REVIEW

nature neuroscience

Disentangling neural cell diversity using single-cell transcriptomics

Jean-Francois Poulin[1], Bosiljka Tasic[2], Jens Hjerling-Leffler[3], Jeffrey M Trimarchi[4] & Rajeshwar Awatramani[1]

Cellular specialization is particularly prominent in mammalian nervous systems, which are composed of millions to billions of neurons that appear in thousands of different 'flavors' and contribute to a variety of functions. Even in a single brain region, individual neurons differ greatly in their morphology, connectivity and electrophysiological properties. Systematic classification of all mammalian neurons is a key goal towards deconstructing the nervous system into its basic components. With the recent advances in single-cell gene expression profiling technologies, it is now possible to undertake the enormous task of disentangling neuronal heterogeneity. High-throughput single-cell RNA sequencing and multiplexed quantitative RT-PCR have become more accessible, and these technologies enable systematic categorization of individual neurons into groups with similar molecular properties. Here we provide a conceptual and practical guide to classification of neural cell types using single-cell gene expression profiling technologies.

A driving force behind the appearance of multicellularity throughout evolution is thought to be the advantage conferred by multiple types of cells, each specialized in the execution of specific tasks[1]. This division of labor between distinct cell types is particularly prominent in the mammalian nervous system, which is thought to contain thousands of neural cell types[2–4]. A cell type is often defined as a group of cells that perform a similar function (**Box 1**)[1–7]. However, as the function of most cell types in the nervous system is unknown, pragmatic approaches for cell classification have been employed that rely on more easily measurable and accessible properties. Given that a cell's function is rooted in its molecular composition, defining neural cell types on the basis of gene expression is one reasonable approach[7].

Cell identity is generally thought to be defined by the expression of a combination of genes[8–10]. To classify cells on a molecular basis, it is imperative to assess the expression of genes at single-cell resolution. The most widely employed methods for molecular characterization at the single-cell level are immunolabeling, RNA *in situ* hybridization (ISH) and transgenic approaches. These techniques have been used in large-scale projects such as the Allen Brain Atlas (ABA), based on RNA ISH, and the Gene Expression Nervous System Atlas (GENSAT), which used GFP-expressing BAC transgenics; they have provided unprecedented insight into the specificity of individual gene expression in different parts of the mouse brain[11,12]. One drawback of these methods is that usually only one or two markers are examined simultaneously. Given that brain regions are composed of numerous intermingled cell types, it is impossible to precisely overlay multiple gene expression patterns obtained from different experiments to determine the combinatorial molecular profile of each individual cell. Thus, simultaneous profiling of the expression of many genes in a single cell has not been possible using these techniques, but is considered to be crucial to assign a cell type identity[4,6,7,13].

Recent technological advancements have facilitated the quantitative analysis of a multitude of markers in a single cell, thereby enabling classification of neural cells into categories. A common approach in recent cell classification studies is to analyze the expression of many genes in individual cells and then, based on gene coexpression patterns, divide cells into groups by clustering. The identity of these cell groups is assigned *post hoc* on the basis of previously known or newly discovered markers. An example of such a cellular classification scheme is shown in **Figure 1**, using genome-wide single-cell gene expression profiling data from the adult mouse visual cortex (adapted from ref. 14). The clustering in this case was performed iteratively: the first step of clustering separated neuronal from non-neuronal cells; subsequent steps separated neuronal cells into GABAergic and glutamatergic neurons, whereas non-neuronal cells were separated into glia and other non-neuronal types. This iterative segregation process ultimately resulted in the identification of 49 cortical cell types, some of them previously uncharacterized and each defined by a combination of expressed genes. This study and others have begun to reveal remarkable cellular complexity in different brain regions. Here we review the transformative influence of single-cell gene expression profiling on neural cell type classification. We will discuss current single-cell gene expression profiling approaches and their limitations, recent single-cell expression profiling studies that have begun to categorize cell types in various regions of the nervous system, implications of neuronal cell classification using single-cell transcriptomics, and future applications of single-cell technologies in neuroscience.

Approaches for single-cell transcriptional profiling

Single-cell gene expression profiling was pioneered more than 20 years ago using exponential amplification of cDNAs by PCR[15] and linear

[1]Department of Neurology, Northwestern University, Chicago, Illinois, USA. [2]Department of Molecular Genetics, Allen Institute for Brain Science, Seattle, Washington, USA. [3]Division of Molecular Neurobiology, Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden. [4]Department of Genetics, Development and Cell Biology, Iowa State University, Ames, Iowa, USA. Correspondence should be addressed to R.A. (r-awatramani@northwestern.edu).

Received 12 February; accepted 22 July; published online 26 August 2016; doi:10.1038/nn.4366

Box 1 What is a neuronal type?

Ramon y Cajal, using the Golgi staining method, revealed an immense diversity of neuron morphologies in the nervous system. His contribution, and those of generations of neuroscientists following him, led to the general agreement that distinct neuronal cell types are the building blocks of the nervous system. However, more than a century later, there is still no clear consensus on how to define neuronal types, and several reviews have provided eloquent discussions of this topic[2–7,89,98]. An often proposed definition relies on neuronal function[3–7], that is, all neurons that perform the same function in a defined circuit belong to the same cell type. Implicit in this definition is that a neuron's identity is linked to the multiple parameters that contribute to its function, such as its dendritic arborization and axonal projections, but also its electrophysiological, synaptic and molecular properties. Relative to other approaches to characterize neuronal types, single-cell gene expression profiling has the advantage of being a high-throughput, quantitative and relatively cost-effective method that generates highly multidimensional data for cell classification. It is also conceptually intuitive, as many of the functional aspects of a particular neuron (for example, electric currents, neurotransmitter synthesis, etc.) are dictated by its molecular components.

