ANNUAL REVIEWS

Annual Review of Neuroscience

Spatial Transcriptomics: Molecular Maps of the Mammalian Brain

Cantin Ortiz,[1,2,*] Marie Carlén,[1] and Konstantinos Meletis[1,*]

[1]Department of Neuroscience, Karolinska Institutet, Stockholm 171 77, Sweden; email: dinos.meletis@ki.se

[2]Current affiliation: Department of Neuroscience, Institut Pasteur, 75015 Paris, France; email: cantin.ortiz@pasteur.fr

Annu. Rev. Neurosci. 2021. 44:547–62

First published as a Review in Advance on April 29, 2021

The *Annual Review of Neuroscience* is online at neuro.annualreviews.org

https://doi.org/10.1146/annurev-neuro-100520-082639

*These authors contributed equally to this article

Keywords

brain, atlas, gene expression, spatial transcriptomics, neuroanatomy

Abstract

Maps of the nervous system inspire experiments and theories in neuroscience. Advances in molecular biology over the past decades have revolutionized the definition of cell and tissue identity. Spatial transcriptomics has opened up a new era in neuroanatomy, where the unsupervised and unbiased exploration of the molecular signatures of tissue organization will give rise to a new generation of brain maps. We propose that the molecular classification of brain regions on the basis of their gene expression profile can circumvent subjective neuroanatomical definitions and produce common reference frameworks that can incorporate cell types, connectivity, activity, and other modalities. Here we review the technological and conceptual advances made possible by spatial transcriptomics in the context of advancing neuroanatomy and discuss how molecular neuroanatomy can redefine mapping of the nervous system.

Contents

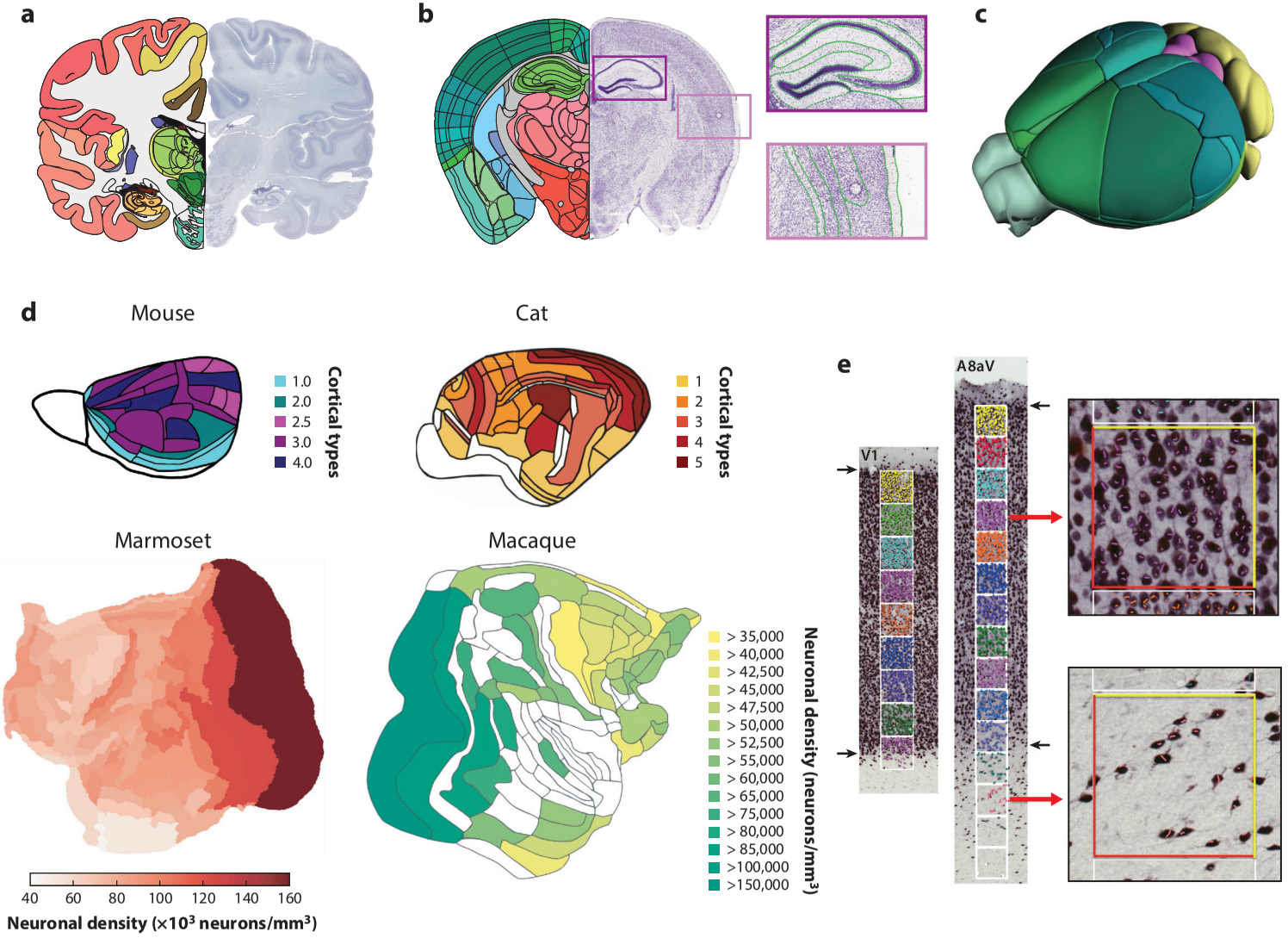
INTRODUCTION

We review here how spatial transcriptomics (ST) can transform the mapping of the nervous systems of animals and humans. Today, advances in molecular biology have formed a new wave of methods that reveal the spatial and molecular organization of the nervous system based on mapping gene expression at a large scale (e.g., whole brain) and at subcellular resolution. Neuroanatomy is an investigation into structure–function relationships in the nervous system and in essence aims to develop a formalized system to describe the spatial division of the brain into discrete regions that represent biologically or functionally meaningful descriptions.

The main principles that guide the mapping of tissue organization in general and of the nervous system in particular can be traced back to the comparative biology approach developed by Aristotle in antiquity and to the fundamental neuroanatomy of Santiago Ramón y Cajal at the turn of the twentieth century. The concepts and definitions widely used today to establish brain atlases stem from the first attempts to annotate the human cerebral cortex based on differences in tissue organization (i.e., cytoarchitecture), which are visible using light microscopy, an approach pioneered and refined a century ago by Brodmann, von Economo, and Koskinas.

