REVIEW

Answering open questions in biology using spatial genomics and structured methods

Siddhartha G. Jena[1†], Archit Verma[2†] and Barbara E. Engelhardt[3*]

†Siddhartha G. Jena and Archit Verma have contributed equally to this work.

*Correspondence:
bengelhardt@stanford.edu

[1] Department of Stem Cell and Regenerative Biology, Harvard, 7 Divinity Ave, Cambridge, MA, USA
[2] Gladstone Institutes, 1650 Owens Street, San Francisco, CA 94158, USA
[3] Stanford University, 1265 Welch Road, x327, Stanford, CA 94305, USA

Abstract

Genomics methods have uncovered patterns in a range of biological systems, but obscure important aspects of cell behavior: the shapes, relative locations, movement, and interactions of cells in space. Spatial technologies that collect genomic or epigenomic data while preserving spatial information have begun to overcome these limitations. These new data promise a deeper understanding of the factors that affect cellular behavior, and in particular the ability to directly test existing theories about cell state and variation in the context of morphology, location, motility, and signaling that could not be tested before. Rapid advancements in resolution, ease-of-use, and scale of spatial genomics technologies to address these questions also require an updated toolkit of statistical methods with which to interrogate these data. We present a framework to respond to this new avenue of research: four open biological questions that can now be answered using spatial genomics data paired with methods for analysis. We outline spatial data modalities for each open question that may yield specific insights, discuss how conflicting theories may be tested by comparing the data to conceptual models of biological behavior, and highlight statistical and machine learning-based tools that may prove particularly helpful to recover biological understanding.

Keywords: Spatial genomics, Biophysics, Cell biology, Machine learning, Statistical models

Introduction

The invention of the microscope allowed for unprecedented glimpses into the micronscale world, and led to the first characterizations of the cell [1]. A subsequent push to discover the constituent components of the cell led to the development of modern biochemical methods, predominantly based around density centrifugation or chemical separation. In this process, cells and tissues are dissociated and then separated by density to study the subcellular interactions between individual biopolymers. This approach progressively revealed the "parts list" of the cell, illuminating the composition of cellular structures such as the rough and smooth endoplasmic reticulum [2] and the Golgi body [3]. These methods, however, lost the spatial context of where biopolymers were located in the cell, not to mention the relative locations of the cells themselves.

Jena *et al. BMC Bioinformatics*   (2024) 25:291

Page 2 of 25

Imaging and biochemical characterization of cells and tissues have both made incredible progress since their initial development. Advances in physics in the second half of the 20th and early 21st century led to the invention of the electron microscope [4], scanning tunneling microscope [5], atomic force microscope [6], and super-resolution microscopy [7]. Together with fluorescent proteins and affinity reagents (such as increasingly specific antibodies), these instruments opened a new frontier of molecular-level imaging. Scientists could interrogate the spatial location of different proteins, nucleic acids, or lipids within a tissue sample, and associate their distribution with particular cell morphologies or phenotypes.

The development of high-throughput genomic sequencing technologies in the early 21st century led to the characterization of biology at base-pair resolution, first with bulk tissue samples as input, and later within single cells [8, 9]. These protocols revealed the molecular composition of nucleic acids within tissues and cells, but without the spatial or visual context of imaging, since these methods required lysing cells to extract nucleic acids for sequencing.

The parallel technologies of sequencing and imaging have continued to increase in quality and resolution, and have complemented one another in important biological findings. A common post-hoc structure for leveraging the two approaches is to use statistical methods to identify correlations between an imaging-based readout and a sequencing-based readout [10], or predict gene expression levels in a sample using histology imaging [11, 12]; one such example is the mapping of somatic mutations, such as those found in cancer, to a cellular phenotype such as the emergence of dense cancerous tissue that is easily identifiable in pathology imaging [13]. More recently, pairing the two measurements in the same sample has become possible as biochemical methods to study genomics have expanded into the spatial realm. Fluorescence *in situ* hybridization (FISH) methods involve probes that directly bind to proteins, RNA, or DNA of interest, allowing them to be imaged while preserving the location of the biomolecule [14]. Alternatively, cells from a particular region of a tissue section can be sequenced together and reassigned to the tissue image afterwards, providing a coarse-grained view of cell-based gene expression across the tissue. Cells may also be optically barcoded prior to sequencing assays to capture their relative location. Using these methods, the high-dimensional genomic state of a single cell can be measured, and the cell can subsequently be mapped back onto its position in its native context, whether in culture or embedded in a tissue [15]; both cellular state and cellular environment is explicit in these approaches.

Spatial genomics have already been used in a number of contexts to characterize genome-wide changes associated with cellular differentiation, development, interventions, and the progression of diseases such as cancer [16–18]. With this new genome-scale spatially-resolved readout, researchers have the opportunity to discover general principles that govern cellular behavior in their environmental contexts. As better experimental methods are developed, of equal importance are the analytic frameworks that we use to understand and interpret the resulting spatial data.

In this review, we take a look at the types of questions to which scientists may apply spatial technologies, with an eye towards the methods appropriate for analyzing experimental results in the context of open questions in cellular biology. We first summarize the different spatial scales of analysis: molecular-, cellular-, and tissue-level data

Jena *et al. BMC Bioinformatics*   (2024) 25:291

Page 3 of 25

resolutions. We then examine four open questions that may be answered using spatial genomics:

1. What is the functional spatial effect size of a cell?
2. How do cell state and expression profile interact with cellular morphology, movement, and behavior?
3. What local effects shape clonal dynamics in dividing and differentiating tissue?
4. How does a cellular environment shape rare events?

We review current and potential methods for answering these questions from spatially-resolved genomic data. As the suite of spatial genomics tools expands, we hope that the approaches discussed here may be generalized to a broad collection of robust, usable tools and data resources.

Spatial scales of biology

Peer through a microscope at a slice of tissue on a slide, and a wide range of cell shapes, sizes, and patterns present themselves. Further antibody staining reveals the location of proteins in specific intracellular compartments and throughout the extracellular matrix [19]. A tissue sample contains biological processes occurring at three scales: *subcellular* processes taking place within a subcompartment of a single cell, *cellular* processes taking place within $1 - 10$ cells, and *multicellular* processes taking place among > 10 cells (Fig. 1). At the subcellular scale, our questions primarily involve interactions between individual molecules in organelles or membranes. At the cellular level, we ask questions about the overall composition of the cell and interactions with nearby cells. Finally, at the multicellular level, we ask how groups of cells of different types come together to form tissues with multifaceted functions. The scales described here map neatly onto the paradigms of autocrine, paracrine, and encodrine signaling that are common parlance in physiology; however, we hope that generalizing these terms to their relevant length scales may lead to deeper insights about systems not currently or commonly studied in medicine.

I. Subcellular resolution

What molecules are in an individual cell and where do they function? Nucleic acids and individual proteins are largely the drivers of cellular morphology and behavior. Using specific affinity reagents, such as antibodies or oligonucleotide probes, one can identify specific RNA and protein species in a fixed sample, providing insight into function. These molecules are often complexed together; one such example is chromatin, which consists of DNA, histone proteins, and often associated RNAs [20]. Here, we will describe some of the promising use cases for investigating these molecules at subcellular resolution.

DNA: Accessibility and structure

DNA acts as the biological blueprint for an organism. With the exception of somatic mutations [21], cells across an organism largely share the same DNA, yet serve vastly different functions. This functional heterogeneity is made possible through epigenetic
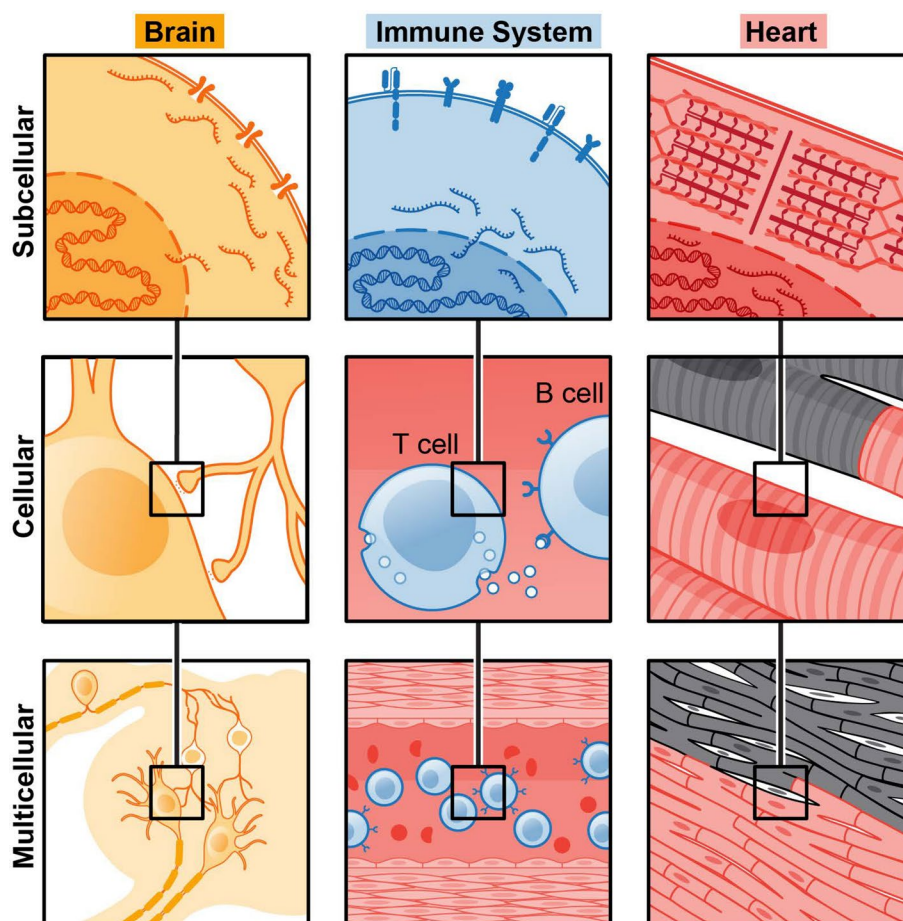
Fig. 1 Distinct scales of organization at different parts of the body. Subcellular localization of receptors, cytosolic proteins, and signaling molecules affects cellular communication between neurons, B and T cells, or cardiac muscle in the heart. Each of these cell types is, further, a components of multicellular assemblies of many neurons, immune cells in the bloodstream, or heart tissue

modifications, which control the genes that are transcribed or repressed in a cell [22]. Structural changes from epigenetic modifications such as DNA or histone methylation [23–25] can lead to differentially accessible regions along the length of the genome. These exposed chromatin regions, which may be read out through methods such as ATAC-seq [26], allow binding of regulatory molecules such as transcription factors and RNA polymerase, leading to transcription. Other modifications, such as histone acetylation, can lead to recruitment of specific transcription factors and result in gene expression [22].