Neuronal classification based on molecular profiles has conceptual limitations. First, it is possible that some cell types might not exist as discrete entities, but rather as parts of phenotypic continua. Second, some features essential to a neuron's identity might be established during development, for example its dendritic arborization or axon trajectory, and might not be reflected in the molecular profile of the mature neuron. Third, several transcriptional cascades in individual neurons change in response to a variety of stimuli such as neuronal activity[147] and hormones[10]. Particularly notable are situations in which cells change even core aspects of their identity, such as neurotransmitter type[148,149], in response to environmental alterations. Recognizing this ambiguity, some have proposed differentiating between 'cell type' versus 'cell state'[7,13]. The molecular definition of a cell type could be based on a unique, stably expressed suite of genes, whose profiles have been firmly established during development and are mostly invariable to stimuli in adulthood. A key component of this cell-type-defining molecular signature likely is a combination of transcription factors, long recognized for their role in early specification, as well as later maintenance of neuron-specific programs[8,9,150]. In contrast to cell type, cell state could be defined by genes that are reversibly regulated by extracellular cues or transitory stimuli[147]. Resolving genes contributing to cell type versus cell state is no easy task and warrants comparing expression profiles in several developmental and physiological scenarios. In summary, cell classification based on gene expression is an effective starting point that should be extended by examination of other neuronal properties.

amplification by *in vitro* transcription[16]. These protocols, as well as the subsequent availability of high-density microarrays, led to the application of these technologies to study the cellular complexity of the nervous system[17–20]. New single-cell capture approaches and nucleic acid amplification protocols, coupled with next-generation sequencing (NGS), have permitted the parallel sequencing of a large number of cDNA molecules and scaled down the amount of starting material required for whole transcriptome analysis[21]. Here, we will summarize the recent coming of age of single-cell gene expression profiling technologies and the technical considerations for each step of this multilayered process.

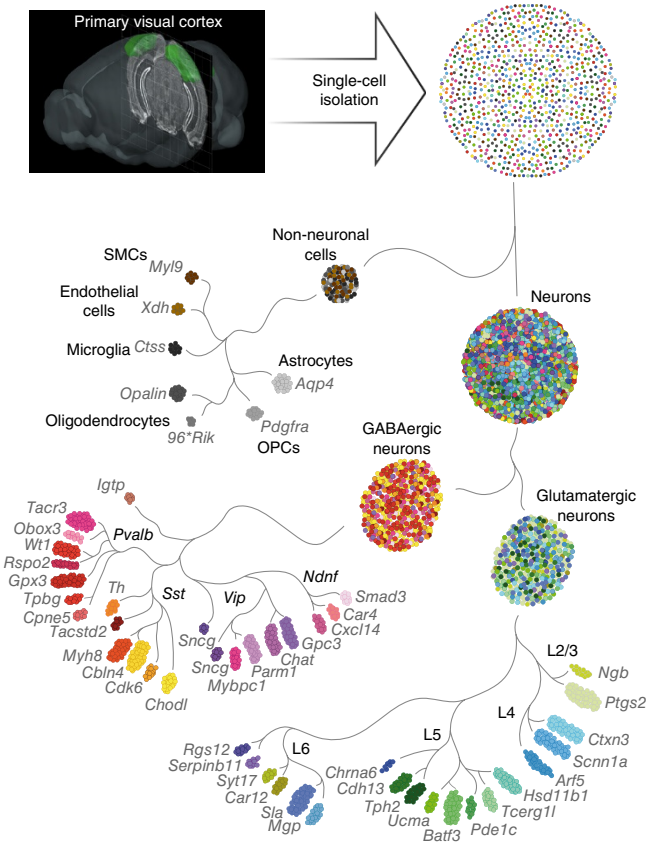

Figure 1 Neural cell diversity in the visual cortex. Schematized example of the process of cell classification based on the analysis of more than 1,600 cells from the adult mouse visual cortex (green region on the brain schematic) by single-cell RNA sequencing[14]. An unbiased iterative clustering analysis first reveals the presence of neuronal and non-neuronal cells. The non-neuronal cells further segregate into several glial types (gray shades; microglia, astrocytes, oligodendrocyte precursor cells (OPCs) and oligodendrocytes) and other cell types (brown shades; smooth muscle cells (SMCs) and endothelial cells). The neurons segregate into two major types: glutamatergic neurons (green and blue) and GABAergic neurons (orange, red and pink). Four major clusters of GABAergic neurons can be readily identified by the expression of the markers parvalbumin (*Pvalb*), somatostatin (*Sst*), vasoactive intestinal polypeptide (*Vip*) and neuron-derived neurotrophic factor (*Ndnf*). These can be further divided and, together with several rare types, make a total of 23 types of interneurons. Excitatory neurons further segregate into 19 types that correspond to cortical layers or sublayers. Overall, the visual cortex was parceled into 49 cell types based on their expression profiles. *96*Rik* is *3630013A20Rik*. Adapted from Brain Explorer and Science Vignette (http://casestudies.brain-map.org/celltax), Allen Institute for Brain Science.

Single-cell capture/isolation. Most single-cell profiling methods involve the generation of a cell suspension that is then subjected to single-cell isolation. Solid tissues are often treated with a proteolytic enzyme (for example, papain) and then dissociated by trituration into a single-cell suspension. Individual cells are isolated from the suspension by manual or automated micropipetting, fluorescence-activated cell sorting (FACS), or microfluidic devices (**Fig. 2**)[22,23]. Manual isolation requires a trained experimentalist and has low throughput, but it can be improved by a robotic micromanipulator setup[24]. FACS-mediated isolation requires proper instrument settings and controls to ensure that only individual cells are collected. Microfluidic systems permit not only single-cell capture, but can also automate some downstream biochemical reactions. The most popular microfluidic systems are the Fluidigm C1 and microdroplet technologies. Fluidigm C1 arrays contain up to 800 single-cell capture sites and permit reverse

