al.* Democratising deep learning for microscopy with ZeroCostDL4Mic. *Nat. Commun.* **12**, 2276 (2021).

110. Moen, E. *et al.* Deep learning for cellular image analysis. *Nat. Methods* **16**, 1233–1246 (2019).

111. Partel, G. & Wählby, C. Spage2vec: Unsupervised representation of localized spatial gene expression signatures. *FEBS J.* **288**, 1859–1870 (2021).

112. Petukhov, V., Khodosevich, K., Soldatov, R. A. & Kharchenko, P. V. Bayesian segmentation of spatially resolved transcriptomics data. 2020.10.05.326777 (2020) doi:10.1101/2020.10.05.326777.

113. Park, J. *et al.* Cell segmentation-free inference of cell types from in situ transcriptomics data. *Cold Spring Harbor Laboratory* 800748 (2020) doi:10.1101/800748.

114. Hu, J. *et al.* Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Cold Spring Harbor Laboratory* 2020.11.30.405118 (2020) doi:10.1101/2020.11.30.405118.

115. Daly, A. C., Geras, K. J. & Bonneau, R. A. A convolutional neural network for common coordinate registration of high-resolution histology images. 2020.09.18.303875 (2020) doi:10.1101/2020.09.18.303875.

116. Biancalani, T. *et al.* Deep learning and alignment of spatially-resolved whole transcriptomes of single cells in the mouse brain with Tangram. 2020.08.29.272831 (2020) doi:10.1101/2020.08.29.272831.

117. Kiemen, A. L. *et al.* In situ characterization of the 3D microanatomy of the pancreas and pancreatic cancer at single cell resolution. *Cold Spring Harbor Laboratory* 2020.12.08.416909 (2020) doi:10.1101/2020.12.08.416909.

118. Lin, J.-R. *et al.* Multiplexed 3D atlas of state transitions and immune interactions in colorectal cancer. *bioRxiv* 2021.03.31.437984 (2021) doi:10.1101/2021.03.31.437984.

119. Qiang, Y. *et al.* A Probabilistic Approach for Registration of Multi-Modal Spatial Transcriptomics Data. *bioRxiv* 2021.10.05.463196 (2021) doi:10.1101/2021.10.05.463196.

120. Ruiz Tejada Segura, M. *et al.* A 3D transcriptomics atlas of the mouse olfactory mucosa. *bioRxiv* 2021.06.16.448475 (2021) doi:10.1101/2021.06.16.448475.

121. Schede, H. H. *et al.* Spatial tissue profiling by imaging-free molecular tomography. 2020.08.04.235655 (2020) doi:10.1101/2020.08.04.235655.

122. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).

123. Amezquita, R. A. *et al.* Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17**, 137–145 (2020).

124. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

125. Dries, R. *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial

expression data. *Genome Biol.* **22**, 78 (2021).

126. Bergenstråhle, J., Larsson, L. & Lundeberg, J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 482 (2020).

127. Righelli, D. *et al.* SpatialExperiment: infrastructure for spatially resolved transcriptomics data in R using Bioconductor. *Cold Spring Harbor Laboratory* 2021.01.27.428431 (2021) doi:10.1101/2021.01.27.428431.

128. Sztanka-Toth, T. R., Jens, M., Karaiskos, N. & Rajewsky, N. Spacemake: processing and analysis of large-scale spatial transcriptomics data. *bioRxiv* 2021.11.07.467598 (2021) doi:10.1101/2021.11.07.467598.

129. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).

130. Andersson, A. *et al.* Spatial mapping of cell types by integration of transcriptomics data. *bioRxiv* 2019.12.13.874495 (2019) doi:10.1101/2019.12.13.874495.

131. Maaskola, J. *et al.* Charting Tissue Expression Anatomy by Spatial Transcriptome Decomposition. *bioRxiv* 362624 (2018) doi:10.1101/362624.

132. Cable, D. M. *et al.* Robust decomposition of cell type mixtures in spatial transcriptomics. *bioRxiv* 2020.05.07.082750 (2020) doi:10.1101/2020.05.07.082750.

133. Aliee, H. & Theis, F. AutoGeneS: Automatic gene selection using multi-objective optimization for RNA-seq deconvolution. *Bioarxiv* (2020) doi:10.1101/2020.02.21.940650.

134. Wang, X., Park, J., Susztak, K., Zhang, N. R. & Li, M. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nat. Commun.* **10**, 380 (2019).

135. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782 (2019).