Although distinct chromatin modifications have been associated with transcriptional activation or repression, the spatial organization of the genome and its link to the expression of specific programs remains less clearly defined. The genome is spatially partitioned, in structures largely driven by these epigenetic modifications, into domains of active or inactive genes called A and B compartments, respectively [27]. Although chromatin conformation capture methods such as Hi-C are able to capture these compartments [28, 29], the link between these compartments and their transcription products is still being uncovered on a spatial level within intact cells. What occurs on the border

Jena *et al. BMC Bioinformatics*   (2024) 25:291

Page 5 of 25

between A and B compartments? Are there features that further distinguish different genomic compartments? A deeper understanding of spatial genome organization and its effect on the transcriptome would provide answers to these questions, as well as potentially addressing epigenetic dysregulation, which has been implicated in aging, response to environmental exposures, and disease progression [30, 31].

RNA: Diversity and function

Given the (generally) shared DNA sequences across cells from a single organism, variation in RNA expression is a major driver of cellular variation. Different cell types and cell states show different patterns of RNA expression, but RNAs spatially confined to compartments in the nucleus or cytoplasm are difficult to capture through conventional RNA sequencing. This is of particular interest since the dynamic organization of mRNAs may produce differential protein gradients in a tissue, driving processes such as metabolic regulation [32], polarization during embryonic development, or synapse formation in neurons [33–35]. Beyond mRNAs, a number of noncoding RNAs (ncRNA), such as long noncoding RNAs (lncRNAs) and microRNAs (miRNAs), have been identified and found to have important regulatory functions [36]. Understanding the relationship between ncRNA function and their localization in specific nuclear and cellular compartments, combined with absolute transcript levels, would provide a more complete characterization of the transcriptional state of single cells [37]. Outside of the cell, RNA in extracellular vesicles may be implicated in inter-cellular signaling [38]. Spatial transcriptomics provides previously unavailable insights that will further scientists' understanding of these RNA molecules.

Proteins: localization, abundance, and modifications

When possible, protein measurements provide the most direct window into active cell function. While the prevailing view is that transcript levels correlate with protein levels, possible discrepancies may arise between the two [39, 40], which may necessitate direct measurement of protein levels depending on cellular context [41, 42]. Inferring protein levels from transcript levels also ignores aspects of protein regulation such as localization or post-translational modifications that may activate, modify, or suppress protein function [43]. Antibodies to common protein modifications have allowed scientists to visualize cell processes such as signaling, while more extensive spatial measurements allow for mapping of specific versions of proteins to subcellular locations within individual cells. Once localization of a particular protein or protein family is established, further analyses such as proximity biotinylation [44] or affinity purification mass spectrometry [45] may be performed in the same sample, allowing for deeper insights as to what tasks the protein was performing in its targeted location.

II. Cellular resolution

Complex life is dependent on the cooperation and communication between many diverse cell types. Cell types are often categorized based on their interactions with other cells and tissues: for instance, neurons transmit signals to one another to form the basis of cognition [46], T cells identify and kill foreign cells [47], and cardiomyocytes coordinate signaling between themselves to drive pacemaker activity in the heart [48]. Recent

Jena *et al. BMC Bioinformatics*   (2024) 25:291

Page 6 of 25

advances in high-throughput single-cell measurements allow us to survey this diversity of cell types and interactions from a transcriptional or protein expression perspective.

Variation also exists within cell type, often inelegantly lumped into the vague term *cell state*. For instance, although a population of T cells are likely more similar to one another than they are to a muscle cell or skin cell, individual cells or subsets within the T cell population may be in different states of proliferation, activation, or quiescence at any given moment [47]. Cell states are controlled in part by local interactions between T cells and their environment, causing their transcriptomes and functional responses to diverge [39, 49]. Even in the absence of different environments, there are many sub-types of T cells each with their own cell state profile, and moreover cell states possess a natural level of variation within a population [49, 50]. Some of this variation is due to the stochastic nature of reactions such as transcription or chromatin dynamics occurring in single cells [51]. However, it is an open question how much of heterogeneity is random and how much is a byproduct of factors that are not measured in transcriptomic studies, such as interactions at the spatial boundaries of the cell population [51, 52].

Cells function together, so questions at the cellular scale must consider the interactions with individual cells in a local neighborhood. Receptor-ligand interactions that activate biochemical signaling pathways allow cells to modulate the transcriptomic state of nearby cells [53]. Cells from the same organisms may work together to perform complementary functions, like Schwann cells coating astrocyte neurons with myelin sheets to improve cell-to-cell signaling [54]. Cells from different organisms may also compete in the same tissue; for instance, immune cells fighting an infection or autoimmune diseases [55].

III. Multicellular resolution

In multicellular organisms, groups of many heterogeneous cells come together to form cellular complexes, tissues, and organs. Repeating patterns of multiple cell types in close proximity in a tissue are referred to as *cellular niches* [56–59]. Beyond the identification of the cell type composition of a particular niche, there is considerable interest in understanding niche sizing, the variation in niche architecture, the developmental trajectory of niches, and the interactions between niches [60, 61]. For instance, stem cell niches are of particular interest as they possess the potential to regrow and regenerate specific tissues [62].

Collections of niches create tissue architectures, and spatial transcriptomics presents the opportunity to bring more context to the organization of tissues from a molecular lens. How a repeated niche differs across the tissue may be explained by chromatin modifications or differences in RNA and protein expression, which lead to cell-type heterogeneity. Differences in structure from genetic defects can be explained causally by linking mutations to specific changes that propagate across the tissue [63].

The ambitions of single-cell studies have grown from defining distinct cell types [64] to the creation of comprehensive atlases— from the tissue level [50, 65] to organs [66, 67] to full organisms [68–70], across age [71, 72] and disease status [73–77]. These atlases have the potential to advance biological discovery, in particular because they may provide a more thorough description of the distinct cell states in a larger population. Future work projecting single-cell atlases to spatial scales will add more context to these cell

states, providing insights into the organization of specific cell states, cell types, and cellular niches in tissues and organs.

Key questions in spatial biology

I. What is the functional spatial effect size of a cell?

In a multicellular context, cells use many modes of communication to convey messages to their surroundings (Fig. 2, Question I). The mechanisms by which and extent to which cells are able to communicate across the body has long captivated biologists. The concept of *morphogens*—hormones that enable cellular communication over distance—is an old one [78]. As biochemical tools grew more sophisticated, numerous signaling molecules and pathways were found to serve this critical role [79–82]. These signaling pathways, which are often conserved across organisms, continue to inform research today; for example, live-cell imaging revealed that the Ras/ERK pathway propagates waves of signaling activity during development in response to processes within the cell as well as directing events such as wound healing that take place outside of the cell [83–85]. As we piece together the toolbox of molecules used for cellular communication, it remains unclear how to best measure coupling between cells in a tissue. For a given cell, how do we know how much of its behavior is due to communication with cells around it? How far does this communication reach?

  Biophysical models of cell interactions form a useful framework within which to ask these questions. Perhaps the earliest of these models is the the French Flag Model of morphogen gradients, where particular levels of a molecule are mapped to distinct outcomes in a tissue. Results from this model provided a conceptual explanation of
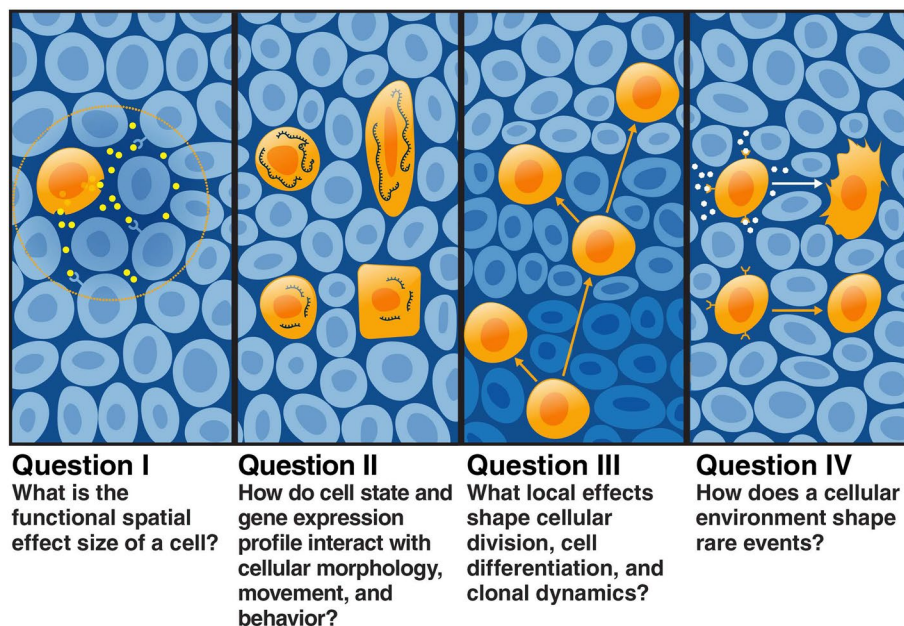

Question I
What is the functional spatial effect size of a cell?

Question II
How do cell state and gene expression profile interact with cellular morphology, movement, and behavior?

Question III
What local effects shape cellular division, cell differentiation, and clonal dynamics?

Question IV
How does a cellular environment shape rare events?

Fig. 2 Four key questions in spatial biology. I. Cells can release ligands that allow them to communicate with Other cells across various, unknown spatial scales. II. Cell location can affect morphology, movement of cells within a tissue and gene expression in unknown ways. III. It remains unclear how dividing clonal cells distribute within a tissue, and how this spatial distribution affects dynamics of gene expression. IV. It remains unclear how rare events in gene expression are influenced and orchestrated in within a tissue

how diffusing gradients of such a chemical could result in patterning along the length of a developing embryo. This class of model led to more mathematically sophisticated descriptions of cellular interactions, including the Turing model, which describes pairwise interactions between two molecules that are capable of generating stable patterns in a tissue [86]. In these early models, the parameters of interest were (i) the geometry of the tissue, (ii) the level of morphogens in space and time, (iii) the feedback, feed-forward, or cooperative interactions occurring between morphogens in space, and (iv) the effects of these morphogens on cell state.

As more interesting biological patterning questions emerged, modeling these behaviors expanded accordingly. Ising models—spin models based on a lattice-structured Markov random field adapted from statistical physics—were used to model cellular interactions [87]. Kuramoto oscillators model coupled cells with continuous states to drive phase differentiation [88]. Alternatively, information-theoretic approaches have been used to understand how small sets of signaling genes encode a rich space of spatial architectures from experimental data, combining mathematical biophysical models with experimentally collected data [89, 90]. All of these modeling paradigms capture the key components of cell connectivity, morphogen levels, and morphogen interactions.

