

REVIEW

Open Access



Single-cell RNA-sequencing of the brain

Raquel Cuevas-Diaz Duran^{1,2†}, Haichao Wei^{1,2†} and Jia Qian Wu^{1,2*}

Abstract

Single-cell RNA-sequencing (scRNA-seq) is revolutionizing our understanding of the genomic, transcriptomic and epigenomic landscapes of cells within organs. The mammalian brain is composed of a complex network of millions to billions of diverse cells with either highly specialized functions or support functions. With scRNA-seq it is possible to comprehensively dissect the cellular heterogeneity of brain cells, and elucidate their specific functions and state. In this review, we describe the current experimental methods used for scRNA-seq. We also review bioinformatic tools and algorithms for data analyses and discuss critical challenges. Additionally, we summarized recent mouse brain scRNA-seq studies and systematically compared their main experimental approaches, computational tools implemented, and important findings. scRNA-seq has allowed researchers to identify diverse cell subpopulations within many brain regions, pinpointing gene signatures and novel cell markers, as well as addressing functional differences. Due to the complexity of the brain, a great deal of work remains to be accomplished. Defining specific brain cell types and functions is critical for understanding brain function as a whole in development, health, and diseases.

Keywords: Single-cell RNA-sequencing, Brain, Heterogeneity, Bioinformatic analyses

Introduction

Single-cells are the fundamental units of unicellular and multicellular organisms. Every single-cell in an organism is unique in its transcriptome, epigenome, and its local microenvironment. Even genetically identical cells display stochastic gene expression due to random fluctuations in the mechanisms driving and regulating transcription and translation [1, 2]. The underlying heterogeneity within cells is a fundamental property of cellular systems for homeostasis and development [3]. Different cell types specialize in the execution of specific tasks [4].

Next-generation sequencing technologies, such as RNA-sequencing, have become a standard for querying gene expression [5, 6]. However, gene expression levels obtained through such ensemble-based approaches yield expression values averaged across large populations of input cells, masking cellular heterogeneity. Recent experimental advances have allowed the isolation of single-cells

and the generation of cDNA libraries from low amounts of RNA. Through scRNA-seq researchers are able to determine expression profiles in single-cell resolution. Since the introduction of scRNA-seq [7], the number of single-cell experiments has greatly increased. scRNA-seq has demonstrated to be a powerful tool to identify and classify cell subpopulations [8], characterize rare or small subpopulations [9], and trace cells along dynamic cellular stages, such as during differentiation [10].

The mammalian brain is a complex tissue that contains a large number of specialized cells with differences in morphology, connectivity, and functions [11–13]. Brain cells have been classified by location, morphology, electrophysiological characteristics, target specificity, molecular markers and gene expression patterns [14–17]. Single-cell analysis is critical for studying the brain since small differences in a seemingly homogeneous population may explain issues relating cells to learning, memory, and other cognitive functions [18]. scRNA-seq makes it possible to understand the heterogeneity and the regulatory networks within brain cells at the transcriptome level.

The general framework of a scRNA-seq experiment consists of: single-cell isolation, cell lysis, mRNA capturing, mRNA reverse transcription into cDNA, cDNA

*Correspondence: Jiaqian.Wu@uth.tmc.edu

†Raquel Cuevas-Diaz Duran and Haichao Wei contributed equally to the work

¹The Vivian L. Smith Department of Neurosurgery, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA

Full list of author information is available at the end of the article

amplification, library preparation, and sequencing [19]. Herein, we will review recent research in brain cells with scRNA-seq. In the first two sections, we will discuss the advances and limitations of the methods for single-cell isolation and library generation. Section three will summarize the analysis methods of scRNA-seq data. Subsequently, we will discuss recent and relevant findings derived from scRNA-seq of brain cells. Finally, we will highlight future applications and challenges of scRNA-seq in brain.

Single-cell isolation protocols

The first important step in scRNA-seq is to isolate single-cells from tissues keeping their expression patterns as accurate as possible. Several technologies have been used, such as: FACS (Fluorescence-activated cell sorting), MACS (Magnetic-activated cell sorting), LCM (Laser capture microdissection), manual cell picking and microfluidics. Depending on the nature of samples, different methods may be more suitable for single-cell isolation in distinct samples. In this section, we will discuss some methods used for isolating brain cells.

Fluorescence-activated cell sorting and MACS are widely used methods to isolate single-cells. FACS can purify single-cells based on cell size, granularity and fluorescence. Surface markers are different in individual cells, so FACS can isolate specific cells stained with different fluorescently-tagged monoclonal antibodies [20]. In brain cell research, cells have been labelled with different markers. For example, Tasic et al. [21] used combinations of *Snap25*, *Slc17a7*, and *Gad1* to find subpopulations in the primary visual cortex as listed in Table 1 and depicted in Fig. 1. Similarly, Llorens-Bobadilla et al. [22] labelled cells with *GLAST/Prom1* and *PSA-NCAM* to dissect populations in the subventricular zone. Although FACS is a highly efficient method to isolate single-cells, it has its limitations: not all cell types have their own specific gene markers [23], and the binding of fluorescently-tagged monoclonal antibodies to cells might alter their function [24]. One major disadvantage of FACS is its low cell throughput rate. Even high-speed sorters will yield a few thousand cells per second [25]. Since many experiments require large number of cells, sorting runs may take long times posing quality issues to sorted cells. MACS is another method used to isolate single-cells [26]. The cells are isolated by biodegradable iron based nanobeads bound with specific cell surface antibodies. Although MACS can produce high yield single-cells and is widely used, one of its main limitations is that antibody-coated magnetic beads are specific only for cell surface markers.

Laser capture microdissection is a useful method to isolate cells using a laser pulse [27]. Microscopy is used to verify the position of cells of interest, and then

a thermoplastic polymer coating is placed on the tissue over a glass slide. The polymer is melted and then the polymer-cell composition is removed from the tissue. Although specific cells in a tissue are captured, there are some limitations. Contrary to FACS and MACS, LCM is a low-throughput technology. Additionally, LCM relies heavily on cell identification. LCM needs an expert pathologist or cytologist, limiting its extensive application. However, the main advantages of LCM are that it allows researchers to study single-cells within their niche or microenvironment and preserves their spatial location. A cell's niche is relevant when studying cells with functional diversity linked to spatial location such as brain cells.

cDNA amplification and sequencing library construction

A single-cell can only supply very limited starting material (about 0.1 pg of mRNA in each cell), so amplification methods are needed to produce high fidelity, high coverage and reliable data [28]. Some of the common reverse transcription and amplification methods used include: SMART-seq/SMART-seq2 (switching mechanism at the 5' end of the RNA transcript) [9, 29], STRT-seq (single-cell tagged reverse transcription sequencing) [30], CEL-seq (cell expression by linear amplification and sequencing) [31], PMA (Phi29 DNA polymerase-based mRNA transcriptome amplification) [32], SMA (semi-random primed PCR-based mRNA transcriptome amplification procedure) [32], and Quartz-seq [33]. Researchers studying brain scRNA-seq typically use SMART-seq, SMART-seq2, and STRT-seq as outlined in Table 1.

SMART-seq is a reverse transcription and amplification method based on template-switching [9]. First strand cDNAs are created by an oligo(dT)-containing primer, and a few untemplated poly(C) nucleotides are added as overhang at the end of cDNA molecules. The second strand is synthesized by an oligonucleotide primer which can hybridize to the poly(C) overhang, generating full length cDNA products. The purified PCR products can then be used for constructing cDNA libraries. SMART-seq2 is an updated version of SMART-seq [29]. It can significantly improve cDNA yield. In SMART-seq2 protocol, similar to SMART-seq, the first strand is synthesized with 2–5 untemplated nucleotides added at the end of cDNA molecules. Then TSO (template-switching oligonucleotides) with two riboguanosines and a modified guanosine are added to the end of cDNAs. Compared with SMART-seq, SMART-seq2 can produce twofold cDNA products for constructing cDNA libraries.