136. Elosua, M., Nieto, P., Mereu, E., Gut, I. & Heyn, H. SPOTlight: Seeded NMF regression to Deconvolute Spatial Transcriptomics Spots with Single-Cell Transcriptomes. *bioRxiv* 2020.06.03.131334 (2020) doi:10.1101/2020.06.03.131334.

137. Danaher, P. *et al.* Advances in mixed cell deconvolution enable quantification of cell types in spatially-resolved gene expression data. 2020.08.04.235168 (2020) doi:10.1101/2020.08.04.235168.

138. Erdmann-Pham, D. D., Fischer, J., Hong, J. & Song, Y. S. RNA-Sieve: A likelihood-based deconvolution of bulk gene expression data using single-cell references. 2020.10.01.322867 (2020) doi:10.1101/2020.10.01.322867.

139. Su, J. & Song, Q. DSTG: Deconvoluting Spatial Transcriptomics Data through Graph-based Artificial Intelligence. *Cold Spring Harbor Laboratory* 2020.10.20.347195 (2020) doi:10.1101/2020.10.20.347195.

140. Welch, J. D. *et al.* Single-Cell Multi-omic Integration Compares and Contrasts Features of Brain Cell Identity. *Cell* **177**, 1873–1887.e17 (2019).

141. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).

142. Argelaguet, R., Cuomo, A. S. E., Stegle, O. & Marioni, J. C. Computational principles and challenges in single-cell data integration. *Nat. Biotechnol.* (2021) doi:10.1038/s41587-021-00895-7.

143. Kleshchevnikov, V. *et al.* Comprehensive mapping of tissue cell architecture via integrated single cell and spatial transcriptomics. *Cold Spring Harbor Laboratory* 2020.11.15.378125 (2020) doi:10.1101/2020.11.15.378125.

144. Lopez, R. *et al.* A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements. *arXiv [cs.LG]* (2019).

145. Abdelaal, T., Mourragui, S., Mahfouz, A. & Reinders, M. SpaGE: Spatial Gene

Enhancement using scRNA-seq. *bioRxiv* 2020.05.08.084392 (2020) doi:10.1101/2020.05.08.084392.

146. Li, Z., Song, T., Yong, J. & Kuang, R. Imputation of Spatially-resolved Transcriptomes by Graph-regularized Tensor Completion. 2020.08.05.237560 (2020) doi:10.1101/2020.08.05.237560.

147. Achim, K. *et al.* High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509 (2015).

148. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

149. Fleck, J. S., He, Z., Boyle, M. J., Gray Camp, J. & Treutlein, B. Resolving brain organoid heterogeneity by mapping single cell genomic data to a spatial reference. *bioRxiv* 2020.01.06.896282 (2020) doi:10.1101/2020.01.06.896282.

150. Nitzan, M., Karaiskos, N., Friedman, N. & Rajewsky, N. Gene expression cartography. *Nature* **576**, 132–137 (2019).

151. Chlis, N.-K., Rausch, L., Brocker, T., Kranich, J. & Theis, F. J. Predicting single-cell gene expression profiles of imaging flow cytometry data with machine learning. *Nucleic Acids Res.* **48**, 11335–11346 (2020).

152. Tan, X., Su, A., Tran, M. & Nguyen, Q. SpaCell: integrating tissue morphology and spatial gene expression to predict disease cells. *Bioinformatics* 2419 (2019) doi:10.1101/837211.

153. He, B. *et al.* Integrating spatial gene expression and breast tumour morphology via deep learning. *Nat Biomed Eng* **4**, 827–834 (2020).

154. Bergenstråhle, L. *et al.* Super-resolved spatial transcriptomics by deep data fusion. *bioRxiv* (2020) doi:10.1101/2020.02.28.963413.

155. Qin, Y. *et al.* Mapping cell structure across scales by fusing protein images and interactions. *Cold Spring Harbor Laboratory* 2020.06.21.163709 (2020)

doi:10.1101/2020.06.21.163709.

156. Newell, A. & Deng, J. How Useful is Self-Supervised Pretraining for Visual Tasks? *arXiv [cs.CV]* (2020).

157. Lu, A. X., Kraus, O. Z., Cooper, S. & Moses, A. M. Learning unsupervised feature representations for single cell microscopy images with paired cell inpainting. *PLoS Comput. Biol.* **15**, e1007348 (2019).

158. Mund, A. *et al.* AI-driven Deep Visual Proteomics defines cell identity and heterogeneity. *bioRxiv* 2021.01.25.427969 (2021) doi:10.1101/2021.01.25.427969.