Spatial sequencing provides a high-dimensional dataset (Fig. 3) to statistically identify the genes involved in intercellular communication in different contexts. Early analysis has focused on identifying and building on known ligand-receptor pairs. In the analysis of the initial seqFISH+ results [91], the authors looked for enrichment of expression of known ligand-receptor pairs in adjacent cells by comparing against a null expression distribution created by permutation shuffling. On the same data, graph convolutional neural networks were trained to predict the probability of two genes interacting given the spatial neighbors graph and expression of the two genes in each cell [92]; known ligand-receptor pairs were used as positive and negative examples. Optimal transport methods were used to identify similar distributions of known receptor and ligand expression patterns in spatial data [93], which captured potential interactions beyond spatially-adjacent cells.

Although limiting analysis to known receptors limits the investigator's ability to discover new signaling motifs, testing pairs or higher-order sets of genes for interactions leads to a combinatorial increase in statistical tests and computational demands. Few experiments have sufficient sample size to adequately power investigations of higher-order interactions. Instead, statistical methods often jointly model all genes together and try to learn subsets of co-varying genes that are associated with spatial patterns. For example, Gaussian process regression can model spatial gene expression with clever kernel composition [94]. Three kernels are used to decompose gene expression variance into intrinsic cell effects, extrinsic or environmental effects, and cell-cell interactions. Comparison to a null model assuming no cell-cell interactions identifies communication-related genes. Related work reconstructs gene expression from given cell-type labels and spatial neighbor graphs using autoencoder architectures [95]. While not explicitly using the expression levels of other genes, the cell-type label serves a similar role in capturing expected nonspatial gene-gene correlations. This work similarly uses a null model without spatial connectivity to test for interacting genes. For both strategies of testing pairs of genes or a gene against all other genes, the number of tests done requires proper null
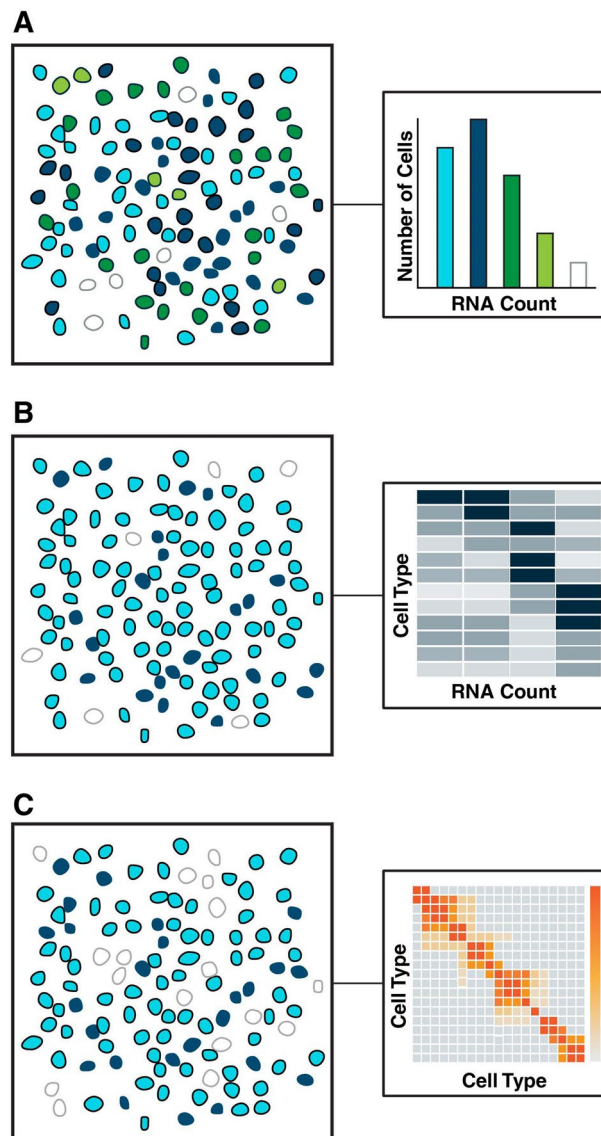
Fig. 3 Essential cellular behaviors assayed in spatial genomics. Distributions of RNA (**A**), cell type clustering from gene expression (**B**) and spatial correlations (**C**) can all be measured from spatially resolved sequencing data

models, multiple hypothesis testing correction, and a reliable way to control for known cell-type heterogeneity and adjacency, which may bias results.

Single snapshots of spatial expression data sets miss important information on the temporal nature of signaling. Parameters such as the responsivity of a cell type to a particular protein, or the pairwise interactions between two genes, may change as a function of time. For example, spatial measurements at two stages of uterine development were used alongside CellPhoneDB [96, 97], a database of know ligand pairs, to identify which cell types had compatible signaling proteins and were likely to be in communication during development [98]. Increasing the resolution of time points will allow the expansion of such statistical techniques to identify interactions over time. Time-series analyses can also help identify more causal relationships in signaling. Fluorescent live-cell imaging

data and point-process models were used to quantify ERK signaling and downstream Fos expression under different drug treatments [99]. Specifically, self-exciting Hawkes processes model expression and signaling among cells over time and space. Newer fluorescence imaging protocols will expand the number of behaviors that can be captured simultaneously in live-cell imaging data [100].

Understandably, we currently have the most confidence in interactions between directly adjacent cells, since long-distance channels or indirect secretory pathways through which cells can send or receive messages are more difficult to study. By incorporating multiomic measurements of cells in space, as well as potentially integrating time-resolved measurements, we may be able to better understand cell communication at a distance.

II. How do gene expression profiles interact with cellular morphology?

A common practice in both basic cell biology and pathology is to use cell morphology to distinguish cell types or states. Cellular structure informs function, and thus cells from different tissues and different cell types in a single tissue vary markedly in their appearance (Fig. 2, Question II). For instance, due to dysregulation in growth pathways, cancerous cells are commonly larger than their healthy counterparts, and are often more motile when imaged over time under a microscope. Physical stress can also change cell state and gene expression; mechanical stretching of fibroblasts was found to dramatically alter their epigenome states to enable cells to prevent damage to the physical structure of the genome [101].

Before spatial single-cell technologies, some methods attempted to jointly model morphology patterns and gene expression from paired bulk tissue samples [10, 102]. The limitation here is the mismatch between resolutions: Images can provide cellular-level phenotypes, but bulk expression cannot. The emergence of spatial single-cell measurement techniques is poised to overcome this limitation.

One open question is how to best represent cellular morphology. While gene expression is conventionally represented by a count matrix, there is not a similar universal tabular form to represent cellular morphology. Recent approaches have attempted to provide solutions to this problem. One strategy is to convert morphological data, generally in the form of images, into tabular data of derived features. One study measured gene expression with the L1000 assay and morphological features such as nuclear area or DNA organization using the Cell Painting assay [103]. Lasso logistic regression and a multi-layer perceptron accurately predict gene expression from ~1000 morphological traits provided by CellProfiler [104]. CellProfiler was also used to create tabular readouts from paired imaging and single-cell CRISPR-Cas9 knockouts, in order to cluster gene knockouts with similar morphological changes and build genetic interaction networks [105].

An alternative approach is to use neural networks to capture the features of an image. For example, MorphNet uses a variational autoencoder to encode the cell state into a lower dimension, and a generative adversarial network (GAN) [106], which jointly optimizes two neural nets–one to generate imaging samples from the encoded gene expression that look like real imaging data, and the other to predict whether the generated imaging sample is real or fake—in order to predict cellular morphology from gene

expression in brain-wide MERFISH data [107]. A recent study collected a paired CRISPR knockout and imaging dataset, as in related work [105], but calculated embeddings of images from a self-supervised vision transformer trained on single-cell Cell Painting images [108]. This approach outperformed classic image featurization for classifying single CRISPR perturbations' mechanism of action and recovering known biological relationships between genes [108].

Each approach to study the relationship between cellular morphology and cell state has different strengths and limitations. Individual features derived from cell painting methods are easier to interpret and can be used in small sample size regimes. Tabular data is amenable to traditional statistical regression methods and the accompanying theoretical guarantees; however, count data and specific experimental designs often require additional structure on the methods that are challenging for non-statisticians. Neural networks provide more flexibility in the morphological variation that they capture, but require both an adequate amount of data for training and some expertise in training and interpreting the models. Both approaches also require identifying which variable should be the output and which variable should be the input.

Biologically, cell morphology and movement are determined largely by membrane contacts; cell membranes are predominantly composed of lipids and proteins, and the abundances of these components are largely dictated by gene expression. However, changes in morphology and motion also drive changes in gene expression as the cell responds to new conditions. Many signaling pathways begin with external influences on membrane proteins. These feedback loops suggest the causal relationship goes both ways, limiting static data to providing mostly correlative findings.

In the near future, decreases in costs and improvements in resolution will allow scientists to better establish the causal relationships between gene expression and morphology. Time-series measurements and live-cell imaging can uncover the temporal ordering between gene or protein expression events and morphological changes. Single molecule tracking will be able to resolve where in the cell proteins are functioning and creating structural features [109]. These improvements will shine further light on the relationship between morphology, motion, and function. With improved experimental methods and proper statistical techniques, a complete understanding of the determinants of cell morphology is within grasp.

III. What local effects shape clonal dynamics of dividing and differentiating cells?

Cell division establishes populations of clones in various tissues around the body (Fig. 2, Question III). Cell division may maintain genomic, transcriptomic, and epigenomic information, but comes with the downside of passing on potentially deleterious properties such as DNA mutations and aberrant epigenomic states. On the other hand, precise maintenance and expansion of particular clones underlies important processes such as the development of adaptive immunity. Some biological processes are "bottlenecked" in the sense that unfit clones die out due to physiological conditions [110]. However, many cell populations, including hematopoetic stem cells that give rise to the entire lineage of circulating blood cells, are comprised of hundreds or thousands of clonal populations, including clones that harbor mutations that decrease proliferation [111], suggesting that clonal heterogeneity may be the rule rather than the exception.

Jena *et al. BMC Bioinformatics*   (2024) 25:291

Page 12 of 25

Biophysical models of clonal dynamics have been studied for many years in the context of stem cells. A primary question for any stem cell population is whether and how the population replenishes. This has been modeled by three parameters representing three distinct probabilities of division outcomes for a stem cell: (1) division into two stem cells, (2) two differentiated cells, or (3) one stem cell and one differentiated cell [112, 113]. Recently, another intriguing possibility has been introduced: rather than remaining in static states, cells may stochastically transition between stemlike and differentiated states with some probability, only fully converting to a distinct fate when faced with a particular signal [114].