STRT-seq is also based on templated-switching methods. In this protocol, single-cells are collected and

Table 1 Overview of recent brain scRNA-seq studies

Brain region	Isolation method	RT and cDNA amplification	# of cells sequenced	Major populations identified	# of cells in population	Subpopulations identified	Refs.
Primary visual cortex	FACS	SMARTer	1600	GABAergic neurons	761	23	[21]
				Glutamatergic neurons	764	19	
Somatosensory cortex, hippocampal CA1	Fluidigm, FACS	STRT/C1	3005	Non-neuronal	103	7	
				S1 pyramidal neurons	390	8	[8]
				CA1 pyramidal neurons	961	4	
				Interneurons	300	16	
				Oligodendrocytes	811	6	
				Astrocytes	210	2	
Hippocampal dentate gyrus	Pipetting	SMART-seq	168	Microglia	90	5	
				Vascular endothelial cells	180	2	
				Mural cells	60	2	
				Ependymal cells	30	2	
				CFP ^{intc} qNSC or NPCs	132	6	[73]
				CFP ^{intc} - non-NPCs	26	5	
				Neurons (D1-MSN, D2-MSN, interneurons)	368	3	[74]
				Astrocytes	107	1	
				Oligodendrocytes: newly formed (NFO), mature (MO)	43	2	
				Stem cells (NSC, OPC)	20	2	
Somatosensory cortex, dentate gyrus, hippocampus CA1, corpus callosum, amygdala, hypothalamus, zona incerta, SN-VTA, dorsal horn	Fluidigm, FACS	STRT/C1	5072	Vascular cells (VSMCs, Endothelial)	43	2	
				Immune cells (microglia, macrophage)	119	2	
				Ependymal (ciliated, secretory)	39	2	
				Oligodendrocyte precursor cells (OPC)	310	1	[75]
				Committed oligodendrocyte precursors (COP)	140	1	
				Newly formed oligodendrocytes (NFOL1, NFOL2)	219, 293	2	
				Myelin-forming oligodendrocytes (MFOL1, MFOL2)	353, 933	2	
				Mature oligodendrocytes (MOL1, ..., MOL6)	126, ..., 835	6	
				Vascular and leptomeningeal cells (VLMC)	76	1	

qNSC quiescent neural stem cells, NPC neural precursor cells, MSN medium spiny neurons, NSC neural stem cell

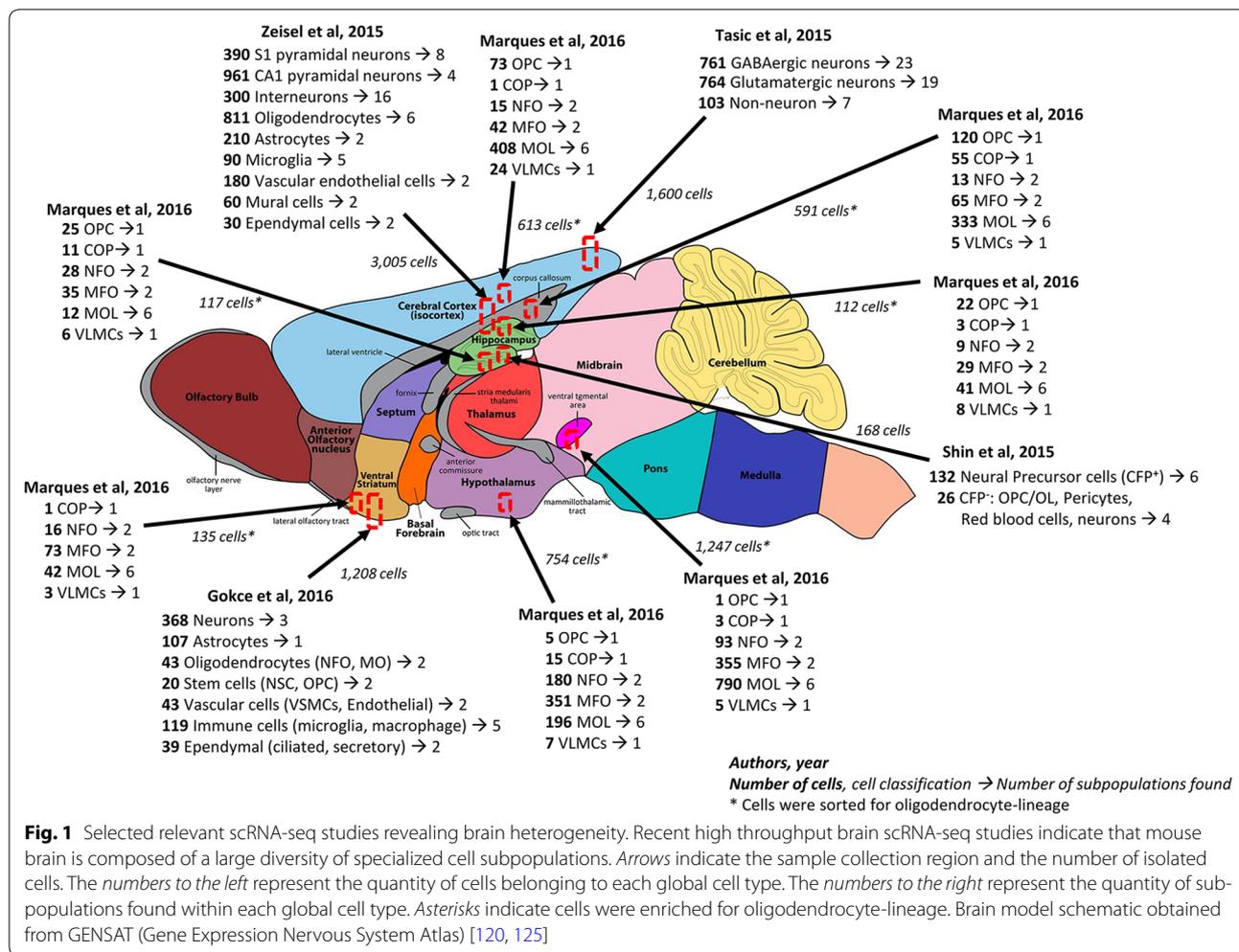


Fig. 1 Selected relevant scRNA-seq studies revealing brain heterogeneity. Recent high throughput brain scRNA-seq studies indicate that mouse brain is composed of a large diversity of specialized cell subpopulations. Arrows indicate the sample collection region and the number of isolated cells. The numbers to the left represent the quantity of cells belonging to each global cell type. The numbers to the right represent the quantity of subpopulations found within each global cell type. Asterisks indicate cells were enriched for oligodendrocyte-lineage. Brain model schematic obtained from GENSAT (Gene Expression Nervous System Atlas) [120, 125]

distributed into 96-well PCR plates [30]. Then the cells are lysed by lysis buffer. The first strand is synthesized using oligo(dT) primer and 3–6 cytosines are added to the end of cDNAs. The secondary strand is created using a primer with a cell specific barcode corresponding to each well. After cDNA synthesis, all the products are pooled and then, cDNAs are amplified by a single-primer PCR.

Although reverse transcription and amplification methods can supply sufficient material, they have different levels of amplification bias which are either over-representing or under-representing certain regions of cDNA [28]. For example, SMART-seq, which can provide full-length coverage of cDNAs, has 3'-end bias; but in SMART-seq2, the bias is decreased [9, 34]. STRT-seq has high 5'-end bias [28, 30]. In order to reduce the amplification bias of STRT-seq, UMI (unique molecular identifiers) are integrated in the sequencing primer used for reverse transcription or template switching [35–37]. UMIs are tens of thousands of short, random DNA

molecules which are used to label mRNA molecules during reverse transcription prior to amplification. They allow for absolute molecule quantification.

After the cDNA amplification, the cDNA library is constructed. cDNA libraries must be compatible with the sequencing platform. Nextera XT is a widely used library preparation kit. Libraries are generally sequenced by Illumina platforms, such as HiSeq, MiSeq and NextSeq.

Single-cell RNA-sequencing data analysis

Two important questions which need to be addressed in scRNA-seq assays are the minimum number of cells to be sequenced and the sequencing depth at which the majority of transcripts in a cell can be detected. The answers depend on the experiment's aims and the nature of the isolated cells. In general, deeper sequencing is required to classify distinct cell types within a homogeneous population of cells [38]. In a sufficiently heterogeneous population, Pollen et al. [39] were able to classify 301 neural cells from the human neural cortex in different

developmental stages with as few as 50,000 reads. With numerous experiments with microliter and nanoliter volumes, Wu et al. [40] concluded that beyond one million reads, the number of detected genes per cell varies less than 5%. However, the main variable which will define sequencing depth is the population's heterogeneity.

Quality control

As with bulk RNA-seq, the first step in data analysis is quality control. Quality control is generally performed before and after sequencing. Before sequencing, the quality of single-cells is addressed through visual inspection or automated imaging and viability dyes. In contrast with bulk RNA-seq, scRNA-seq protocols result in cells isolated in microwell plates, droplets, or chambers in microfluidic devices. Using microfluidics of droplet technologies, hundreds to thousands of cells can be sequenced in a single run [41, 42]. Due to massive and parallel processing, capture sites may be empty or contain either single or multiple cells. Furthermore, captured cells may be healthy, stressed, broken, or even damaged due to handling. Low quality sites and cells need to be excluded from the experiment since their data may be misleading. Several approaches have been proposed for filtering low quality sites and cells [29, 35, 43–46]. They may be classified into microscopic imaging of individual cells and staining cells with viability dyes.