It is worth noting that the actual definition of a brain region or area is surprisingly opaque, and variations in experimental and classification approaches have resulted in conflicting atlases. Furthermore, the natural variability in brain structure between individuals complicates the establishment of strict neuroanatomical definitions, raising the question of what features should be considered to define a brain region. In line with this, the establishment of the type of information (i.e., modality) and level of detail (i.e., resolution) relevant for accurate and reproducible description of the tissue organization is a central challenge in brain mapping. Four characteristics have been suggested to best define brain regions: function, architecture, connectivity, and topography (Van Essen & Glasser 2018). Maps of the primate visual areas illustrate the challenges of neuroanatomical definitions, since the tissue architecture and functional signals form fluid borders (**Figure 1**).

These challenges are also reflected in uncertainties regarding the parcellation of the human cerebral cortex, with suggestions that there are 50–200 discrete brain regions (Fischl et al. 2004,



(*Caption appears on following page*)

Figure 1 (*Figure appears on preceding page*)

Maps of the mammalian brain. (*a*) A coronal section of the human brain showing anatomical delineations (*left hemisphere*) and Nissl staining of the tissue (*right hemisphere*). Panel *a* adapted from Ding et al. (2016) and the Allen Adult Human Brain Atlas (**http://atlas.brain-map.org/**). (*b*) A coronal section of the mouse brain showing anatomical delineations (*left hemisphere*) and Nissl staining of the tissue (*right hemisphere*). Boxes show magnification of the stained tissue, with outline (*green*) of subregions in the hippocampal region (*dark purple box*) and the isocortex (*light purple box*). Panel *b* adapted from Lein et al. (2007) and the Allen Adult Mouse Brain Atlas (**http://atlas.brain-map.org/**). (*c*) 3D rendering of anatomical delineations of the adult mouse brain. Panel *c* adapted from Wang et al. (2020) and the Allen Adult Mouse Brain Atlas (**http://connectivity.brain-map.org/3d-viewer**). (*d*) Anatomical delineation of the isocortex in different mammalian species based on various tissue characteristics. Image of mouse brain adapted with permission from Goulas et al. (2017), image of cat brain adapted from Beul et al. (2015), image of marmoset brain adapted with permission from Atapour et al. (2019), and image of macaque brain adapted from Beul et al. (2017). (*e*) Example of the sampling method for stereological analysis of tissue composition, here NeuN-stained sections from primary visual cortex (V1) and ventral subdivision of area 8a (A8aV) in the marmoset cortex. Boxes show magnification of the cell-counting method for calculation of neuronal density, which is used, for example, for the estimation of areal boundaries. Panel *e* adapted with permission from Atapour et al. (2019).

Glasser et al. 2016). Recent human brain atlases incorporate multimodal definitions of brain regions, based on features defined by histology and connectivity, and probabilistic models (Amunts et al. 2020, Ding et al. 2016, Glasser et al. 2016).

Mouse and rat brain atlases have been constructed using basic histochemical tissue staining (e.g., Nissl, acetylcholinesterase) for outlining anatomical structures (Lein et al. 2007, Paxinos & Franklin 2004, Swanson 2018) and are sometimes combined with measurements of tissue structure and composition using MRI (Ma et al. 2005, MacKenzie-Graham et al. 2004). The most recent mouse brain reference atlas from the Allen Institute for Brain Science [the Allen mouse brain common coordinate framework version 3 (CCFv3)] outlines 658 individual regions parcellated in 3D. The anatomical template is of unprecedented detail (isotropic resolution of 10 μm) and represents the population averages of 1,675 specimens (mice) (Wang et al. 2020). The template was constructed by imaging tissue autofluorescence of a very large number of sections, and multiple reference data sets were overlaid with the average template to generate the final CCFv3 reference atlas. Importantly, the availability of coordinate frameworks like the CCFv3 enables the scientific community to register image data sets into a reference atlas, for example, for anatomical annotation or comparison of data sets.

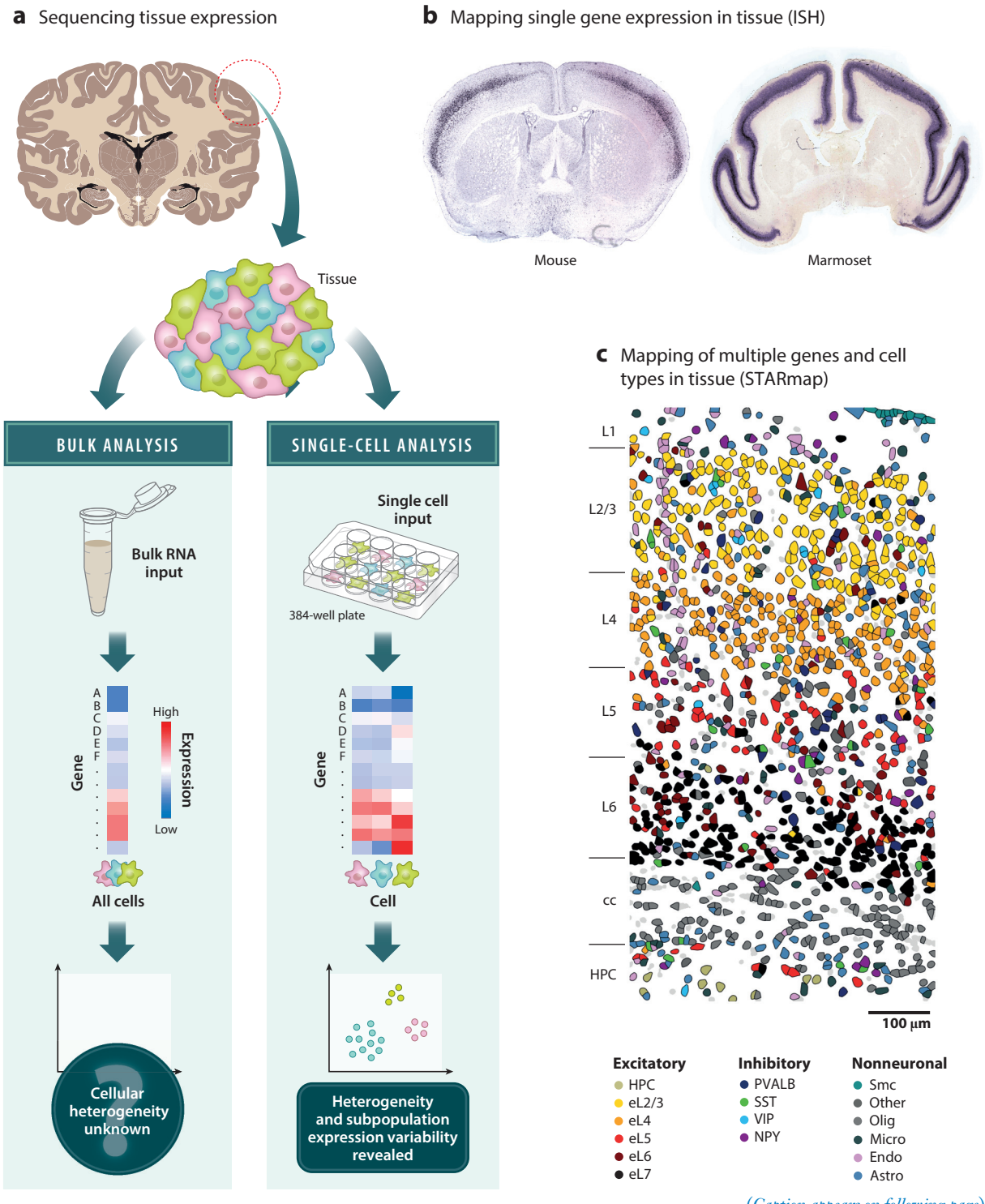
We propose that ST offers a powerful solution to uncover and describe brain organization: It introduces an unbiased classification of spatial territories based on biological signals, opening up new possibilities for comparative neuroanatomy and functional classification of brain circuits.