These relatively simple "state transition" models, applied to well-mixed or spatially structured populations, have been used to great effect to predict behavior of stem cell populations. Crucially, certain regimes representing distinct probabilities of differentiation or division can be distinguished from one another experimentally through the resulting predicted distributions of clone sizes. One early method for marking clones involves dosing subsets of cells with a dye that subsequently becomes diluted over time, a method that is commonly used to monitor T cell proliferation [115]. While this can accurately mark the generation, it does not provide direct linkage across generations. Another method involves inducing fluorescent protein expression in a subset of cells, using this to identify groups of fluorescent cells that all originated from the same clone. Similarly, this approach does not allow for identification of mother-daughter cell relationships, but can be used to measure clone size distributions by quantifying the size of distinct groups.

Experimental methods to identify mother-daughter relationships between single cells within a clonal population, on the other hand, were difficult until CRISPR-Cas9 was developed. DNA-based barcodes for clonal tracking are an attractive technological development towards addressing clone-related questions. Static barcodes can be introduced into the genomes of cells in a random fashion, so that some distribution of barcodes is introduced into the first generation and subsequently passed on to each cell's progeny [116, 117]. Through subsequent DNA sequencing, the barcode for each cell can be recovered to establish clusters of cells that arise from the same clone. More recently, dynamic barcoding can be used to establish precise lineages: in this method, CRISPR-Cas9 randomly edits a barcode as it is passed on from cell to cell, allowing researchers to reconstruct lineages through the introduction of random SNPs [118, 119].

Combining imaging-based methods with barcoding offers an opportunity to build models of clonal expansion in a spatial context. In particular, work extending clonal dynamics models to the mammalian epidermis exposed complex emergent clonal behavior that arises when cells are confined to a tissue [112, 113]. The epidermis is highly stratified, and, within a layer, clonal populations of stem cells can often be visualized as spatially defined clusters of mitotically active or inactive cells. Post-mitosis, differentiated cells that arise from a stem cell on a basal layer will often emerge on a suprabasal layer, giving rise to complex geometries of clone dispersion spanning three dimensions [120–123]. Specific functional geometries of tissues, such as the crypts of the stomach or the cylindrical structure of vasculature, likely involve similarly unique geometric patterning of clones.

We expect that interrogating clonal populations in their native tissue through a combination of imaging, barcoding, and transcriptomics will allow for a broader range of clonal behaviors to be defined. In particular, although clonal populations tend to be "coarse-grained," as observed in the epidermis as well as in metastatic clones using spatial DNA sequencing [18], it remains to be seen how fluid individual clonal populations are within a tissue.

Are there definable subclones within a clone that occupy their own spatial niche? In the case of cancers, cells from one clone may metastasize to form their own population elsewhere. In what ways is this subclone distinct from the original? Prior work used variance decomposition of Slide-seq data to identify gene signatures that explained differences between distinct clones as well as subclones within cancerous tissue [124]; similarly, constrained regression and covariance estimation were used to study clonal populations using copy number variation [125]. Related work jointly identified copy number polymorphisms in spatial transcriptomics and inferred cellular clones in tissues using a hidden Markov random field [126]. Extending spatial experiments using dynamic barcoding would allow for fine-grained resolution of subclone emergence in the future; analytic methods to reconstruct the full clonal trajectories would add specific mother-daughter relationships in space.

IV. How does a cellular environment shape rare events?

The first single-cell RNA-seq datasets confirmed what many biologists had already suspected: that substantial expression heterogeneity exists between cells in a tissue, and that this heterogeneity underlies a wide range of diseases. For instance, cancers often arise not as a function of cellular collectives, but as a function of one particular cell. A dominating paradigm in cancer is that single cells experience a perfect storm of factors that lead to them becoming *jackpot cells*, or clones that express a specific mRNA at extremely high levels while their sister clones express none [127] (Fig. 2, Question IV). In some cases, these rare cellular states are transient; jackpot cells may not always express combinations of genes throughout their lifetime, and may not pass on their phenotypes to their progeny. In BRAF melanoma, jackpot cells fail to follow Luria-Delbruck behavior and do not pass on their properties unless challenged with the addition of a drug, which then stably locks in the resistant state [127]. This implies that every time a population of cancer cells is challenged with a drug, a constant but small fraction of the population experiences stochastic resistance. Other rare cellular phenotypes are more consistent with Luria-Delbruck dynamics; for instance, rare mutations causing cancerous growth are passed on from mother cell to daughter cell to create large colonies and eventually solid tumors [128].

While we are beginning to understand the factors affecting jackpot cell emergence in culture, the environmental factors (e.g., tissue niche, surrounding cellular milieu, position in the tissue) that regulate the cell states giving rise to heterogeneous gene expression events are still unknown. Leveraging spatial genomics to identify these rare events such as jackpot cells among other cells in a tissue may lead to a better understanding of these rare events. However, a major limiting factor in studying such rare events is statistical confidence in detecting such events. In studies performed on melanoma cells, jackpot cells were detected using RNA-FISH with probes targeting a pool of pre-identified

drug resistance genes [127]; this allows for high-confidence calls of jackpot cells that may not be possible in standard single-cell sequencing workflows. The total number of mRNA transcripts per cell is typically much lower than the mRNA counts collected using FISH methods, especially in such a small pool of target genes. In this particular case, bulk RNA-seq was used to identify a set of high-confidence genes for RNA-FISH probing. However, the candidate genes designating jackpot cells may not always be so well defined, and using a sparse readout such as single-cell transcriptomics to identify novel jackpot cells presents a circular problem.

Methods opportunities in spatial biology

These four open questions in spatial biology—along with the existing or forthcoming technologies to observe the corresponding biological phenomena in tissues—require the development of statistical approaches to arrive at precise and reproducible answers. The opportunity here is in building models that incorporate additional structure—time, space, or environmental context. Here, we outline opportunities for methods development in each of the four areas, focusing on methods that are most likely to be successful given the constraints of the data and sample size (Figure 4).

To illustrate the structure of potential novel and existing methods, we assume that we start with one of two structured datasets. The first dataset is two tables, a cell by gene (or other feature) count matrix $X \in \mathcal{R}^{N \times G}$ and cell by spatial coordinate matrix $C \in \mathcal{R}^{N \times D}$, where N is the total number of cells assayed, G is the number of genes assayed, and D is the number of spatial dimensions (this will generally be 2 or 3). We will use $x_{i,j}$ to refer to the count of gene j in cell i, $x_{i,-j}$ to represent the gene counts in cell i of genes other than j, x_i to refer to the vector of all gene counts in cell i, and x_{-i} to refer to all gene expression in cells other than i, with similar subscripts for the coordinates.

Alternatively, we may have a more granular set of observations of the identity of each of M molecules observed (e.g., spatially localized RNA transcripts), with a location for each molecule $c_m \in R^D$, and the cell it belongs to $o_m \in 1, 2, \ldots, N$. We will use c_i to loosely refer to the coordinates of all molecules in cell i and m_i for the identity of all molecules in a cell i. In the following section, question-specific data and notation will also be introduced to illuminate the modeling approaches proposed. For each opportunity, we try to identify challenges across data collection, model architectures, and model inference and evaluation.

I. Methods to characterize the functional spatial effect size of a cell

A spatial experiment observes an instance from some distribution over the expression and spatial coordinates of the cells, $p(X, C)$. Signaling between cells implies there is some conditional relationship of a cell's state on other cell's state. A model to identify spatial signaling assumes that the variability of cell state can be decomposed into factors from other cell states (*extrinsic factors*) and cell-specific factors (*intrinsic factors*) [129, 130]. This may look like a model with form:

$$x_{i,j}|x_{i,-j}, x_{-i}, C \sim f_1(x_{i,-j}) + f_2(c_i) + f_3(x_{-i}, C) + \epsilon,$$

where ϵ is some noise distribution. We use f_1 to represent how cellular state feature j is dependent on the other state features in the cell (i.e., intrinsic factors). Often, *cell type*
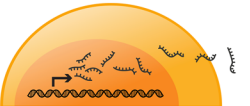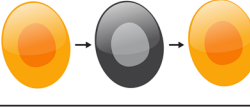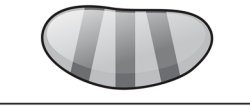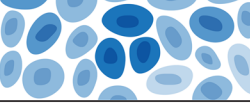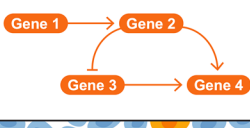
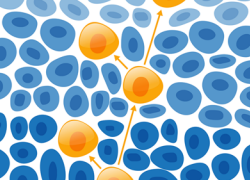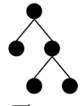|		
Diffusion *how far do RNA molecules or proteins move in space after they are synthesized?*		$\frac{\partial \phi(r,t)}{\partial t} = \boldsymbol{D}\nabla^2\phi(r,t)$ learn the diffusion constant D given spatial measurements ϕ
Dynamics *what is the oscillatory behavior of a signaling pathway measured over time?*		$x(t) = A\cos(\omega t)$ learn the amplitude A and frequency ω of the oscillator
Communication *how is a cell's behavior controlled by the behavior of cells around it?*		$P(x_i)\|P(x_j)$ learn the dependency of a certain gene expression profile as a function of adjacent cell(s)
Patterning *what interactions give rise to a particular set of patterns in a multicellular tissue or organism?*		$\partial_t q = D\nabla^2 q + R(q)$ Learn the set of intermolecular reactions R and diffusion D
Movement *how do cells move within a tissue, and how is this affected by their environment?*		$\frac{\partial x}{\partial t},\ \frac{\partial^2 x}{\partial t^2}$ directly measure velocity and acceleration from timecourse data
Heterogeneity *what governs differences in cell gene expression within a tissue?*		$P(x_i\|c,x)$ learn the probability of a rare state from locations and gene profile of surrounding cells
Gene Regulatory Networks *how do genes interact to produce diverse outcomes in a cell?*	Gene 1 → Gene 2 Gene 3 → Gene 4	learn the adjacency matrix of genes 1 through J $\begin{bmatrix} a_{11} & a_{12} & \cdots \\ \vdots & \ddots & \\ a_{J1} & & a_{JJ} \end{bmatrix}$
Clonal Behavior *how do cells from the same single origin distribute themselves in a tissue?*		$P(\mathcal{T}\|x,c)$ learn the lineage tree $\mathcal{T}$ given coordinates and cell states

Fig. 4 Methodological opportunities for spatial genomics. We describe distinct "classes" of biological and biophysical measurements that fall within our four key areas of interest. These include diffusion of RNA away from the site of transcription, establishment of patterning in a multicellular tissue or organism, and gene regulatory networks giving rise to particular behaviors. For each, we describe how the underlying processes may be directly measured, or indirectly inferred, from spatial genomics data

is used as a proxy for intrinsic cell state. Methods for dimension reduction, such factor analysis, fit without spatial information can also be used to find the intra-cellular covariance between features of cell state.