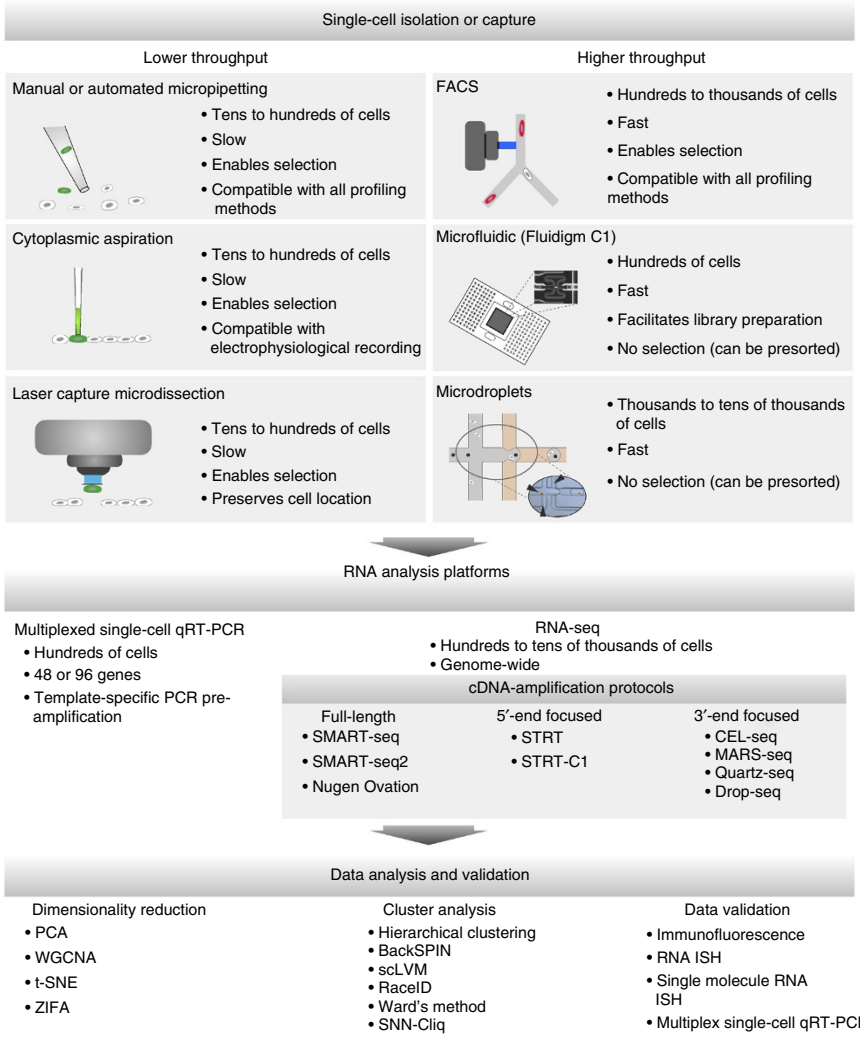

Figure 2 Experimental approaches commonly used for single-cell gene expression profiling. Several methods have been developed to isolate and capture cells for single-cell analysis, including manual or automated micropipetting, cytoplasmic aspiration, laser capture microdissection, FACS, and microfluidic and microdroplet devices. Following library generation, a limited number of genes can be assayed with multiplex qRT-PCR or the transcriptome of a cell can be characterized by RNA-seq. Several bioinformatic approaches allow clustering of cells and assigning identities on the basis of their gene expression profiles.

transcription and cDNA amplification in nano-chambers enclosing individual captured cells. New technologies (for example, ICELL8 from Wafergen) will likely further increase the throughput of microfluidic cell capture. Microdroplet systems, such as the Drop-seq[25], In-drop[26] or commercial 10X Genomics devices, use microfluidic technologies to isolate single cells in aqueous droplets in a non-aqueous suspension. These droplets also serve as individual nanoliter-scale aqueous reaction chambers for reverse transcription or PCR[27].

Cell-capture approaches based on dissociation have inherent caveats. First, the dissociation of tissue itself might lead to transcriptional changes as a result of mechanical stress, loss of cell adhesion and/or aberrant neuronal excitation. Second, some cell types might be underrepresented as a result of a low propensity to dissociate from other cells or low viability during the dissociation process. Third, these procedures result in damage to neuronal processes, thereby likely depleting RNAs that are enriched in these cellular compartments and potentially contaminating other cells with cellular remnants containing RNA. Finally, as a result of experimental imperfections, two cells, referred to as 'doublets', may be harvested in a single sample. The percentage of doublets varies considerably among different methods and studies. For example, 0.36–11.3% of doublets were observed using Drop-seq depending on cell concentration[25], 0–2.3% of doublets were reported using FACS[14,28] and up to 11–44% were observed using the Fluidigm C1 microfluidic device[25,29]. When employing Fluidigm C1, doublet-containing wells can be removed from analysis after visual inspection of the cell-capture array and, in an optimized version of microfluidic chips, the doublets have been reduced to ~3% (Fluidigm white paper).

Methods that do not require tissue dissociation are more difficult to implement and generally have a lower throughput, but can provide additional information about the cells examined (for example, location in the brain, morphology, electrophysiological properties; **Fig. 2**). Aspiration of cellular content by patch pipet has been used to acquire RNA for quantitative RT-PCR (qRT-PCR)[30] or microarray analysis[31], and has recently been optimized for RNA-seq[32,33]. Laser capture microdissection can be used to collect cells from thinly sectioned tissue, but is prone to sample contamination[23,34,35]. Recently, a novel approach was designed to harvest mRNA from cells in their natural environment based on a light-activated mRNA-capture reagent but this has yet to be broadly applied[36].

cDNA amplification and generation of libraries. All presently available technologies that analyze RNA from single cells require reverse transcription and several rounds of nucleic acid amplification. The first genome-wide single-cell RNA-sequencing protocol for mammalian cells was developed in 2009 (ref. 37). Currently, several protocols exist and they differ in the full-length coverage of mRNAs, the ability to pool samples at the early stage of the process and the ability to minimize amplification biases (reviewed in refs. 38,39). SMART-seq approaches provide full-length coverage with some 3′-end bias[40], which has been decreased in a more recent SMART-seq2 protocol[41]. Other methods deliberately generate 5′-end focused STRT[42]) or 3′-end focused (i.e., CEL-seq[43], Quartz-seq[44], MARS-seq[28] and Drop-seq[25]) RNA-sequencing libraries. The end-focused methods can reduce technical variability and bias introduced during amplification by the incorporation of unique molecular identifiers (UMIs) in primers used for reverse transcription or template switching[45–47]. In some procedures, to prepare for sequencing by NGS devices, additional fragmentation steps may need to be employed (for example, sonication and adaptor ligation or transposase-based fragmentation and tagging using Nextera kits)[48]. After library preparation, samples are typically sequenced using NGS (for example, Illumina HiSeq platform).