Microscopic cell imaging has proven to identify a high proportion of low quality cells, however this approach is not compatible with all platforms, it is time-consuming, and its automation is challenging. Automated imaging systems rely on visual inspection derived metrics, such as morphology, pixel intensity and frequency. As with other imaging systems, their automation requires a training set of images and machine-learning algorithms, such as Support Vector Machines to discriminate between low and high quality cells. Figure 2 shows representative wide field images captured with an automated imaging device. Staining of dead or viable cells is an effective and relatively fast method, however it can modify a cell's

transcriptional state and alter the experiment's outcome. After staining cells, an imaging system can determine the cell's viability by determining pixel intensities as depicted in Fig. 2a.

After sequencing, quality control is performed on raw reads, aligned reads, and across the collection of cells to identify low quality cells. Relevant quality control metrics, similar to those used for bulk RNA-seq, include: per base sequence quality, sequence duplication levels, over-represented sequences, sequence length distribution, and GC content, among others. Quality control metrics should be calculated for raw reads, as well as for aligned reads. Popular tools for assessing these metrics are FastQC, Kraken [47], and RNA-SeQC [44]. Additionally, parameters such as depth of coverage and library complexity should be addressed. Comparing quality control metrics across all cells is helpful in identifying outliers.

Filtering thresholds are also commonly used for identifying low quality cells after sequencing. Thresholds are typically based on the number of mapped reads and/or on the proportion of detected genes. A comprehensive analysis on low quality cells was published by Ilic et al. [48]. The authors obtained a set of technical and biological measures useful for discriminating low quality cells. Researchers demonstrated that broken cells have a downregulation of genes enriched in gene ontology terms “cytoplasm”, “metabolism”, and “membrane” and an upregulation of genes related to “mitochondrially encoded genes” and “mitochondrially localized proteins”. Due to a compromised cell membrane, broken cells have most likely lost cytoplasmic mRNA while maintaining mRNA enclosed in the mitochondrial membrane, thus resulting in the upregulation of mitochondrially encoded genes. Ilic et al. also proved that empty capture sites and broken cells display lower number of total reads yielding a decreased number of detected genes. Similarly, they concluded that the proportion of duplicated reads is higher in multiple captured cells than in single-cells. Their work was implemented in R and Python libraries available in GitHub repositories. Islam et al. [35] used the

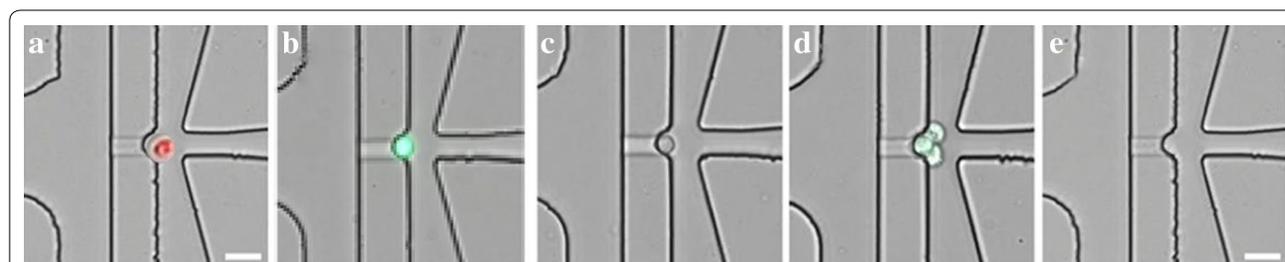


Fig. 2 Single-cell widefield representative images acquired by an automated device (C1 Fluidigm chip). **a** Cell stained with ethidium homodimer-1 (EthD-1, red) labeling unhealthy or dead cells. **b** Single GFP⁺ cell. **c** Single GFP⁻ cell. **d** Capture site containing three cells. **e** Empty capture site (Figure adapted from [126])

total number of detected genes greater than 5000 and at least 85% of cytoplasmic genes (non-mitochondrial and non-ribosomal RNA) as criteria for selecting high quality

cells. Figure 3 outlines the various processes involved in scRNA-seq quality control assessment required for discriminating between high and low quality cells. Another

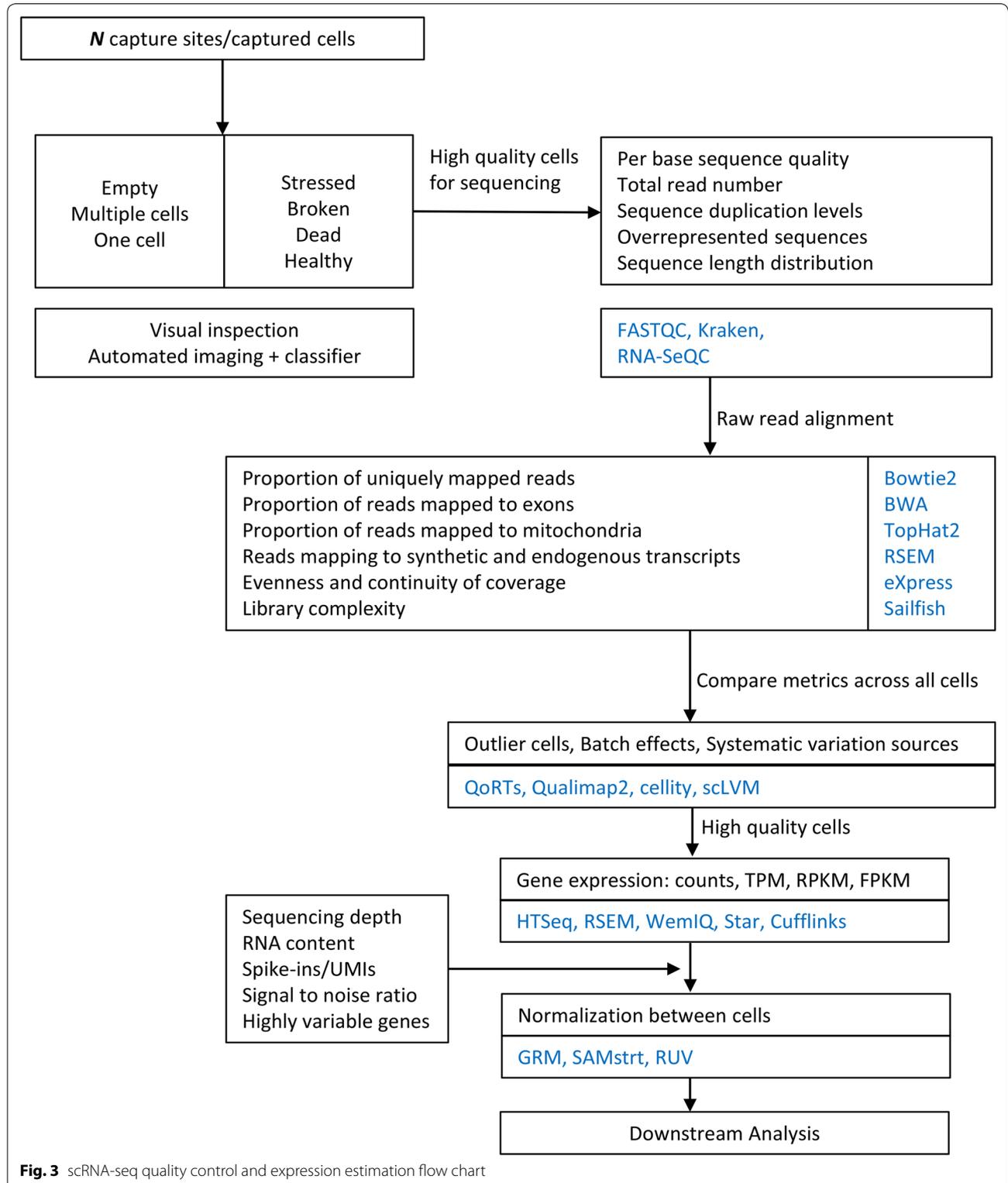


Fig. 3 scRNA-seq quality control and expression estimation flow chart

useful approach for discriminating low-quality cells is to apply principal component analysis (PCA) to gene expression. The underlying premise of this is that good-quality cells will cluster together and low-quality cells will appear as outliers.

Gene expression estimation

To quantify gene expression, sequencing reads from high quality cells are aligned to a reference genome and gene counts are computed. If UMIs were used, transcript molecules may be counted directly since the number of UMIs linked to each gene accounts for the number of cDNA molecules associated with it. For non-UMI data, expression may be obtained as counts using tools such as HTSeq [49], RSEM [50], WemIQ [51], and featureCounts [52], among others. Expression is also addressed as relative expression with metrics including transcripts per million mapped reads (TPM), counts per million mapped reads (CPM), reads per kilobase per million mapped reads (RPKM) or fragments per kilobase per million mapped reads (FPKM). Popular tools for assessing relative expression include Cufflinks [53–55], and STAR [56].