GENE EXPRESSION IN NEUROANATOMY

Mapping of gene expression in the past decades has transformed our approaches to define cell identify and tissue organization and has introduced a new dimension in neuroanatomy (**Figure 2**). For example, early studies used microarray profiling and histological analysis for generating genome-wide maps of transcript distributions in the adult human brain, mapping the global and regional transcriptional architectures (Hawrylycz et al. 2012), and revealing region-specific gene expression signals in the adult mouse brain (Ng et al. 2009, Sandberg et al. 2000).

Methods to Map Gene Expression in Cells

The development of transcriptomic technologies (i.e., techniques used to study an organism's transcriptome, the set of RNA transcripts), particularly single-cell RNA sequencing (scRNA-seq), has greatly facilitated and accelerated the study of neuron diversity, and advancements in computational tools have transformed how large-scale sequencing data are analyzed (Sandberg 2014,



(Caption appears on following page)

Figure 2 (*Figure appears on preceding page*)

Gene expression redefines neuroanatomy. (*a*) Illustration of RNA sequencing of tissue. Bulk RNAseq analysis reveals the combined gene expression of the cells in the tissue, while single-cell RNAseq can capture the gene expression profile of individual cells and reveal the cellular heterogeneity within a tissue. Regardless, the spatial organization of the cells is lost in both methods. Panel *a* adapted with permission from 10x Genomics (LIT000027 Rev D; **https://pages.10xgenomics.com/3p-getting-started-guide-single-cell-gene-expression.html**), and coronal section of the human brain adapted from Ding et al. (2016) and the Allen Adult Human Brain Atlas (**http://atlas.brain-map.org/**). (*b*) ISH detecting the expression of the *Rorb* gene in the adult mouse brain (*left*) and in the marmoset brain (*right*), showing similar laminar expression patterns in the isocortex. Image of mouse brain adapted from Lein et al. (2007) and the Allen Adult Mouse Brain Atlas (**https://mouse.brain-map.org/**), and image of marmoset brain adapted from Shimogori et al. (2018) and the Marmoset Gene Atlas (**https://gene-atlas.bminds.brain.riken.jp**). (*c*) In situ transcriptomics in intact tissue using STARmap. Gene expression reveals a spatial map of cell types, here excitatory, nonneuronal, and inhibitory cell types in visual cortex. Panel *c* adapted with permission from Wang et al. (2018). Abbreviations: Astro, astrocyte; cc, corpus callosum; eL, excitatory layer; Endo, endothelial cell; HPC, hippocampus; ISH, in situ hybridization; L, layer; Micro, microglia; NPY, neuropeptide Y; Olig, oligodendrocyte; PVALB, parvalbumin; RNAseq, RNA sequencing; Smc, smooth muscle cell; SST, somatostatin; STARmap, spatially resolved transcript amplicon readout mapping; VIP, vasointestinal peptide.

Stegle et al. 2015, Stuart et al. 2019). Studies using scRNA-seq of the adult mouse brain have established a basic catalog of the cell types (Saunders et al. 2018, Zeisel et al. 2018) and also revealed aspects of gene expression that reflect the spatial organization in the brain (Märtin et al. 2019, Saunders et al. 2018). The main limitation of single-cell sequencing approaches is the requirement of tissue dissociation for harvesting single cells, invariably separating spatial and molecular mappings. Accordingly, the characterization of the gene expression profile of single cells is missing information on how the cells were organized in the tissue. The Patch-seq method accounts for the relationship between gene expression and the spatial location of neurons by combining mapping of electrophysiological characteristics, morphology, localization, and gene expression but at the expense of scalability (Cadwell et al. 2016, Fuzik et al. 2016).

Methods to Visualize Gene Expression in Tissue

The power of applying transcriptomic technologies to neuroanatomy lies in the possibility of mapping gene expression signals in the tissue at very high resolution. Inspiration for most methods that map the transcriptome in intact tissue comes from in situ hybridization (ISH) (Singer & Ward 1982), a method that introduced the possibility of visualizing how RNA is distributed in a cell or tissue (i.e., in situ) based on the hybridization of probes to predetermined and selected RNA molecules. The most extensive ISH mapping of the adult mouse brain is compiled in the gene expression atlas generated by the Allen Institute for Brain Science (Lein et al. 2007). This resource catalogs the expression of approximately 20,000 genes in coronal and sagittal sections from the entire adult mouse brain and established a new standard for visualizing and exploring gene expression data (**http://mouse.brain-map.org/**). The massive resources required for whole-brain ISH mapping of several thousand genes limit the possibilities of applying this approach to mapping neuroanatomy in multiple animals or across many experiments.

Over time, increasingly sensitive methods to quantify RNA molecules in tissue have been developed, providing the possibility of mapping the expression of any gene (or multiple genes) at cellular resolution (Asp et al. 2020, Strell et al. 2019). For example, semiquantitative methods with cellular resolution have been introduced with some multiplexing capability [e.g., RNAscope (Wang et al. 2012)]. Further improving in situ mapping, single-molecule fluorescent in situ hybridization (smFISH) detects individual RNA molecules, provides information about the cell-to-cell abundance of transcripts, and builds on imaging of multiple hybridized complementary probes conjugated with fluorescent dye (Femino et al. 1998). The advantage of smFISH is the combination of sensitive detection of RNA with subcellular resolution and the possibility of mapping the spatial distribution of the RNA, resulting in high-resolution maps of how each RNA molecule

is distributed in a tissue. A number of sensitive in situ methods for visualization of gene expression in tissue based on the detection of known RNA sequences have been developed, including in situ sequencing (ISS) (Ke et al. 2013), multiplexed error-robust fluorescence in situ hybridization (MERFISH) (Chen et al. 2015), sequential fluorescence in situ hybridization (seqFISH+) (Eng et al. 2019), ouroboros single-molecule fluorescent in situ hybridization (osmFISH) (Codeluppi et al. 2018), and spatially-resolved transcript amplicon readout mapping (STARmap) (Wang et al. 2018). These methods have in common the sensitive mapping of multiple known targets (i.e., preselected genes). In contrast, fluorescent in situ sequencing (FISSEQ) is based on sequencing amplified probes directly in situ and was developed for transcriptome-wide quantitative visualization of RNA, allowing mapping of unknown RNA sequences (Lee et al. 2014). In the methods mentioned above, multiplexing is achieved by multiple rounds of hybridization and imaging and therefore depends on accurate registration of detected features in a single reference image. The imaging time remains a significant limitation and restricts application of the most sensitive approaches to only small tissue samples (Moffitt et al. 2018). While these limitations could be resolved by using automated microfluidics for tissue and probe handling together with imaging speed, the use of such approaches to comprehensively map cell types and regions at the whole-brain scale remains a major challenge.