Quality control and technical variability. When attempting to classify cells on the basis of the biologically relevant differences in their transcriptomes, it is essential to be able to account for technical noise in experimental data. The noise can be introduced at various steps of the procedure (for example, cell isolation, different processing batches, etc.) and it should be estimated and controlled as much as possible. No technique currently enables 'complete' and fully unbiased detection of all mRNA molecules in a cell. Variations in sensitivity and consistency of RNA amplification among single-cell samples are most commonly estimated by the detection of introduced external spike-in RNAs of known concentrations (for example, External RNA Controls Consortium (ERCC) mix[49]). Amplification biases can also be corrected by inclusion of UMIs[45], although not every amplification protocol is conducive to their inclusion. To date, published studies have established diverse, but frequently overlapping, quality control criteria including total number of reads, percentage of reads that map to transcriptome, number of UMIs and/or genes detected, ERCC read percentage and ERCC amplification linearity. Thoughtful quality control criteria should be employed based on the method used and types of cells collected, as the same criteria may not be valid even for all cells in a single data set. For example, glial cells express fewer genes than neurons and may appear as low-quality samples in a neuron-dominated data set if the gene number per sample is used as the key criterion for quality control[14,50]. Superimposed on the technical noise is the fact that, even among cells of the same type, RNA expression levels can vary substantially as a result of the pulsatile nature of the transcriptional process[51], an aspect that cannot be controlled for experimentally.

Cell number and sequencing depth. Two key variables to consider when designing neural cell classification experiments based on single-cell RNA-seq are the number of cells to be sampled and the sequencing depth (typically presented as the number of reads per cell or the number of transcripts per cell). These two variables are intertwined, as *in silico* simulations have shown that sampling more cells[25,50] or deeper sequencing of individual cells[14] results in better resolution of cell types. Thus, an ideal experimental design would continue to increase the number of cells and/or the depth at which each cell is sequenced until the number of cell types identified plateaus[52]. This is rarely possible in practice, as it becomes prohibitively expensive. Given the lack of standards in this new field and continuous evolution of approaches and techniques, we provide broad recommendations based on published examples.

In a piece of neural tissue, distinguishing major neuronal and non-neuronal types only requires a relatively shallow sequencing of hundreds of cells without enrichment strategies for particular cell type. For example, 301 cells sequenced to about 50,000 reads per cell were sufficient to distinguish neural progenitors, radial glia, and newborn and mature neurons from a developing human brain[53]. When cell availability is not limiting, Drop-seq is a good alternative for sequencing thousands of cells to a low depth. For example, when Drop-seq was used to study retinal cell types, rare amacrine subpopulations, which represented between 0.1–0.9% of the cells analyzed, were identified[25]. To determine how cell number influences classification, the authors of that study randomly selected increasing numbers of cells (500–44,808) from their data set and found that, as the number of sampled cells increased, the distinctions between related clusters became clearer, stronger and finer in resolution. When an enrichment strategy is possible (for example, FACS of a genetically labeled population), smaller sample sizes may be sufficient.

As there is a tradeoff between sequencing depth and cost, an important question is to what depth should every cell be sequenced. The number of reads obtained per cell have varied widely between different studies, with some studies suggesting that low coverage is sufficient, whereas others suggest that greater depth is beneficial. Consistent with the former, reevaluation of cortical single-cell transcriptomes[50] concluded that as little as 100 transcripts per cell were sufficient to distinguish neurons from oligodendrocytes, but 1,000 transcripts per cell were required to differentiate two different kinds of pyramidal neurons[54]. This study suggests that low depths are sufficient, in part because of the natural 'effective low dimensionality' of gene expression data that can be attributed to gene co-regulation by underlying transcriptional networks. In contrast, many recent reports of neuronal classification used ~1–30 million reads per cell (**Table 1**)[14,24,32,52,55,56], and one study demonstrated that subsampling of data to lower depths resulted in fewer cell types being detected[14]. Thus, highly disparate cell types can likely be separated by low coverage sequencing, but deeper sequencing appears to be beneficial for more complete classification of closely related cells. In terms of maximal gene detection, some studies have suggested that ~2 million reads are required to reliably detect >90% of all genes expressed[57,58]. Another study suggested that the number of genes detected plateaued at 30 million reads[52]. Ultimately, gene coverage is likely to be dependent on the cell types being investigated, as the number of genes expressed has been shown to vary substantially among cell types[14,50].

Data analysis. The goal of cell-type categorization experiments based on single-cell transcriptomic data is to group cells according to the similarity of their transcriptomic signatures. The final output of this data analysis is a list of cell clusters, and a set of marker genes that can robustly differentiate these clusters. This analysis involves multiple steps and can be performed by a variety of approaches, which we summarize below. We also refer readers to review articles dedicated to this topic[39,57,59].

The mRNA expression profile of a cell can be represented as a highly multidimensional vector, with each dimension corresponding to the expression of a single gene. As a result of the inherently high dimensionality of transcriptomic data, all currently used classification methods reduce the number of genes, or gene-based features, to calculate distances among cells based on their transcriptomic data. These techniques either select correlated genes directly[50] or combine genes with similar expression pattern across the data set into new dimensions[14,24,25].

Reducing the dimensionality of single-cell transcriptomic data facilitates both data visualization and cluster analysis. The genes that are used as input into dimensionality reduction are typically selected on the basis of their variance across the single-cell data set being higher than technical noise. Technical noise can be estimated by the variance of ERCCs spike-ins[60] or a statistical noise model of read count variance[61]. Most frequently used dimensionality reduction techniques include principal component analysis (PCA)[14,24], t-distributed stochastic neighbor embedding (t-SNE)[25,62] and weighted gene coexpression network analysis (WGCNA)[14,63]. More recently, a new dimensionality-reduction method, zero inflated factor analysis (ZIFA), has been developed to account for the 'dropouts' in single-cell data (that is, false quantification of a gene as being absent, possibly as a result of the corresponding transcript not being reverse-transcribed or not being detected because of limitations in sequencing depth)[64]. In addition to providing input for clustering, the dimensionality reduction techniques also allow researchers to