Normalization of scRNA-seq counts is a critical step which allows for expression values to be comparable among cells [57]. Variability between cells may be due to differences in sequencing depth, RNA concentration, GC content, and amplification biases, among others. Normalization methods differ depending on the incorporation of quantitative standards used during library preparation. One approach commonly used in scRNA-seq experiments is adding extrinsic spike-in molecules. Spike-ins are RNA molecules which are either artificially synthesized or obtained from a distant species. Their sequences are known and they are added in a constant concentration to individual cell lysates making them ideal to serve as internal controls. Since the number of spike-in molecules is theoretically the same across all single-cell

libraries, they can be used to calculate scaling factors to normalize for differences in RNA concentration between individual cells. The most commonly used artificial spike-in is the External RNA Controls Consortium (ERCC), a set of 96 synthetic RNA molecules based on bacterial sequences [58]. If the ratio between reads mapped to the genome and the number of reads mapped to spike-ins is low, then that cell must be filtered out since this is indicative of low RNA concentration and will bias the results. Normalization approaches are outlined in Fig. 4.

Normalization in the absence of spike-ins or UMIs is generally performed using bulk RNA-seq methods. Several scRNA-seq studies have normalized for sequencing depth by calculating TPM [39, 59] and FPKM/RPKM [60–62]. More sophisticated between-cell normalization approaches include methods where scaling factors are computed, such as in DESeq [63], and edgeR [64]. Median-based normalization methods [43, 65–68] are also widely used. They calculate global scaling factors based on the identification of stable house-keeping genes. Their main premise is that variations in house-keeping gene expression are due to technical sources, however, this is not always valid due to variations in RNA content. The amount of RNA contained in each cell varies intrinsically due to cell-cycle, cell size, and transcriptional gene dynamics [69]. If spike-ins are available, they can be used to estimate individual cell’s RNA content and normalize expression estimates more accurately.

Low amounts of RNA in single-cells are one of the main challenges in scRNA-seq data analysis. There is a negative correlation between the RNA concentration and the number of genes affected by technical noise [43]. Technical noise is generally addressed with the coefficient of variation (CV) in gene expression across control samples, including spike-ins. Technical noise must be accounted for since it may be confounded with biological noise. Determining technical noise is challenging

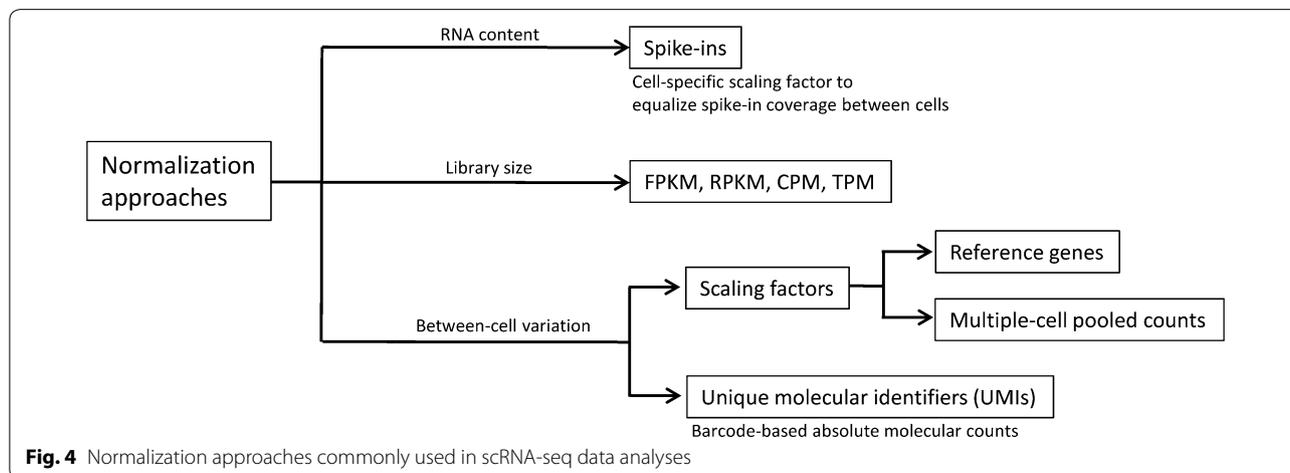


Fig. 4 Normalization approaches commonly used in scRNA-seq data analyses

because even housekeeping genes from genetically identical cells may have noisy gene expression [70]. Technical noise may be modeled with a log-normal function to adjust gene expression estimates. Low amounts of RNA present in a single-cell also yield numerous genes with zero or near-zero values. The high frequency of genes with zero counts may affect normalization methods. To overcome this problem, a recent approach, specific for scRNA-seq normalization without spike-ins, proposed a deconvolution method based on pooled counts of genes across multiple cells [71].

In summary, including synthetic spike-ins or unique molecular identifiers with known concentrations (UMIs)

has advantages in normalization and expression estimation, however their use still needs to be standardized.

Downstream analysis

The most common applications for scRNA-seq experiments are: identification of cell types, pseudo-temporal ordering, and network inference. The normalized gene expression count matrix is used for these downstream analyses. A good review on bioinformatics tools useful for single-cell data analysis was published by Poirion et al. [72]. Typical downstream analyses are depicted in Fig. 5. Algorithms used in recent brain scRNA-seq studies are listed in Table 2.

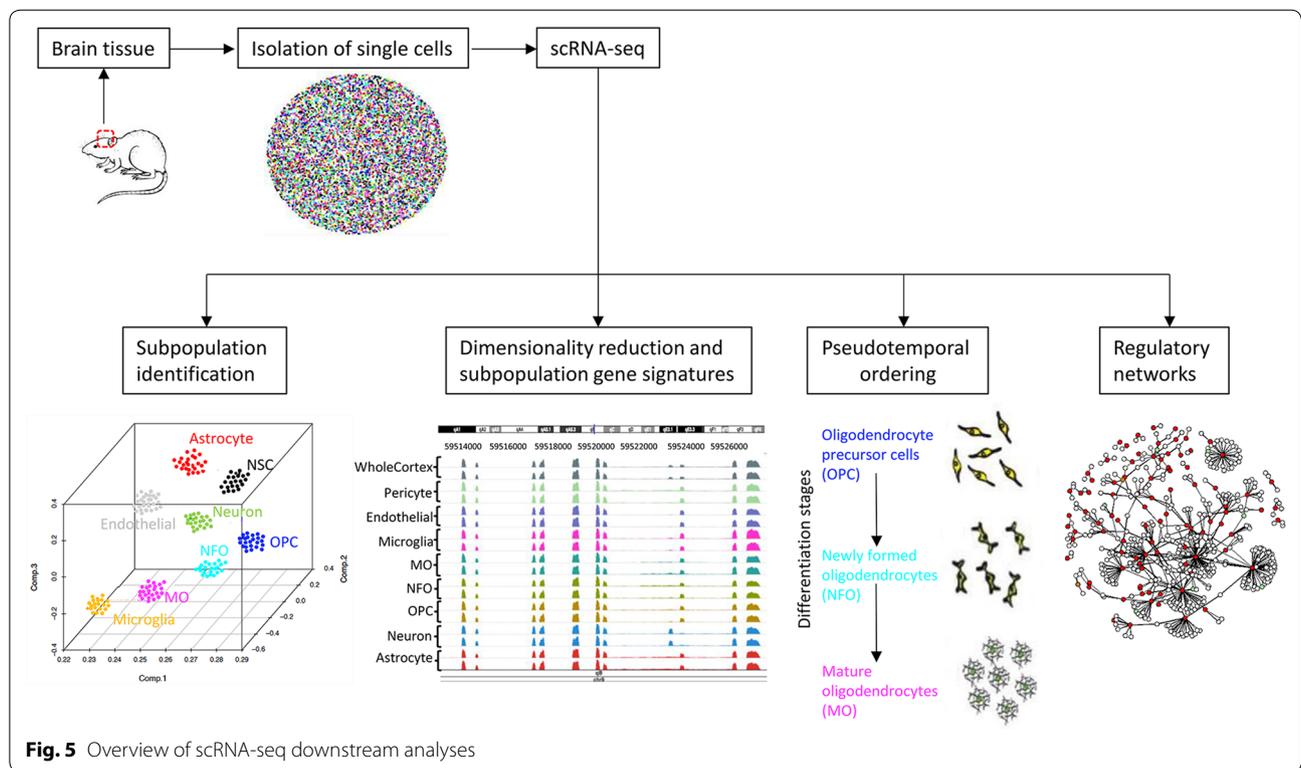


Fig. 5 Overview of scRNA-seq downstream analyses

Table 2 Data analysis methods used in recent brain scRNA-seq studies

Brain region	Expression measure	Clustering	Refs.
Primary visual cortex (L1, 2/3, 4, 5, 6)	RPKM and counts	PCA, WGCNA, random forests	[21]
Somatosensory cortex and hippocampal CA1	CPM	BackSPIN	[8]
Hippocampal dentate gyrus	TPM	Hierarchical clustering, PCA, Waterfall	[73]
Striatum	CPM	2D-tSNE, rPCA	[74]
Somatosensory cortex, striatum, dentate gyrus, hippocampus CA1, corpus callosum, amygdala, hypothalamus, zona incerta, SN-VTA, dorsal horn	CPM	BackSPINv2	[75]