SPATIAL TRANSCRIPTOMICS

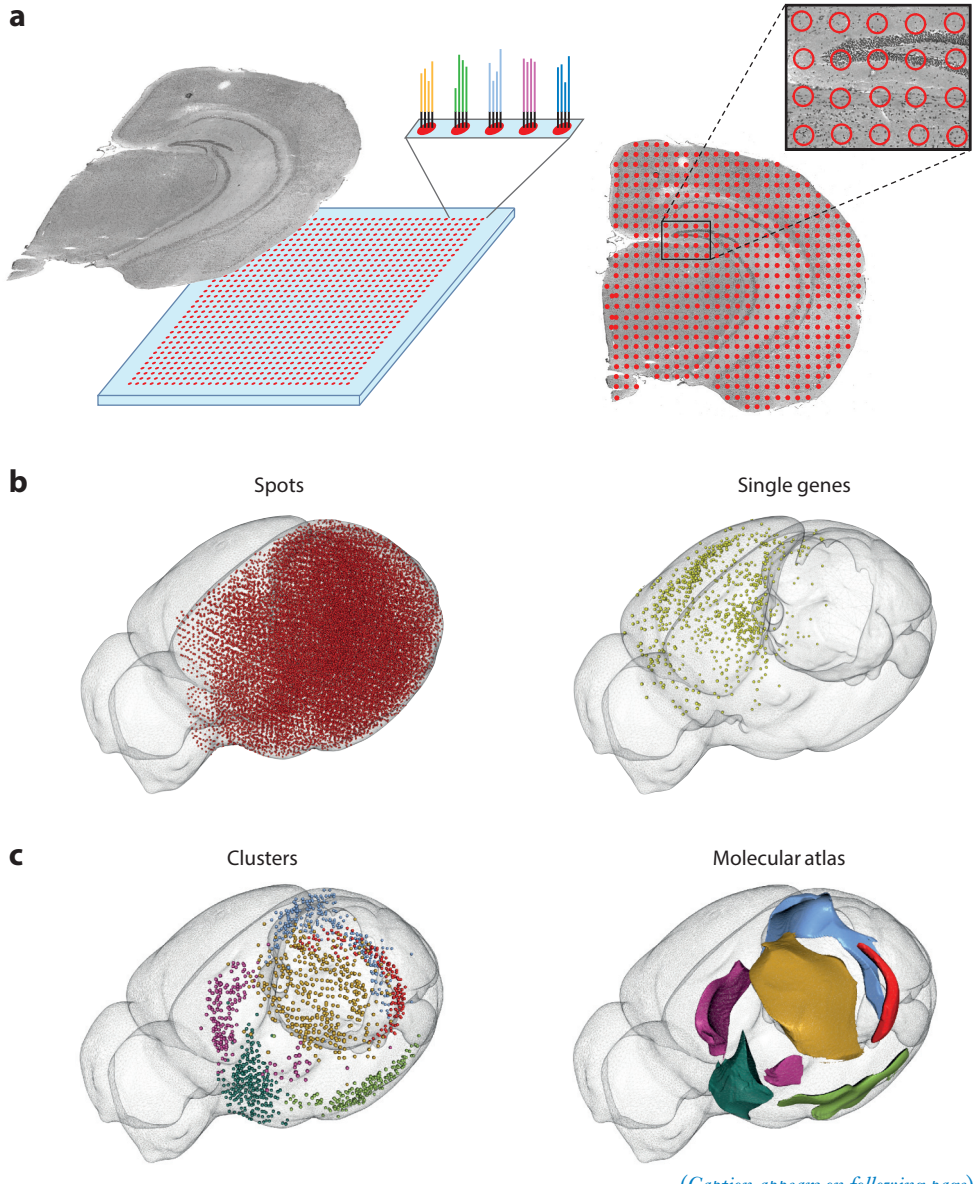
First-Generation Glass Array Spatial Transcriptomics

With advances in mapping gene expression in tissue based on ISH methods, it is possible to map the expression of a selected number of genes at very high resolution. However, transcriptome-wide quantification across a whole brain, for example, the adult mouse brain, requires a method that is scalable and captures gene expression with high spatial resolution. The ST method introduced an innovative approach to quantitatively map gene expression in tissue sections through unbiased high-throughput RNA sequencing (i.e., without selecting candidate genes) at the supracellular scale (Ståhl et al. 2016) (**Figure 3**). ST builds on the development of glass slides with printed spots holding oligonucleotide (capture) probes. The first-generation ST arrays consisted of 100-μm-diameter spots arranged in a rectangular pattern (6.2 mm × 6.6 mm with 1,007 spots), but the spot configuration can be designed according to specific experimental requirements using oligo printing technologies. The probes include an oligo-dT sequence to capture polyadenylated RNAs [messenger RNA (mRNA)], and probes in the same spot are assigned a unique spatial identity with a spatial barcode (i.e., a unique probe sequence). Thin tissue sections are placed on the glass slides with ST spots, and the spots are imaged together with a histochemical stain of the tissue (e.g., hematoxylin and eosin) using bright-field microscopy to map the exact position of the spots relative to tissue landmarks. This allows registration of the spot locations into a common reference framework. The ST spots capture the RNA in the tissue, and after in situ complementary DNA (cDNA) synthesis, cDNA-RNA hybrids (including barcodes) are cleaved off the glass slides. The unique barcodes can then be identified in the sequencing libraries and are used to identify the location in tissue (i.e., ST spot) where the RNA was captured. A key advantage of the ST method is that the probes capture all polyadenylated RNAs, establishing ST as a uniquely unbiased approach to tissue transcriptome mapping.