Table 1 Recent neural cell classification studies in various regions of the mouse nervous system

Study	Brain structure	Cell isolation or capture	Library generation	Gene detection	Number of cells	Reads per cell	Genes per cell	Bioinformatics
Poulin *et al.*[84]	Midbrain (dopamine neurons)	FACS	TSA	qRT-PCR	159	NA	NA	PCA, HCA
Spaethling *et al.*[55]	Hindbrain (serotonin neurons)	Cytoplasm aspiration	LARA	RNA-seq	20	>20 million (mapped reads per cell)	NP	NP
Okaty *et al.*[56]	Hindbrain (serotonin neurons)	FACS	SPIA (Nugen)	RNA-seq	56	18 million (AVG reads per cell)	~8,000	HCA
Chiu *et al.*[87]	DRG (neurons)	FACS	TSA	qRT-PCR	334	NA	NA	PCA, HCA
Usoskin *et al.*[24]	DRG (neurons)	Automated micropipetting	STRT	RNA-seq	622	1,14 million (AVG mapped reads per cell)	3,574 ± 2,010	PCA, SCDE
Li *et al.*[52]	DRG	Manual micropipetting	SMARTer	RNA-seq	203	58.2 million (AVG mapped reads per cell)	10,950 ± 1,218	PCA, HCA, WGCNA
Zeisel *et al.*[50]	Somatosensory cortex and hippocampus	Microfluidic; Fluidigm C1	STRT	RNA-seq	3,005	1.2 million (AVG mapped reads per cell)	1,865–4,760	BackSPIN
Tasic *et al.*[14]	Visual cortex	FACS	SMARTer	RNA-seq	1,679	8.7 million (median reads per cell)	~7,250 neurons, 4,274 non-neuronal cells	PCA, Ward's method, WGCNA
Cadwell *et al.*[33]	Cortex	Cytoplasm aspiration; Patch-seq	Smart-Seq2	RNA-seq	58	NP	~7,000	PCA, HCA
Fuzik *et al.*[32]	Cortex (CCK$^+$ neurons)	Cytoplasm aspiration; Patch-seq	STRT	RNA-seq	83	1.6 million (AVG reads per cell)	2,068	NP
Macosko *et al.*[25]	Retina	Microdroplets; Drop-seq	STAMP	RNA-seq	44,808	NP	>900 genes for 13,155 cells, >900 genes for 36,145 cells	PCA, HCA

HCA, hierarchical clustering analysis; LARA, linear antisense RNA amplification; NA, not applicable; NP, not provided; SPIA, single primer isothermal amplification; SCDE, single-cell differential expression; TSA, template-specific amplification.

visualize the complex RNA-seq data by focusing on major gene expression trends in the data set. Additional methods have been developed specifically for visualization purposes (for example, sorting points into neighborhood, SPIN)[65].

Clustering methods aim to separate cells into groups on the basis of distances in the reduced-feature space, so that cells in a group are more similar to each other than cells between groups. Many approaches exist to evaluate within- and between-cluster similarity: hierarchical clustering, centroid-based clustering (K-means)[39], BackSPIN[50], scLVM[66], RaceID[67] and SNN-Cliq[68]. The latter approach, SNN-Cliq, employs secondary similarities that are shared by nearest neighbor to attribute cells to specific clusters and has been shown to be effective at identifying rare cell types. An inherent problem in clustering is that no algorithm that is completely unsupervised exists; all algorithms rely on the investigator to select statistical limits or criteria (often in retrospect, after preliminary clustering with initially chosen parameters) that will label two clusters as different. Regardless of the strategy for single-cell transcriptome analysis, dimensionality reduction and clustering approaches can be applied iteratively to allow the use of different sets of differentially expressed genes for clustering at different levels of cell relatedness. Finally, additional tests are usually performed to confirm the robustness of produced clusters. This can be done with machine-learning algorithms trained on subsets of the data[14], by comparing expression between clusters[69] or by broken-stick regression[70].

Once the clusters are defined, further analyses can be used to confirm and identify differentially expressed genes. For example, Tasic *et al.*[14] applied the DESeq package[71], whereas Macosko *et al.*[25] used a package designed for single-cell qRT-PCR data to compute *P* values and assess significance. Given that single-cell transcriptomic data contain many 'dropouts', methods based on Bayesian statistics that take dropouts into account may be particularly well suited for defining differentially expressed genes between single-cell clusters[50,72,73]. Recently, software packages have been published that unite many steps of the bioinformatics analysis of single-cell transcriptomics data sets[74,75]. In general, although there are several existing methods and packages to perform these analyses, few of the commonly used packages were written specifically for single-cell RNA-seq, and a consensus on which method to use has yet to emerge.

Data validation using complementary approaches. The specific gene coexpression patterns discovered by single-cell RNA-seq can be confirmed using a variety of methods, either at the RNA level (qRT-PCR and RNA ISH) or at the protein level by immunolabeling (**Fig. 2**). qRT-PCR technologies have markedly advanced in the last decade, mostly as a result of the advancement of microfluidic technologies[76]. For example, the high-throughput Fluidigm Biomark system employs a microfluidic chip to simultaneously examine expression of 96 genes from 96 cells. The classical RNA ISH methods have limited multiplexing ability and sensitivity, but preserve the location of the cell in the tissue. Recently, highly sensitive single-molecule RNA ISH protocols have been employed to validate the transcriptomic data[50,51,56]. Co-immunolabeling has also been widely used, but discrepancies with RNA-based analyses could exist as a result of translational or post-translational regulation[77].

Examples of neural cell type classification

Many studies have used single-cell gene expression profiling methods to classify cell types in multiple tissues[28,29,67,78–82]. Here we will review recent studies that have exploited these technologies to classify neural cell types in various regions of the mouse nervous system.

Monoaminergic systems. The dopaminergic (DA) system has been implicated in a wide range of behaviors, including movement, reward and learning[83]. Recent studies have suggested the presence of physiologically diverse DA neurons in the neuroanatomical boundaries of the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA)[83]. To explore the molecular underpinnings of DA neuron heterogeneity, one group of researchers isolated individual fluorescently labeled DA neurons by FACS and analyzed them using a high-throughput qRT-PCR Fluidigm Biomark system for the expression of 96 genes[84], which were selected from population-based microarray studies[85]. Six types of DA neurons were identified, five of which could be localized by immunofluorescence and RNA ISH and had unique anatomical distributions in the VTA or SNc[84]. One of the neuron types, defined by the expression of *Vip*, was shown to send a specific projection to nuclei in the extended amygdala, suggesting that molecularly defined DA neuron types have unique anatomical properties.

