nature neuroscience

Review article

https://doi.org/10.1038/s41593-024-01806-0

Opportunities and challenges of single-cell and spatially resolved genomics methods for neuroscience discovery

Received: 22 November 2023

Accepted: 23 September 2024

Published online: 3 December 2024

Check for updates

Boyan Bonev[1,2], Castelo-Branco Gonçalo[3], Fei Chen[4], Simone Codeluppi[5], M. Ryan Corces[6,7,8], Jean Fan[9], Myriam Heiman[10,11], Kenneth Harris[12], Fumitaka Inoue[13], Manolis Kellis[4,14], Ariel Levine[15], Mo Lotfollahi[16,17], Chongyuan Luo[18], Kristen R. Maynard[19,20,21], Mor Nitzan[22,23,24], Vijay Ramani[7,25,26], Rahul Satija[27,28], Lucas Schirmer[29], Yin Shen[8,30,31], Na Sun[4,14], Gilad S. Green[32], Fabian Theis[17,18], Xiao Wang[4,33,34], Joshua D. Welch[35], Ozgun Gokce[36,37] ✉, Genevieve Konopka[38,39] ✉, Shane Liddelow[40,41,42,43] ✉, Evan Macosko[4,34,44,45] ✉, Omer Bayraktar[46] ✉, Naomi Habib[32] ✉ & Tomasz J. Nowakowski[31,47,48,49,50] ✉

Over the past decade, single-cell genomics technologies have allowed scalable profiling of cell-type-specific features, which has substantially increased our ability to study cellular diversity and transcriptional programs in heterogeneous tissues. Yet our understanding of mechanisms of gene regulation or the rules that govern interactions between cell types is still limited. The advent of new computational pipelines and technologies, such as single-cell epigenomics and spatially resolved transcriptomics, has created opportunities to explore two new axes of biological variation: cell-intrinsic regulation of cell states and expression programs and interactions between cells. Here, we summarize the most promising and robust technologies in these areas, discuss their strengths and limitations and discuss key computational approaches for analysis of these complex datasets. We highlight how data sharing and integration, documentation, visualization and benchmarking of results contribute to transparency, reproducibility, collaboration and democratization in neuroscience, and discuss needs and opportunities for future technology development and analysis.

Cells in our bodies contain roughly the same genomic information encoded within the DNA, but develop remarkably different properties as a consequence of intrinsic regulation of gene expression and intercellular communication. Nowhere is this clearer than in the mammalian brain, where hundreds of molecularly distinct cell subpopulations have recently been mapped using a combination of single-cell technologies and shown to be organized into neighborhoods and circuits that can be visualized using spatially resolved technologies[1–9]. Intrinsic regulation of gene expression and cell–cell interactions (CCIs) represent two orthogonal, and yet interrelated, axes of biological variation in complex tissues that frequently become altered in disease states. Emerging technologies for mapping these modalities create exciting opportunities for uncovering disease-related changes with fewer a priori assumptions than has been possible before. In turn, unbiased profiling of diseased tissues has the potential to uncover new disease-relevant changes that could be targeted therapeutically.

A full list of affiliations appears at the end of the paper. ✉e-mail: ozguengoekce@gmail.com; genevieve.konopka@utsouthwestern.edu; shane.liddelow@nyulangone.org; emacosko@broadinstitute.org; ob5@sanger.ac.uk; naomi.habib@mail.huji.ac.il; tomasz.nowakowski@ucsf.edu

https://doi.org/10.1038/s41593-024-01806-0

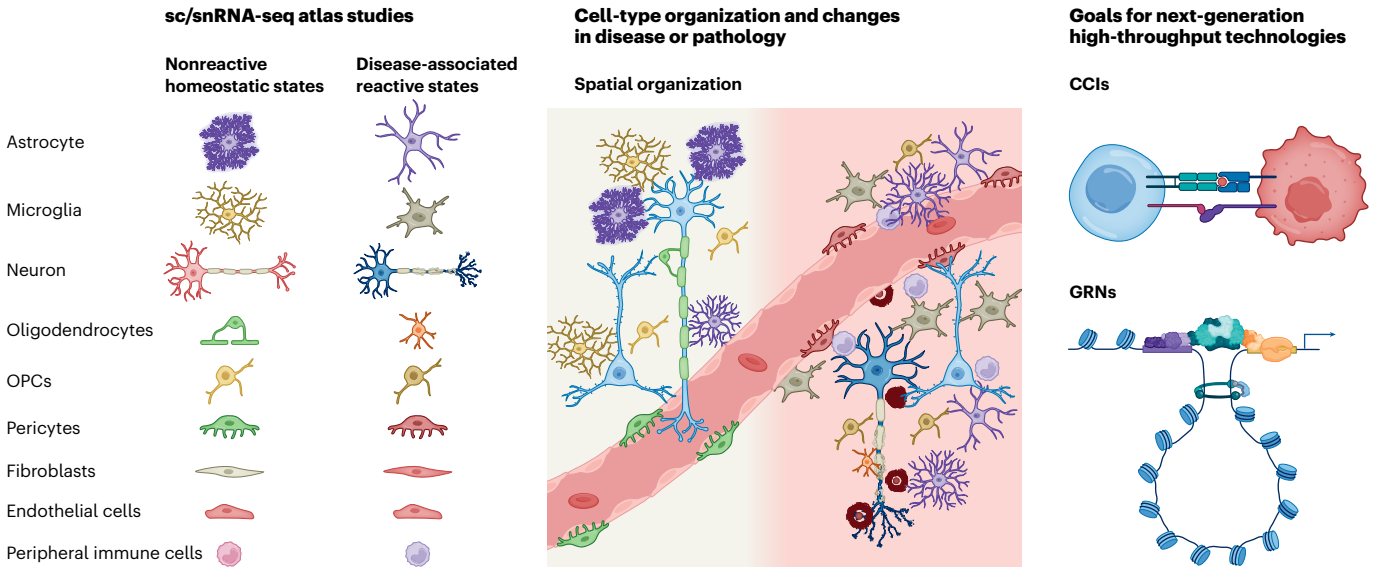


Fig. 1 | Biological considerations: cellular architecture and research questions. Emerging single-cell atlas studies have created reference resources for defining cell types in normal and pathological brain tissue. Cell types and states defined by these studies exist in complex and dynamic communities in vivo, and more disease-associated states and types may emerge in the future. The advent of ST technologies helps to define cellular neighborhoods and identify candidate networks of molecular interactions, while advanced single-cell genomic technologies can provide insights into dynamic intracellular pathways underlying cellular transitions. OPCs, oligodendrocyte precursor cells.

Here, we strive to provide an overview of the main technologies and approaches currently present in single-cell epigenomics and spatially resolved transcriptomics, as well as discussing various strategies for data analysis and considerations in experimental design. In particular, we highlight the proliferation of single-cell epigenomic data collection that has provided exciting opportunities to reveal gene regulatory networks (GRNs), while highlighting the paucity of methods for functional validation of these predictions. By contrast, spatially resolved transcriptomics approaches vary widely depending on the specific tissue preservation method, size and resolution needed. When coupled with single-cell transcriptomics and rigorous data analysis, such as deconvolution, trajectory analysis and CCI prediction, such experiments can provide invaluable insights into tissue biology (Fig. 1).

We recommend that technology choices and computational schemes should be motivated by the biological questions being investigated, while balancing discovery, analysis and validation wherever possible to maximize biological insights. Best practices in computational analysis should guide experimental design and be considered before data generation, taking into consideration the required number of samples, coverage of cells per sample and design of experimental batches, to facilitate accurate analysis. In turn, the design of the computational scheme for the data analysis should be tailored to the specific features of the dataset as well as to the biological questions, guiding choices such as de novo versus reference-based annotations and discrete versus continuous analysis of cell states.

Technical considerations in study design

High-throughput genomic technologies have created unprecedented opportunities for data-driven discovery of biological processes underlying normal tissue structure and function and changes in disease. Considering the cost of many such studies, responsible experimental design is often required to maximize biological insights, and should start with considerations of best practices in data analysis (Fig. 2). We recommend considering the following components for single-cell and spatial genomic studies.

Sample size

Evaluation of the necessary sample size for robust analysis is essential, especially when testing changes in cell abundance, differential genes or trait associations with experimental or clinical conditions. Computational methods should be tailored to the sample size and to the corresponding statistical power of the dataset. There are tools and resources that can be used to estimate the necessary number of biological replicates and technical replicates ideally required for single-cell or single-nucleus RNA-sequencing (scRNA-seq or snRNA-seq) studies[10–14]. As fewer spatial transcriptomics (ST) studies have been conducted so far, recommendations of sample size estimation are driven largely by theoretical and statistical considerations[15,16]. Adhering to the power estimations discussed can greatly increase the confidence in biological findings derived from scRNA-seq studies, and we predict that the rapid increase in the number of ST and epigenomic datasets will lead to better understanding of technical variation in the data and inform new methods for quantifying effect sizes that may be specific to the assay or data generation platform. Beyond the number of samples, the number and design of batches should be carefully considered as well. Specifically, ensuring balanced batches across different experimental groups can greatly facilitate batch correction to mitigate technical artifacts. It is important to note that the required sample size is contingent on the intrinsic variation between samples, which may vary depending on the genetic diversity of the study population and could furthermore be influenced by the technical idiosyncrasies of individual platforms or dataset quality.

Underpowered studies may still provide biologically meaningful insights, but they require specific considerations. We advise incorporating strategies for orthogonal validation using the methods discussed in ref. 17. Alternatively, leveraging large cohorts of bulk datasets can augment the sample size, enabling the correlation of gene and pathway signatures to conditions and traits. For example, different strategies have been suggested for using scRNA-seq and snRNA-seq data to deconvolve signatures of cell-type abundance hidden within bulk tissue measurements using algorithmic approaches based on deconvolution[18,19] and increasingly deep learning[20–22]. This strategy reduces the time and cost of experiments, and may be essential when obtaining sufficient biological replicates of tissue specimens is challenging. In essence, acknowledging and accounting for sample size and diversity among samples is crucial for ensuring the reliability and validity of research conclusions.

https://doi.org/10.1038/s41593-024-01806-0

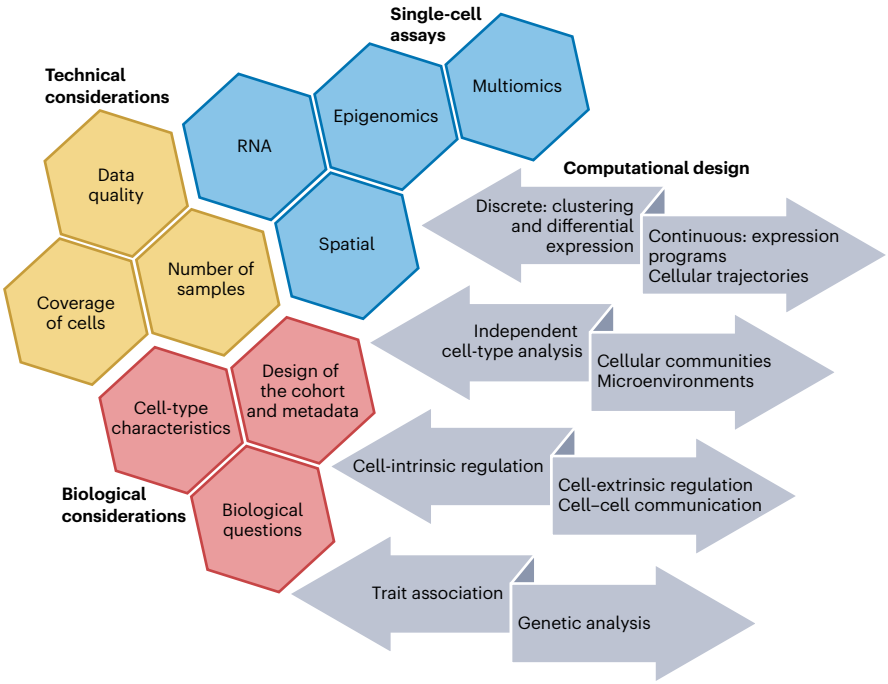


Fig. 2 | Outline of key considerations involved in designing high-throughput single-cell and ST studies. Experimental design considerations are essential for successful studies utilizing high-throughput single cell and spatial genomics. The cartoon outlines key features of successful study design, highlighting synergies between biological questions, technical design, and analytical considerations.

Cellular coverage

Each profiling experiment involves a decision step to profile a subset of cells present in the tissue, and, owing to financial and tissue availability constraints, every study will balance the numbers of biological or technical replicates involved with the number of cells that will be profiled. Understandably, heterogeneous tissues such as the brain pose an additional challenge because cell types are not present in equal proportions, and thus the effective cell numbers involved in a study will vary from cell type to cell type.