RPKM reads per kilobase of million mapped reads, CPM counts per million mapped reads, TPM transcripts per million mapped reads, PCA principal component analysis, WGCNA weighted gene coexpression network analysis, t-SNE t-distributed stochastic neighbor embedding, rPCA robust principal component analysis

Subpopulation identification

Mapping cells individually, rather than in aggregated components as in bulk RNA-seq, makes it feasible to assess the uniqueness of cell subpopulations. Therefore, some of the most popular applications of scRNA-seq is the identification of subpopulations, novel cell subtypes, and rare cell species in a tissue or biological condition [76]. Clustering algorithms are used for grouping cells which have similar gene expression. Cells in each group or cluster are believed to belong to a specific cell subpopulation or cell state. De novo identification of cell-types may be modeled as an unsupervised clustering problem since prior information regarding the number of clusters or marker genes is unknown. Unsupervised clustering methods extensively used to identify cell subpopulations from scRNA-seq samples include PCA and its variants (e.g. Kernel PCA, tPCA) [21, 73, 74], k -means, and other distance-based algorithms, such as hierarchical clustering [73]. Common similarity metrics used for distance-based methods are Euclidean distance, Pearson, and Spearman correlation coefficients [39, 77]. A recently developed and frequently used hierarchical clustering method is BackSPIN [8], which allows for biclustering of both genes and cells. The non-linear unsupervised clustering method, t-SNE [78], has also been widely used in scRNA-seq samples [42, 74]. Clustering methods are generally applied to highly variable genes [41, 42], differentially expressed genes (DE) [59, 79], or to highly expressed genes [80]. More sophisticated machine learning methods have been used to overcome the limitations in conventional methods due to the frequency of genes with zero counts. An interesting example is the zero inflated factor analysis (ZIFA), which implements a dimension-reduction approach and uses a latent variable factor model to accommodate zeros [81].

The majority of computational methods for subpopulation identification only address abundant cell types. Therefore, rare cell type identification is a challenging application. Grün et al. [76] developed RaceID, an algorithm for the identification of rare and abundant cell types based on transcript counts obtained with UMIs. RaceID first identifies large clusters defined through k -means clustering of the expression correlation matrix of genes. Next, rare cell types are identified within each cluster by detecting cells whose transcript counts do not display cluster specific expression.

Pseudotemporal ordering

scRNA-seq data may be useful for understanding dynamic cellular processes, such as development, reprogramming, differentiation, and disease progression. The underlying premise is that a collection of single-cells will most likely contain cells at different stages during

a dynamic process (e.g. differentiation) and profiling their gene expression will allow for the reconstruction of cascades of gene expression changes placing cells in a pseudotemporal order. Pseudotemporal ordering applies machine learning methods to scRNA-seq data to reconstruct cells' trajectories as they undergo a dynamic biological process. Different algorithms have been implemented for inferring pseudotemporal ordering of single-cells. The first step performed by most temporal ordering algorithms is a dimension reduction such as PCA. For scRNA-seq data, as for bulk RNA-seq, the number of variables or dimensions corresponds to the number of genes. After dimension reduction, if there is prior knowledge of the key marker genes driving the transition between states, methods such as Wanderlust [82] will use graph-based trajectory detection algorithms to order cells along a path. The key marker genes selected for defining a path's distance may be previously known genes (e.g. genes known to be involved in a differentiation process) or differentially expressed genes. Single-cells may be clustered into subpopulations before temporal ordering.

Several methods which do not require prior knowledge of marker genes have been developed [61, 83, 84]. These methods reconstruct trajectory paths in reduced spaces using several algorithms such as minimum spanning trees (MST), and principal curves. Monocle, developed by Trapnell et al. [61] uses independent component analysis (ICA) for dimension reduction and then constructs an MST to find the paths based on Euclidean distance. Authors achieved a more robust temporal cell ordering when using differentially expressed genes. Monocle2 [85] was recently implemented to overcome the accuracy challenges in trajectory reconstruction. Monocle2 applies reversed graph embedding (RGE) [86] to reconstruct complex single-cell trajectories.

Another popular method for pseudotemporal ordering is Waterfall [73]. Waterfall uses k -means and PCA to cluster cells before constructing an MST for ordering cell subpopulations.

Finding regulatory networks

Important applications of gene expression profiling have been the identification of co-regulated groups of genes and inferring gene regulatory network dynamics. In co-expression analysis, pairs of genes with similar expression profiles are assumed to be co-regulated and may be part of a signaling cascade. Computational methods have been developed to identify correlated genes or modules [87]. Weighted gene co-expression network analysis (WGCNA) has been a popular network reconstruction tool used for bulk RNA-seq [88]. Xue et al. [89] applied WGCNA to scRNA-seq data obtained from single-cells derived from human and mouse embryos. The authors

found functional modules of co-expressed genes for each developmental stage indicating sequential order of transcriptional changes in relevant pathways.

Several mathematical methods such as ordinary differential equations (ODE)-based models and stochastic models have been developed for understanding the dynamics of gene regulation. However, such methods require time-series gene expression profiling, which, for scRNA-seq is unlikely due to sequencing costs. To overcome the lack of temporal data, Ocone et al. [90] proposed a framework which allows the reconstruction of regulatory network dynamics through the combination of dimensionality reduction using diffusion maps [91], pseudo-time single-cell ordering implementing Wanderlust [82], and the generation of ODE-based mathematical transcriptional models. Through their framework, authors were able to reconstruct transcriptional dynamics of specific genes during differentiation of hematopoietic stem cells.

The application of scRNA-seq in the brain

The mammalian brain is considered to be the most complex organ due to its cellular diversity, the variety and scope of its functions and its transcriptional regulation [92]. Previous studies have aimed at studying the diversity of brain cells through RNA-seq samples from purified populations of cerebral cortex [93, 94]. Recently, scRNA-seq is being used as a tool to assess the brain's complexity and to identify new cell subpopulations, specific gene signatures, and underlying regulatory networks. This section will provide an overview of relevant scRNA-seq studies related to different types of brain cells. A more detailed description of selected studies is listed in Tables 1 and 2 and depicted in Fig. 1.

The identification of brain cell types

The brain contains highly complex neural cell types/subtypes. Traditionally, neural cells were identified by morphology, excitability, connectivity and the cell's location [95]. Recently, scRNA-seq was used to identify different neural types and subtypes, and to discover novel cell-specific markers. For instance, Amit Zeisel et al. [8] sequenced 3005 single-cells and revealed 9 major classes of cells (S1 and CA1 pyramidal neurons, interneurons, oligodendrocytes, astrocytes, microglia, vascular endothelial cells, mural cells and ependymal cells). The authors identified specific novel gene markers for different cell types, for example, S1 pyramidal cells were characterized by *Gm11549* (a long noncoding RNA), hippocampal pyramidal cells by *Spink8* (a serine protease inhibitor), and interneurons by *Pnoc* (prepronociceptin).

Striatum is a subcortical part of the forebrain. The striatal dysfunction can cause many neuropsychiatric

disorders, for instance, Parkinson's and Huntington's disease, obsessive-compulsive disorder, and autism [96, 97]. Traditionally, the neuronal composition of the striatum has been defined by mostly medium spiny neurons (MSN) and a small population of interneurons [74]. MSNs have been classified anatomically and functionally into D1 and D2 MSNs [98] however, striatal diversity has not been assessed.

Ozgun Gokce et al. [74] used two approaches: microfluidic single-cell RNA sequencing (MIC-scRNA-seq) and single-cell isolation by fluorescence-activated cell sorting (FACS-scRNA-seq) to analyze the transcriptomes of 1028 single striatal cells. The transcriptomes revealed ten different cell subpopulations including neurons, astrocytes, oligodendrocytes, stem cells, immune, ependymal, and vascular cells. Through robust PCA, novel gene markers were found to discriminate between D1 and D2 MSN cells.