Serotoninergic neurons have been implicated in a variety of behaviors, including temperature homeostasis, breathing, aggression and anxiety[86]. To uncover serotonergic neuron diversity, two studies analyzed these neurons at the single-cell level. One study compared the transcriptomes of cells located in different regions of the dorsal raphe (DR) nucleus and demonstrated differences in ventromedial versus lateral DR neurons, especially with respect to the expression of G-protein-coupled receptors[55]. A second study exploited the fact that, during development, serotonergic neurons arise from distinct transverse domains in the hindbrain, known as rhombomeres[56]. This study used an intersectional genetic approach to label subpopulations of serotonergic neurons derived from five specific hindbrain rhombomeres, performed RNA-seq on pooled cells and demonstrated unique expression profiles for neurons from different rhombomeres. To complement these experiments, the authors analyzed individual cells by RNA-seq. The single-cell data confirmed the pooled-cell analysis, but also revealed additional types in the rhombomere-2-derived serotonergic neurons. The existence of these two subpopulations was confirmed by characterizing different electrophysiological responses to cognate ligands of the differentially expressed receptors Tacr3 and Oxtr. In summary, serotonergic neuron types correlate with the rhombomere of origin, but single-cell analyses have provided additional resolution for neuronal types that did not depend on prior knowledge of embryonic origins.

Dorsal root ganglia. The somatosensory nervous system is composed of multiple neuronal subtypes with distinct conduction and innervation properties that are responsible for tactioception, nociception, thermoception, proprioception and pruriception. Three studies have used single-cell profiling approaches to examine the diversity of somatosensory neurons from the dorsal root ganglia (DRG) and have obtained partially overlapping conclusions[24,52,87]. In one study[87], the authors used the Pvalb-Cre and SNS-Cre transgenes,in conjunction with isolectin B4 labeling (IB4), to initially isolate three populations of DRG neurons by FACS. They determined the transcriptomes of pooled neurons, which revealed new markers with differential expression. Then, using these differentially expressed markers, they performed single-cell high-throughput qRT-PCR, which resulted in the identification of seven distinct subgroups (**Fig. 3a**). In another study[24], DRG cells were collected using an in-house robotic cell-picking setup. The analysis initially revealed one non-neuronal and four neuronal clusters based on the expression of several known marker genes. The four neuronal clusters were further subdivided by iterative PCA into 11 neuronal types. The most recent study revealed ten broad groups of somatosensory neurons by single-cell RNA-seq, which could be further subdivided into a total of 17 distinct cell types[52]. The authors also analyzed the relationship between cell types and neuron size and found that neuron types segregated according to size: small and large neurons could be sorted into 11 and 6 types, respectively. In addition, they ascribed functional phenotypes to neuron types by patch-clamp recordings of single DRG neurons after various stimuli, combined with single-cell qRT-PCR analysis for key diagnostic markers. They were therefore able to develop a classification scheme based on multiple parameters. In summary, these three studies have identified some concordant cell types, but some types may not have clear counterparts, likely because of different cell sampling, gene expression profiling and/or data-processing approaches.

Cortex. Cortical neuron diversity has been explored using traditional approaches[88–90], but recent studies have used single-cell RNA-seq to reveal the heterogeneity in different cortical regions[14,32,33,50]. To define diversity in the somatosensory cortex and hippocampus, the first study used a microfluidic cell-capture platform (Fluidigm C1) to harvest individual cells from these regions for RNA-seq (**Fig. 3b**)[50]. Nine major clusters were revealed by biclustering, accounting for cortical and hippocampal pyramidal neurons, interneurons, oligodendrocytes, astrocytes, microglia, vascular endothelial cells, mural cells, and ependymal cells. Biclustering was repeated on these major classes, revealing the existence of 47 molecularly distinct cell types. This study demonstrated the extent of non-neuronal diversity, including six subpopulations of oligodendrocytes at different stages of maturation and two types of astrocytes that could be discriminated by differential expression of *Gfap* and *Mfge8*. The most heterogeneous class of cells consisted of interneurons, with 16 types being identified in the somatosensory cortex and hippocampus, although they comprised only 10% of the cells analyzed. With regard to pyramidal cells, seven subclasses were observed, which were mainly layer specific, in accordance with previous literature[91]. The expression of many transcription factors was specific to distinct cell types, consistent with their role in the establishment and maintenance of cellular identity[8,9]. A second study examined diversity in the mouse primary visual cortex[14], using 24 different transgenic recombinase lines to fluorescently label discrete groups of cells for single-cell isolation by FACS. The use of transgenic lines allowed access to some rare cell types and deliberately biased the sampling toward neuronal types. The authors identified 49 clusters, of which 23 were GABAergic, 19 glutamatergic and 7 non-neuronal (**Fig. 1**). Overall, the types described agreed well with previously suggested major divisions: glutamatergic neurons were divided by layer, while GABAergic neurons were divided by previously known markers (*Sst*, *Pvalb*, and *Vip*). For all major neuronal types, subtypes were revealed, some of which had not been previously reported.