Thus, consideration for cellular coverage and sequencing depth should guide the computational strategy applied for cell annotations as well as for association analysis, depending on the biological question[23,24]. Data downsampling can offer a data-driven approach to determining whether a given observation or conclusion is robust, and has been effectively used to analyze the saturation of cell cluster discovery in mouse brain scRNA-seq data[7]. Unfortunately, similar considerations have not been developed yet for ST or epigenomic studies. Analysis of ST data in particular will require deep assessment as datasets become more common, and we predict that lessons from stereology[25] might be helpful in interpreting the results of ST-based experiments.

Ideally, when experimental design limitations prohibit obtaining adequate cellular coverage to comprehensively profile cells in a given tissue, strategies that enrich for a desired cell population can be used (see ref. 26 for methods of mining rare cells). Importantly, the computational study design should be guided by the coverage of the dataset, as different clustering algorithms have different sensitivities for detecting rare cell types[27]. Moreover, rare cells can be mis-assigned to transcriptomically similar populations if not enough cells are sampled. Mapping cells against a well-powered reference atlas dataset can help to overcome this limitation[28], and such atlases are increasing in availability for the brain across various species[1–9], ages and conditions, offering an important resource that can be leveraged to annotate cell types in smaller studies (see ref. 26).

Data quality and sequencing depth

The sequencing coverage (3′-end, 5′-end, whole coding region), data quality and sequencing depth (number of unique reads or unique molecular identifiers (UMIs) and number of genes detected per cell) represent important metrics of underlying data quality and should ideally be compared to published studies from the same tissue or cell type, and reported across technical and biological replicates in a study (see Box 1 for an overview of quality-control (QC) data analysis). For scRNA-seq and snRNA-seq assays, high-quality data are important to ensure that differences in expression programs within specific cell types can be robustly detected. Insufficient coverage might mean that biological insights are overlooked, a factor that must be considered during the differential expression analysis of genes and pathways. While the specific number of sequencing reads per cell will depend on both the technology and the nature of the sample, for the most common experimental platform from 10x Genomics, gene expression libraries sequenced at or above 25,000 reads per cell or nucleus would generally be considered as reasonable to identify individual subtypes, but these may be too shallow for some cell-type-specific responses to disease and infection (for example, in microglia and astrocytes) where 50,000 reads per cell or nucleus are more often required to detect subtle gene expression changes. Published reference atlases can provide approximate numbers of genes detected across brain cell types, while data derived from whole dissociated cells typically yield higher numbers of genes detected per cell compared to nuclei.

Epigenetic assays cover a larger sample space (whole genome versus transcriptome); consequently, the sequencing depth per cell should be appropriately higher, with a recommended minimal sequencing depth of 50,000 reads per cell for single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq) libraries. For reliable discovery of gene regulatory elements from snATAC-seq data, the number of cells depends on data quality and the analytical context, yet we advise at least 200 cells or nuclei per cell population with biological replicates to ensure reproducibility. These metrics are based on experience and lack systematic and quantitative

https://doi.org/10.1038/s41593-024-01806-0

BOX 1

Guidelines for data processing and quality controls in scRNA-seq and snRNA-seq studies

QC is an essential step in the analysis of scRNA-seq and snRNA-seq data. Nevertheless, although too permissive thresholds might lead to technical artifacts such as false positives in differential gene expression analysis and misclassification of cell types, too stringent parameters can lead to false negatives and failure to detect relevant biology. Thus, it is crucial to implement robust QC workflows before further downstream analysis. For cell QC, begin by excluding cell barcodes that are likely to represent dead cell debris or free-floating RNA (ambient RNA), as they don't correspond to intact individual cells. A straightforward approach for assessing cell quality is calculating metrics such as the number of transcripts (UMIs), or detected genes. In practice, the number of UMIs and genes detected varies notably between brain cell types such as microglia and neurons, and between datasets depending on quality and sequencing depth. Hence, assigning a cell-type-specific and dataset-specific threshold for filtering low-quality cells is important. This can be done by initial classification of cells to broad cell classes using classifiers trained on existing datasets and assessing the appropriate thresholds on the basis of the distribution of the number of detected genes within each cell class.

Another frequently used cell quality measure is the proportion of mitochondrial RNA, with high proportions possibly indicating damaged cells. Nonetheless, cells with high mitochondrial content should not be automatically excluded, as they might indicate metabolic changes such as increased mitochondrial activity or be informative in the context of neurodegenerative diseases. We recommend that mitochondrial content should be taken into account, but not used as the only exclusion criterion for low-quality cells.

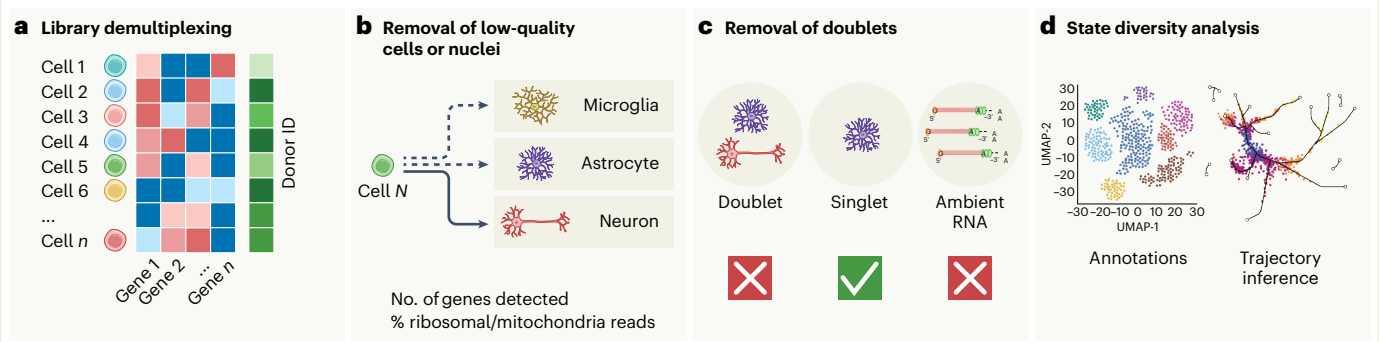
Conversely, cells with unexpectedly high counts may indicate doublets (or multiplets), where cell barcodes correspond to multiple cells. It is essential to remove doublets, as they can constitute a substantial portion of cell barcodes in high-throughput scRNA-seq or snRNA-seq methods. However, we caution against frequently used filtration methods based solely on the number of detected transcripts, especially in complex tissues such as the brain, as they are not accurate enough; specialized algorithms that model doublet cells (for example, Scrublet[176], DoubletFinder[177] and scds[178]) are much more robust. Also, transitory cell states, which might present cell signatures of different cell populations and are frequent in development and disease[26], might also be mistaken for doublets.

Ambient RNA transcripts, which are free-floating and barcoded with the cell or nucleus, can impact the cellular expression profile and potentially bias cell annotations and functional interpretations. Given the varying extent of ambient RNA, dependent on tissue quality and cell or nuclei isolation protocols, it is essential to evaluate each dataset individually. The ambient RNA can be corrected as necessary using methods such as CellBender[179], SoupX[180] and DecounX[181]. We also recommend extracting signatures of ambient RNA directly from the data by compiling abundant transcripts in empty droplets, to assess the contamination within each dataset and to ensure that the corrected expression profiles remain undistorted.

Furthermore, QC can be applied at the gene level, although this is recommended only if the computational resources are constrained or the noise in the dataset is high. One can filter out genes with limited expression (that is, detected in a small number of cells) by choosing an appropriate filtration threshold considering the number of cells expected from the smallest cell population. In addition, methods to identify informative genes and focus the analysis on this gene subset can be applied (such as variance-stabilizing transformation).

Finally, data normalization and corrections for technical features such as batch and sample quality are crucial to exclude technical artifacts from the downstream analysis, but overcorrection and normalization could result in loss of the biological signal. Therefore, careful examination of the data after correction is necessary, and comparison to reference atlases can be used as a benchmark.

It is important to note the evolving guidelines, especially for newer data types such as single-cell epigenomics and ST methods, which have unique preprocessing and QC challenges that vary between platforms. As technologies mature, these challenges are expected to diminish. For example, in multiplexed in situ methods (for example, Stomics, MERFISH and Xenium), the main challenge is image analysis, specifically cell segmentations and data-to-noise ratio. For ISS-based methods, the main challenge lies in data resolution and coverage, which require more advanced computational techniques to deconvolute the expression signal to cell types and states, or to find expression patterns based on communities of cells or based on the spatial gradient to the center of pathology.



Box Fig. 1 | QC for scRNA-seq and snRNA-seq data involves evaluating both cells and genes.

https://doi.org/10.1038/s41593-024-01806-0

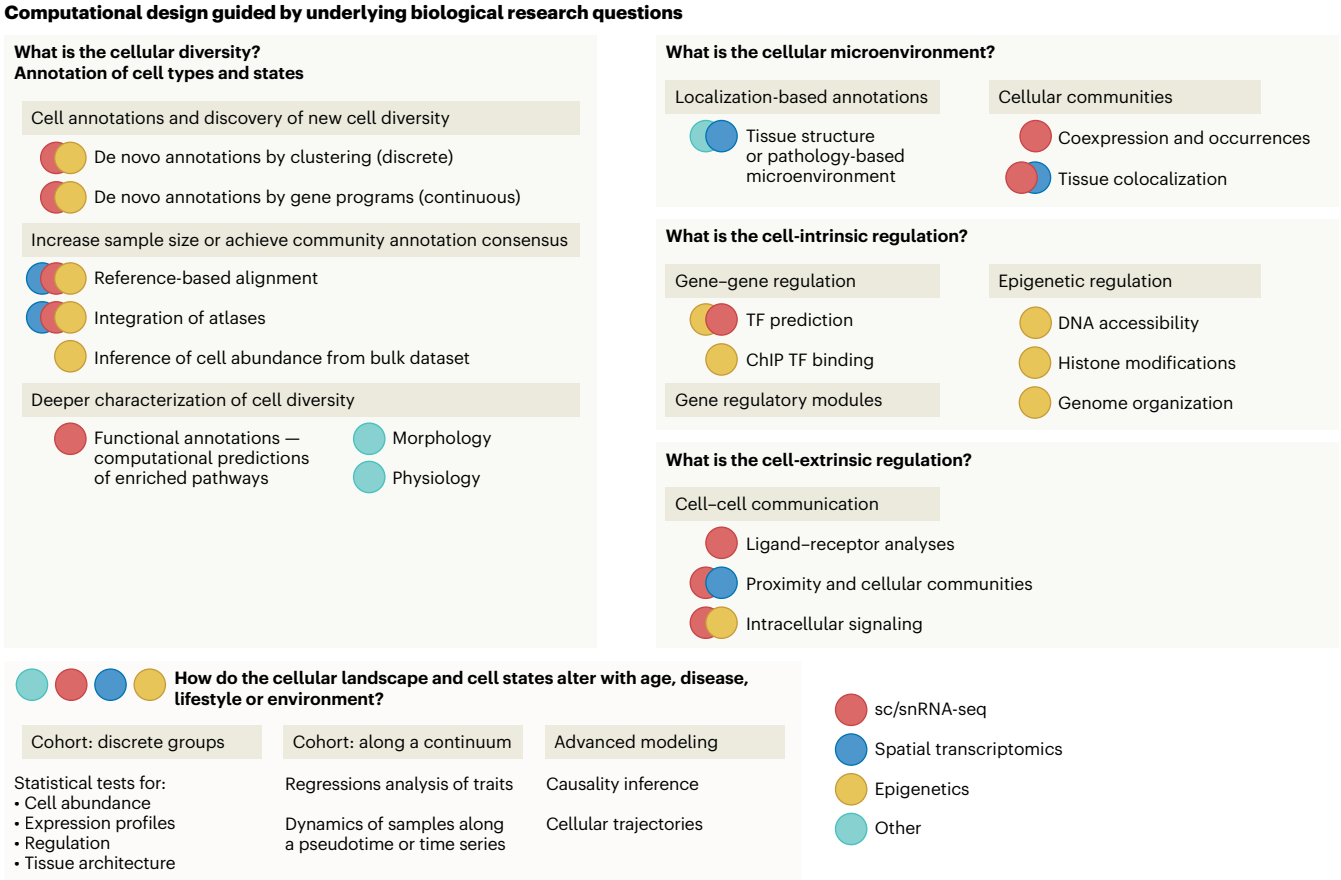


Fig. 3 | Outline of computational design of high-throughput single-cell or single-nucleus and spatial omics studies. ChIP, chromatin immunoprecipitation.

metanalysis and, therefore, should be taken as general guidelines as opposed to prescriptive guidelines. Large-scale consortia efforts will probably define these parameters in increasing detail.

These estimates are intended to provide general guidance, and we recommend consulting several published studies before embarking on experimental data generation.

While analysis of scRNA-seq and snRNA-seq data has reached a point of relative consensus, and broadly applicable recommendations can be proposed, guidelines for analysis and assessing the quality of data from more recent technologies, such as epigenomics and ST, are likely to emerge as the number of datasets increases.