Next, f_2 represents a spatial pattern of cell state that is a function of location but not environment. This variability may reflect some global architecture of cell types and niche organization. It accounts for variation in cell state that is not part of the signaling pattern that we are attempting to find. For example, a tissue sample might be organized with different cell types separated in distinct regions of space, which creates a spatial pattern of gene expression that is not the result of short term signaling behavior. We can think of a model like nonnegative spatial factorization [131] as decomposing the variance among these two terms: non-spatial factors capture the

intra-cellular covariance while spatial factors learn the spatial archetypes for each feature of cell state.

The final term, f_3, represents perhaps the most interesting behavior—the dependence of cell state on local cells. We are looking for repeated patterns of variation in cell state that cannot be explained by the other features in a cell or by global patterns of expression. Of critical importance is correctly teasing out this relationship from our spatial term f_2. This can be done by restricting the cell's dependence to only neighboring or nearby cells, making f_3 represent the unique local covariance of gene expression.

As observed before, spatial factor models tend to capture only the first two functions while missing local signaling effects [131, 132]. Looking at the correlation between known ligand-receptor pairs expression across neighbors uses cell type to control for the effects f_1 and proximity to zero out f_2, testing specifically for the existence of f_3.

Thus the opportunity remains to fully model all three factors simultaneously. Data with distinct local and global signals are essential for a model to learn the desired patterns. The appropriate functional forms of each term will be required to accurately capture biological processes; nonlinear functions will likely be necessary for an accurate model, although they will increase the difficulty of inference and also the data requirements for adequate power. For f_3, given the most obvious adjacency heuristics, models can estimate signaling between adjacent cells, but more complex communication across larger spatial scales may be hard to detect effectively. Ideally, these models can look at signaling across all features, though computational complexity may require low-dimensional latent factor representations to tractably model complex signaling. Bayesian representations can provide proper uncertainty quantification and identify multiple parameter optima that explain the data equally well, but also require more expensive computation for posterior distributions.

Using the specific location of molecules, our second data representation allows for increased granularity and ability to look for causal signals. Here, models can explicitly condition on the location of a molecule inside or outside a cell as a proxy for determining its contribution to signaling behavior. Proteins near the membrane, for example, are more likely to be involved in some extracellular signaling than nuclear proteins. In these cases, the coordinate of a molecule might be considered rather than the cell center:

$$m_i, c_i | m_{-i}, c_{-i}, o_i, o_{-i} \sim f_1(o_i) + f_2(c_{n/-i}, m_{n/-i}) + f_3(m_{-n}, o_{-n}, c_{-n}) + \epsilon,$$

where $c_{n/-i}$ and $m_{n/-i}$ represent the coordinates and identities of molecules in cells other than i, and ϵ is white noise. Here, f_1 is dependent on the cell type, positing some shared spatial organization across cells of the same type. f_2 accounts for some variation from the organization of the other molecules in the cell and f_3 accounts for variation from molecules outside the cell. In this setup, f_2 captures intra-cellular signaling, perhaps of some cascading pathway, and f_3 captures inter-cellular signaling.

Like the models based on cell count tables, opportunities exist to model local and global effects at molecular resolution. Data with both intra-cellular and inter-cellular behavior measurements will be needed to calibrate the effectiveness of such a model, though the identification of known pathways serves as a model evaluation metric. Similar computational challenges in terms of dimension of possible gene-gene interactions

will plague these kinds of models, which may require the development of latent variable models of single-cell spatial organization.

For both approaches, we are missing an important variable in the time dependency of signaling. A spatial measurement only provides a snapshot of the present; some molecules may be moving towards their destinations while others with important interaction effects may have just degraded. The current location may not be entirely informative about the relevant signaling actors. As multiple spatial snapshots and live-cell imaging become more affordable and widespread, models that explicitly include dynamic behaviors will be invaluable for establishing causality in biological signaling processes.

II. Methods to investigate the relationship between morphology and expression

When biologists study the relationship between morphology and expression, they require measurements of cell shape and molecular counts. These may come from paired histology and sequencing or a combination of cell segmentation and count measurements from *in situ* fluorescence. In addition to our count matrix X from earlier, let $\mathcal{S}$ containing $s_i \in 1, 2, \ldots, N$ cells represent the measurements of morphology, generally images or derived features. An experiment captures one realization of the distribution over morphology, cell position, and cell molecular state $p(\mathcal{S}, X, C)$.

The analysis methods that currently exist that connect cell morphology and state make two strong assumptions: first, that observations from each cell are independent, and, second, that the position of the cell in space does not affect the morphology: $p(s_i, x_i | s_{-1}, x_{-1}, C) = p(s_i, x_i)$. Then, one set of measurements is defined as a function of the other; shape as a function of gene expression, $s_i | x_i \sim f(x_i)$, or gene expression as a function of shape, $x_i | s_i \sim f(s_i)$. This is a reasonable assumption to make with current data and suggests a tractable class of model, but it obscures the complexity of the underlying mechanobiology that considers both intrinsic cell state and extrinsic environmental factors in cell morphology [133].

A natural opportunity in this space is to jointly model morphology and expression together, possibly by representing morphology using functional data analyses [134] or an autoencoder. Within a latent variable model, we may learn a shared representation of both cellular state and some encoding of cellular morphology Z given some form $s_i, x_i | z_i \sim f(z_i)$. Canonical correlation analysis, for example, has been used to jointly learn embeddings of gene expression and histology images for bulk RNA-seq data [10]. Given sufficient single-cell data for network training, similar methods could be used to capture the relationship at single-cell resolution without causality assumptions.

More intriguing are models that are able to capture the effect of nearby cell morphology and expression, similar to the signaling models explored before. A simple model would decompose the likelihood of expression and morphology, $p(x_i, s_j | x_{-i}, s_{-i}, C)$, into terms representing the intrinsic cell morphology and deviations induced by environmental effects. With appropriate data, one could imagine more sophisticated models that are able to account for the organization of cells alongside their shapes and expression, a full joint model of $p(X, \mathcal{S}, C)$. Models of this type will likely require multiple replicates, both technical and biological, of spatial experiments to accurately estimate these distributions. But the increased use of spatial experiments and expanded fields-of-view in each sample will open these avenues for investigation.

Returning to our second data representation—the list of molecular locations, identities, and cellular groupings—the opportunity exists to model the molecular level effects on morphology and, conversely, the change in spatial distribution of molecules given morphology. A simple model might rely on the assumption of independence between cells, and posit that $s_i | m_i, c_i \sim f(m_i, c_i)$. The correct functional form will depend on the representation of the shapes in $\mathcal{S}$; some tabular featurization can take advantage of regression models while a full image might require a neural network or other non-linear model. Most exciting would be a model that can capture biophysical properties of the molecular interaction, learning how specific proteins or RNA molecules lead to the formation and warping of individual cell parts such as membrane structures within and across cells. Natural extensions would jointly consider cellular niches, to model $p(\mathcal{S}, M, O, C)$.

For biologists who study dynamic processes such as development or cell response, time-dependent models will be the key to answering those scientific questions. The desired model will include the evolution of expression and morphology as a function of time, $p(\mathcal{S}, X, C | t)$ or $p(\mathcal{S}, M, O, C | t)$. These models, coupled with appropriate data, may untangle the order in which morphology changes drive expression or expression changes morphology. Fitting models with clear biophysical structure—combined with hypothesis testing—may be one strategy to obtain interpretable and quantifiable results, e.g., estimating the mechanical forces contributed from membrane proteins on maintaining rigidity. Combining flexible machine learning methods with a biophysical interpretation will likely be required to fully capture the complexity of these dynamic morphological processes.

III. Methods to investigate how cellular environment shapes cellular state, cellular division, cell differentiation, and clonal dynamics

An exciting future direction is to map existing lineage-tracing methods onto spatial coordinates to better understand the spatial distribution and behavior of clones. Within our hypothetical framework, let us imagine that we have a count matrix X and spatial coordinates C, as well as some additional data structure Q that defines the relationship between cells (i.e., mother-daughter relationships in cells or cells that are part of one clonal population). One way to represent the ancestry of cells is by making Q an adjacency matrix that represents a directed tree, where $Q_{ij} = 1$ if cell i is a daughter of cell j. Connected components in this graph represent clonal outgrowth, and can be traced back to a single progenitor.

Although current analyses can identify clonal population sizes, it remains an open question whether these sizes are governed by cell-intrinsic or cell-extrinsic factors. If a set of cells Y represent a connected component of Q, we can identify generations at which clonal expansion slowed or halted, and ask whether clonal size (the cardinality of Y, $|Y|$) is a function of expression in surrounding cells, $|Y| = f_1(x_{-Y}, c_{-Y}) + f_2(x_Y)$, where x_Y, x_{-Y} are the expression profiles of cells in Y and all cells not in Y, respectively. Similar to our spatial signaling framework, this treatment decouples the effects on clonal population size into clonal effects and the effects of environment around the clone.

Using this framework, we can also ask spatial questions about cells within a single clonal lineage: single-cell sequencing is able to resolve these populations but, before

spatial sequencing, was unable to resolve their location. In some tissues, clones of cells remain close to each other in space and share a common niche. However, it is also possible for clones to split, migrate away from each other, and otherwise disperse in space. Given a set Y of clones originating from a single cell, we can study their dispersion patterns across space. Taking inspiration from our discussion of spatial signaling limits in cells, we can define a radius r and compute the probability of a given clone lying within radius r from other clones in the population, $p(\text{dist}(c_i, c_j) \leq r | i, j \in Y)$.

We can also ask whether this clonal colocalization is more, less, or equally likely if cells come from the same clone. This value can be calculated and tested for multiple clonal populations $q_1, q_2, \ldots$ to identify clone-specific spatial distributions and behaviors of daughter cells to stay close to their mother or intentionally disperse. If there are members of a clonal lineage that are separated in space, we can then ask how this stratification may have occurred as a function of cell state as well as the local cell population: $\text{dist}(c_i, c_j) \leq r | i, j \in Y \sim f(x_{i,j}, x_{-i,-j}, c)$.

With sufficient spatial genomic data, learning the function f would most likely give higher weight to cells closer to the clones of interest, while also capturing environmental factors that define spatial clonal heterogeneity. The driving factors behind this spatial segregation may also be differentiation in the clones themselves; for instance, in the layers of the epidermis, cells from a single clone differentiate as they stratify from basal to apical [135]. In this case, spatial segregation may largely be a function of the intrinsic cell state within clones $x_{i,j}$, and these effects, too, can be decoupled from effects from local cells.