Advancing Spatial Transcriptomics

An important step in improving ST is to increase both sensitivity (i.e., genes detected and dynamic range) and resolution (i.e., size of spot). The commercially available Visium array (10x Genomics)

offers an improved spatial resolution, allowing detection of gene expression signals from smaller spots (55 μm in diameter). In an attempt to approach single-cell resolution, Slide-seq (Rodriques et al. 2019) and high-density spatial transcriptomics (HDST) (Vickovic et al. 2019) have been developed based on barcoded beads deposited on a surface. Slide-seq and HDST offer an improvement in cellular resolution compared to ST due to the small bead size; nevertheless, this seems to come at the cost of reduced detection efficiency (Asp et al. 2020). To circumvent the reduction in detection efficiency, it is possible to computationally bin beads together to form virtual spatial features at the expense of spatial resolution. An improved Slide-seq version (Slide-seq2) was developed by increasing the sensitivity by a factor of 10 and reaching 550 median transcripts



(*Caption appears on following page*)

Figure 3 (*Figure appears on preceding page*)

Spatial transcriptomics (ST) to map the molecular organization of the adult mouse brain. (*a*, *left*) A single coronal tissue section from the adult mouse brain is placed on a ST array. Each ST spot (*red*) holds barcoded oligonucleotide (*capture*) probes (*black vertical bars*). The probes include an oligo-dT sequence that captures polyadenylated RNAs (messenger RNA, *colored vertical bars*), and probes with the same barcode are located in the same spot, providing the probes with a spatial identity. (*Right*) The spots are imaged together with a histochemical stain of the tissue (e.g., hematoxylin and eosin) using bright-field microscopy to map the exact position of the spots relative to tissue landmarks. The magnification shows each spot capturing the RNA of multiple cells. (*b*, *left*) 3D rendering of all spots in 75 coronal sections, registered into a common reference framework. (*Right*) Example of visualization of the spatial expression of a single gene. Spots show areas with high expression of *Rorb* (top 1% of the spots with *Rorb* expression across the data set). (*c*, *left*) Six clusters of spots (color coded) identified by their respective gene expression pattern. (*Right*) The molecularly defined clusters can be used to delineate the spatial organization of the brain and for the ultimate establishment of a brain atlas. Data used to generate figure from Ortiz et al. (2020).

per bead while maintaining high spatial resolution using the same bead diameter (Stickels et al. 2020). The accuracy and sensitivity are sufficient to detect mRNA transcripts localized to subcellular domains (e.g., the dendrite of a neuron). Finally, in addition to improving resolution, it will be important to increase the size of the ST arrays for use in mammalian species with larger brains, as the dimensions of commercially available ST arrays can maximally fit one coronal section of one hemisphere of the adult mouse brain per array (6.5 mm $\times$ 6.5 mm). Scaling the ST technology to match the dimension of human brain regions such as the human cerebral cortex would require extending the array surface by at least two orders of magnitude.

COMPUTATIONAL METHODS IN SPATIAL TRANSCRIPTOMICS

Data Analysis Pipelines

ST experiments can generate very large sequencing data sets, introducing the same computational challenges as large-scale single-cell RNA sequencing (scRNA-seq) experiments. Since ST and scRNA-seq data are similarly structured, the same methods can be applied by inputting gene expression from single spots instead of single cells. Standardized computational pipelines such as Seurat (Satija et al. 2015) and Scanpy (Wolf et al. 2018) are widely used to analyze and explore scRNA-seq data. These pipelines have been expanded to analyze and visualize ST data. In addition, dedicated ST pipelines have been developed for sequencing analysis together with visualization of spatial information (Dries et al. 2020, Fernández Navarro et al. 2019). To integrate multiple tissue sections into a common reference space, several semiautomated pipelines can be used to register brain sections into preexisting atlases, also allowing for further manual annotation (Fürth et al. 2018, Niedworok et al. 2016, Renier et al. 2016). In the following sections, we briefly present the key steps for the analysis of sequencing data from ST experiments.

Data Normalization

Gene expression experiments often include technical variability that interferes with the quantification of biological signals (Hicks et al. 2018). Therefore, a critical first step in the ST analysis workflow is to normalize the data by transforming the absolute gene counts into a new expression that is normalized across the whole data set. Importantly, the aim of normalization is to correct the technical bias while preserving meaningful biological signals. A broad range of methods correcting different sources of technical noise are available and being updated to improve this step (Tran et al. 2020, Vallejos et al. 2017).

An important source of technical noise is the differences in sequencing depth between samples, which results in differences in the number of detected reads per spot. Relative counts normalization is the most straightforward solution to account for this variability and simply consists of dividing each gene count by the total number of reads per spot. Technical noise also stems from the systematic difference in expression levels of tissue sections that have been processed in different batches (Tung et al. 2017). Batch normalization algorithms can be applied to remove systematic differences in gene expression between batches (Haghverdi et al. 2018, Polański et al. 2020). Finally, the difference in the number of reads detected per gene can partially be explained by the fact that the number of reads also depends on transcript length (Oshlack & Wakefield 2009). To compensate for this bias, gene counts are usually converted into reads per kilobase per million reads (Mortazavi et al. 2008) or transcripts per million (Wagner et al. 2012). This normalization does not remove biases in downstream analyses that tend to overemphasize the contribution of highly expressed genes, a bias that can be reduced by applying a scaling factor and a logarithmic function (Satija et al. 2015). For genes that are sparsely or lowly expressed, showing a high incidence of missing values (zeroes) in the data set, it is challenging to determine whether missing values represent selective expression, missing data, or low sensitivity (Hicks et al. 2018). Some algorithms such as scran (Lun et al. 2016) or SCnorm (Bacher et al. 2017) implement strategies to correct for zero values arising due to low sensitivity.

Dimensionality Reduction

After normalization of gene expression, it is necessary to reduce the data set into a compact representation that preserves the key biological information. This dimensionality reduction is necessary since downstream analyses require a smaller number of dimensions and cannot handle data structures holding many thousands of genes (Beyer et al. 1999). Dimensionality reduction replaces the expression levels across all genes with the expression on a reduced set of new variables.

The most common dimensionality reduction method is principal component analysis (PCA) (Abdi & Williams 2010). PCA calculates a reduced set of components that replace the large number of genes while capturing the largest possible degree of variance. As a result, a lower-dimensional representation of the gene expression data is captured by a small number of PCA components that explain most of the data variance. PCA forces each component to be orthogonal to all others to ensure that the variance accounted for by one component is not captured by any other component. Instead, independent component analysis (ICA) is a method for dimensionality reduction where the calculated components represent the independent sources of variance in the data. Compared to PCA, ICA relaxes the orthogonality requirement and allows shared variance across the dimensions (Hyvarinen 1999). ICA tends to generate more interpretable components, allowing researchers to differentiate components carrying technical signals from those representing biological phenomena. As a result, the biological components can be specifically selected for further processing, which also helps in reducing batch effects. Among its broad range of applications, ICA has been successfully used to analyze various types of gene expression data from microarrays (Lee & Batzoglou 2003), scRNA-seq (Saunders et al. 2018), and ST (Ortiz et al. 2020).