Biological considerations in study design

Cellular architecture guiding study design

Apart from technical aspects, the computational study design should also be informed by the distinct characteristics of each cell type and the specific research questions of the study. These elements should guide the choice between a discrete clustering-based analysis of cellular diversity with a case–control differential expression analysis to uncover changes due to conditions and traits, and a continuous analysis of gene programs to describe cellular diversity and alignment of cells along continuous trajectories (Fig. 3). The particular research questions will further shape subsequent downstream analysis steps, which could encompass regulatory networks, cellular interactions, intersection with genetics and other modalities, and more.

Beyond the diversity of cell types, studies have revealed the vast diversity of cell subpopulations and cell states within the brain. Conventional approaches identify cellular diversity by clustering, commonly applying nearest-neighbor graphs and community detection algorithms, to subset cells of a specific cell type to subclusters that capture transcriptionally distinct cell subsets. Yet, cellular diversity might not always be adequately captured by a discrete model, such as microglia and astrocyte cells that rapidly respond to the changing environment to maintain brain homeostasis, or oligodendrocyte lineage cells that change along the maturation process. Alternative methods that model continuous variation in gene expression have been developed. For example, inference of gene expression programs enables us to model the complexity of cellular functions and response to diverse stimuli, by modeling cells as a combination of expression programs (for example, topic modeling[29], nonnegative matrix factorization[30] and weighted gene coexpression matrix analysis[31]). Algorithms that align cells along continuous trajectories of change (for example, Palantir[32] and Monocle[33]), as typically applied in developmental datasets[26], can also be used to study transitions between cell states in the adult brain, specifically along aging or disease processes. Importantly, as physiological processes involve cooperation of multiple cell types within the brain, new frameworks expand the analysis from the traditional focus on the diversity of individual cell types to multicellular environments[4].

Addressing complex cellular microenvironments in study design

The complex spatial arrangement of cells in the CNS is what enables it to execute its numerous, highly specialized functions, as physiological processes involve multiple cell types working in cooperation. Investigation of changes in cellular communities or microenvironments, instead of independent investigation of individual cell types, could be a more effective and accurate approach, despite its added complexity. Furthermore, future development of effective drugs and therapies is likely to require targeting a community of tightly co-regulated cells that together provide the necessary environment for the required

healthy function of the brain. In experimental study design, considering the different properties of brain cell types should guide the choice of methods for cell dissociation, nucleus isolation and tissue processing to capture the diversity of cells. In computational study design, beyond the analysis of individual cell types and cellular abundance, communities of cells with coordinated abundance and/or activity can be predicted using single-cell omics. For example, exploring changes in cell state and cell abundance across cell types in aging human brains uncovered an Alzheimer's disease-associated cellular community that captured coordinated changes in glial, endothelial and neuronal cell types[4,8]. Algorithms for the identification of coordinated cell programs across cell types, co-occurring cellular communities, co-regulated cells or communities with shared dynamics have been developed[8,34–36]. To further advance our understanding of the brain's dynamic tissue architecture, we will need to expand these computational methods. In particular, we will need to integrate predictions of the co-regulation and cross-talk between subsets of cells using multiple data modalities.

ST refers to recently developed technologies that make it possible to probe cellular microenvironments in situ. ST methods can simultaneously spatially position cells and quantify their transcriptomic profiles, with most methods applicable in histological tissue sections consisting of 1–2 cell layers. There are two broad classes of ST methods that use sequencing or imaging-based readouts. Many ST technologies were first applied to the mouse brain and have been used to spatially map cell types across entire brain regions[37–39], with recent efforts extending this approach to the whole mammalian brain[9,40–42]. ST can enrich cell-type and cell-state annotations from scRNA-seq and snRNA-seq data with spatial information and ascribe meaning to gene expression gradients, and even entire clusters, by identifying their spatial correlates, as exemplified in continuous expression gradients in neurons across the medial–lateral and superior–inferior axes of the striatum[43,44]. Broadly, ST methods fall into two categories: in situ RNA-seq-based technologies aimed at unbiased profiling of whole transcriptomes within tissues, and imaging-based technologies aimed at probing a multiplexed defined set of hundreds of genes. Sequencing or imaging-based ST methods each offer distinct advantages and have different challenges in balancing among resolution, sensitivity and scalability. Therefore, their advantages and respective limitations need to be considered carefully when choosing the right method to use. See Box 2 for a detailed description of technical consideration and limitations of imaging-based and sequencing-based ST methods.

The computational integration of spatial and single-cell transcriptomics provides a practical approach to construct multimodal brain atlases. The integration methodologies have matured dramatically in the past few years, so that now even relatively fine cell-type distinctions can be accurately mapped and resolved on high-quality ST datasets[45–48]. For imaging-based ST, integration of spatial and single-cell measurements enables transcriptome-wide imputation of unmeasured genes in space[40,49]. There are also emerging computational integration benchmarks for single-cell and spatial data, although we lack methods to quantitatively assess the accuracy and robustness of integration between scRNA-seq or snRNA-seq and ST datasets[50,51].

Cohort design and analytical considerations in disease studies

Increasing amounts of brain scRNA-seq and snRNA-seq data from different species, individuals, developmental stages and pathological states have revealed the diversity of neuronal and non-neuronal cells[26]. Such studies use a variety of cohort designs and computational schemes to identify transcriptional changes that occur in disease (Fig. 3). A carefully balanced case–control cohort is a conventional and powerful approach for identifying gene expression differences. For such a discrete cohort design, applying a statistical test (with an appropriate noise model[52]) with correction for multiple testing can be used to link cellular changes to the studied trait, correcting for technical (for example, batch, library quality) and biological (for example, interindividual variability, age and sex) confounders. Alternatively, pseudobulk differential gene expression methods (for example, DESeq[53], edgeR[54]) account for variability between biological replicates and thus avoid false discovery and overcome gene dropouts and noise in scRNA-seq or snRNA-seq data[55].

Newer approaches to cohort design and analytical methods consider continuous variation, as well as mixed sources of variation, ranging from sampling along the continuum of disease stages to random sampling of the population in abundant pathologies. For such a continuous cohort design, different analytical methods have been developed. For instance, linear mixed models can simultaneously take into account orthogonal sources of transcriptional variation and rank genes according to variance explained by specific variables[56,57]. Additionally, manifold learning and trajectory inference enable us to align individuals along a pseudotime of disease progression and infer the intricate cellular dynamics underlying the disease process[8,32,58]. These approaches provide a more nuanced perspective, offering insights into the temporal and spatial aspects of disease progression that might be missed in traditional case–control studies. To provide further confidence in the rigor and robustness of case–control datasets, transcriptomic changes predicted to occur in disease states should be validated using orthogonal methodologies. For an extensive discussion of validation approaches, see ref. 17.

Many neurological diseases are defined by specific histological lesions. ST methods provide an opportunity to bridge our historical understanding of these diseases with modern, hypothesis-generating genomics experiments of the same tissue and cells, which will pave the way to linking newly discovered disease-associated states and pathways with histopathological disease phenotypes, such as aggregates or multiple sclerosis lesions. Such discoveries may be facilitated by imaging-based ST technologies such as in situ sequencing (ISS), MERFISH or STARmap PLUS[59], but the limited genes detected by these technologies may limit the discovery of novel cell states that are uniquely associated with histopathology, without the accompanying transcriptome-wide profiling by scRNA-seq or snRNA-seq. Efforts to further expand the multiplexing (that is, to increase the number of genes and isoforms that can be measured in imaging ST) is an important area of future technology development. Sequencing-based ST approaches, including the original ST or DBiT-Seq[60,61], largely lack the resolution needed to precisely pair these histological lesions with gene expression in single cells, but neighborhood-based analyses are feasible. New sequencing-based approaches to capture individual cells with high spatial resolution and sufficient gene coverage, such as Stereo-seq, Slide-seq, Slide-Tags or the recently developed Visium HD, could help to overcome this challenge[62,63].

Emerging multi-omics spatial technologies as a bridge between modalities

The benefits of recent ST technological advancements provide an opportunity to combine ST and scRNA-seq or snRNA-seq to create a standardized cell atlas of the nervous system across diverse organisms by bridging anatomical, functional and molecular analyses of neural cell types. A key opportunity associated with ST is its ability to serve as a bridge between the fields of cellular–molecular genomics and systems neuroscience. Specifically, technological approaches are increasingly enabling genomic measurements of cells to be directly paired with measurements of connectivity and neural activity. For example, BARseq combines sequencing-based barcoding of neuronal projections with ISS-based spatial mapping of gene expression and neuronal cell typing[64]. Other studies have combined two-photon calcium imaging and cFOS-staining with ST to probe the molecular identities of neurons activated in different behaviors[38,65]. Furthermore, electrophysiological recordings have been coupled to STARmap in vitro in other systems[66]. These developments provide exciting avenues to marry ST with functional studies of neural circuits.

https://doi.org/10.1038/s41593-024-01806-0

BOX 2

Technical considerations and current limitations of spatial transcriptomics methods

Sequencing-based ST

Sequencing-based ST technologies use RNA-seq to enable unbiased profiling of whole transcriptomes in tissues. The various methods differ in their cell or transcript capture approach, spatial resolution, throughput (sample size, number) and sensitivity. For example, whereas some methods (for example, Visium[182], Slide-seq[37]) directly capture transcripts from tissues, others spatially barcode cells or nuclei (for example, Slide-Tags[62]) or tissue areas (for example, DBiT-seq[60], Nanostring CosMX/GeoMX[183]) before sequencing. The major benefit of sequencing-based ST is the discovery-based analysis of cellular transcriptomes in situ. This can be applied to healthy or diseased neural tissue samples with little prior information about tissue architecture and without target gene selection or differential gene analysis from prior scRNA-seq or snRNA-seq. Many methods (for example, Visium) are readily scalable as they require minimal specialized equipment and rely on standard histological methods and commercially available kits and sequencing reagents. The major limitations of these methods are summarized below.

Spatial resolution. Most techniques do not offer true single-cell resolution, as they profile multiple cells (for example, Visium with 55-µm resolution) or transcripts from neighboring cells (for example, Slide-seq with 10-µm resolution) in tissues. Hence, to perform cell-specific analysis akin to single-cell transcriptomics, they require computational deconvolution of cell-type-specific information. This is often based on cell-type-specific gene expression signatures extracted from reference scRNA-seq or snRNA-seq studies[45,46,48]. Hence, it is important to choose a reference that matches the biological characteristics of the ST dataset, such as brain region, cell-type composition and disease states. For disease studies, paired single-cell or single-nucleus and spatial datasets might be necessary for accurate deconvolution. Recent developments, such as Visium HD and Stereo-seq, provide higher spatial resolution and could address this limitation, although computational pipelines that can segment these data to single cells are not well established. Some low-level ST methods can have computationally enhanced resolution using Bayesian statistical tools like BayesSpace[184].

Tissue quality and assay performance. Many sequencing-based ST methods are best applied to fresh frozen tissue samples with high RNA integrity, which has been limiting for disease studies based on archival patient-derived samples. Yet, recent developments (for example, Visium) extend sequencing-based ST to formalin-fixed and paraffin-embedded samples through targeted sequencing of probes. There is no strong consensus yet on key tissue QC metrics: RNA integrity, histological stains and correlation of ST data with bulk or single-cell or single-nucleus RNA-seq are generally used in the field. Given the variations in human brain biopsy quality and autopsy protocols, standardization and benchmarks are needed to assess tissue quality and compare different technologies. While many of the sequencing-based ST methods aim to reach whole-transcriptome and gene coverage, in many cases the number of features per pixel or area is limited to a few thousand unique reads and/or genes, which might give limited insights into the cellular and molecular composition of the areas investigated.

Imaging-based ST

Imaging-based ST technologies use imaging to enable targeted analysis of transcripts in tissues. To image transcripts at high resolution, most methods use probe-based detection approaches derived from single-molecule fluorescence in situ hybridization (ISH) or custom sequencing chemistries. As with sequencing-based ST, there are many imaging-based methods, and they provide different levels of target gene multiplexing, detection sensitivity and specificity. For example, RNAscope ISH[185] can detect the expression of a few genes at high sensitivity, whereas high-multiplexed methods such as MERFISH[38], ISS[186], STARmap[187] and Xenium[188] use iterative cycles of labeling and combinatorial barcoding to simultaneously distinguish transcripts from hundreds or thousands of genes.

A major benefit of imaging-based ST is its high spatial resolution, which can resolve single cells in tissues and even subcellular localizations of targeted transcripts. Imaging-based methods are truly orthogonal to sequencing-based scRNA-seq and snRNA-seq technologies for validation of transcripts of interest, such as novel cell-type markers or genes that are differentially expressed in disease. Another benefit of imaging-based ST is direct 3D intact-tissue imaging of thick samples when combined with hydrogel tissue-clearing techniques[187]. The major limitations of these methods are summarized below.

Ease of use. High-multiplexed methods such as MERFISH and ISS require specialized automated microscopy equipment and extensive image analysis (for example, barcode decoding and cell segmentation) expertise[38,187]. Hence, their community uptake has been limited compared to more convenient methods such as RNAscope ISH or sequencing-based Visium. However, several imaging-based ST methods have recently been commercialized as end-to-end workflows with automated data collection and low-level image analysis, such as the MERSCOPE system based on MERFISH[5,9,88], the Xenium system based on ISS[189,190] and the Plexa system based on STARmap[187]. These commercial solutions will probably play important roles in democratizing access to these technologies.