IV. Methods to understand the relationship between cellular environment and rare events

A number of methods are needed to resolve the relationship between cellular environment and rare events. First, identification of rare events is essential but challenging given current pipelines. Currently, rare events are often filtered or overlooked in spatial transcriptomic data. Rare transcription events may not be captured without sufficient cells [136], and even when present may not be detected [137], often inseparable from poorly detected gene expression patterns [138]. For example, jackpot cells likely will not be identified because of the large numbers of zeros in marker transcripts of rare cell types across all cells, leading to marker genes being removed from the analysis and preventing identification of rare cell types. The opportunity here is to work with the mapped but unfiltered data to identify rare cell types through rare marker gene profiles.

Second, understanding the environmental characteristics that lead to rare events requires phenotyping a cellular environment and testing for enrichment of rare events within specific types of cellular environments. A number of methods perform related analyses, quantifying differential cell-type adjacency across a tissue [139], functional cellular collectives [140–142], and identifying de novo spatial domains [131, 132, 143].

Third, identifying enrichment of specific rare events within a cellular environment may be challenging given the paucity of these rare events and the complexity of a cellular environment. Outlier detection methods may be useful in this space, but these methods are broad; in the context of probabilistic models, identifying cells that have a low probability of being generated from a foundational model or latent space model of diverse cells may suggest a rare cell type or cell state [144–146]. A marked Poisson process may

be useful to identify enrichment of specific environments in which these rare cell types arise. Marked Poisson processes consider specific events (here, a rare cell type) in the context of time or space with a "mark" or an identifier; then specific marks will filter up as enriched for rare events.

Concluding remarks

The rapid development of spatial genomics technologies, for the first time combining spatial imaging of cells and tissues with an analysis of their state and genomic profiles, provides an opportunity to revisit the types of questions we are able to ask and the quantitative methods we may use to answer those questions.

Here, we present four fundamental biological questions, each with profound implications for health and disease, that can now be addressed using spatial genomics technologies combined with appropriate machine learning methods. Future work will build on existing spatial genomics technologies and tailored analyses through the integration of time series data, better predictions of short-range and long-range correlations in multi-omic spatial datasets, and the ability to reason about biological processes across many scales.

Declarations

Ethics approval and consent to participate
No human subjects were involved in this paper.

Consent for publication
Not applicable.

Competing interests
BEE is on the SAB of Creyon Bio, Arrepath, and Freenome; BEE consults for Neumora. AV consults for NE47 Bio.

References
1. Mazzarello P. A unifying concept: the history of cell theory. Nat Cell Biol. 1999;1(1):13–5.
2. Fujiki Y, Hubbard AL, Fowler S, Lazarow PB. Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum. J Cell Biol. 1982;93(1):97–102.
3. Ehrenreich J, Bergeron J, Siekevitz P, Palade G. Golgi fractions prepared from rat liver homogenates: I. isolation procedure and morphological characterization. J Cell Biol. 1973;59(1):45–72.
4. Koster AJ, Klumperman J. Electron microscopy in cell biology: integrating structure and function. Nat Rev Mol Cell Biol. 2003;4(9):6–9; SUPP.
5. Hansma PK, Tersoff J. Scanning tunneling microscopy. J Appl Phys. 1987;61(2):1–24.
6. Alonso JL, Goldmann WH. Feeling the forces: atomic force microscopy in cell biology. Life Sci. 2003;72(23):2553–60.
7. Schermelleh L, Ferrand A, Huser T, Eggeling C, Sauer M, Biehlmaier O, Drummen GP. Super-resolution microscopy demystified. Nat Cell Biol. 2019;21(1):72–84.
8. McGettigan PA. Transcriptomics in the rna-seq era. Curr Opin Chem Biol. 2013;17(1):4–11.

9. Kulkarni A, Anderson AG, Merullo DP, Konopka G. Beyond bulk: a review of single cell transcriptomics methodologies and applications. Curr Opin Biotechnol. 2019;58:129–36.

10. Ash JT, Darnell G, Munro D, Engelhardt BE. Joint analysis of expression levels and histological images identifies genes associated with tissue morphology. Nat Commun. 2021;12(1):1609.

11. Schmauch B, Romagnoni A, Pronier E, Saillard C, Maillé P, Calderaro J, Kamoun A, Sefta M, Toldo S, Zaslavskiy M, et al. A deep learning model to predict RNA-Seq expression of tumours from whole slide images. Nat Commun. 2020;11(1):3877.

12. Comiter C, Vaishnav ED, Ciapmricotti M, Li B, Yang Y, Rodig SJ, Turner M, Pfaff KL, Jané-Valbuena J, Slyper M, et al. Inference of single cell profiles from histology stains with the single-cell omics from histology analysis framework (schaf). BioRxiv, 2023;2023–03.

13. Rios Velazquez E, Parmar C, Liu Y, Coroller TP, Cruz G, Stringfield O, Ye Z, Makrigiorgos M, Fennessy F, Mak RH, et al. Somatic mutations drive distinct imaging phenotypes in lung cancersomatic mutations and radiomic phenotypes. Can Res. 2017;77(14):3922–30.

14. Levsky JM, Singer RH. Fluorescence in situ hybridization: past, present and future. J Cell Sci. 2003;116(14):2833–8.

15. Bouwman BA, Crosetto N, Bienko M. The era of 3d and spatial genomics. Trends Genet 2022

16. Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, Vanderburg CR, Welch J, Chen LM, Chen F, Macosko EZ. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. Science. 2019;363(6434):1463–7.

17. Payne AC, Chiang ZD, Reginato PL, Mangiameli SM, Murray EM, Yao C-C, Markoulaki S, Earl AS, Labade AS, Jaenisch R, et al. In situ genome sequencing resolves dna sequence and structure in intact biological samples. Science. 2021;371(6532):3446.

18. Zhao T, Chiang ZD, Morriss JW, LaFave LM, Murray EM, Del Priore I, Meli K, Lareau CA, Nadaf NM, Li J, et al. Spatial genomics enables multi-modal study of clonal heterogeneity in tissues. Nature. 2022;601(7891):85–91.

19. Thul PJ, Åkesson L, Wiking M, Mahdessian D, Geladaki A, Ait Blal H, Alm T, Asplund A, Björk L, Breckels LM, et al. A subcellular map of the human proteome. Science. 2017;356(6340):3321.

20. Van Holde KE. Chromatin. Berlin: Springer Science & Business Media; 2012.

21. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015;349(6255):1483–9.

22. Goldberg AD, Allis CD, Bernstein E. Epigenetics: a landscape takes shape. Cell. 2007;128(4):635–8.

23. Jones PA, Takai D. The role of dna methylation in mammalian epigenetics. Science. 2001;293(5532):1068–70.

24. Das PM, Singal R. Dna methylation and cancer. J Clin Oncol. 2004;22(22):4632–42.

25. Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES, Jaenisch R. Reduced representation bisulfite sequencing for comparative high-resolution dna methylation analysis. Nucleic Acids Res. 2005;33(18):5868–77.

26. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. Atac-seq: a method for assaying chromatin accessibility genome-wide. Curr Protoc Mol Biol. 2015;109(1):21–9.

27. Belmont AS. Nuclear compartments: an incomplete primer to nuclear compartments, bodies, and genome organization relative to nuclear architecture. Cold Spring Harb Perspect Biol. 2022;14(7): 041268.

28. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289–93.

29. Fortin J-P, Hansen KD. Reconstructing a/b compartments as revealed by hi-c using long-range correlations in epigenetic data. Genome Biol. 2015;16(1):1–23.

30. Jones MJ, Goodman SJ, Kobor MS. Dna methylation and healthy human aging. Aging Cell. 2015;14(6):924–32.

31. Zampieri M, Ciccarone F, Calabrese R, Franceschi C, Bürkle A, Caiafa P. Reconfiguration of dna methylation in aging. Mech Ageing Dev. 2015;151:60–70.

32. Arceo XG, Koslover EF, Zid BM, Brown A. Translation kinetics and diffusive timescales regulate mitochondrial localization of mrnas in yeast and mammalian cells. Biophys J. 2023;122(3):300.

33. Holt CE, Bullock SL. Subcellular mrna localization in animal cells and why it matters. Science. 2009;326(5957):1212–6.

34. Lécuyer E, Yoshida H, Parthasarathy N, Alm C, Babak T, Cerovina T, Hughes TR, Tomancak P, Krause HM. Global analysis of mrna localization reveals a prominent role in organizing cellular architecture and function. Cell. 2007;131(1):174–87.

35. Little SC, Tkačik G, Kneeland TB, Wieschaus EF, Gregor T. The formation of the bicoid morphogen gradient requires protein movement from anteriorly localized mrna. PLoS Biol. 2011;9(3):1000596.

36. Laurent GS, Wahlestedt C, Kapranov P. The landscape of long noncoding rna classification. Trends Genet. 2015;31(5):239–51.

37. McKellar DW, Mantri M, Hinchman MM, Parker JS, Sethupathy P, Cosgrove BD, De Vlaminck I. Spatial mapping of the total transcriptome by in situ polyadenylation. Nat Biotechnol. 2023;41(4):513–20.

38. Mateescu B, Jones JC, Alexander RP, Alsop E, An JY, Asghari M, Boomgarden A, Bouchareychas L, Cayota A, Chang H-C, et al. Phase 2 of extracellular rna communication consortium charts next-generation approaches for extracellular rna research. Iscience. 2022;25(8)

39. Li J, Zhang Y, Yang C, Rong R. Discrepant mrna and protein expression in immune cells. Curr Genom. 2020;21(8):560–3.

40. Wang D. Discrepancy between mrna and protein abundance: insight from information retrieval process in computers. Comput Biol Chem. 2008;32(6):462–8.

41. Gry M, Rimini R, Strömberg S, Asplund A, Pontén F, Uhlén M, Nilsson P. Correlations between rna and protein expression profiles in 23 human cell lines. BMC Genom. 2009;10(1):1–14.

42. Liu Y, Beyer A, Aebersold R. On the dependency of cellular protein levels on mrna abundance. Cell. 2016;165(3):535–50.

43. Krishna RG, Wold F. Post-translational modifications of proteins. Methods in protein sequence analysis, 1993;167–172.

44.   Gingras A-C, Abe KT, Raught B. Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles. Curr Opin Chem Biol. 2019;48:44–54.

45.   Dunham WH, Mullin M, Gingras A-C. Affinity-purification coupled to mass spectrometry: Basic principles and strategies. Proteomics. 2012;12(10):1576–90.

46.   Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci. 2005;9(10):474–80.

47.   Shevach EM. Regulatory t cells in autoimmmunity. Annu Rev Immunol. 2000;18(1):423–49.

48.   Woodcock EA, Matkovich SJ. Cardiomyocytes structure, function and associated pathologies. Int J Biochem Cell Biol. 2005;37(9):1746–51.