Dimensionality reduction can also be used to project high-dimensional data directly into 2D or 3D space for visualization. Nonlinear methods perform better compared to linear dimensionality-reduction methods (e.g., PCA) to position data points that share similar molecular signatures in close distance in 2D or 3D plots. The disadvantage is that coordinates in the low-dimensional space are harder to interpret and can lead to erroneous conclusions regarding the presence of

clusters, their size, or the biological significance of relative distances (Wattenberg et al. 2016). The most popular nonlinear algorithms are t-stochastic neighbor embedding (t-SNE) (van der Maaten & Hinton 2008) and uniform manifold approximation and projection (UMAP) (McInnes et al. 2018). UMAP has been suggested as a superior method compared to t-SNE for sequencing data, especially in its ability to preserve the global data organization and in terms of execution time (Becht et al. 2019). This conclusion has been challenged by authors arguing that the difference is mostly due to suboptimal use of t-SNE (Kobak & Berens 2019).

Clustering

In order to identify brain regions that share common molecular signatures, clustering algorithms can be applied to the PCA- or ICA-calculated components to identify ST spots that show similar gene expression patterns. A broad range of clustering methods are available, although most of them are built on top of a few core algorithms. Among them, *k*-means is probably the most popular clustering technique (Lloyd 1982), and it is suitable for very large sequencing data sets since the algorithm scales linearly with the amount of data. A disadvantage of *k*-means clustering is the tendency to identify equal-sized clusters (Kiselev et al. 2019), which could lead to the misclassification of clusters formed by data from a few spots.

Another widely used method is hierarchical clustering, based on establishing a hierarchy between spots. A major benefit of using hierarchical clustering is that results can be graphically represented as dendrograms, which offer an intuitive and compact visualization of the data set. However, the algorithm has a quadratic complexity in terms of both time and memory requirements, making it computationally impractical for very large data sets (Day & Edelsbrunner 1984).

Another alternative is graph-based clustering, where sequencing data sets of up to tens of millions of data points can be represented as a graph of interconnected nodes (Waltman & van Eck 2013). Clusters are identified in the graph by detecting groups of highly interconnected nodes (Blondel et al. 2008). In graph-based clustering, the number of clusters is not directly imposed. Instead, the number of clusters is adjusted mostly via two parameters: *k*, which is used for constructing the graph, and the resolution, which is an important setting for the community detection. A possible disadvantage of this approach is its applicability to small data sets and the identification of rare clusters (Fortunato & Barthélemy 2007). Nevertheless, graph-based clustering was able to identify small spatial clusters in a whole-brain ST data set, with some clusters consisting of less than 0.1% of the total number of spots (Ortiz et al. 2020). Ultimately, and regardless of the algorithm applied, the clustering approach offers a conversion of the data into possibly biologically meaningful groups and represents an important first step toward converting the ST data into applicable knowledge.

Downstream Analysis and Applications

Following the identification of molecular clusters in ST data, it is possible to investigate the gene expression differences that drive the assignment of spots to specific clusters. Differential expression analysis (DEA) is a powerful tool to systematically detect genes whose expression levels differ between multiple groups of spots (Anders & Huber 2010, Robinson & Oshlack 2010). Using DEA, one can apply specific selection criteria, for example, the anatomical location of spots or their association with a selected tissue marker, to compare gene expression between subsets of spots. For instance, ST was used to identify a subset of genes associated with the formation of Aβ plaques in an animal model of Alzheimer's disease by grouping spots according to the density of Aβ (Chen et al. 2020).

Another application of ST is to support computational methods that aim to recreate and predict the spatial organization of isolated single cells based on their gene expression (e.g., from scRNA-seq experiments). Importantly, accurate predictions of spatial identity rely heavily on the access to a preexisting tissue reference map outlining spatial gene expression patterns (Achim et al. 2015, Karaiskos et al. 2017, Nitzan et al. 2019, Satija et al. 2015). Offering a complete solution, the spatially defining gene expression patterns found in ST maps can form the basis for improving computational predictions on the spatial origin of single cells (Ortiz et al. 2020).

SPATIAL TRANSCRIPTOMICS IN NEUROANATOMY

A data-driven approach to define brain organization, void of potential biases introduced by expert-based definition, would be an ideal framework for establishing brain reference atlases. The potential of a molecular-based classification of the brain was demonstrated in a recent study, which used unbiased ST mapping of gene expression to develop the concept of a data-driven molecular atlas (Ortiz et al. 2020). In this study, a whole-brain molecular atlas was built by capturing the expression of approximately 15,000 genes from 34,000 ST spots across 75 coronal tissue sections. The position of each spot in the tissue was determined by imaging all sections followed by image registration in a common coordinate framework of the mouse brain (CCFv3), resulting in a whole-brain view of the expression of each gene in every spot. To reveal the relationship between the molecular and neuroanatomical organization in the tissue, unsupervised algorithms were used to identify spots with a shared molecular profile and the molecular clusters were thereafter visualized in the reference coordinate system. To facilitate 3D visualization of the molecular clusters, machine learning was used to define continuous volumes based on the identity and distribution of the spots, generating a 3D atlas with molecularly defined brain regions. The 3D atlas allows for further visualization of 2D sections of any virtual angle. Altogether, the computational analyses unbiasedly classified the adult mouse brain into 181 unique molecular domains (**https://www.molecularatlas.org/**). Supporting the value of the ST-based classification in generating new atlases, the molecular domains showed a high level of agreement with current structure-based reference atlases and, importantly, also established new and finer subdivisions in several brain areas, including the isocortex and striatum. For example, the molecular atlas revealed a new classification of subregions in the dorsal striatum and provided evidence for the separate molecular identity of different parts of layer 2/3 in isocortex. Furthermore, the molecular atlas suggests a new classification of isocortical subregions in the anteroposterior and mediolateral dimension. The molecular annotation of isocortical domains showed that gene expression defines the isocortical layers in a domain-specific manner, thereby providing a molecular identity that identifies both layer and position (e.g., distinguishing layer 6 in frontal versus posterior domains).

It has become clear from scRNA-seq experiments that neuron identity cannot be accurately defined by the expression of a single gene, and instead the expression of several genes must be combined to capture the diversity of neuron subtypes. Similarly, even if expression of single genes can reveal some aspect of brain organization, it is unlikely that the full neuroanatomical complexity can be represented with a small number of genes. The full molecular atlas was based on the expression of the 7,000 most variable genes, demonstrating the power of gene expression as a tool to map the adult brain, and the number of genes is beyond what is feasible to map using cellular resolution methods such as STARmap or MERFISH. As a possible solution, Ortiz et al. (2020) showed that most of the whole-brain neuroanatomical classification details could be recapitulated using only a small subset of genes by demonstrating classification of the brain into 181 molecular clusters using a restricted set of 266 spatially defining genes (i.e., the brain palette). This suggests that the whole-brain ST data are a valuable resource for extraction of region-specific palettes,

palettes that could guide the establishment of regional identities using in situ expression mapping of cellular resolution. In addition, spatially defining genes are also important to machine learning approaches used to predict the spatial origin of cell types identified by single-cell transcriptomics.