Method of choice. Different methods present different trade-offs. MERFISH provides high detection sensitivity that requires high-resolution imaging, and in turn, long image acquisition times for large tissue samples[38]. By contrast, ISS detects fewer transcripts per cell, probably owing to the enzymatic steps used for signal amplification, but it can be performed at low resolution in a more scalable manner[187]. STARmap has a higher signal-to-noise ratio than single-molecule fluorescence ISH and higher detection efficiency than ISS, but it requires a high-end confocal microscope for 3D imaging. Finally, imaging-based technologies provide different error correction and detection capabilities[38,187], which greatly influence the specificity of transcript assignment and false discovery rates.

Probe selection. The curation of the probe panel is a critical step. Ideally, probe selection is guided by a scRNA-seq or snRNA-seq dataset of the specific tissue or condition, to avoid optical crowding by highly expressing genes (that is, the labeling of numerous

(continued from previous page)

transcripts in a given cell that obstructs optical identification of individual RNA spots), and includes markers of diverse cell types or pathological cell states[38]. In the absence of prior scRNA-seq or snRNA-seq data, probe selection could involve trial and error. Furthermore, this process also needs to be tailored to the sensitivity of the given ST technology. Although there are several computational tools to automate probe selection from scRNA-seq or snRNA-seq data[190,191], there is no strong consensus on panel curation approaches, and it is often done in a hybrid fashion involving both manual and automated curation.

Protocol optimization. Although it is cumbersome, optimizing ISH protocols (for example, proteinase treatment, autofluorescence removal) on new tissue types and sample sources is important. The study of human brain tissue, given wide variability in tissue quality and high autofluorescence, could be challenging[192].

Epigenomic technologies such as CUT&Tag, which is discussed below, have recently been successfully applied at the spatial level using the DBiT-Seq technology, enabling the genome-wide mapping of histone modifications in the mouse and human brain at different stages of development at a resolution approaching that of single cells (20–50 µm; spatial CUT&Tag)[67] or at hundreds of loci at subcellular resolution[68]. In addition, the recent developments of spatial ATAC-seq[69,70] and spatially resolved single-cell translatomics (RIBOmap)[71] potentiate spatial multi-omics mapping of epigenome, transcriptome and translatome data from the same brain samples to understand gene regulation mechanisms at both transcriptional and post-transcriptional levels. Moreover, multi-omic approaches combining RNA and CUT&Tag (and ATAC) have also been developed at a spatial level[72], while combined ST–lipidomics[73] and ST–metabolomics[74] are emerging. Thus, the simultaneous probing of several modalities might become standard in the spatial omics area, as it has in the single-cell and single-nucleus arena.

The Allen Brain Atlas Common Coordinate Framework[75] is a three-dimensional (3D) average map of the adult mouse brain, and provides an anatomical reference to standardize spatial measurements of neural activity and connectivity. The integration of ST-based cell maps with such common coordinate frameworks provides an opportunity for charting brain atlases. This is an active area of computational development[76], where a major challenge is the accurate mapping of 2D ST datasets to 3D coordinates at cellular resolution. Furthermore, it is likely that ST data will challenge some traditionally defined neuroanatomical boundaries[40,77]. New computational approaches will be needed to learn cytoarchitectural features from the spatial data and use them to improve our understanding of regional boundaries. Ultimately, functional perturbations of cells may also be required to refine these regional definitions.

The ability to map ST data to common coordinate frameworks, as discussed above, is likely to open opportunities to correspond molecular histopathology measurements to in vivo phenotypes. Functional magnetic resonance imaging can measure correlates of neural activity of a given brain region in the context of a particular behavioral task, and alterations in such activity have been observed in a wide range of neurological and psychiatric disorders[78,79]. For example, 7T magnetic resonance imaging can resolve disease structures such as iron-positive lesion rims in multiple sclerosis[80], and PET imaging can resolve metabolic tissue states, with ST measurements.

With the rapid proliferation of ST, systematic benchmarking of different methods is needed. These efforts should formally evaluate the consequences of spatial resolution, sensitivity and multiplexing levels for cell typing and CCI analysis of brain tissue. An emerging computational challenge is the integration of ST datasets across experimental batches, studies and technologies[81,82]. Although scRNA-seq and snRNA-seq methods are applicable to a certain extent, new methods to integrate imaging-based and sequencing-based ST as well as formal benchmarks to evaluate such methods will be necessary.

Finally, the current cost and throughput of both sequencing-based and imaging-based ST are prohibitive for mapping the whole human brain at the coverage of the mouse brain atlases. Technological or platform investments to increase the feasibility of large-scale ST and its application to 3D brain volumes, prioritizing human brain regions relevant to diseases and disorders such as autism or schizophrenia, and developing computational approaches to approximate full brain maps from incomplete 3D volumes or to predict ST profiles from magnetic resonance and metabolic positron emission tomography imaging provide future avenues for exploration. For example, myeloid cell subtypes—visualized by cell-type-specific molecular RNA and protein tagging—can be mapped to inflamed tissue areas on the basis of iron-sensitive magnetic resonance imaging[83] or metabolic tracers, as used in positron emission tomography imaging[84].

Considering cell regulation in the study design

Inference of CCIs. CCIs play key roles in the specification and function of the nervous system. Myriads of neuronal cell types interact to form synaptic connections and neural circuits across multiple scales, whereas glial and vascular cell-derived signals are important regulators of neuronal synapses and brain development and homeostasis. Furthermore, neuro-immune interactions are prominent in many neurological disorders. CCIs can be predicted using scRNA-seq or snRNA-seq by coupled expression of known ligand–receptor pairs. As RNA abundance plays only a stoichiometric role in mediating signaling activity, there is potential risk of false positive results, and multiple algorithms have been developed with diverse computational strategies and statistical frameworks to limit the sources of noise, such as CellPhoneDB[85], CellChat[86] and NicheNet[87].

ST can identify neuroglial tissue microenvironments that consist of spatially co-localized cell types that are specialized to support specific neural circuits and underlie neural pathologies. As ST jointly resolves the spatial positions and transcriptomes of cell types in tissues, it is uniquely suited for inferring CCIs by identifying cells that are in close proximity as well as mapping complementary expression of receptor–ligand pairs in neighboring cells. This can resolve short-range interactions (autocrine, juxtracrine or paracrine) such as neuronal–glial interactions at a cellular and molecular level[88,89]. As with cell-type mapping, integration with scRNA-seq or snRNA-seq data can enhance CCI analysis in ST by leveraging whole-transcriptome information from the former[90]. Current limitations of ST-based mapping of CCIs should be considered, as detailed in Box 3. Finally, CCIs can be linked to predicted downstream cellular phenotypes by inferring biological pathway activities or GRNs from single-cell or spatial data[87]. However, no existing approaches can trace CCIs from spatial information to gene expression to epigenomic profiles and other modalities.

Considering cell-intrinsic regulation in the experimental study design. Uncovering the major regulators that drive distinct transcriptional programs is a key step toward understanding brain function and dysfunction and can inform therapeutic strategies and drug discovery efforts. New analytical approaches and emerging technologies that allow researchers to profile epigenetic states of single cells can provide pivotal information related to the intrinsic mechanisms that drive expression programs and should be considered in study designs.

The most mature technology for identifying regulatory DNA measures chromatin accessibility—a state in which genomic DNA is

https://doi.org/10.1038/s41593-024-01806-0

BOX 3

Shortcomings of ST approaches to inferring CCIs

Resolution of sequencing-based ST. For CCI analysis, it is important to consider that most sequencing-based ST methods do not provide cellular resolution and profile multiple cells at each spot. Hence, these data are not equivalent to scRNA-seq or snRNA-seq data for interaction analysis and should be treated as such. Cell-type deconvolution preceding CCI detection (that is, where interactions are inferred from spatially co-located cell types from scRNA-seq or snRNA-seq profiles) or focused analysis of receptor–ligand transcript spatial co-localization are more appropriate analysis avenues for these datasets[45,87].

Multiplexing levels of imaging-based ST. While these methods provide single-cell resolution in situ, targeted probe panel selection often imposes limits on CCI analysis, as often only selected receptors and ligands are profiled in these experiments. In addition, targeting of a limited number of genes means that the data should not be treated as equivalent to scRNA-seq or snRNA-seq data, so subsequent interaction analysis will also need to be treated as such. For example, CCI analyses that rely on the assumption that most genes are not interacting are best applicable to single-cell full transcriptome datasets, where null distributions can be directly generated from the data. However, for imaging-based ST experiments, users may specifically select for receptors, ligands, hormones and other proteins that are expected to be interacting.

Computational models. Most CCI analysis tools were originally developed for suspension scRNA-seq or snRNA-seq data and do not incorporate true spatial information. Whereas ST can be used to prioritize CCI analysis results from scRNA-seq or snRNA-seq by identifying spatially co-located cell-type pairs[90], an active area of computational development is focused on inferring spatial effects of CCIs in ST, such as the identification of neighbor-dependent gene expression patterns[193]. Last, almost all models focus on discovering correlations of gene expression. However, models that build in causality (that is, GRNs) may enable the generation of more accurate hypotheses.

Long-range cellular interactions. Short-range interactions can be robustly captured by ST, but medium to long-range interactions, especially at axonal or dendritic processes of neurons or oligodendrocyte processes, complicate cell communication analysis in the nervous system. This challenge could be addressed by integration of ST with viral tracing methods.

actively accessed by macromolecules. Chromatin accessibility thus reflects whether proteins, typically transcription factors (TFs) and a high density of nucleosomes, are bound at a particular genomic locus[91]. Chromatin accessibility often correlates with gene expression, but it is not a direct proxy for it. As the chromatin accessibility landscape is highly cell-type specific, it provides a robust epigenomic measurement to identify cell types and states. It can be measured by commercially available assays such as snATAC-seq. Analysis of chromatin accessibility has been mainly used for identifying putative gene regulatory elements (enhancers, promoters, silencers and insulators), although it cannot necessarily distinguish among classes of gene regulatory elements. It also retains signals related to nucleosome occupancy, although it is less tailored to infer nucleosome positioning than techniques such as MNase-seq.

Moreover, available multi-omic assays enable joint profiling of RNA abundance (snRNA-seq) and chromatin accessibility (snATAC-seq) within a single nucleus, linking changes in gene regulatory element accessibility to changes in gene expression. One of the largest effects of locating cell-type-specific and context-specific gene regulatory elements is the identification of disease-associated noncoding variants that are predicted to influence gene regulation, enabling fine-mapping of thousands of genetic risk loci. Furthermore, chromatin accessibility-based mapping of quantitative trait loci (QTLs) provides more direct observation of the effect of a variant compared to the more indirect effects observed in expression QTL mapping due to linkage disequilibrium[92,93]. Variants that are predicted to be functional on the basis of residence in a gene regulatory element or the existence of a chromatin accessibility QTL can be further prioritized in validation experiments and linked to nearby genes via co-accessibility or 3D chromatin contacts. Of note, regulatory elements predicted by chromatin accessibility are putative and require downstream functional validation. For information on approaches that can validate the functional role of a specific element, see ref. 17.

Each cell has only two copies of each genomic locus (alleles), leading to unique challenges for single-nucleus epigenomics in: (i) sparsity, (ii) scale and (iii) cell-type specificity[94,95]. To partially address the challenges of data sparsity and specificity, snATAC-seq data are typically converted after clustering to pseudobulk to increase reliability by summing the information over hundreds to thousands of cells, each with a few thousands of fragments captured. Unlike scRNA-seq and snRNA-seq, which often fail to capture genes with the lowest levels of expression, the dropout in snATAC-seq is likely to be stochastic across the genome[96]. Moreover, snATAC-seq has a larger feature space compared to scRNA-seq or snRNA-seq, as a typical multicell-type dataset could have >1 million regulatory elements, in comparison to scRNA-seq and snRNA-seq, which have <30,000 unique transcripts. Because of this, snATAC-seq often requires a higher sequencing depth per cell and the larger size of the cell × feature space presents unique challenges for the analysis.

The epigenetic landscape of a cell is altered along differentiation, maturation, age and disease states; thus, it is interesting to compare chromatin accessibility and other epigenetic features between cases and controls or along biological processes (Fig. 3). Such comparisons from snATAC-seq data should be performed at the pseudobulk level[55] to reduce noise, while maintaining biological and cellular diversity by aggregating all cells of a single cell type from a given individual into a single pseudobulk profile. The use of pseudobulk profiles circumvents issues with sparsity by combining signal across many individual cells, and thus we recommend a minimum of 100 cells per profile on the basis of current studies. Once pseudobulk objects are created, differential testing should be performed using one of many tools that have been carefully benchmarked previously[97], for example, DESeq2 (ref. 53) or edgeR[98]. As tools vary in the prevalence of false positives and false negatives, the analyses should be tailored to the specific application and the tolerance for false positives versus false negatives, as discussed previously[97,99].