Overall, the two studies partially overlap in their identification of cell types, especially where highly expressed unique markers were detected. A major difference between these two studies is the computational approaches that were employed to reveal cell type identity. In one study[14], cells were allowed to have more than one identity, and were ultimately divided into cells that were always classified into the same type (named core cells) and cells that could have more than one identity (named intermediate cells). The distribution of intermediate cells (~15% of all cells) paints the overall phenotypic landscape of cortical cell types as a combination of continuity and discreteness. These two studies also differed in a number of experimental and data analysis parameters: cell isolation, RNA-seq procedures, sequencing depths, and the genetic background and age of the analyzed mice. The comparison between the two studies[14] exemplifies the fact that

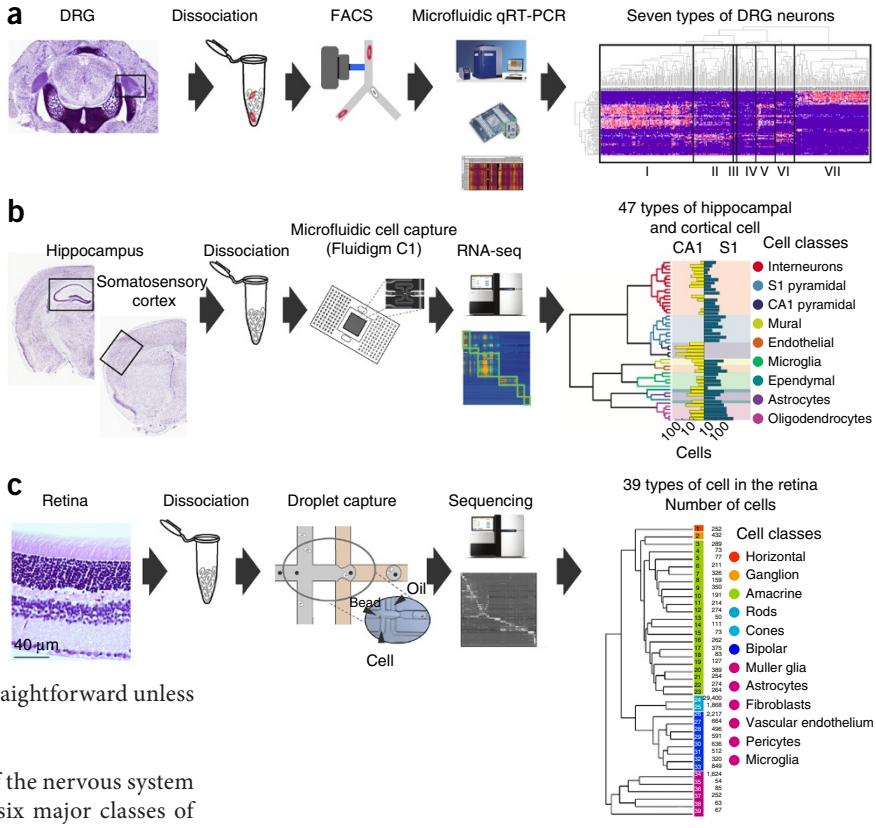

Figure 3 Examples of approaches for neural cell classification in various regions of the nervous system. (**a**) Chiu and colleagues[87] dissected DRG from two distinct Cre lines and isolated single neurons by FACS. Using single-cell qRT-PCR on a microfluidic array, they revealed the presence of seven types of DRG neurons. Adapted from ref. 87. (**b**) Zeisel and colleagues[50] dissociated the hippocampus and somatosensory cortex of adult mouse and used a microfluidic chip to capture single cells. After cell lysis, a library was generated for RNA-seq, sequenced, and cells were clustered into nine major classes and a total of 47 hippocampal and cortical cell types. Adapted from ref. 50, AAAS. (**c**) Macosko and colleagues[25] dissociated mouse retina and used a droplet-based method to capture cells. The RNA of each cell was captured with microbeads and libraries were generated. They identified 12 major cell classes that could be subdivided into 39 retinal cell types. Adapted from ref. 25, Cell Press. Image of the retina adapted from ref. 146.

establishment of cell-type concordance is not straightforward unless the same combinations of markers are detected.

Retina. The retina is a highly specialized part of the nervous system that processes visual information. It contains six major classes of neurons: rod, cone, ganglion, horizontal, amacrine and bipolar cells, many of which can be further subdivided[92]. Several pioneering studies used manually collected single cells from the developing retina and genome-wide transcriptional microarrays to examine the molecular diversity of its cell types[20,93]. Many new molecular markers were discovered, but the small number of cells analyzed in these studies prevented a comprehensive characterization of cellular diversity.

Recently, retinal cell heterogeneity was investigated by Drop-seq (**Fig. 3c**)[25]. Based on the analysis of 44,808 cells, the authors defined 39 transcriptionally distinct clusters. Rods, cones, retinal ganglion cells (RGCs) and horizontal cells each accounted for one type. Bipolar cells and amacrine cells displayed more diversity, accounting for 8 and 21 types, respectively. The remaining cell types corresponded to Müller glia, astrocytes, resident microglia, pericytes, endothelial cells and fibroblasts. The clusters varied in size, from 50 to 29,400 cells, and the proportion of each type observed generally matched anatomical estimates from the literature. However RGCs, which are thought to represent ~30 subtypes[94], initially clustered as a single entity despite the large numbers of cells sequenced. Although supervised clustering could divide this group into two subsets, isolating RGCs from transgenic mice in which this broad cell class is specifically labeled[95] would be more suitable for resolving their molecular diversity. Overall, through sequencing a very large number of cells at a limited sequencing depth, Drop-seq deconstructed the retina into its major types.

The implications of cell-type classification

Inferring biological properties of neural cell types. The key output of single-cell transcriptomics experiments is a list of molecularly distinct cell types along with a set of genes expressed in each type. The list usually contains some previously identified and some new cell types, both revealed in unprecedented molecular detail. For instance, two cortical single-cell profiling studies[14,50] revealed the existence of a new transcriptomically defined interneuron type expressing, uniquely in the cortex, the marker gene chondrolectin, *Chodl* (named Sst-Chodl or Int1 type). These studies also showed that this cell type specifically expresses many other genes known to be cell surface receptors that likely contribute to the unique physiology of these cells (for example, *Hcrtr1*, *Chrm2* and *Gpr126*). Indeed, the molecular profiles of distinct neuronal types provide a plethora of information that can be used to formulate a specific hypothesis on the ability of a particular type of neuron to integrate, process and respond to synaptic and extracellular stimuli. For example, single-cell transcriptomics of DRG neurons revealed a particular cell type uniquely expressing the *Il31r* receptor[24], which is responsible for pruritus. In contrast, other functionally relevant molecules such as the *Trpa1* channel, which mediates responses to temperature, may be expressed in multiple cell types[24,52]. In this case, although different types of neurons may be activated by the same stimulus, it is conceivable that integration of the signal may differ depending on the unique molecular milieu or connectivity of each neuronal type. Eventually, coupling electrophysiological neuronal recording with single-cell gene expression profiling will be instrumental to understanding the molecular correlates of complex biophysical and electrical properties of distinct neurons[32,33]. Indeed, as computational models keep evolving, we can envision that specific physiological properties of a neuron type could be inferred directly from its molecular profile[96,97].