159. Mahdessian, D. *et al.* Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* **590**, 649–654 (2021).

160. Viana, M. P. *et al.* Robust integrated intracellular organization of the human iPS cell: where, how much, and how variable? *Cold Spring Harbor Laboratory* 2020.12.08.415562 (2020) doi:10.1101/2020.12.08.415562.

161. Heinrich, L. *et al.* Automatic whole cell organelle segmentation in volumetric electron microscopy. *Cold Spring Harbor Laboratory* 2020.11.14.382143 (2020) doi:10.1101/2020.11.14.382143.

162. Zhu, Q., Shah, S., Dries, R., Cai, L. & Yuan, G.-C. Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nat. Biotechnol.* (2018) doi:10.1038/nbt.4260.

163. Edsgärd, D., Johnsson, P. & Sandberg, R. Identification of spatial expression trends in single-cell gene expression data. *Nat. Methods* **15**, 339–342 (2018).

164. Svensson, V., Teichmann, S. A. & Stegle, O. SpatialDE: identification of spatially variable genes. *Nat. Methods* **15**, 343–346 (2018).

165. Sun, S., Zhu, J. & Zhou, X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat. Methods* **17**, 193–200 (2020).

166. Zhu, J. & Sabatti, C. Integrative Spatial Single-cell Analysis with Graph-based Feature

Learning. 2020.08.12.248971 (2020) doi:10.1101/2020.08.12.248971.

167. Kueckelhaus, J. *et al.* Inferring spatially transient gene expression pattern from spatial transcriptomic studies. *Cold Spring Harbor Laboratory* 2020.10.20.346544 (2020) doi:10.1101/2020.10.20.346544.

168. Kats, I., Vento-Tormo, R. & Stegle, O. SpatialDE2: Fast and localized variance component analysis of spatial transcriptomics. *bioRxiv* 2021.10.27.466045 (2021) doi:10.1101/2021.10.27.466045.

169. BinTayyash, N. *et al.* Non-parametric modelling of temporal and spatial counts data from RNA-seq experiments. *Cold Spring Harbor Laboratory* 2020.07.29.227207 (2021) doi:10.1101/2020.07.29.227207.

170. Li, Q., Zhang, M., Xie, Y. & Xiao, G. Bayesian Modeling of Spatial Molecular Profiling Data via Gaussian Process. *arXiv [stat.AP]* (2020).

171. Zhao, E. *et al.* Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* (2021) doi:10.1038/s41587-021-00935-2.

172. Chidester, B., Zhou, T. & Ma, J. SPICEMIX: Integrative single-cell spatial modeling for inferring cell identity. *Cold Spring Harbor Laboratory* 2020.11.29.383067 (2020) doi:10.1101/2020.11.29.383067.

173. Oh, S., Park, H. & Zhang, X. Hybrid Clustering of single-cell gene-expression and cell spatial information via integrated NMF and k-means. *Cold Spring Harbor Laboratory* 2020.11.15.383281 (2020) doi:10.1101/2020.11.15.383281.

174. Anderson, A. & Lundeberg, J. sepal: Identifying Transcript Profiles with Spatial Patterns by Diffusion-based Modeling. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab164.

175. Schapiro, D. *et al.* histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* **14**, 873–876 (2017).

176. Li, D., Ding, J. & Bar-Joseph, Z. Identifying signaling genes in spatial single cell expression data. 2020.07.27.221465 (2020) doi:10.1101/2020.07.27.221465.

177. Schmid, K. T. *et al.* Design and power analysis for multi-sample single cell genomics experiments. *bioRxiv* 2020.04.01.019851 (2020) doi:10.1101/2020.04.01.019851.

178. Vieth, B., Ziegenhain, C., Parekh, S., Enard, W. & Hellmann, I. powsimR: power analysis for bulk and single cell RNA-seq experiments. *Bioinformatics* **33**, 3486–3488 (2017).

179. Martin, K. C. & Ephrussi, A. mRNA localization: gene expression in the spatial dimension. *Cell* **136**, 719–730 (2009).

180. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015).

181. Burgess, D. J. Spatial transcriptomics coming of age. *Nature reviews. Genetics* vol. 20 317 (2019).

182. Perkel, J. M. Starfish enterprise: finding RNA patterns in single cells. *Nature* **572**, 549–551 (2019).

183. Asp, M., Bergenstråhle, J. & Lundeberg, J. Spatially Resolved Transcriptomes-Next Generation Tools for Tissue Exploration. *Bioessays* e1900221 (2020).