Additional technologies in the field of epigenomics are available, enabling the measurement of histone modifications, DNA methylation and chromatin contacts (as described in more detail below). Integration of these methods will provide a more complete view of the intrinsic mechanisms that underlie gene expression regulation within and across cell types.

Analytical approaches to infer GRNs. GRNs, the interconnected set of molecular regulators and their targets, together orchestrate the biological programs responsible for specific gene functions and ultimately govern all cellular and biological activities by controlling

https://doi.org/10.1038/s41593-024-01806-0

Table 1 | Common GRN tools

	Method description	Advantages	Pitfalls	Available tools
Correlation-based	Based on calculating the correlation coefficient (for example, Pearson correlation or Spearman rank correlation) between pairs of genes across multiple samples	Simple and computationally efficient	Cannot capture complex regulatory relationships and cannot differentiate between direct and indirect interactions	SCENIC[35]/SCENIC+[108], GENIE3 (ref. 157), PPCOR[158], LEAP[159]
Regression-based	Model the expression level of a gene as a function of the expression levels of other genes. Techniques include linear regression, LASSO and ridge regression	Can capture direct interactions	May not handle nonlinear relationships	GRNBoost2 (ref. 160), SINGE[161]
Information theory-based	Use measures such as mutual information to infer relationships between genes	Can handle nonlinear relationships and can differentiate direct from indirect interactions	May have difficulty with high-dimensional data	ARACNE[162], PIDC[163], SCRIBE[164] and CLR[165]
Bayesian network	Probabilistic graphical models that represent the dependencies among a set of variables	Can model complex relationships and differentiate direct from indirect interactions	May have difficulty with large networks as they are computationally intensive	GRNVBEM[166], BANJO[167] and BNFinder[168,169]
Boolean network	Model gene expression as on/off states and gene interactions as logical functions	Computationally efficient and can handle large networks	Oversimplify gene expression and cannot capture graded changes in expression levels	SCNS[170]
Traditional deep learning	Interpret GRNs from scRNA-seq or snRNA-seq data and deduce causal relationships between genes	Can handle complex, nonlinear relationships and high-dimensional data	Require large amounts of data and can be computationally intensive	CNNC (convolutional neural networks for coexpression)[171], foundation models[172]
Emerging deep learning	Based on foundation models, that is, deep learning models trained on vast amounts of data in an self-supervised fashion[173]. Emerging in single-cell transcriptomics as well[172], based on transfer learning from reference atlas and expand beyond	Attention mechanisms in the transformer architecture reflect the underlying GRN structure	Only in early stages. Require large amounts of data and can be computationally intensive	A single-cell transcriptomics transfer learning-based GRN model[172]
Differential equations	Model gene interactions as a system of differential equations	Can model time-dependent changes in gene expression	Require time-series data and can be computationally intensive	SCODE[174], GRISLI[175]

the activation and deactivation of individual genes. GRN analysis has become an increasingly important tool in neuroscience research. The complexity of the nervous system, characterized by its diverse cell types and intricate interconnections, makes it especially suitable for the GRN analysis. GRN analysis has uncovered the pivotal role of GRNs in multiple brain processes, such as differentiation of neuron and glial cells.

Regulators of gene expression are not restricted to regulatory proteins, and include various noncoding RNAs such as long, short and antisense noncoding RNAs. Regulators can act as activators or repressors via different regulatory elements (such as enhancers and promoters). Each stage of gene expression is regulated, including post-transcriptional regulation (splicing, translation, transport and degradation) and post-translational regulation (modifications, transport, localization and degradation). As the targets of each regulator include additional regulators, a network of interconnected regulators is formed, represented as a graph, where nodes capture regulators and their relationships are represented as directed and weighted edges.

As many of these regulatory modalities are not measured with single-cell resolution owing to the lack of scalable technologies, GRN inference is mainly focused on TF regulation, inferred from available single-cell omics measurements[100,101]. The inference of TF regulation is highly challenging for several reasons: (i) TFs can regulate expression of target genes at large genomic distances by binding to remote enhancer regions, often by folding to form close contacts proximal to their targets, limiting our ability to link a DNA element to its target gene; (ii) we lack reliable species-specific mapping of TFs to the DNA sequence motifs they can bind; (iii) physical binding of a TF to a DNA element can lead to activation, repression or no modulation, under different cellular contexts; (iv) TF binding often depends on the chromatin state. Thus, beyond relying on scRNA-seq or snRNA-seq data to infer GRNs[102], the use of single-cell multi-omic data, which includes both epigenomic (DNA contact, accessibility and methylation, or histone modification) and transcriptomic information, can enhance the accuracy and precision of identifying these regulatory relationships. This is achieved through integrative analysis that examines the correlations and colocalizations of open regions, methylation sites or regulatory motifs with gene expression.

Multiple computational methods have been developed to infer and to analyze GRNs at single-cell resolution from scRNA-seq or snRNA-seq data, including methods based on correlation, regression, information theory, Bayesian, Boolean and deep learning. An overview of such methods is provided in Table 1. Of course, different methods may produce divergent results, posing challenges in determining the most accurate representation of the network.

Identifying differentially accessible regions (DARs) between biological conditions is fundamental for pinpointing differentially active regulatory elements and regulators. Many methods for differential expression have been repurposed for finding DARs (for example, DESeq2 (ref. 53), edgeR[98] and limma voom[103]). Pairing DARs with differential expression analysis can enable the inference of context-associated GRNs using packages such as cisTopic[104], Signac[105], DORCs[106], FigR[107], SCENIC+[108], ArchR[109], MIRA[110], scBPGRN[111] and Symphony[112].

Additionally, CRISPR–Cas9-based screens, assessed at the single-cell level via Perturb-seq or genome and transcriptome sequencing

https://doi.org/10.1038/s41593-024-01806-0

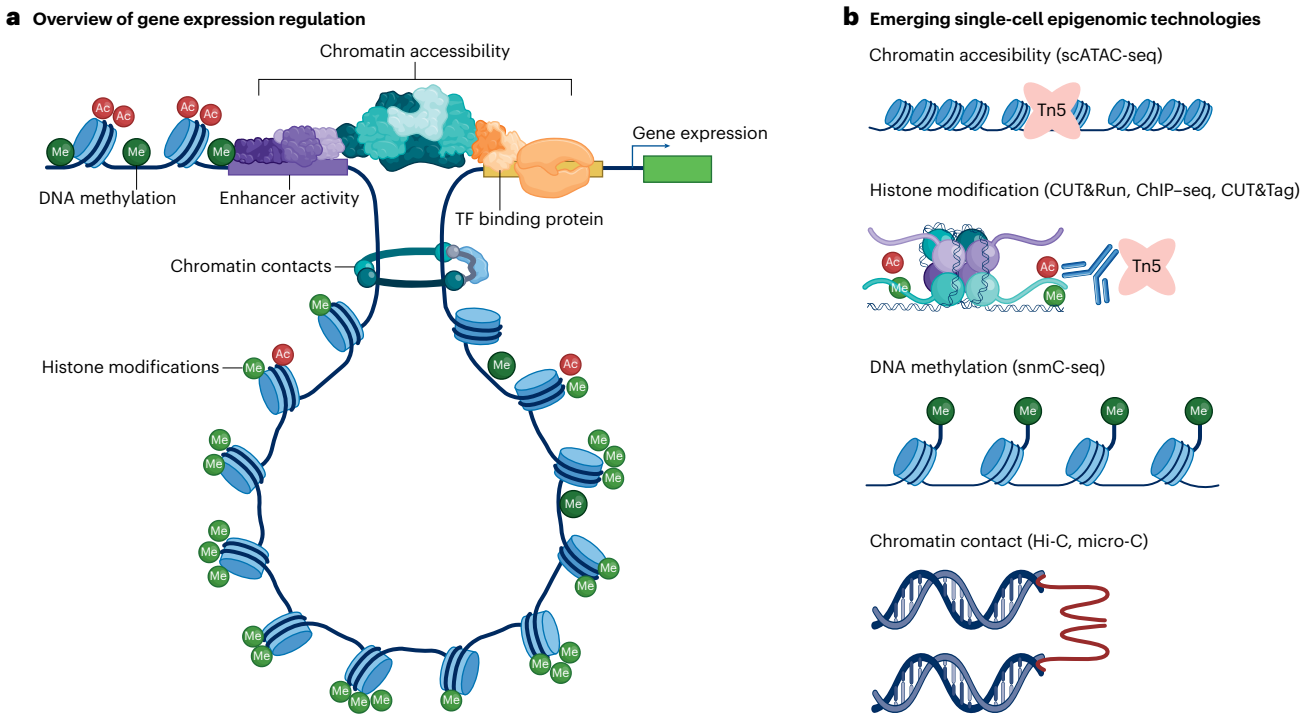


Fig. 4 | Epigenomic technologies. a, Overview of the major modalities of gene regulation that are currently studied using high-throughput single-cell technologies. **b**, Summary of key single-cell epigenomics technologies. Tn5, hyperactive Tn5 transposase; Me, methyl group; Ac, acetyl group.

(G&T-seq)[113], can also inform the inference of GRNs or experimentally validate them.

Computational models of GRNs are valuable for modeling complex data, generating hypotheses and directing future research efforts. However, the field is still evolving, and several challenges lie ahead. First, efficient analysis of GRNs requires large-scale high-quality data, often constrained by technical limitations, high cost, and computational complexity. Next, biological variation and technical noise might hinder the distinction between true regulatory interactions and fluctuations in gene expression. Thus, there is a need for further development of specialized statistical methods with heightened sensitivity for accurate DAR identification in future research efforts. Furthermore, gene expression regulation encompasses multiple layers of molecular interactions, making it difficult to accurately model and analyze GRNs. Finally, gene expression is a dynamic process, but experimental measurements provide only a static snapshot, greatly complicating the temporal dynamic analysis of GRNs. As a consequence of these shortcomings, our understanding of GRNs in the brain and the available tools to analyze them are lacking, yet future research in this field, empowered by new technologies, improved methodologies and the accumulation of large-scale datasets, is expected to yield important insights into brain function and dysfunction.

Emerging technologies for single-cell epigenomics

Various types of genetic and environmental perturbation can lead to changes in cell state driven by modulation of epigenetic states. To uncover the underlying regulatory mechanisms of such perturbations requires large-scale single-cell epigenetic data. Emerging technologies now allow researchers to profile epigenetic states of single cells and, in this section, we highlight emerging data modalities beyond DNA accessibility, discuss the main technologies and summarize their strengths and weaknesses (Fig. 4).

Histone modifications

Histones undergo multiple post-translational modifications that are involved in transcription regulation by affecting TF binding and RNA polymerase activity. Different histone modification patterns define specific genomic features, such as enhancers, promoters or coding regions, and some can be associated with transcriptional states (for example, activation, repression, elongation or poised[114]). Therefore, profiling histone modifications at single-cell resolution represents an emerging area of technological innovation that could facilitate the profiling of *cis*-regulatory elements that are likely to be functional, as opposed to all accessible DNA loci.

Profiling histone modifications was traditionally done by chromatin immunoprecipitation followed by sequencing (ChIP–seq), which requires large amounts of input material, hampering single-cell applications. CUT&Run[115] and CUT&Tag[116] methods, related to the ChiC[117] method and based on a fusion of MNase/Tn5 transposase with protein A, have been recently developed and successfully applied to single cells, allowing individual[116,118–122] or a combination of[123–126] histone modifications to be probed. Methods to investigate chromatin dynamics during cell-state transitions have recently been developed, taking advantage of multi-omic single-cell chromatin accessibility and transcriptomic data[106,127] or single-cell histone modification data, such as chromatin velocity[123,125].

Chromatin contact

Genome organization and DNA methylation represent additional layers of gene expression regulation[128,129]. Genome topology can modulate enhancer–promoter communication[130], and there is considerable evidence linking disease-relevant structural variants or epigenetic changes to changes in 3D genome organization[131].

Genome organization has been traditionally studied using chromatin conformation capture assays such as Hi-C or micro-C in bulk or in populations subjected to fluorescence-activated cell sorting, yet recent advances in technologies have enabled profiling in single

cells/nuclei[132–134]. Moreover, recent technologies allow co-profiling of chromatin architecture and transcriptomes at a single-cell level[135]. However, current methods require high sequencing depth compared with ATAC-seq and lack robust experimental and computational benchmarking and validations.

Methylation

The most common DNA base modification in mammalian species is cytosine 5-methylation (5mC) and its oxidated derivatives including 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[136]. The mammalian CNS is associated with unique patterns of cytosine modifications, including unusually high levels of non-CpG methylation (5mCH) and 5hmC[136–138]. For the majority of somatic tissues in mammalian species, 5mC is almost exclusively found in the CpG dinucleotide context (5mCG). However, in mature human cortical neurons, 5mCH could account for more than half of cytosine methylation, with between 2% and 8% of CH sites methylated depending on neuronal types[139].