49.   Domínguez Conde C, Xu C, Jarvis L, Rainbow D, Wells S, Gomes T, Howlett S, Suchanek O, Polanski K, King H, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science. 2022;376(6594):5197.

50.   Verhoeven BM, Mei S, Olsen TK, Gustafsson K, Valind A, Lindström A, Gisselsson D, Fard SS, Hagerling C, Kharchenko PV, et al. The immune cell atlas of human neuroblastoma. Cell Rep Med **3**(6)2022;

51.   Raj A, Van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008;135(2):216–26.

52.   Snijder B, Pelkmans L. Origins of regulated cell-to-cell variability. Nat Rev Mol Cell Biol. 2011;12(2):119–25.

53.   Almet AA, Cang Z, Jin S, Nie Q. The landscape of cell-cell communication through single-cell transcriptomics. Curr Opin Syst Biol. 2021;26:12–23.

54.   Nave K-A, Werner HB. Myelination of the nervous system: mechanisms and functions. Annu Rev Cell Dev Biol. 2014;30:503–33.

55.   Baker NE. Emerging mechanisms of cell competition. Nat Rev Genet. 2020;21(11):683–97.

56.   Liu S, Iorgulescu JB, Li S, Borji M, Barrera-Lopez IA, Shanmugam V, Lyu H, Morriss JW, Garcia ZN, Murray E, et al. Spatial maps of t cell receptors and transcriptomes reveal distinct immune niches and interactions in the adaptive immune response. Immunity. 2022;55(10):1940–52.

57.   Marrahi AE, Lipreri F, Alber D, Hausser J. Four tumor micro-environmental niches explain a continuum of inter-patient variation in the macroscopic cellular composition of breast tumors. bioRxiv, 2022;2022–03.

58.   Medaglia C, Giladi A, Stoler-Barak L, De Giovanni M, Salame TM, Biram A, David E, Li H, Iannacone M, Shulman Z, et al. Spatial reconstruction of immune niches by combining photoactivatable reporters and scrna-seq. Science. 2017;358(6370):1622–6.

59.   Tikhonova AN, Lasry A, Austin R, Aifantis I. Cell-by-cell deconstruction of stem cell niches. Cell Stem Cell. 2020;27(1):19–34.

60.   Mayer AT, Holman DR, Sood A, Tandon U, Bhate SS, Bodapati S, Barlow GL, Chang J, Black S, Crenshaw EC, et al. A tissue atlas of ulcerative colitis revealing evidence of sex-dependent differences in disease-driving inflammatory cell types and resistance to tnf inhibitor therapy. Sci Adv. 2023;9(3):1166.

61.   Pelka K, Hofree M, Chen JH, Sarkizova S, Pirl JD, Jorgji V, Bejnood A, Dionne D, William HG, Xu KH, et al. Spatially organized multicellular immune hubs in human colorectal cancer. Cell. 2021;184(18):4734–52.

62.   Lane SW, Williams DA, Watt FM. Modulating the stem cell niche for tissue regeneration. Nat Biotechnol. 2014;32(8):795–803.

63.   Tepass U, Theres C, Knust E. crumbs encodes an egf-like protein expressed on apical membranes of drosophila epithelial cells and required for organization of epithelia. Cell. 1990;61(5):787–99.

64.   Regev A, Teichmann S, Rozenblatt-Rosen O, Stubbington M, Ardlie K, Amit I, Arlotta P, Bader G, Benoist C, Biton M, et al. The human cell atlas white paper. arXiv preprint arXiv:1810.05192 2018;

65.   Shahan R, Hsu C-W, Nolan TM, Cole BJ, Taylor IW, Greenstreet L, Zhang S, Afanassiev A, Vlot AHC, Schiebinger G, et al. A single-cell arabidopsis root atlas reveals developmental trajectories in wild-type and cell identity mutants. Dev Cell. 2022;57(4):543–60.

66.   Sikkema L, Ramírez-Suástegui C, Strobl DC, Gillett TE, Zappia L, Madissoon E, Markov NS, Zaragosi L-E, Ji Y, Ansari M, et al. An integrated cell atlas of the lung in health and disease. Nat Med 2023;1–15.

67.   Nguyen QH, Pervolarakis N, Blake K, Ma D, Davis RT, James N, Phung AT, Willey E, Kumar R, Jabart E, et al. Profiling human breast epithelial cells using single cell rna sequencing identifies cell diversity. Nat Commun. 2018;9(1):2028.

68.   Consortium TM, coordination Schaum Nicholas 1 Karkanias Jim 2 Neff Norma F. 2 May Andrew P. 2 Quake Stephen R. quake@ stanford. edu 2 3 f Wyss-Coray Tony twc@ stanford. edu 4 5 6 g Darmanis Spyros spyros. darmanis@ czbiohub. org 2 h, O., coordination Batson Joshua 2 Botvinnik Olga 2 Chen Michelle B. 3 Chen Steven 2 Green Foad 2 Jones Robert C. 3 Maynard Ashley 2 Penland Lolita 2 Pisco Angela Oliveira 2 Sit Rene V. 2 Stanley Geoffrey M. 3 Webber James T. 2 Zanini Fabio 3, L., data analysis Batson Joshua 2 Botvinnik Olga 2 Castro Paola 2 Croote Derek 3 Darmanis Spyros 2 DeRisi Joseph L. 2 27 Karkanias Jim 2 Pisco Angela Oliveira 2 Stanley Geoffrey M. 3 Webber James T. 2 Zanini Fabio 3, C.: Single-cell transcriptomics of 20 mouse organs creates a tabula muris. Nature 562(7727), 2018;367–372 .

69.   Hung R-J, Hu Y, Kirchner R, Liu Y, Xu C, Comjean A, Tattikota SG, Li F, Song W, Ho Sui S, et al. A cell atlas of the adult drosophila midgut. Proc Natl Acad Sci. 2020;117(3):1514–23.

70.   Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, Adey A, Waterston RH, Trapnell C, Shendure J. Comprehensive single-cell transcriptional profiling of a multicellular organism. Science. 2017;357(6352):661–7. https://doi.org/10.1126/science.aam8940.

71.   Taylor DM, Aronow BJ, Tan K, Bernt K, Salomonis N, Greene CS, Frolova A, Henrickson SE, Wells A, Pei L, et al. The pediatric cell atlas: defining the growth phase of human development at single-cell resolution. Dev Cell. 2019;49(1):10–29.

72.   Chen D, Wang W, Wu L, Liang L, Wang S, Cheng Y, Zhang T, Chai C, Luo Q, Sun C, et al. Single-cell atlas of peripheral blood mononuclear cells from pregnant women. Clin Transl Med. 2022;12(5):821.

73.   Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martínez-Colón GJ, McKechnie JL, Ivison GT, Ranganath T, Vergara R, Hollis T, et al. A single-cell atlas of the peripheral immune response in patients with severe covid-19. Nat Med. 2020;26(7):1070–6.

74. Schiller HB, Montoro DT, Simon LM, Rawlins EL, Meyer KB, Strunz M, Vieira Braga FA, Timens W, Koppelman GH, Budinger GS, et al. The human lung cell atlas: a high-resolution reference map of the human lung in health and disease. Am J Respir Cell Mol Biol. 2019;61(1):31–41.

75. Zhou Y, Xu J, Hou Y, Bekris L, Leverenz JB, Pieper AA, Cummings J, Cheng F. The alzheimer's cell atlas (taca): A single-cell molecular map for translational therapeutics accelerator in alzheimer's disease. Alzheimer's & Dementia: Transl Res Clin Interve. 2022;8(1):12350.

76. Grubman A, Chew G, Ouyang JF, Sun G, Choo XY, McLean C, Simmons RK, Buckberry S, Vargas-Landin DB, Poppe D, et al. A single-cell atlas of entorhinal cortex from individuals with alzheimer's disease reveals cell-type-specific gene expression regulation. Nat Neurosci. 2019;22(12):2087–97.

77. Winkler EA, Kim CN, Ross JM, Garcia JH, Gil E, Oh I, Chen LQ, Wu D, Catapano JS, Raygor K, et al. A single-cell atlas of the normal and malformed human brain vasculature. Science. 2022;375(6584):7377.

78. Starling E. Discussion on the therapeutic value of hormones. Proceedings of the Royal Society of Medicine 7(Ther_Pharmacol). 1914;29–31.

79. Nair A, Chauhan P, Saha B, Kubatzky KF. Conceptual evolution of cell signaling. Int J Mol Sci. 2019;20(13):3292.

80. Levi-Montalcini R, Hamburger V. Selective growth stimulating effects of mouse sarcoma on the sensory and sympathetic nervous system of the chick embryo. J Exp Zool. 1951;116(2):321–61.

81. Hokin MR, Hokin LE. Enzyme secretion and the incorporation of p32 into phospholipides of pancreas slices. J Biol Chem. 1953;203(2):967–77.

82. Krebs EG, Fischer EH. The phosphorylase b to a converting enzyme of rabbit skeletal muscle. Biochem Biophys Acta. 1956;20:150–7.

83. Aoki K, Kumagai Y, Sakurai A, Komatsu N, Fujita Y, Shionyu C, Matsuda M. Stochastic erk activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. Mol Cell. 2013;52(4):529–40.

84. McFann SE, Shvartsman SY, Toettcher JE. Putting in the erk: Growth factor signaling and mesoderm morphogenesis. Curr Top Dev Biol. 2022;149:263–310.

85. Marmion RA, Simpkins AG, Barrett LA, Denberg DW, Zusman S, Schottenfeld-Roames J, Schüpbach T, Shvartsman SY. Stochastic phenotypes in ras-dependent developmental diseases. Curr Biol. 2023;33(5):807–16.

86. Green JB, Sharpe J. Positional information and reaction-diffusion: two big ideas in developmental biology combine. Development. 2015;142(7):1203–11.

87. Merle M, Messio L, Mozziconacci J. Turing-like patterns in an asymmetric dynamic ising model. Phys Rev E. 2019;100(4): 042111.

88. Breakspear M, Heitmann S, Daffertshofer A. Generative models of cortical oscillations: neurobiological implications of the kuramoto model. Front Hum Neurosci. 2010;4:190.

89. Petkova MD, Tkačik G, Bialek W, Wieschaus EF, Gregor T. Optimal decoding of cellular identities in a genetic network. Cell. 2019;176(4):844–55.

90. Tkačik G, Callan CG Jr, Bialek W. Information flow and optimization in transcriptional regulation. Proc Natl Acad Sci. 2008;105(34):12265–70.

91. Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, Yun J, Cronin C, Karp C, Yuan G-C, et al. Transcriptome-scale super-resolved imaging in tissues by rna seqfish+. Nature. 2019;568(7751):235–9.