184. Smith, E. A. & Hodges, H. C. The Spatial and Genomic Hierarchy of Tumor Ecosystems Revealed by Single-Cell Technologies. *Trends Cancer Res.* **5**, 411–425 (2019).

185. Wilbrey-Clark, A., Roberts, K. & Teichmann, S. A. Cell Atlas technologies and insights into tissue architecture. *Biochem. J* **477**, 1427–1442 (2020).

186. Zhang, M. *et al.* Spatial molecular profiling: platforms, applications and analysis tools. *Brief. Bioinform.* (2020) doi:10.1093/bib/bbaa145.

187. Zhuang, X. Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat. Methods* **18**, 18–22 (2021).

188. Larsson, L., Frisén, J. & Lundeberg, J. Spatially resolved transcriptomics adds a new dimension to genomics. *Nat. Methods* **18**, 15–18 (2021).

189. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat. Methods* **18**, 9–14

(2021).

190. Close, J. L., Long, B. R. & Zeng, H. Spatially resolved transcriptomics in neuroscience. *Nat. Methods* **18**, 23–25 (2021).

191. Moses, L. & Pachter, L. Museum of Spatial Transcriptomics. *bioRxiv* 2021.05.11.443152 (2021) doi:10.1101/2021.05.11.443152.

192. Qian, X. *et al.* Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* **17**, 101–106 (2020).

193. Prabhakaran, S., Nawy, T. & Pe'er', D. Sparcle: assigning transcripts to cells in multiplexed images. *Cold Spring Harbor Laboratory* 2021.02.13.431099 (2021) doi:10.1101/2021.02.13.431099.

194. He, Y. *et al.* ClusterMap: multi-scale clustering analysis of spatial gene expression. *Cold Spring Harbor Laboratory* 2021.02.18.431337 (2021) doi:10.1101/2021.02.18.431337.

195. He, Y. *et al.* ClusterMap for multi-scale clustering analysis of spatial gene expression. *Nat. Commun.* **12**, 5909 (2021).

196. Caicedo, J. C. *et al.* Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849–863 (2017).

197. Baranski, A. *et al.* MAUI (MBI Analysis User Interface)-An image processing pipeline for Multiplexed Mass Based Imaging. *PLoS Comput. Biol.* **17**, e1008887 (2021).

198. Roberts, K. *et al.* Transcriptome-wide spatial RNA profiling maps the cellular architecture of the developing human neocortex. *bioRxiv* 2021.03.20.436265 (2021) doi:10.1101/2021.03.20.436265.

199. Bahry, E. *et al.* RS-FISH: Precise, interactive, fast, and scalable FISH spot detection. *bioRxiv* 2021.03.09.434205 (2021) doi:10.1101/2021.03.09.434205.

200. Wang, Y. *et al.* De-noising Spatial Expression Profiling Data Based on in situ Position and Image Information. *bioRxiv* 2021.11.03.467103 (2021) doi:10.1101/2021.11.03.467103.

201. Andersson, A. *et al.* A Landmark-based Common Coordinate Framework for Spatial Transcriptomics Data. *bioRxiv* 2021.11.11.468178 (2021) doi:10.1101/2021.11.11.468178.

202. Schapiro, D. *et al.* MCMICRO: A scalable, modular image-processing pipeline for multiplexed tissue imaging. *Cold Spring Harbor Laboratory* 2021.03.15.435473 (2021) doi:10.1101/2021.03.15.435473.

203. Bergenstråhle, J., Bergenstråhle, L. & Lundeberg, J. SpatialCPie: an R/Bioconductor package for spatial transcriptomics cluster evaluation. *BMC Bioinformatics* **21**, 161 (2020).

204. Fan, Z., Chen, R. & Chen, X. SpatialDB: a database for spatially resolved transcriptomes. *Nucleic Acids Res.* **48**, D233–D237 (2020).

205. Queen, R., Cheung, K., Lisgo, S., Coxhead, J. & Cockell, S. Spaniel: analysis and interactive sharing of Spatial Transcriptomics data. *bioRxiv* 619197 (2019) doi:10.1101/619197.

206. Solorzano, L., Partel, G. & Wählby, C. TissUUmaps: interactive visualization of large-scale spatial gene expression and tissue morphology data. *Bioinformatics* **36**, 4363–4365 (2020).

207. Li, D., Ding, J. & Bar-Joseph, Z. Identifying signaling genes in spatial single cell expression data. doi:10.1101/2020.07.27.221465.

208. Dimitrov, D. *et al.* Comparison of Resources and Methods to infer Cell-Cell Communication from Single-cell RNA Data. *bioRxiv* 2021.05.21.445160 (2021) doi:10.1101/2021.05.21.445160.