The single-nucleus profiling of cytosine modifications reveals epigenetic states at different genomic scales[138]. While the depletion of 5mCG and/or enrichment of 5hmCG at regulatory elements indicates a local epigenomically permissive state, the cell-type and developmental specificity of 5mCH can shed light on the epigenomic states of mega-scale regions, such as topologically associating domains, or intermediate-scale regions such as gene bodies. Existing single-cell methylome techniques provide sparse measurements at the individual cell level, typically covering 5% to 10% of the genome[139]. The aggregation of single-cell methylome profiles for identified cell types (pseudobulk) has been applied to reconstruct cell-type-specific profiles that are reminiscent of traditional bulk methylome profiles. The pseudobulk approach provides a reasonable approximation for steady-state cell populations in the adult brain but inevitably leads to the under-appreciation of ongoing cellular dynamics in developing brains. We anticipate that data imputation tools and trajectory inference algorithms that can effectively use sparse single-cell methylome profiles will expand the knowledge of methylation dynamics during brain development[140].

The development of single-cell profiling techniques for 5mC and 5hmC has revealed the cell-type-specific patterns of cytosine modifications in mammalian brains. The snmC-seq family of assays were based upon earlier methods such as scBS-seq and provided enhanced throughput, enabling the generation of hundreds of thousands of single-cell methylome profiles[139,141,142]. SnmC-seq has also been extended to multi-omic approaches, including the joint profiling of chromatin conformation and DNA methylation by snm3C-seq and the simultaneous profiling of transcriptomes, DNA methylation and chromatin accessibility by snmCAT-seq[143,144]. A method for single-cell profiling of 5hmC, snhmC-seq, was recently developed by integrating chemical protection of 5hmC by bisulfite conversion and selective deamination by APOBEC3A[145]. The absence of easy-to-use and commercially available assays has severely impeded the adoption of single-cell methylome methods. A recently developed combinatorial indexing-based approach, sciMETv2, provides a feasible route toward commercialization[146]. Last, methods that can generate high-coverage methylomes from a single cell, but probably from a smaller number of cells, could be useful for analyzing highly specific cell populations such as those associated with a neural circuit in adult brains or daughter cells derived from asymmetric divisions during neural development.

Single-molecule epigenomic assays

Finally, single-molecule epigenomic assays use high-throughput long-read sequencing technologies (for example, Pacific Biosciences or Oxford Nanopore) to make high-throughput, single-molecule genomic measurements of chromatin accessibility[147–151], as well as single-molecule sequencing of intact RNA isoforms[152], at the resolution of single cells[153]. Single-molecule chromatin accessibility profiling approaches allow one to 'deconvolute' the population averages provided by approaches such as DNase-seq and ATAC-seq—that is, one can explicitly map the presence of nucleosomes, TFs and their respective co-occupancy patterns on individual DNA molecules.

For instance, it is now possible to identify heterogeneity of nucleosome positioning and TF binding[147,148], and this has been informative in dissecting complex epigenome and regulatory pathways[154,155]. Key challenges are currently the throughput of long-read sequencing and its cost, which are expected to be reduced in the coming years.

Data sharing, dissemination and visualization

Data sharing, dissemination and effective visualization are crucial aspects of modern research, especially in the realm of single-cell data analysis, given the speed of data generation and the need for large-scale datasets for capturing the full complexity and diversity of cells in the brain, to ensure transparency and reproducibility.

One noteworthy step toward enhancing transparency is the open sharing of code. Platforms such as Jupyter notebooks allow researchers to share not only their results but also the entire analysis pipeline. This practice is particularly valuable for highly customized data analysis that extends beyond standard packages. It enables others to reproduce and validate complex analyses, fostering trust and collaboration within the scientific community. It is essential to follow best practices in documenting data analysis methods and parameters when using complex analysis packages. Ideally, there is a convergence to shared data formats and structures such as AnnData in single-cell analysis, and first efforts for extensions via the Open Microscopy Environment (OME) Standard are recommendable[156]. Benchmarking algorithms is another vital aspect of advancing single-cell data analysis. By rigorously evaluating the performance of analysis methods, researchers can identify the most reliable and efficient tools for their specific research questions. This process contributes to the continuous improvement of analytical techniques and ensures the validity of scientific findings.

To further enhance the accessibility and usability of data, adhering to the principles of FAIR (findable, accessible, interoperable and reusable) data is essential. Proper curation of patient and experimental metadata within datasets ensures that critical context accompanies the data, making it more valuable for researchers and promoting data reuse.

Data visualization plays a pivotal role in translating complex datasets into understandable insights. While shiny apps have initially been used extensively to allow exploration of single-cell data, various single-cell data portals, such as UCSC Cell Browser, Cell Annotation Platform and CELLxGENE, have made important strides in democratizing data access and navigation. However, for these efforts to thrive and expand, there is a pressing need for broader community engagement and sustained financial support. Visualizing spatial genomic data, particularly datasets integrated with single-cell omics data, presents unique challenges. These datasets offer critical insights into the spatial organization of cells within tissues. To make this information more accessible to the research community, concerted efforts are required to develop online resources that facilitate data exploration and visualization.

Overall, we see effective data sharing, documentation, visualization and benchmarking as integral to the progress of single-cell data analysis. These practices promote transparency, reproducibility and collaborations, and ultimately will lead to better understanding of complex biological systems.

Limitations

Emerging genomic technologies have the potential to transform our understanding of nervous system development, structure and function. This Review aims to discuss the challenges of designing studies that are rigorous, well powered and informative, and outlines several key

applications including building atlases, uncovering disease processes and predicting gene regulatory relationships. Clearly, applications of these technologies extend beyond those specific use cases, and space limitations required us to omit some details, particularly related to development and cross-species comparisons[17,26]. These applications involve their own sets of opportunities and challenges that we regretfully could not cover in this article.

References

1. Siletti, K. et al. Transcriptomic diversity of cell types across the adult human brain. *Science* **382**, eadd7046 (2023).
2. Kim, S. S. et al. Leveraging single-cell ATAC-seq and RNA-seq to identify disease-critical fetal and adult brain cell types. *Nat. Commun.* **15**, 563 (2024).
3. Sun, N. et al. Single-nucleus multiregion transcriptomic analysis of brain vasculature in Alzheimer's disease. *Nat. Neurosci.* **26**, 970–982 (2023).
4. Cain, A. et al. Multicellular communities are perturbed in the aging human brain and Alzheimer's disease. *Nat. Neurosci.* **26**, 1267–1280 (2023).
5. Kim, C. N., Shin, D., Wang, A. & Nowakowski, T. J. Spatiotemporal molecular dynamics of the developing human thalamus. *Science* **382**, eadf9941 (2023).
6. Pineda, S. S. et al. Single-cell dissection of the human motor and prefrontal cortices in ALS and FTLD. *Cell* **187**, 1971–1989 (2024).
7. Zeisel, A. et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
8. Green, G. S. et al. Cellular communities reveal trajectories of brain ageing and Alzheimer's disease. *Nature* https://doi.org/10.1038/s41586-024-07871-6 (2024).
9. Yao, Z. et al. A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature* **624**, 317–332 (2023).
10. Davis, A., Gao, R. & Navin, N. E. SCOPIT: sample size calculations for single-cell sequencing experiments. *BMC Bioinformatics* **20**, 566 (2019).
11. Schmid, K. T. et al. scPower accelerates and optimizes the design of multi-sample single cell transcriptomic studies. *Nat. Commun.* **12**, 6625 (2021).
12. Su, K., Wu, Z. & Wu, H. Simulation, power evaluation and sample size recommendation for single-cell RNA-seq. *Bioinformatics* **36**, 4860–4868 (2020).
13. Phipson, B. et al. Propeller: testing for differences in cell type proportions in single cell data. *Bioinformatics* **38**, 4720–4726 (2022).
14. Lin, Y. et al. scClassify: sample size estimation and multiscale classification of cells using single and multiple reference. *Mol. Syst. Biol.* **16**, e9389 (2020).
15. Jeon, H. et al. Statistical power analysis for designing bulk, single-cell, and spatial transcriptomics experiments: review, tutorial, and perspectives. *Biomolecules* **13**, 221 (2023).
16. Ryaboshapkina, M. & Azzu, V. Sample size calculation for a NanoString GeoMx spatial transcriptomics experiment to study predictors of fibrosis progression in non-alcoholic fatty liver disease. *Sci. Rep.* **13**, 8943 (2023).
17. Colonna, M. et al. Implementation and validation of single-cell genomics experiments in neuroscience. *Nat. Neurosci.* https://doi.org/10.1038/s41593-024-01814-0 (2024).
18. Zhang, Y. et al. Deconvolution algorithms for inference of the cell-type composition of the spatial transcriptome. *Comput. Struct. Biotechnol. J.* **21**, 176–184 (2023).
19. Im, Y. & Kim, Y. A comprehensive overview of RNA deconvolution methods and their application. *Mol. Cells* **46**, 99–105 (2023).
20. Charytonowicz, D., Brody, R. & Sebra, R. Interpretable and context-free deconvolution of multi-scale whole transcriptomic data with UniCell deconvolve. *Nat. Commun.* **14**, 1350 (2023).
21. Chen, Y. et al. Deep autoencoder for interpretable tissue-adaptive deconvolution and cell-type-specific gene analysis. *Nat. Commun.* **13**, 6735 (2022).
22. Liao, J. et al. De novo analysis of bulk RNA-seq data at spatially resolved single-cell resolution. *Nat. Commun.* **13**, 6498 (2022).
23. Heimberg, G., Bhatnagar, R., El-Samad, H. & Thomson, M. Low dimensionality in gene expression data enables the accurate extraction of transcriptional programs from shallow sequencing. *Cell Syst.* **2**, 239–250 (2016).
24. Haque, A., Engel, J., Teichmann, S. A. & Lönnberg, T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. *Genome Med.* **9**, 75 (2017).
25. Boyce, R. W., Dorph-Petersen, K. -A., Lyck, L. & Gundersen, H. J. G. Design-based stereology: introduction to basic concepts and practical approaches for estimation of cell number. *Toxicol. Pathol.* **38**, 1011–1025 (2010).
26. Adameyko, I. et al. Applying single-cell/nucleus genomics to studies of cellular heterogeneity and cell fate transitions in the nervous system. *Nat. Neurosci.* https://doi.org/10.1038/s41593-024-01827-9 (2024).
27. Yu, L., Cao, Y., Yang, J. Y. H. & Yang, P. Benchmarking clustering algorithms on estimating the number of cell types from single-cell RNA-sequencing data. *Genome Biol.* **23**, 49 (2022).
28. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587 (2021).
29. Swapna, L. S., Huang, M. & Li, Y. GTM-decon: guided-topic modeling of single-cell transcriptomes enables sub-cell-type and disease-subtype deconvolution of bulk transcriptomes. *Genome Biol.* **24**, 190 (2023).
30. Zhang, S., Yang, L., Yang, J., Lin, Z. & Ng, M. K. Dimensionality reduction for single cell RNA sequencing data using constrained robust non-negative matrix factorization. *NAR Genom. Bioinform.* **2**, lqaa064 (2020).
31. Morabito, S., Reese, F., Rahimzadeh, N., Miyoshi, E. & Swarup, V. hdWGCNA identifies co-expression networks in high-dimensional transcriptomics data. *Cell Rep. Methods* **3**, 100498 (2023).
32. Setty, M. et al. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).
33. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
34. Biancalani, T. et al. Deep learning and alignment of spatially resolved single-cell transcriptomes with Tangram. *Nat. Methods* **18**, 1352–1362 (2021).
35. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
36. Cortal, A., Martignetti, L., Six, E. & Rausell, A. Gene signature extraction and cell identity recognition at the single-cell level with Cell-ID. *Nat. Biotechnol.* **39**, 1095–1102 (2021).
37. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
38. Moffitt, J. R. et al. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, eaau5324 (2018).
39. Bayraktar, O. A. et al. Astrocyte layers in the mammalian cerebral cortex revealed by a single-cell in situ transcriptomic map. *Nat. Neurosci.* **23**, 500–509 (2020).
40. Shi, H. et al. Spatial atlas of the mouse central nervous system at molecular resolution. *Nature* **622**, 552–561 (2023).
41. Zhang, M. et al. Molecularly defined and spatially resolved cell atlas of the whole mouse brain. *Nature* **624**, 343–354 (2023).
42. Langlieb, J. et al. The molecular cytoarchitecture of the adult mouse brain. *Nature* **624**, 333–342 (2023).