Neural stem cells (NSCs) can self-renew and produce neural cell types, including neurons, astrocytes and oligodendrocytes [99, 100]. NSCs maintain a balance between quiescent and activated states [101, 102]. If the brain is injured, endogenous NSC will be activated to repair brain tissue [103]. Previous works were limited by small number of factors analyzed and mixed cell populations. It was not completely understood how the NSCs became activated. Recently, two studies have used single-cell methods to examine the activation of dormant neuron stem cells after injury. In one study, Llorens-Bobadilla and colleagues investigated the characteristics of the activation of dormant NSCs after brain injury [22]. The authors identified NSCs in quiescent and active states and uncovered the progression of activation using single-cell sequencing. They identified new gene markers of NSCs subpopulations and they found that, during brain ischemia, dormant NSCs proceed to activation via interferon gamma signaling. Another study also showed that central nervous system (CNS) injury could activate CD133⁺ quiescent NSCs. Luo et al. [104] demonstrated that vascular endothelial growth factor (VEGF) could activate CD133⁺ ependymal neural stem cell (NSCs), and together with basic fibroblast growth factor, elicit neural lineage differentiation and migration. In a recent study, Dulken et al. [57] sequenced 329 high quality single-cells sorted by FACS from four different populations [astrocytes, quiescent neural stem cells (qNSC), activated neural stem cells (aNSCs), and neural precursor cells (NPCs)] within the sub-ventricular zone of adult mice. Through PCA, authors were able to discriminate quiescent cell types (astrocytes and qNSCs) from active and proliferative cell types (aNSCs and NPCs). Interestingly, authors compared their single-cell transcriptomes with those from similar cells [NSCs and transit

amplifying progenitors (TAPs)] sorted with different cell markers [22]. To be able to compare single-cell datasets processed in different batches and thus with dissimilar library preparations and sequencing depths, Dulken et al. mapped Llorens-Bobadilla and colleagues' datasets using their own pipeline and then performed PCA with the most variable genes. Additionally, they performed pseudo-time ordering using Monocle with their consensus-ordering genes and found similar dynamic gene expression related to quiescence and activation of NSCs. Through this meta-analysis, authors were able to observe a high correlation between NSCs from both studies in spite of divergent isolation methods and batch effects.

Oligodendrocytes were considered an important functionally homogeneous population in the CNS, however these cell's morphologies are diverse [105]. It is unclear whether the diversity in morphology is due to oligodendrocytes interacting with the local environment during maturation or due to their intrinsic functional heterogeneity [106, 107]. Marques et al. [75] isolated single-cells from 10 different regions of juvenile and adult mice CNS by FACS and sequenced 5072 oligodendrocytes by scRNA-seq. The authors identified 13 distinct subpopulations from which 12 represent differentiation stages from oligodendrocyte precursor cells to mature oligodendrocytes. The fine differentiation stages were identified using t-SNE for dimensionality reduction and the biclustering tool BackSPIN2 for pseudo-time analysis. Thereby, using scRNA-seq methods, the authors revealed the dynamics of the differentiation and maturation of oligodendrocytes.

It is difficult to interrogate the underlying transcription landscape of individual neurons. Previously, many studies of single adult human neurons were dependent on the availability of freshly isolated neurosurgical tissues from limited regional samples [109]. Although freshly isolated neurosurgical tissues are better for analyzing single neurons, postmortem tissues can provide more input sample. Lake and colleagues developed a new method which can sequence and quantify RNA in isolated neuron nuclei from postmortem brains [108]. They dissected six distinct regions of the cerebral cortex, and produced 3227 sets of single-neuron RNA-seq data. After clustering and classification, 16 neuronal subtypes were identified and were evaluated by known markers and cortical cytoarchitecture.

The regulation of brain developments by long non-coding RNAs (lncRNAs)

Studies have revealed thousands of lncRNAs in mammalian transcriptomes [110]. lncRNAs are not well conserved during evolution [111], but the promoters of lncRNAs are more conserved than protein coding genes [112, 113]. lncRNAs have tissue specific expression in

human brain [114, 115] and have been shown to be involved in the regulation of brain diseases and neurodevelopmental disorders [116, 117]. Previous studies based on bulk tissues suggested that the expression levels of lncRNAs are lower than those of protein coding genes [114, 118]; however, it is unknown whether lncRNAs are expressed at low levels in all cells [119].

Researchers have studied the expression of lncRNAs in purified mouse brain cells and found their role in fate determination of oligodendrocyte precursor cells (OPC) [120]. Recent approaches are now aiming at addressing lncRNAs in brain scRNA-seq samples. Liu et al. [119] used scRNA-seq to analyze lncRNAs in the developing human neocortex. The authors isolated total RNA from 276 single-cells of different stages of human neocortex development and analyzed their transcriptomes. To evaluate if lncRNAs were expressed at high levels in subpopulations of cells, the authors used the lncRNA:mRNA median ratio which compares the median expression of lncRNAs to the median expression of mRNA. Compared with lncRNAs from bulk tissue (the median lncRNA:mRNA ratio was 0.31), many lncRNAs were abundantly expressed in individual cells (in single-cells, 32.2% of cells' median lncRNA:mRNA ratio exceeded 1.0). The authors found that lncRNA *LOC646329* was enriched in the ventricular zone, where most radial glia reside. When *LOC646329* was knocked down, the propagation of U87 cells was reduced. Results suggest that lncRNAs might regulate cell proliferation.

Future perspectives

In summary, scRNA-seq is a powerful tool that will allow researchers to address human brain complexity by identifying cell subpopulations and elucidating specific functions. scRNA-seq has a higher resolution than bulk RNA-seq and allows us to better understand cellular heterogeneity and how it changes during dynamic processes, such as development, differentiation and disease progression. Major resolution, however, makes samples more vulnerable to disturbances and confounding effects. Experimental and computational methods are being developed to overcome challenges posed by detecting single-cell signal in the presence of intrinsic noise and technical variability. Recently, chromatin accessibility [121, 122], chromatin conformation [123], and DNA methylation [124] with single-cell resolution were successfully implemented. Single-cell DNA/RNA-seq approaches will allow scientists to simultaneously assess the genomic, epigenomic, and transcriptomic states of individual cells in biological processes. Single-cell sequencing will be expanded to also address metabolomics in order to construct a more complete picture of a cell.

Abbreviations

CEL-seq: cell expression by linear amplification and sequencing; CNS: central nervous system; COP: committed oligodendrocyte precursors; CPM: counts per million mapped reads; CV: coefficient of variation; ERCC: External RNA Controls Consortium; FACS: fluorescence-activated cell sorting; FPKM: fragments per kilobase per million mapped reads; GFP: green fluorescent protein; GRM: gamma regression model; ICA: independent component analysis; LCM: laser capture microdissection; lncRNA: long non-coding RNA; MACS: magnetic-activated cell sorting; MO: mature oligodendrocytes; MSN: medium spiny neurons; MST: minimum spanning tree; NFO: newly formed oligodendrocytes; NPC: neural precursor cells; NSC: neural stem cells; OPC: oligodendrocyte precursor cells; PCA: principal component analysis; PMA: Phi29 DNA polymerase-based mRNA transcriptome amplification; qNSC: quiescent neural stem cells; rPCA: robust principal component analysis; RPKM: reads per kilobase per million mapped reads; RUV: removed unwanted variation; scRNA-seq: single-cell RNA-sequencing; SMA: semi-random primed PCR-based mRNA transcriptome amplification procedure; SMART-seq: switching mechanism at the 5' end of the RNA transcript sequencing; STRT-seq: single-cell tagged reverse transcription sequencing; TPM: transcripts per million mapped reads; TSO: template-switching oligonucleotides; UMI: unique molecular identifiers; VEGF: vascular endothelial growth factor; VLMC: vascular and leptomenigeal cells; VSMCs: vascular smooth muscle cells; WGCNA: weighted gene co-expression network analysis; ZIFA: zero inflated factor analysis.

Authors' contributions

JQW, HW, and RCDD conceived the project and participated in writing the manuscript. RCDD designed the figures. All authors read and approved the final manuscript.

Author details

¹ The Vivian L. Smith Department of Neurosurgery, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA. ² Center for Stem Cell and Regenerative Medicine, UT Brown Foundation Institute of Molecular Medicine, Houston, TX 77030, USA.

Acknowledgements

The authors would like to thank Dr. Eva M. Zsigmond for editing the manuscript.

Competing interests

The authors declare that they have no competing interests.