In summary, the first application of whole-brain ST clearly demonstrates the potential of data-driven molecular neuroanatomy on the whole-brain scale, and it serves as a roadmap toward the unbiased classification of mammalian brain organization.

CONCLUSION

Pioneering work used cyto- and chemoarchitecture to establish the main principles of tissue organization in the nervous system, and the concepts emerging from neuroanatomical investigations have deeply influenced how we study the brain. Today, we are witnessing an evolution of neuroanatomical definitions and an intense investigation of the modalities that can optimally capture principles of brain organization in a reference map. This new era capitalizes on high-resolution molecular tools that enable large-scale mapping of markers, which, in an unbiased and unsupervised fashion, capture the complexity of the nervous system. In this vein, ST offers a roadmap to define and study brain regions in any species, importantly, independent of prior neuroanatomical knowledge. ST gives a new view of brain organization but does not provide in itself a final map. It will be essential to develop experimental and computational tools that integrate different definitions of tissue organization into a single reference map of the brain.

For future efforts, it will be important to consider how detailed the definitions need to be to best serve the purpose of developing neuroanatomical frameworks that support understanding of brain function. We need to consider at what point the addition of an increasing number of borders and small subregions benefits the functional investigation and understanding. For example, it remains unclear whether mapping single cells in high detail on a whole-brain scale in tissue is necessary to establish a neurobiologically relevant and reproducible subdivision of brain regions.

We expect that spatially restricted gene expression signatures are conserved across mammalian species, and the ST approach can guide the development of new atlases in several species, including the human brain. The large-scale imaging of RNA distribution in intact brain tissue will be an important challenge to solve in the coming years, as current state-of-the-art methods realistically are restricted to imaging small tissue volumes. The combination of a molecular atlas (Ortiz et al. 2020) with high-definition, cell type–specific wiring diagrams (Huang et al. 2020) and brain-wide activity maps can inform the building of new reference atlases and provide a united framework for understanding how circuits are organized to shape behavior.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

Abdi H, Williams LJ. 2010. Principal component analysis. *WIREs Comput. Stat.* 2:433–59

Achim K, Pettit J-B, Saraiva LR, Gavriouchkina D, Larsson T, et al. 2015. High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* 33:503–9

Amunts K, Mohlberg H, Bludau S, Zilles K. 2020. Julich-Brain: a 3D probabilistic atlas of the human brain's cytoarchitecture. *Science* 369:988–92

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11:R106

Asp M, Bergenstråhle J, Lundeberg J. 2020. Spatially resolved transcriptomes—next generation tools for tissue exploration. *BioEssays* 42:1900221

Atapour N, Majka P, Wolkowicz IH, Malamanova D, Worthy KH, Rosa MGP. 2019. Neuronal distribution across the cerebral cortex of the marmoset monkey *(Callithrix jacchus). Cereb. Cortex* 29:3836–63

Bacher R, Chu L-F, Leng N, Gasch AP, Thomson JA, et al. 2017. SCnorm: robust normalization of single-cell RNA-seq data. *Nat. Methods* 14:584–86

Becht E, McInnes L, Healy J, Dutertre C-A, Kwok IWH, et al. 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37:38–44

Beul SF, Barbas H, Hilgetag CC. 2017. A predictive structural model of the primate connectome. *Sci. Rep.* 7:43176

Beul SF, Grant S, Hilgetag CC. 2015. A predictive model of the cat cortical connectome based on cytoarchitecture and distance. *Brain Struct. Funct.* 220:3167–84

Beyer K, Goldstein J, Ramakrishnan R, Shaft U. 1999. When is "nearest neighbor" meaningful? In *Database Theory—ICDT'99*, ed. C Beeri, P Buneman, pp. 217–35. Berlin: Springer

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* 2008:P10008

Cadwell CR, Palasantza A, Jiang X, Berens P, Deng Q, et al. 2016. Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* 34:199–203

Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X. 2015. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090

Chen W-T, Lu A, Craessaerts K, Pavie B, Sala Frigerio C, et al. 2020. Spatial transcriptomics and in situ sequencing to study Alzheimer's disease. *Cell* 182:976–91.e19

Codeluppi S, Borm LE, Zeisel A, La Manno G, van Lunteren JA, et al. 2018. Spatial organization of the somatosensory cortex revealed by osmFISH. *Nat. Methods* 15:932–35

Day WHE, Edelsbrunner H. 1984. Efficient algorithms for agglomerative hierarchical clustering methods. *J. Classif.* 1:7–24

Ding S-L, Royall JJ, Sunkin SM, Ng L, Facer BAC, et al. 2016. Comprehensive cellular-resolution atlas of the adult human brain. *J. Comp. Neurol.* 524:3127–481

Dries R, Zhu Q, Dong R, Eng C-HL, Li H, et al. 2020. Giotto, a toolbox for integrative analysis and visualization of spatial expression data. bioRxiv 701680. **https://doi.org/10.1101/701680**

Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, et al. 2019. Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* 568:235–39

Femino AM, Fay FS, Fogarty K, Singer RH. 1998. Visualization of single RNA transcripts in situ. *Science* 280:585–90

Fernández Navarro J, Lundeberg J, Ståhl PL. 2019. ST viewer: a tool for analysis and visualization of spatial transcriptomics datasets. *Bioinformatics* 35:1058–60

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, et al. 2004. Automatically parcellating the human cerebral cortex. *Cereb. Cortex* 14:11–22

Fortunato S, Barthélemy M. 2007. Resolution limit in community detection. *PNAS* 104:36–41

Fürth D, Vaissière T, Tzortzi O, Xuan Y, Märtin A, et al. 2018. An interactive framework for whole-brain maps at cellular resolution. *Nat. Neurosci.* 21:139–49

Fuzik J, Zeisel A, Máté Z, Calvigioni D, Yanagawa Y, et al. 2016. Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat. Biotechnol.* 34:175–83

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, et al. 2016. A multi-modal parcellation of human cerebral cortex. *Nature* 536:171–78

Goulas A, Uylings HBM, Hilgetag CC. 2017. Principles of ipsilateral and contralateral cortico-cortical connectivity in the mouse. *Brain Struct. Funct.* 222:1281–95

Haghverdi L, Lun ATL, Morgan MD, Marioni JC. 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* 36:421–27

Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, et al. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489:391–99

Hicks SC, Townes FW, Teng M, Irizarry RA. 2018. Missing data and technical variability in single-cell RNA-sequencing experiments. *Biostatistics* 19:562–78