43. Stanley, G., Gokce, O., Malenka, R. C., Südhof, T. C. & Quake, S. R. Continuous and discrete neuron types of the adult murine striatum. *Neuron* **105**, 688–699 (2020).
44. Muñoz-Manchado, A. B. et al. Diversity of interneurons in the dorsal striatum revealed by single-cell RNA sequencing and PatchSeq. *Cell Rep.* **24**, 2179–2190 (2018).
45. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
46. Cable, D. M. et al. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nat. Biotechnol.* **40**, 517–526 (2022).
47. Ghazanfar, S., Guibentif, C. & Marioni, J. C. Stabilized mosaic single-cell data integration using unshared features. *Nat. Biotechnol*. https://doi.org/10.1038/s41587-023-01766-z (2023).
48. Mages, S. et al. TACCO unifies annotation transfer and decomposition of cell identities for single-cell and spatial omics. *Nat. Biotechnol.* **41**, 1465–1473 (2023).
49. Lohoff, T. et al. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.* **40**, 74–85 (2022).
50. Li, B. et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670 (2022).
51. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2022).
52. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* **16**, 278 (2015).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
54. Baldoni, P. L. et al. Dividing out quantification uncertainty allows efficient assessment of differential transcript expression with edgeR. *Nucleic Acids Res.* **52**, e13 (2024).
55. Squair, J. W. et al. Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).
56. Hoffman, G. E. et al. Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. Preprint at *bioRxiv* https://doi.org/10.1101/2023.03.17.533005 (2023).
57. Hoffman, G. E. & Schadt, E. E. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics* **17**, 483 (2016).
58. Gabitto, M. I. et al. Integrated multimodal cell atlas of Alzheimer's disease. *Nat. Neurosci.* https://doi.org/10.1038/s41593-024-01774-5 (2024).
59. Zeng, H. et al. Integrative in situ mapping of single-cell transcriptional states and tissue histopathology in a mouse model of Alzheimer's disease. *Nat. Neurosci.* **26**, 430–446 (2023).
60. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **185**, 1665–1681 (2020).
61. Ke, R. et al. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
62. Russell, A. J. C. et al. Slide-tags enables single-nucleus barcoding for multimodal spatial genomics. *Nature* **625**, 101–109 (2024).
63. Chen, A. et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).
64. Chen, X. et al. High-throughput mapping of long-range neuronal projection using in situ sequencing. *Cell* **179**, 772–786 (2019).
65. Condylis, C. et al. Dense functional and molecular readout of a circuit hub in sensory cortex. *Science* **375**, eabl5981 (2022).
66. Li, Q. et al. Multimodal charting of molecular and functional cell states via in situ electro-sequencing. *Cell* **186**, 2002–2017 (2023).
67. Deng, Y. et al. Spatial-CUT&Tag: spatially resolved chromatin modification profiling at the cellular level. *Science* **375**, 681–686 (2022).
68. Lu, T., Ang, C. E. & Zhuang, X. Spatially resolved epigenomic profiling of single cells in complex tissues. *Cell* **185**, 4448–4464 (2022).
69. Deng, Y. et al. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature* **609**, 375–383 (2022).
70. Llorens-Bobadilla, E. et al. Solid-phase capture and profiling of open chromatin by spatial ATAC. *Nat. Biotechnol.* **41**, 1085–1088 (2023).
71. Zeng, H. et al. Spatially resolved single-cell translatomics at molecular resolution. *Science* **380**, eadd3067 (2023).
72. Zhang, D. et al. Spatial epigenome-transcriptome co-profiling of mammalian tissues. *Nature* **616**, 113–122 (2023).
73. Sans, M. et al. Integrated spatial transcriptomics and lipidomics of precursor lesions of pancreatic cancer identifies enrichment of long chain sulfatide biosynthesis as an early metabolic alteration. Preprint at *bioRxiv* https://doi.org/10.1101/2023.08.14.553002 (2023).
74. Vicari, M. et al. Spatial multimodal analysis of transcriptomes and metabolomes in tissues. *Nat. Biotechnol*. https://doi.org/10.1038/s41587-023-01937-y (2023).
75. Wang, Q. et al. The Allen Mouse Brain Common Coordinate Framework: a 3D reference atlas. *Cell* **181**, 936–953 (2020).
76. Clifton, K. et al. STalign: alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nat. Commun.* **14**, 8123 (2023).
77. Ortiz, C. et al. Molecular atlas of the adult mouse brain. *Sci. Adv.* **6**, eabb3446 (2020).
78. Kumar, K. et al. Subcortical brain alterations in carriers of genomic copy number variants. *Am. J. Psychiatry* **180**, 685–698 (2023).
79. Moreau, C. A. et al. Brain functional connectivity mirrors genetic pleiotropy in psychiatric conditions. *Brain* **146**, 1686–1696 (2023).
80. Bruschi, N., Boffa, G. & Inglese, M. Ultra-high-field 7-T MRI in multiple sclerosis and other demyelinating diseases: from pathology to clinical practice. *Eur. Radiol. Exp.* **4**, 59 (2020).
81. Tang, Z. et al. Search and match across spatial omics samples at single-cell resolution. *Nat. Methods* **21**, 1818–1829 (2024).
82. Xia, C. -R., Cao, Z. -J., Tu, X. -M. & Gao, G. Spatial-linked alignment tool (SLAT) for aligning heterogenous slices. *Nat. Commun.* **14**, 7236 (2023).
83. Hofmann, A. et al. Myeloid cell iron uptake pathways and paramagnetic rim formation in multiple sclerosis. *Acta Neuropathol.* **146**, 707–724 (2023).
84. Sucksdorff, M. et al. Brain TSPO-PET predicts later disease progression independent of relapses in multiple sclerosis. *Brain* **143**, 3318–3330 (2020).
85. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
86. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
87. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
88. Fang, R. et al. Conservation and divergence of cortical cell organization in human and mouse revealed by MERFISH. *Science* **377**, 56–62 (2022).
89. Wu, S. J. et al. Cortical somatostatin interneuron subtypes form cell-type-specific circuits. *Neuron* **111**, 2675–2692 (2023).
90. Garcia-Alonso, L. et al. Mapping the temporal and spatial dynamics of the human endometrium in vivo and in vitro. *Nat. Genet.* **53**, 1698–1711 (2021).

91. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
92. Tehranchi, A. et al. Fine-mapping *cis*-regulatory variants in diverse human populations. *Elife* **8**, e39595 (2019).
93. Kosoy, R. et al. Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nat. Genet.* **54**, 1145–1154 (2022).
94. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
95. Cusanovich, D. A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
96. Buenrostro, J. D. et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
97. Gontarz, P. et al. Comparison of differential accessibility analysis strategies for ATAC-seq data. *Sci. Rep.* **10**, 10150 (2020).
98. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
99. Grandi, F. C., Modi, H., Kampman, L. & Corces, M. R. Chromatin accessibility profiling by ATAC-seq. *Nat. Protoc.* **17**, 1518–1552 (2022).
100. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
101. Macneil, L. T. & Walhout, A. J. M. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Res.* **21**, 645–657 (2011).
102. Pratapa, A., Jalihal, A. P., Law, J. N., Bharadwaj, A. & Murali, T. M. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nat. Methods* **17**, 147–154 (2020).
103. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
104. Bravo González-Blas, C. et al. cisTopic: cis-regulatory topic modeling on single-cell ATAC-seq data. *Nat. Methods* **16**, 397–400 (2019).
105. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
106. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. *Cell* **183**, 1103–1116 (2020).
107. Kartha, V. K. et al. Functional inference of gene regulation using single-cell multi-omics. *Cell Genom.* **2**, 100166 (2022).
108. Bravo González-Blas, C. et al. SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat. Methods* **20**, 1355–1367 (2023).
109. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
110. Lynch, A. W. et al. MIRA: joint regulatory modeling of multimodal expression and chromatin accessibility in single cells. *Nat. Methods* **19**, 1097–1108 (2022).
111. Xuan, C. et al. scBPGRN: integrating single-cell multi-omics data to construct gene regulatory networks based on BP neural network. *Comput. Biol. Med.* **151**, 106249 (2022).
112. Kang, J. B. et al. Efficient and precise single-cell reference atlas mapping with Symphony. *Nat. Commun.* **12**, 5890 (2021).
113. Macaulay, I. C. et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 (2015).
114. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).
115. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, e21856 (2017).
116. Kaya-Okur, H. S. et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).
117. Schmid, M., Durussel, T. & Laemmli, U. K. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol. Cell* **16**, 147–157 (2004).
118. Grosselin, K. et al. High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat. Genet.* **51**, 1060–1066 (2019).
119. Wu, S. J. et al. Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat. Biotechnol.* **39**, 819–824 (2021).
120. Wang, Q. et al. CoBATCH for high-throughput single-cell epigenomic profiling. *Mol. Cell* **76**, 206–216 (2019).
121. Handa, T. et al. Chromatin integration labeling for mapping DNA-binding proteins and modifications with low input. *Nat. Protoc.* **15**, 3334–3360 (2020).
122. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* **39**, 825–835 (2021).
123. Bartosovic, M. & Castelo-Branco, G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nat. Biotechnol.* **41**, 794–805 (2023).
124. Gopalan, S., Wang, Y., Harper, N. W., Garber, M. & Fazzio, T. G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol. Cell* **81**, 4736–4746 (2021).
125. Tedesco, M. et al. Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. *Nat. Biotechnol.* **40**, 235–244 (2022).
126. Meers, M. P., Llagas, G., Janssens, D. H., Codomo, C. A. & Henikoff, S. Multifactorial profiling of epigenetic landscapes at single-cell resolution using MulTI-Tag. *Nat. Biotechnol.* **41**, 708–716 (2023).
127. Li, C., Virgilio, M. C., Collins, K. L. & Welch, J. D. Multi-omic single-cell velocity models epigenome-transcriptome interactions and improves cell fate prediction. *Nat. Biotechnol.* **41**, 387–398 (2023).
128. Schoenfelder, S. & Fraser, P. Long-range enhancer-promoter contacts in gene expression control. *Nat. Rev. Genet.* **20**, 437–455 (2019).
129. Noack, F. et al. Multimodal profiling of the transcriptional regulatory landscape of the developing mouse cortex identifies Neurog2 as a key epigenome remodeler. *Nat. Neurosci.* **25**, 154–167 (2022).
130. Zuin, J. et al. Nonlinear control of transcription through enhancer-promoter interactions. *Nature* **604**, 571–577 (2022).
131. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. *Nat. Rev. Genet.* **19**, 453–467 (2018).
132. Nagano, T. et al. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61–67 (2017).
133. Tan, L., Xing, D., Chang, C. -H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).
134. Tan, L. et al. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. *Cell* **184**, 741–758 (2021).
135. Zhou, T. et al. GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. *Nat. Genet.* https://doi.org/10.1038/s41588-024-01745-3 (2024).
136. Wu, H. & Zhang, Y. Charting oxidized methylcytosines at base resolution. *Nat. Struct. Mol. Biol.* **22**, 656–661 (2015).
137. Lister, R. et al. Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).