209. Gao, C. *et al.* Iterative single-cell multi-omic integration using online learning. *Nat. Biotechnol.* 1–8 (2021).

210. Geuenich, M. J., Hou, J., Lee, S., Jackson, H. W. & Campbell, K. R. Automated assignment of cell identity from single-cell multiplexed imaging and proteomic data. *Cold*

Spring Harbor Laboratory 2021.02.17.431633 (2021) doi:10.1101/2021.02.17.431633.

211. Dong, R. & Yuan, G.-C. SpatialDWLS: accurate deconvolution of spatial transcriptomic data. *Cold Spring Harbor Laboratory* 2021.02.02.429429 (2021) doi:10.1101/2021.02.02.429429.

212. Äijö, T. *et al.* Splotch: Robust estimation of aligned spatial temporal gene expression data. *bioRxiv* 757096 (2019) doi:10.1101/757096.

213. Biancalani, T. *et al.* Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).

214. Chen, S., Zhang, B., Chen, X., Zhang, X. & Jiang, R. stPlus: a reference-based method for the accurate enhancement of spatial transcriptomics. *bioRxiv* 2021.04.16.440115 (2021) doi:10.1101/2021.04.16.440115.

215. Fleck, J. S. *et al.* Resolving organoid brain region identities by mapping single-cell genomic data to reference atlases. *Cell Stem Cell* (2021) doi:10.1016/j.stem.2021.02.015.

216. Okochi, Y., Sakaguchi, S., Nakae, K., Kondo, T. & Naoki, H. Model-based prediction of spatial gene expression via generative linear mapping. *bioRxiv* 2020.05.21.107847 (2020) doi:10.1101/2020.05.21.107847.

217. Mattiazzi Usaj, M. *et al.* Systematic genetics and single-cell imaging reveal widespread morphological pleiotropy and cell-to-cell variability. *Mol. Syst. Biol.* **16**, e9243 (2020).

218. Govek, K. W. *et al.* Single-cell transcriptomic analysis of mIHC images via antigen mapping. *Sci Adv* **7**, (2021).

219. Gu, C. & Liu, Z. A network regularized linear model to infer spatial expression pattern for single cells. *Cold Spring Harbor Laboratory* 2021.03.07.434296 (2021) doi:10.1101/2021.03.07.434296.

220. Xu, Y. & Mc Cord, R. P. CoSTA: Unsupervised Convolutional Neural Network Learning for Spatial Transcriptomics Analysis. doi:10.1101/2021.01.12.426400.

221. Hao, M., Hua, K. & Zhang, X. SOMDE: A scalable method for identifying spatially variable genes with self-organizing map. *Cold Spring Harbor Laboratory* 2020.12.10.419549 (2020) doi:10.1101/2020.12.10.419549.

222. Allen, C. *et al.* A Bayesian Multivariate Mixture Model for Spatial Transcriptomics Data. *bioRxiv* 2021.06.23.449615 (2021) doi:10.1101/2021.06.23.449615.

223. Teng, H., Yuan, Y. & Bar-Joseph, Z. Cell Type Assignments for Spatial Transcriptomics Data. *Cold Spring Harbor Laboratory* 2021.02.25.432887 (2021) doi:10.1101/2021.02.25.432887.

224. Verma, A. & Engelhardt, B. A Bayesian nonparametric semi-supervised model for integration of multiple single-cell experiments. *Cold Spring Harbor Laboratory* 2020.01.14.906313 (2020) doi:10.1101/2020.01.14.906313.

225. Sottosanti, A. & Risso, D. Co-clustering of Spatially Resolved Transcriptomic Data. *arXiv [stat.ME]* (2021).

226. Teng, H., Yuan, Y. & Bar-Joseph, Z. Clustering Spatial Transcriptomics Data. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab704.

227. Stroup, W. W. Power analysis based on spatial effects mixed models: A tool for comparing design and analysis strategies in the presence of spatial variability. *J. Agric. Biol. Environ. Stat.* **7**, 491–511 (2002).

228. Ellis, M. M., Ivan, J. S., Tucker, J. M. & Schwartz, M. K. rSPACE : Spatially based power analysis for conservation and ecology. *Methods Ecol. Evol.* **6**, 621–625 (2015).

229. Buller, I. D., Brown, D. W., Myers, T. A., Jones, R. R. & Machiela, M. J. sparrpowR: a flexible R package to estimate statistical power to identify spatial clustering of two groups and its application. *Int. J. Health Geogr.* **20**, 13 (2021).