Huang L, Kebschull JM, Fürth D, Musall S, Kaufman MT, et al. 2020. BRICseq bridges brain-wide interregional connectivity to neural activity and gene expression in single animals. *Cell* 182:177–88.e27
Hyvarinen A. 1999. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans. Neural Netw.* 10:626–34
Karaiskos N, Wahle P, Alles J, Boltengagen A, Ayoub S, et al. 2017. The *Drosophila* embryo at single-cell transcriptome resolution. *Science* 358:194–99
Ke R, Mignardi M, Pacureanu A, Svedlund J, Botling J, et al. 2013. In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* 10:857–60
Kiselev VY, Andrews TS, Hemberg M. 2019. Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat. Rev. Genet.* 20:273–82
Kobak D, Berens P. 2019. The art of using t-SNE for single-cell transcriptomics. *Nat. Commun.* 10:5416
Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, et al. 2014. Highly multiplexed subcellular RNA sequencing in situ. *Science* 343:1360–63
Lee S-I, Batzoglou S. 2003. Application of independent component analysis to microarrays. *Genome Biol.* 4:R76
Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, et al. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–76
Lloyd S. 1982. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* 28:129–37
Lun ATL, Bach K, Marioni JC. 2016. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* 17:75
Ma Y, Hof PR, Grant SC, Blackband SJ, Bennett R, et al. 2005. A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. *Neuroscience* 135:1203–15
MacKenzie-Graham A, Lee E-F, Dinov ID, Bota M, Shattuck DW, et al. 2004. A multimodal, multidimensional atlas of the C57BL/6J mouse brain. *J. Anat.* 204:93–102
Märtin A, Calvigioni D, Tzortzi O, Fuzik J, Wärnberg E, Meletis K. 2019. A spatiomolecular map of the striatum. *Cell Rep.* 29:4320–33.e5
McInnes L, Healy J, Melville J. 2018. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426 [stat.ML]
Moffitt JR, Bambah-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, et al. 2018. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* 362:eaau5324
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-seq. *Nat. Methods* 5:621–28
Ng L, Bernard A, Lau C, Overly CC, Dong H-W, et al. 2009. An anatomic gene expression atlas of the adult mouse brain. *Nat. Neurosci.* 12:356–62
Niedworok CJ, Brown APY, Cardoso MJ, Osten P, Ourselin S, et al. 2016. aMAP is a validated pipeline for registration and segmentation of high-resolution mouse brain data. *Nat. Commun.* 7:11879
Nitzan M, Karaiskos N, Friedman N, Rajewsky N. 2019. Gene expression cartography. *Nature* 576:132–37
Ortiz C, Navarro JF, Jurek A, Märtin A, Lundeberg J, Meletis K. 2020. Molecular atlas of the adult mouse brain. *Sci. Adv.* 6:eabb3446
Oshlack A, Wakefield MJ. 2009. Transcript length bias in RNA-seq data confounds systems biology. *Biol. Direct* 4:14
Paxinos G, Franklin KBJ. 2004. *The Mouse Brain in Stereotaxic Coordinates.* Houston, TX: Gulf Prof. Publ.
Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park J-E. 2020. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics* 36:964–65
Renier N, Adams EL, Kirst C, Wu Z, Azevedo R, et al. 2016. Mapping of brain activity by automated volume analysis of immediate early genes. *Cell* 165:1789–802
Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11:R25
Rodriques SG, Stickels RR, Goeva A, Martin CA, Murray E, et al. 2019. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* 363:1463–67
Sandberg R. 2014. Entering the era of single-cell transcriptomics in biology and medicine. *Nat. Methods* 11:22–24
Sandberg R, Yasuda R, Pankratz DG, Carter TA, Rio JAD, et al. 2000. Regional and strain-specific gene expression mapping in the adult mouse brain. *PNAS* 97:11038–43

Satija R, Farrell JA, Gennert D, Schier AF, Regev A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33:495–502

Saunders A, Macosko EZ, Wysoker A, Goldman M, Krienen FM, et al. 2018. Molecular diversity and specializations among the cells of the adult mouse brain. *Cell* 174:1015–30.e16

Shimogori T, Abe A, Go Y, Hashikawa T, Kishi N, et al. 2018. Digital gene atlas of neonate common marmoset brain. *Neurosci. Res.* 128:1–13

Singer RH, Ward DC. 1982. Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog. *PNAS* 79:7331–35

Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, et al. 2016. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353:78–82

Stegle O, Teichmann SA, Marioni JC. 2015. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16:133–45

Stickels RR, Murray E, Kumar P, Li J, Marshall JL, et al. 2020. Sensitive spatial genome wide expression profiling at cellular resolution. bioRxiv 2020.03.12.989806. **https://doi.org/10.1101/2020.03.12.989806**

Strell C, Hilscher MM, Laxman N, Svedlund J, Wu C, et al. 2019. Placing RNA in context and space—methods for spatially resolved transcriptomics. *FEBS J.* 286:1468–81

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, et al. 2019. Comprehensive integration of single-cell data. *Cell* 177:1888–902.e21

Swanson LW. 2018. Brain maps 4.0—Structure of the rat brain: an open access atlas with global nervous system nomenclature ontology and flatmaps. *J. Comp. Neurol.* 526:935–43

Tran HTN, Ang KS, Chevrier M, Zhang X, Lee NYS, et al. 2020. A benchmark of batch-effect correction methods for single-cell RNA sequencing data. *Genome Biol.* 21:12

Tung P-Y, Blischak JD, Hsiao CJ, Knowles DA, Burnett JE, et al. 2017. Batch effects and the effective design of single-cell gene expression studies. *Sci. Rep.* 7:39921

Vallejos CA, Risso D, Scialdone A, Dudoit S, Marioni JC. 2017. Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nat. Methods* 14:565–71

van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9:2579–605

Van Essen DC, Glasser MF. 2018. Parcellating cerebral cortex: how invasive animal studies inform noninvasive mapmaking in humans. *Neuron* 99:640–63

Vickovic S, Eraslan G, Salmén F, Klughammer J, Stenbeck L, et al. 2019. High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* 16:987–90

Wagner GP, Kin K, Lynch VJ. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131:281–85

Waltman L, van Eck NJ. 2013. A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* 86:471

Wang F, Flanagan J, Su N, Wang LC, Bui S, et al. 2012. RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* 14:22–29

Wang Q, Ding S-L, Li Y, Royall J, Feng D, et al. 2020. The Allen Mouse Brain Common Coordinate Framework: a 3D reference atlas. *Cell* 181:936–53.e20

Wang X, Allen WE, Wright MA, Sylwestrak EL, Samusik N, et al. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science* 361:eaat5691

Wattenberg M, Viégas F, Johnson I. 2016. How to use t-SNE effectively. *Distill*, Oct. 13. **http://doi.org/10.23915/distill.00002**

Wolf FA, Angerer P, Theis FJ. 2018. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19:15

Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, et al. 2018. Molecular architecture of the mouse nervous system. *Cell* 174:999–1014.e22