Funding

JQW, HW, and RCDD were supported by grants from the National Institutes of Health R01s NS088353 and NS091759; the Staman Ogilvie Fund-Memorial Hermann Foundation; the UTHealth BRAIN Initiative and CTSA UL1 TR000371; and a grant from the University of Texas System Neuroscience and Neurotechnology Research Institute (Grant #362469).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 14 February 2017 Accepted: 19 May 2017

Published online: 08 June 2017

References

- Li GW, Xie XS (2011) Central dogma at the single-molecule level in living cells. *Nature* 475(7356):308–315
- Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135(2):216–226
- Altschuler SJ, Wu LF (2010) Cellular heterogeneity: do differences make a difference? *Cell* 141(4):559–563
- Arendt D (2008) The evolution of cell types in animals: emerging principles from molecular studies. *Nat Rev Genet* 9(11):868–882
- Schuster SC (2008) Next-generation sequencing transforms today's biology. *Nat Methods* 5(1):16–18
- Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 10(1):57–63
- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6(5):377–382
- Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A et al (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347(6226):1138–1142
- Ramskold D, Luo S, Wang YC, Li R, Deng Q, Faridani OR et al (2012) Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat Biotechnol* 30(8):777–782
- Tang F, Barbacioru C, Bao S, Lee C, Nordman E, Wang X et al (2010) Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* 6(5):468–478
- Nelson SB, Sugino K, Hempel CM (2006) The problem of neuronal cell types: a physiological genomics approach. *Trends Neurosci* 29(6):339–345
- Luo L, Callaway EM, Svoboda K (2008) Genetic dissection of neural circuits. *Neuron* 57(5):634–660
- Bota M, Swanson LW (2007) The neuron classification problem. *Brain Res Rev* 56(1):79–88
- Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD (2007) Neuronal subtype specification in the cerebral cortex. *Nat Rev Neurosci* 8(6):427–437
- DeFelipe J, Lopez-Cruz PL, Benavides-Piccione R, Bielza C, Larranaga P, Anderson S et al (2013) New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat Rev Neurosci* 14(3):202–216
- Gelman DM, Marin O (2010) Generation of interneuron diversity in the mouse cerebral cortex. *Eur J Neurosci* 31(12):2136–2141
- Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D et al (2011) A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol* 9(1):e1000582
- Kepecs A, Fishell G (2014) Interneuron cell types are fit to function. *Nature* 505(7483):318–326
- Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA (2015) The technology and biology of single-cell RNA sequencing. *Mol Cell* 58(4):610–620
- Basu S, Campbell HM, Dittel BN, Ray A (2010) Purification of specific cell population by fluorescence activated cell sorting (FACS). *J Vis Exp* 41:e1546
- Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z et al (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 19(2):335–346
- Llorens-Bobadilla E, Zhao S, Baser A, Saiz-Castro G, Zwadlo K, Martin-Villalba A (2015) Single-cell transcriptomics reveals a population of dormant neural stem cells that become activated upon brain injury. *Cell Stem Cell* 17(3):329–340
- Ponten F, Gry M, Fagerberg L, Lundberg E, Asplund A, Berglund L et al (2009) A global view of protein expression in human cells, tissues, and organs. *Mol Syst Biol* 5:337
- Orfao A, Ruiz-Arguelles A (1996) General concepts about cell sorting techniques. *Clin Biochem* 29(1):5–9
- Arnold LW, Lannigan J (2010) Practical issues in high-speed cell sorting. *Current protocols in cytometry*. Chapter 1: Unit 1. pp 1–30
- Almeida M, Garcia-Montero AC, Orfao A (2014) Cell purification: a new challenge for biobanks. *Pathobiol J Immunopathol Mol Cell Biol* 81(5–6):261–275
- Espina V, Milia J, Wu G, Cowherd S, Liotta LA (2006) Laser capture microdissection. *Methods Mol Biol* 319:213–229
- Liu N, Liu L, Pan X (2014) Single-cell analysis of the transcriptome and its application in the characterization of stem cells and early embryos. *Cell Mol Life Sci CMLS* 71(14):2707–2715
- Picelli S, Bjorklund AK, Faridani OR, Sagasser S, Winberg G, Sandberg R (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat Methods* 10(11):1096–1098
- Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P et al (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Res* 21(7):1160–1167
- Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* 2(3):666–673

32. Pan X, Durrett RE, Zhu H, Tanaka Y, Li Y, Zi X et al (2013) Two methods for full-length RNA sequencing for low quantities of cells and single cells. *Proc Natl Acad Sci USA* 110(2):594–599
33. Sasagawa Y, Nikaido I, Hayashi T, Danno H, Uno KD, Imai T et al (2013) Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol* 14(4):R31
34. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9(1):171–181
35. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M et al (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat Methods* 11(2):163–166
36. Kivioja T, Vaharautio A, Karlsson K, Bonke M, Enge M, Linnarsson S et al (2011) Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* 9(1):72–74
37. Grun D, Kester L, van Oudenaarden A (2014) Validation of noise models for single-cell transcriptomics. *Nat Methods* 11(6):637–640
38. Streets AM, Huang Y (2014) How deep is enough in single-cell RNA-seq? *Nat Biotechnol* 32(10):1005–1006
39. Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH et al (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol* 32(10):1053–1058
40. Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME et al (2014) Quantitative assessment of single-cell RNA-sequencing methods. *Nat Methods* 11(1):41–46
41. Macosko EZ, Basu A, Satija R, Nemes J, Shekhar K, Goldman M et al (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161(5):1202–1214
42. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V et al (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161(5):1187–1201
43. Brennecke P, Anders S, Kim JK, Kolodziejczyk AA, Zhang X, Proserpio V et al (2013) Accounting for technical noise in single-cell RNA-seq experiments. *Nat Methods* 10(11):1093–1095
44. DeLuca DS, Levin JZ, Sivachenko A, Fennell T, Nazaire MD, Williams C et al (2012) RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28(11):1530–1532
45. Wang L, Wang S, Li W (2012) RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28(16):2184–2185
46. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I et al (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343(6172):776–779
47. Davis MP, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63(1):41–49
48. Ilicic T, Kim JK, Kolodziejczyk AA, Bagger FO, McCarthy DJ, Marioni JC et al (2016) Classification of low quality cells from single-cell RNA-seq data. *Genome Biol* 17:29
49. Anders S, Pyl PT, Huber W (2015) HTSeq—a python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169
50. Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform* 12:323
51. Zhang J, Kuo CC, Chen L (2015) WemlQ: an accurate and robust isoform quantification method for RNA-seq data. *Bioinformatics* 31(6):878–885
52. Liao Y, Smyth GK, Shi W (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30(7):923–930
53. Ghosh S, Chan CK (2016) Analysis of RNA-Seq data using TopHat and Cufflinks. *Methods Mol Biol* 1374:339–361
54. Pollier J, Rombauts S, Goossens A (2013) Analysis of RNA-Seq data with TopHat and Cufflinks for genome-wide expression analysis of jasmonate-treated plants and plant cultures. *Methods Mol Biol* 1011:305–315
55. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR et al (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 7(3):562–578
56. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S et al (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21
57. Dulken BW, Leeman DS, Boutet SC, Hebestreit K, Brunet A (2017) Single-cell transcriptomic analysis defines heterogeneity and transcriptional dynamics in the adult neural stem cell lineage. *Cell Rep* 18(3):777–790
58. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M et al (2011) Synthetic spike-in standards for RNA-seq experiments. *Genome Res* 21(9):1543–1551
59. Shalek AK, Satija R, Shuga J, Trombetta JJ, Gennert D, Lu D et al (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 510(7505):363–369
60. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH et al (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509(7500):371–375
61. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M et al (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32(4):381–386
62. Deng Q, Ramskold D, Reinius B, Sandberg R (2014) Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343(6167):193–196
63. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):550
64. McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40(10):4288–4297
65. Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ et al (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* 33(2):155–160
66. Kim JK, Marioni JC (2013) Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol* 14(1):R7
67. Dueck H, Khaladkar M, Kim TK, Spaethling JM, Francis C, Suresh S et al (2015) Deep sequencing reveals cell-type-specific patterns of single-cell transcriptome variation. *Genome Biol* 16:122
68. Mahata B, Zhang X, Kolodziejczyk AA, Proserpio V, Haim-Vilmovsky L, Taylor AE et al (2014) Single-cell RNA sequencing reveals T helper cells synthesizing steroids de novo to contribute to immune homeostasis. *Cell Reports* 7(4):1130–1142
69. Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. *Nat Rev Genet* 16(3):133–145
70. Kalisky T, Quake SR (2011) Single-cell genomics. *Nat Methods* 8(4):311–314
71. Lun AT, Bach K, Marioni JC (2016) Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol* 17:75
72. Poirion OB, Zhu X, Ching T, Garmire L (2016) Single-cell transcriptomics bioinformatics and computational challenges. *Front Genet* 7:163
73. Shin J, Berg DA, Zhu Y, Shin JY, Song J, Bonaguidi MA et al (2015) Single-cell RNA-Seq with waterfall reveals molecular cascades underlying adult neurogenesis. *Cell Stem Cell* 17(3):360–372
74. Gokce O, Stanley GM, Treutlein B, Neff NF, Camp JG, Malenka RC et al (2016) Cellular taxonomy of the mouse striatum as revealed by single-cell RNA-Seq. *Cell Rep* 16(4):1126–1137
75. Marques S, Zeisel A, Codeluppi S, van Bruggen D, Mendanha Falcao A, Xiao L et al (2016) Oligodendrocyte heterogeneity in the mouse juvenile and adult central nervous system. *Science* 352(6291):1326–1329
76. Grun D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N et al (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525(7568):251–255
77. Rotem A, Ram O, Shores N, Sperling RA, Goren A, Weitz DA et al (2015) Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* 33(11):1165–1172
78. LvdMG Hinton (2008) Visualizing data using t-SNE. *J Mach Learn Res* 9:2579–2605
79. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublotte JT, Raychowdhury R et al (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498(7453):236–240
80. Cann GM, Gulzar ZG, Cooper S, Li R, Luo S, Tat M et al (2012) mRNA-Seq of single prostate cancer circulating tumor cells reveals recapitulation of gene expression and pathways found in prostate cancer. *PLoS ONE* 7(11):e49144