138. Luo, C., Hajkova, P. & Ecker, J. R. Dynamic DNA methylation: in the right place at the right time. *Science* **361**, 1336–1340 (2018).
139. Luo, C. et al. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science* **357**, 600–604 (2017).
140. Iqbal, W. & Zhou, W. Computational methods for single-cell DNA methylome analysis. *Genomics Proteomics Bioinformatics* **21**, 48–66 (2023).
141. Smallwood, S. A. et al. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* **11**, 817–820 (2014).
142. Luo, C. et al. Robust single-cell DNA methylome profiling with snmC-seq2. *Nat. Commun.* **9**, 3824 (2018).
143. Lee, D. -S. et al. Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat. Methods* **16**, 999–1006 (2019).
144. Luo, C. et al. Single nucleus multi-omics identifies human cortical cell regulatory genome diversity. *Cell Genom.* **2**, 100107 (2022).
145. Fabyanic, E. B. et al. Quantitative single cell 5hmC sequencing reveals non-canonical gene regulation by non-CG hydroxymethylation. Preprint at *bioRxiv* https://doi.org/10.1101/2021.03.23.434325 (2021).
146. Nichols, R. V. et al. High-throughput robust single-cell DNA methylation profiling with sciMETv2. *Nat. Commun.* **13**, 7627 (2022).
147. Abdulhay, N. J. et al. Massively multiplex single-molecule oligonucleosome footprinting. *Elife* **9**, e59404 (2020).
148. Stergachis, A. B., Debo, B. M., Haugen, E., Churchman, L. S. & Stamatoyannopoulos, J. A. Single-molecule regulatory architectures captured by chromatin fiber sequencing. *Science* **368**, 1449–1454 (2020).
149. Shipony, Z. et al. Long-range single-molecule mapping of chromatin accessibility in eukaryotes. *Nat. Methods* **17**, 319–327 (2020).
150. Lee, I. et al. Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. *Nat. Methods* **17**, 1191–1199 (2020).
151. Wang, Y. et al. Single-molecule long-read sequencing reveals the chromatin basis of gene expression. *Genome Res.* **29**, 1329–1342 (2019).
152. Sharon, D., Tilgner, H., Grubert, F. & Snyder, M. A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* **31**, 1009–1014 (2013).
153. Gupta, I. et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202 (2018).
154. Isaac, R. S. et al. Single-nucleoid architecture reveals heterogeneous packaging of mitochondrial DNA. *Nat. Struct. Mol. Biol.* **31**, 568–577 (2024).
155. Abdulhay, N. J. et al. Nucleosome density shapes kilobase-scale regulation by a mammalian chromatin remodeler. *Nat. Struct. Mol. Biol.* **30**, 1571–1581 (2023).
156. Marconato, L. et al. SpatialData: an open and universal data framework for spatial omics. *Nat. Methods* https://doi.org/10.1038/s41592-024-02212-x (2024).
157. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. *PLoS ONE* **5**, e12776 (2010).
158. Kim, S. Ppcor: an R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **22**, 665–674 (2015).
159. Specht, A. T. & Li, J. LEAP: constructing gene co-expression networks for single-cell RNA-sequencing data using pseudotime ordering. *Bioinformatics* **33**, 764–766 (2017).
160. Moerman, T. et al. GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).
161. Deshpande, A., Chu, L. -F., Stewart, R. & Gitter, A. Network inference with Granger causality ensembles on single-cell transcriptomics. *Cell Rep.* **38**, 110333 (2022).
162. Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* **7**, S7 (2006).
163. Chan, T. E., Stumpf, M. P. H. & Babtie, A. C. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell Syst.* **5**, 251–267 (2017).
164. Qiu, X. et al. Inferring causal gene regulatory networks from coupled single-cell expression dynamics using Scribe. *Cell Syst.* **10**, 265–274 (2020).
165. Faith, J. J. et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* **5**, e8 (2007).
166. Sanchez-Castillo, M., Blanco, D., Tienda-Luna, I. M., Carrion, M. C. & Huang, Y. A Bayesian framework for the inference of gene regulatory networks from time and pseudo-time series data. *Bioinformatics* **34**, 964–970 (2018).
167. Yu, J., Smith, V. A., Wang, P. P., Hartemink, A. J. & Jarvis, E. D. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* **20**, 3594–3603 (2004).
168. Dojer, N., Bednarz, P., Podsiadlo, A. & Wilczynski, B. BNFinder2: faster Bayesian network learning and Bayesian classification. *Bioinformatics* **29**, 2068–2070 (2013).
169. Wilczyński, B. & Dojer, N. BNFinder: exact and efficient method for learning Bayesian networks. *Bioinformatics* **25**, 286–287 (2009).
170. Woodhouse, S., Piterman, N., Wintersteiger, C. M., Göttgens, B. & Fisher, J. SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. *BMC Syst. Biol.* **12**, 59 (2018).
171. Yuan, Y. & Bar-Joseph, Z. Deep learning of gene relationships from single cell time-course expression data. *Brief. Bioinform.* **22**, bbab142 (2021).
172. Theodoris, C. V. et al. Transfer learning enables predictions in network biology. *Nature* **618**, 616–624 (2023).
173. Polychronidou, M. et al. Single-cell biology: what does the future hold? *Mol. Syst. Biol.* **19**, e11799 (2023).
174. Matsumoto, H. et al. SCODE: an efficient regulatory network inference algorithm from single-cell RNA-seq during differentiation. *Bioinformatics* **33**, 2314–2321 (2017).
175. Aubin-Frankowski, P. -C. & Vert, J. -P. Gene regulation inference from single-cell RNA-seq data with linear differential equations and velocity inference. *Bioinformatics* **36**, 4774–4780 (2020).
176. Wolock, S. L., Lopez, R. & Klein, A. M. Scrublet: computational identification of cell doublets in Single-cell transcriptomic data. *Cell Syst.* **8**, 281–291 (2019).
177. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: doublet detection in single-cell RNA sequencing data using artificial nearest neighbors. *Cell Syst.* **8**, 329–337 (2019).
178. Bais, A. S. & Kostka, D. scds: computational annotation of doublets in single-cell RNA sequencing data. *Bioinformatics* **36**, 1150–1158 (2020).
179. Fleming, S. J. et al. Unsupervised removal of systematic background noise from droplet-based single-cell experiments using CellBender. *Nat. Methods* **20**, 1323–1335 (2023).
180. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *Gigascience* **9**, giaa151 (2020).

181. Mohammad, N. S., Nazli, R., Zafar, H. & Fatima, S. Effects of lipid based Multiple Micronutrients Supplement on the birth outcome of underweight pre-eclamptic women: a randomized clinical trial. *Pak. J. Med. Sci. Q.* **38**, 219–226 (2022).
182. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
183. Merritt, C. R. et al. Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat. Biotechnol.* **38**, 586–599 (2020).
184. Zhao, E. et al. Spatial transcriptomics at subspot resolution with BayesSpace. *Nat. Biotechnol.* **39**, 1375–1384 (2021).
185. Wang, F. et al. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).
186. Qian, X. et al. Probabilistic cell typing enables fine mapping of closely related cell types in situ. *Nat. Methods* **17**, 101–106 (2020).
187. Wang, X. et al. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* **361**, eaat5691 (2018).
188. Janesick, A. et al. High resolution mapping of the breast cancer tumor microenvironment using integrated single cell, spatial and in situ analysis of FFPE tissue. Preprint at *bioRxiv* https://doi.org/10.1101/2022.10.06.510405 (2022).
189. Langseth, C. M. et al. Single cell-resolution in situ sequencing elucidates spatial dynamics of multiple sclerosis lesion and disease evolution. Preprint at *bioRxiv* https://doi.org/10.1101/2023.06.29.547074 (2023).
190. Missarova, A. et al. geneBasis: an iterative approach for unsupervised selection of targeted gene panels from scRNA-seq. *Genome Biol.* **22**, 333 (2021).
191. Nelson, M. E., Riva, S. G. & Cvejic, A. SMaSH: a scalable, general marker gene identification framework for single-cell RNA-sequencing. *BMC Bioinformatics* **23**, 328 (2022).
192. Borm, L. E. et al. Scalable in situ single-cell profiling by electrophoretic capture of mRNA using EEL FISH. *Nat. Biotechnol.* **41**, 222–231 (2023).
193. Fischer, D. S., Schaar, A. C. & Theis, F. J. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.* **41**, 332–336 (2023).

Acknowledgements

We thank I. Adameyko, R. Awatramani, T. Bakken, H. S. Bateup, A. Bhaduri, C. R. Cadwell, E. Caglayan, J. L. Chen, C. Chhatbar, M. G. Filbin, D. Gate, J. Gillis, H. Hochgerner, M. Kampmann, C. N. Kim, F. Krienen, J. Krull, G. La Manno, S. Marsh, M. Monje, Q. Li, Sten Linnarsson, Q. Ma, C. Mayer, V. Menon, P. Nano, M. R O'Dea, R. Patani, A. Pollen, M. Prinz, S. Quake, F. J. Quintana, M. Scavuzzo, M. Schmitz, S. Sloan, P. Tesar, J. Tollkuhn, M. Antonietta Tosches, M. E. Urbanek, C. Walsh, J. Werner and J. Yang for the insightful feedback on this work. C.-B.G. was supported by the Swedish Research Council (grant 2019-01360), Knut and Alice Wallenberg Foundation (grants 2019-0107 and 2019-0089), The Swedish Brain Foundation (FO2023-0032), The Swedish Society for Medical Research (SSMF, grant JUB2019) and the Göran Gustafsson Foundation for Research in Natural Sciences and Medicine. We acknowledge National Institutes of Health (NIH) grants U01AG072573, P01AG073082 and UM1HG012076 (to M.R.C.) and U01DA052713, RF1AG079557, RF1NS128908 and R01AG079291 (to Y.S.). We also acknowledge NIH awards R01MH125516, R01NS123263, R01MH128364 and U01MH130962, the Esther A & Joseph Klingenstein Fund (to T.J.N.), the Shurl and Kay Curci Foundation (T.J.N.) and the Sontag Foundation (to T.J.N.) and a gift from the William K. Bowes Jr Foundation. T.J.N. is a New York Stem Cell Foundation Robertson Neuroscience Investigator. We acknowledge awards from the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Human Cognition (220020467), the Simons Foundation (947591), NHGRI (HG011641), NINDS (NS115821, NS126143) and NIMH (MH126481, MH103517) to G.K. M.H. received a JPB Foundation Picower Institute Innovation Fund award. Work in the group of B.B. was supported by the Helmholtz Center Munich, DFG priority program SPP2202 (BO 5516/1-1), ERA-NET Neuron (MOSAIC) and European Research Council Consolidator grant to B.B. (EpiCortex, 101044469). M.N. acknowledges support from the Israel Science Foundation (grant no. 1079/21), and the European Union (ERC, DecodeSC, 101040660). O.A.B. acknowledges Wellcome Sanger core and Wellcome LEAP funding. J.F. acknowledges support from the National Institute of General Medical Sciences of the NIH under award number R35-GM142889. We acknowledge NINDS Intramural funds through 1ZIA NS003153 (to A.J.L.). K.R.M. acknowledges support from the Lieber Institute for Brain Development and NIH (R01DA055823). L.S. acknowledges support from the European Research Council (ERC StG 'DecOmPress', 950584), the National Multiple Sclerosis Society (PA-2022-36405 and RFA-2203-39300) and the German Research Foundation through collaborative research projects (SPP 2395, FOR 2690 and GRK 2727), individual research grants (SCHI 1330/4-1 and SCHI 1330/10-1) and a Heisenberg Fellowship (SCHI 1330/6-1). N.H. acknowledges support from the Israel Science Foundation research grant no. 1709/19 and the European Research Council grant 853409. S.A.L. acknowledges support from the NIH (R01EY033353), the Cure Alzheimer's Fund, the Belfer Neurodegeneration Consortium, HHMI Emerging Pathogens Initiative and the Carol and Gene Ludwig Family Foundation.

Competing interests

S.A.L. declares a financial interest in AstronauTx and Synapticure. All other authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Ozgun Gokce, Genevieve Konopka, Shane Liddelow, Evan Macosko, Omer Bayraktar, Naomi Habib or Tomasz J. Nowakowski.

Peer review information *Nature Neuroscience* thanks Christine Cheng and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

[1]Helmholtz Pioneer Campus, Helmholtz Zentrum München, Neuherberg, Germany. [2]Physiological Genomics, Biomedical Center, Ludwig-Maximilians-Universität München, Munich, Germany. [3]Laboratory of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [4]The Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Moleculent, Stockholm, Sweden. [6]Gladstone Institute of Neurological Disease, San Francisco, CA, USA. [7]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA. [8]Department of Neurology, University of California, San Francisco, San Francisco, CA, USA. [9]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA. [10]Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA. [11]The Picower Institute for Learning and Memory, MIT, Cambridge, MA, USA. [12]UCL Queen Square Institute of Neurology, University College London, London, UK. [13]Institute for the Advanced Study of Human Biology

(WPI-ASHBi), Kyoto University, Kyoto, Japan. [14]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, Cambridge, MA, USA. [15]Spinal Circuits and Plasticity Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA. [16]Institute of Computational Biology, Helmholtz Center Munich – German Research Center for Environmental Health, Neuherberg, Germany. [17]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge, UK. [18]Department of Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA. [19]Lieber Institute for Brain Development, Baltimore, MD, USA. [20]Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD, USA. [21]Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD, USA. [22]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel. [23]Racah Institute of Physics, The Hebrew University of Jerusalem, Jerusalem, Israel. [24]Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel. [25]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA. [26]Bakar Computational Health Sciences Institute, San Francisco, CA, USA. [27]New York Genome Center, New York, NY, USA. [28]Center for Genomics and Systems Biology, New York University, New York, NY, USA. [29]Department of Neurology, Mannheim Center for Translational Neuroscience, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [30]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. [31]Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA. [32]The Edmond and Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel. [33]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA. [34]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [35]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [36]German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany. [37]Department of Neurodegenerative Diseases and Geriatric Psychiatry, University Hospital Bonn, Bonn, Germany. [38]Department of Neuroscience, UT Southwestern Medical Center, Dallas, TX, USA. [39]Peter O'Donnell Jr. Brain Institute, UT Southwestern Medical Center, Dallas, TX, USA. [40]Neuroscience Institute, NYU Grossman School of Medicine, New York, NY, USA. [41]Department of Neuroscience & Physiology, NYU Grossman School of Medicine, New York, NY, USA. [42]Parekh Center for Interdisciplinary Neurology, NYU Grossman School of Medicine, New York, NY, USA. [43]Department of Ophthalmology, NYU Grossman School of Medicine, New York, NY, USA. [44]Department of Neurobiology, Harvard Medical School, Boston, MA, USA. [45]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [46]Wellcome Sanger Institute, Hinxton, UK. [47]Department of Neurological Surgery, University of California, San Francisco, San Francisco, CA, USA. [48]Department of Anatomy, University of California, San Francisco, San Francisco, CA, USA. [49]Department of Psychiatry and Behavioral Sciences, University of California, San Francisco, San Francisco, CA, USA. [50]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, USA. ✉e-mail: ozguengoekce@gmail.com; genevieve.konopka@utsouthwestern.edu; shane.liddelow@nyulangone.org; emacosko@broadinstitute.org; ob5@sanger.ac.uk; naomi.habib@mail.huji.ac.il; tomasz.nowakowski@ucsf.edu