92. Yuan Y, Bar-Joseph Z. Gcng: graph convolutional networks for inferring gene interaction from spatial transcriptomics data. Genome Biol. 2020;21(1):1–16.

93. Cang Z, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nat Commun. 2020;11(1):2084.

94. 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. 2019;29(1):202–11.

95. Fischer DS, Schaar AC, Theis FJ. Modeling intercellular communication in tissues using spatial graphs of cells. Nat Biotechnol 2022;1–5.

96. Vento-Tormo R, Efremova M, Botting RA, Turco MY, Vento-Tormo M, Meyer KB, Park J-E, Stephenson E, Polański K, Goncalves A, et al. Single-cell reconstruction of the early maternal-fetal interface in humans. Nature. 2018;563(7731):347–53.

97. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. Cellphonedb: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020;15(4):1484–506.

98. Garcia-Alonso L, Handfield L-F, Roberts K, Nikolakopoulou K, Fernando RC, Gardner L, Woodhams B, Arutyunyan A, Polanski K, Hoo R, et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. Nat Genet. 2021;53(12):1698–711.

99. Verma A, Jena SG, Isakov DR, Aoki K, Toettcher JE, Engelhardt BE. A self-exciting point process to study multicellular spatial signaling patterns. Proc Natl Acad Sci. 2021;118(32):2026123118.

100. Borjini N, Paouri E, Tognatta R, Akassoglou K, Davalos D. Imaging the dynamic interactions between immune cells and the neurovascular interface in the spinal cord. Exp Neurol. 2019;322: 113046.

101. Nava MM, Miroshnikova YA, Biggs LC, Whitefield DB, Metge F, Boucas J, Vihinen H, Jokitalo E, Li X, Arcos JMG, et al. Heterochromatin-driven nuclear softening protects the genome against mechanical stress-induced damage. Cell. 2020;181(4):800–17.

102. GTEx Consortium: Genetic effects on gene expression across human tissues. Nature 550(7675):204–213.

103. Haghighi M, Caicedo JC, Cimini BA, Carpenter AE, Singh S. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. Nat Methods. 2022;1–8.

104. Stirling DR, Swain-Bowden MJ, Lucas AM, Carpenter AE, Cimini BA, Goodman A. Cell Profiler 4: improvements in speed, utility and usability. BMC Bioinf. 2021;22:1–11.

105. Ramezani M, Bauman J, Singh A, Weisbart E, Yong J, Lozada ME, Way GP, Kavari SL, Diaz C, Haghighi M, et al. A genome-wide atlas of human cell morphology. bioRxiv, 2023;2023–08.

106. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. Commun ACM. 2020;63(11):139–44.

107. Lee H, Welch JD. Morphnet predicts cell morphology from single-cell gene expression. bioRxiv, 2022;2022–10.

108. Sivanandan S, Leitmann B, Lubeck E, Sultan MM, Stanitsas P, Ranu N, Ewer A, Mancuso JE, Phillips ZF, Kim A, et al. A pooled cell painting crispr screening platform enables de novo inference of gene function by self-supervised deep learning. bioRxiv, 2023;2023–08.

109. McSwiggen DT, Liu H, Tan R, Agramunt Puig S, Akella LB, Berman R, Bretan M, Chen H, Darzacq X, Ford K, et al. High-throughput single molecule tracking identifies drug interactions and cellular mechanisms. bioRxiv, 2023;2023–01.

110. Sun J, Ramos A, Chapman B, Johnnidis JB, Le L, Ho Y-J, Klein A, Hofmann O, Camargo FD. Clonal dynamics of native haematopoiesis. Nature. 2014;514(7522):322–7.

111. Fabre MA, de Almeida JG, Fiorillo E, Mitchell E, Damaskou A, Rak J, Orrù V, Marongiu M, Chapman MS, Vijayabaskar M, et al. The longitudinal dynamics and natural history of clonal haematopoiesis. Nature. 2022;606(7913):335–42.

112. Klein AM, Doupé DP, Jones PH, Simons BD. Kinetics of cell division in epidermal maintenance. Phys Rev E. 2007;76(2): 021910.

113. Klein AM, Doupé DP, Jones PH, Simons BD. Mechanism of murine epidermal maintenance: Cell division and the voter model. Phys Rev E. 2008;77(3): 031907.

114. Parigini C, Greulich P. Universality of clonal dynamics poses fundamental limits to identify stem cell self-renewal strategies. Elife. 2020;9:56532.

115. Gudmundsdottir H, Wells AD, Turka LA. Dynamics and requirements of t cell clonal expansion in vivo at the single-cell level: effector function is linked to proliferative capacity. J Immunol. 1999;162(9):5212–23.

116. Gerrits A, Dykstra B, Kalmykowa OJ, Klauke K, Verovskaya E, Broekhuis MJ, de Haan G, Bystrykh LV. Cellular barcoding tool for clonal analysis in the hematopoietic system. Blood, J Am Soc Hematol. 2010;115(13):2610–8.

117. Nguyen LV, Pellacani D, Lefort S, Kannan N, Osako T, Makarem M, Cox CL, Kennedy W, Beer P, Carles A, et al. Barcoding reveals complex clonal dynamics of de novo transformed human mammary cells. Nature. 2015;528(7581):267–71.

118. Spanjaard B, Hu B, Mitic N, Olivares-Chauvet P, Janjuha S, Ninov N, Junker JP. Simultaneous lineage tracing and cell-type identification using crispr-cas9-induced genetic scars. Nat Biotechnol. 2018;36(5):469–73.

119. Ishiguro S, Ishida K, Sakata RC, Mori H, Takana M, King S, Bashth O, Ichiraku M, Masuyama N, Takimoto R, et al.: A multi-kingdom genetic barcoding system for precise target clone isolation. BioRxiv, 2023;2023–01.

120. Koster MI, Roop DR. Mechanisms regulating epithelial stratification. Annu Rev Cell Dev Biol. 2007;23:93–113.

121. Colom B, Alcolea MP, Piedrafita G, Hall MW, Wabik A, Dentro SC, Fowler JC, Herms A, King C, Ong SH, et al. Spatial competition shapes the dynamic mutational landscape of normal esophageal epithelium. Nat Genet. 2020;52(6):604–14.

122. Colom B, Herms A, Hall M, Dentro S, King C, Sood R, Alcolea M, Piedrafita G, Fernandez-Antoran D, Ong S, et al. Mutant clones in normal epithelium outcompete and eliminate emerging tumours. Nature. 2021;598(7881):510–4.

123. Blanpain C, Fuchs E. Plasticity of epithelial stem cells in tissue regeneration. Science. 2014;344(6189):1242281.

124. Ma Y, Zhou X. Spatially informed cell-type deconvolution for spatial transcriptomics. Nat Biotechnol. 2022;40(9):1349–59.

125. Ru B, Huang J, Zhang Y, Aldape K, Jiang P. Estimation of cell lineages in tumors from spatial transcriptomics data. Nat Commun. 2023;14(1):568.

126. Elyanow R, Zeira R, Land M, Raphael BJ. Starch: copy number and clone inference from spatial transcriptomics data. Phys Biol. 2021;18(3): 035001.

127. Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, Beqiri M, Sproesser K, Brafford PA, Xiao M, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature. 2017;546(7658):431–5.

128. Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, Tokheim CJ, Brown A, DeBlasio RM, Niyazov J, et al. Minimal functional driver gene heterogeneity among untreated metastases. Science. 2018;361(6406):1033–7.

129. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297(5584):1183–6.

130. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. Proc Natl Acad Sci. 2002;99(20):12795–800.

131. Townes FW, Engelhardt BE. Nonnegative spatial factorization applied to spatial genomics. Nat Methods. 2023;20(2):229–38.

132. Velten B, Braunger JM, Argelaguet R, Arnol D, Wirbel J, Bredikhin D, Zeller G, Stegle O. Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO. Nat Methods. 2022;19(2):179–86.

133. Yeung T, Georges PC, Flanagan LA, Marg B, Ortiz M, Funaki M, Zahir N, Ming W, Weaver V, Janmey PA. Effects of substrate stiffness on cell morphology, cytoskeletal structure, and adhesion. Cell Motil Cytoskelet. 2005;60(1):24–34.

134. Meng K, Wang J, Crawford L, Eloyan A. Randomness and statistical inference of shapes via the smooth Euler characteristic transform. arXiv preprint arXiv:2204.12699 2022;

135. Lechler T, Fuchs E. Asymmetric cell divisions promote stratification and differentiation of mammalian skin. Nature. 2005;437(7056):275–80.

136. 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. Proc Natl Acad Sci. 2016;113(39):11046–51.

137. Groiss S, Pabst D, Faber C, Meier A, Bogdoll A, Unger C, Nilges B, Strauss S, Föderl-Höbenreich E, Hardt M, et al. Highly resolved spatial transcriptomics for detection of rare events in cells. bioRxiv, 2021;2021–10.

138. Torre E, Dueck H, Shaffer S, Gospocic J, Gupte R, Bonasio R, Kim J, Murray J, Raj A. Rare cell detection by single-cell rna sequencing as guided by single-molecule rna fish. Cell Syst. 2018;6(2):171–9.

139. Cable DM, Murray E, Shanmugam V, Zhang S, Zou LS, Diao M, Chen H, Macosko EZ, Irizarry RA, Chen F. Cell type-specific inference of differential expression in spatial transcriptomics. Nat Methods. 2022;19(9):1076–87.

140. Schürch CM, Bhate SS, Barlow GL, Phillips DJ, Noti L, Zlobec I, Chu P, Black S, Demeter J, McIlwain DR, et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. Cell. 2020;182(5):1341–59.

141. Xu C, Jin X, Wei S, Wang P, Luo M, Xu Z, Yang W, Cai Y, Xiao L, Lin X, et al. Deepst: identifying spatial domains in spatial transcriptomics by deep learning. Nucleic Acids Res. 2022;50(22):131–131.

142. Shi X, Zhu J, Long Y, Liang C. Identifying spatial domains of spatially resolved transcriptomics via multi-view graph convolutional networks. Brief Bioi 2023;278.

143. Shang L, Zhou X. Spatially aware dimension reduction for spatial transcriptomics. Nat Commun. 2022;13(1):7203.

144. Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, Mantineo H, Brydon EM, Zeng Z, Liu XS, et al. Transfer learning enables predictions in network biology. Nature. 2023;1–9.

145. Verma A, Engelhardt BE. A robust nonlinear low-dimensional manifold for single cell rna-seq data. BMC Bioinf. 2020;21(1):1–15.

146. Lopez R, Regier J, Cole MB, Jordan MI, Yosef N. Deep generative modeling for single-cell transcriptomics. Nat Methods. 2018;15(12):1053–8.

Publisher's Note