81. Pierson E, Yau C (2015) ZIFA: dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biol* 16:241
82. Bendall SC, Davis KL, el Amir AD, Tadmor MD, Simonds EF, Chen TJ et al (2014) Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* 157(3):714–725
83. Marco E, Karp RL, Guo G, Robson P, Hart AH, Trippa L et al (2014) Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. *Proc Natl Acad Sci USA* 111(52):E5643–E5650
84. Ji Z, Ji H (2016) TSCAN: pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. *Nucleic Acids Res* 44(13):e117
85. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pfliner H et al (2017) Reversed graph embedding resolves complex single-cell developmental trajectories. *BioRxiv*. doi:10.1101/110668
86. Rodriguez A, Laio A (2014) Machine learning. Clustering by fast search and find of density peaks. *Science* 344(6191):1492–1496
87. Segal E, Shapira M, Regev A, Pe'er D, Botstein D, Koller D et al (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 34(2):166–176
88. Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol*. 4:1128
89. Xue Z, Huang K, Cai C, Cai L, Jiang CY, Feng Y et al (2013) Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500(7464):593–597
90. Ocone A, Haghverdi L, Mueller NS, Theis FJ (2015) Reconstructing gene regulatory dynamics from high-dimensional single-cell snapshot data. *Bioinformatics* 31(12):i89–i96
91. Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F et al (2005) Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc Natl Acad Sci USA* 102(21):7426–7431
92. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA et al (2012) An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489(7416):391–399
93. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keefe S et al (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci Off J Soc Neurosci*. 34(36):11929–11947
94. Yan Q, Weyn-Vanhenhenryck SM, Wu J, Sloan SA, Zhang Y, Chen K et al (2015) Systematic discovery of regulated and conserved alternative exons in the mammalian brain reveals NMD modulating chromatin regulators. *Proc Natl Acad Sci USA* 112(11):3445–3450
95. Fuzik J, Zeisel A, Mate Z, Calvigioni D, Yanagawa Y, Szabo G et al (2016) Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat Biotechnol* 34(2):175–183
96. Maia TV, Frank MJ (2011) From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci* 14(2):154–162
97. Kreitzer AC, Malenka RC (2008) Striatal plasticity and basal ganglia circuit function. *Neuron* 60(4):543–554
98. DeLong M, Wichmann T (2009) Update on models of basal ganglia function and dysfunction. *Parkinsonism Relat Disord*. 15(Suppl 3):S237–S240
99. Kornblum HI (2007) Introduction to neural stem cells. *Stroke* 38(2 Suppl):810–816
100. Bartlett PF, Berninger B (2012) Introduction to the special issue on neural stem cells: regulation and function. *Dev Neurobiol*. 72(7):953–954
101. Lugert S, Basak O, Knuckles P, Haussler U, Fabel K, Gotz M et al (2010) Quiescent and active hippocampal neural stem cells with distinct morphologies respond selectively to physiological and pathological stimuli and aging. *Cell Stem Cell* 6(5):445–456
102. Li L, Clevers H (2010) Coexistence of quiescent and active adult stem cells in mammals. *Science* 327(5965):542–545
103. Benner EJ, Luciano D, Jo R, Abdi K, Paez-Gonzalez P, Sheng H et al (2013) Protective astrogenesis from the SVZ niche after injury is controlled by Notch modulator Thbs4. *Nature* 497(7449):369–373
104. Luo Y, Coskun V, Liang A, Yu J, Cheng L, Ge W et al (2015) Single-cell transcriptome analyses reveal signals to activate dormant neural stem cells. *Cell* 161(5):1175–1186
105. Kessar N, Fogarty M, Iannarelli P, Grist M, Wegner M, Richardson WD (2006) Competing waves of oligodendrocytes in the forebrain and postnatal elimination of an embryonic lineage. *Nat Neurosci* 9(2):173–179
106. Tomassy GS, Berger DR, Chen HH, Kasthuri N, Hayworth KJ, Vercelli A et al (2014) Distinct profiles of myelin distribution along single axons of pyramidal neurons in the neocortex. *Science* 344(6181):319–324
107. Bechler ME, Byrne L, Ffrench-Constant C (2015) CNS myelin sheath lengths are an intrinsic property of oligodendrocytes. *Curr Biol CB*. 25(18):2411–2416
108. Lake BB, Ai R, Kaeser GE, Salathia NS, Yung YC, Liu R et al (2016) Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* 352(6293):1586–1590
109. Darmanis S, Sloan SA, Zhang Y, Enge M, Caneda C, Shuer LM et al (2015) A survey of human brain transcriptome diversity at the single cell level. *Proc Natl Acad Sci USA* 112(23):7285–7290
110. Johnsson P, Lipovich L, Grander D, Morris KV (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochem Biophys Acta* 1840(3):1063–1071
111. Ponjavic J, Ponting CP, Lunter G (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res* 17(5):556–565
112. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H et al (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22(9):1775–1789
113. Pang KC, Frith MC, Mattick JS (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet*. 22(1):1–5
114. Ramos AD, Diaz A, Nellore A, Delgado RN, Park KY, Gonzales-Roybal G et al (2013) Integration of genome-wide approaches identifies lncRNAs of adult neural stem cells and their progeny in vivo. *Cell Stem Cell* 12(5):616–628
115. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A et al (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev* 25(18):1915–1927
116. Lipovich L, Dacht F, Cai J, Bagla S, Balan K, Jia H et al (2012) Activity-dependent human brain coding/noncoding gene regulatory networks. *Genetics* 192(3):1133–1148
117. van de Vondervoort II, Gordebeke PM, Khoshab N, Tiesinga PH, Buitelaar JK, Kozicz T et al (2013) Long non-coding RNAs in neurodevelopmental disorders. *Front Mol Neurosci*. 6:53
118. Necsulea A, Soumillon M, Warnefors M, Liechti A, Daish T, Zeller U et al (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505(7485):635–640
119. Liu SJ, Nowakowski TJ, Pollen AA, Lui JH, Horlbeck MA, Attenello FJ et al (2016) Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol* 17:67
120. Dong X, Chen K, Cuevas-Diaz Duran R, You Y, Sloan SA, Zhang Y et al (2015) Comprehensive identification of long non-coding RNAs in purified cell types from the brain reveals functional lncRNA in OPC fate determination. *PLoS Genet* 11(12):e1005669
121. Cusanovich DA, Daza R, Adey A, Pfliner HA, Christiansen L, Gunderson KL et al (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348(6237):910–914
122. Buenostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP et al (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523(7561):486–490
123. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W et al (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502(7469):59–64
124. Smallwood SA, Lee HJ, Angermueller C, Krueger F, Saadeh H, Peat J et al (2014) Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat Methods* 11(8):817–820
125. GENSAT (Gene Expression Nervous System Atlas). http://www.gensat.org/daily_showcase.jsp. Accessed Dec 2016
126. Burns JC, Kelly MC, Hoa M, Morell RJ, Kelley MW (2015) Single-cell RNA-Seq resolves cellular complexity in sensory organs from the neonatal inner ear. *Nat Commun*. 6:8557