Mapping connectivity of neuronal types. Single-cell gene expression profiling will facilitate the mapping of the mammalian connectome in many important ways. First, a neuronal taxonomy will provide a comprehensive list of the fundamental components of the nervous system, whose connectivity needs to be established. Without such a list, it is impossible to assess whether a connectivity pattern associated with a group of neurons could be further divided as a result of the heterogeneity in that group. Second, a neuronal taxonomy will expand

and refine the list of markers that enable specific access to cell types in the nervous system. These markers can serve as valuable genetic entry points for generating Cre drivers based on specific genomic loci[3,98]. With such drivers to target specific neuronal types, it is possible to elucidate their connectivity using genetically modified viruses harboring fluorescent proteins or transynaptic labeling capabilities[3,99–101]. Finally, completely different methodological approaches for resolving the mammalian connectome, as proposed by Zador and collaborators, might convert the brain connectivity into a problem that can be solved by single-cell sequencing[102]. One key to such approaches is generating barcodes unique to each neuron, or even to each synapse formed between two neuron types, and resolving these barcodes by sequencing. Whatever the approach used, a comprehensive neuronal taxonomy will be essential for resolving and interpreting the connectivity matrix of neuronal types.

Determining the contribution of neuronal types to behavior. To understand the function of neuronal types, researchers have developed a wide range of genetically encoded tools to visualize neuronal activity[103] and manipulate neuronal excitability[104,105]. These genetic tools are most useful if precisely targeted to a unique neuronal type. A thorough cell-type classification is essential for assessing the specificity of Cre and other recombinase lines being used to target optogenetic and chemogenetic effectors. For instance, of the 24 recombinase lines in the single-cell RNA-seq study of the visual cortex, only one appears to be specific to a single transcriptomically defined cell type[14]. Thus, individual recombinase driver lines might not be sufficient to unambiguously target some neuronal types, and an increased specificity can be achieved by employing intersectional genetic paradigms using additional recombinase drivers, such as Flp and Dre[106–110]. Only with the list of the cellular components of the nervous system, and the markers to genetically access them that are provided by single-cell profiling experiments, can we hope to assign specific behavioral contributions of neuron types.

Understanding selective neurodegeneration. Transcriptomic profiles of cell types would provide an inroad toward understanding selective neuronal vulnerability observed in many neurodegenerative disorders. In Parkinson's disease (PD), for example, stereotypic neuron loss is observed in several areas of the brain, including the loss of midbrain DA neurons[111,112]. Even in the DA neuron clusters, neurons located in the ventral tier of the SNc appear to be more vulnerable than dorsal tier SNc or VTA neurons. Single-cell profiling of DA neurons revealed a neuronal type defined by the coexpression of *Sox6* and *Aldh1a1* that is located in the ventral tier of the SNc and appears to be selectively vulnerable in a toxin model of PD[84], as well as in post mortem PD brains[113]. The selective vulnerability of specific neuron types is likely a result of the unique molecular properties of the affected cells. Thus, obtaining the complete profiles of the vulnerable cell types will provide insights into why certain types are prone to degeneration[114]. Identifying the molecular culprits of vulnerability could lead to therapeutic approaches to prevent or reduce selective neurodegeneration.

Identifying new therapeutic targets for neuropsychiatric disorders. Complete molecular profiles of specific neuronal types will provide a foundation for the development of targeted pharmacological therapies for neuropsychiatric diseases. For many such diseases, current treatment regimens are often associated with adverse side effects, likely as a result of the effect of therapeutics on cell types not related to the disease. For example, major depressive disorder is commonly treated with selective serotonin reuptake inhibitors, and adverse effects such as akathisia, suicidal ideation, sexual dysfunction and photosensitivity have been reported[115]. With the revelation of the molecular profiles of several serotonergic types[55,56], it is conceivable that pharmacological targeting of specific types of serotonergic neurons, or their cognate targets, may result in more effective antidepressant drugs with reduced side effects. Thus, understanding neuronal taxonomy in depth would not only expedite the study of circuits disrupted in neuropsychiatric diseases, but could also provide molecular targets for circuit-specific interventions.

Modeling diseases using iPS- and iN-derived neurons. Neurons differentiated *in vitro* from patient-derived induced pluripotent stem (iPS) cells or fibroblasts (iN) have been used to model various nervous system diseases, including amyotrophic lateral sclerosis, PD and Alzheimer's disease[116–118]. A key shortcoming of these approaches is that iPS- or iN-derived neuronal cultures are heterogeneous, making it challenging to validate the authenticity of these neurons by bulk RNA expression profiling techniques. Single-cell transcriptional profiling will inform differentiation protocol design and enable more accurate neuronal authentication. For example, single-cell profiling of primary neuronal tissues at different stages of development would reveal key transcriptional cascades underpinning cell fate decisions, and the factors involved could be used to improve protocols to generate specific neuronal types. Once neurons of a certain type have been generated *in vitro*, single-cell profiling can be used to validate the homogeneity and authenticity of the cultured neurons. These approaches will also be instrumental in characterizing the cellular complexity of stem cell derived organoids of various brain regions[119].

Future applications of single-cell technologies

In the next few years, innovative approaches that can provide single-cell transcriptomic or highly multiplexed gene expression information *in situ* will be transformative. They will circumvent the need for tissue dissociation and will allow profiling of individual cells in accurate proportions while preserving spatial information[120,121]. For instance, highly multiplexed single molecule fluorescence ISH[122,123] or *in situ* RNA sequencing[121,124] have been used to assess the expression of up to several thousands of genes while preserving spatial information. Another approach could be to exploit nanopore-based sequencing devices to directly sequence DNA or RNA molecules from a tissue section[120]. For instance, an electrophoretic system could force nucleic acid molecules from a brain section directly into high-density nanopores and could theoretically preserve cellular resolution while providing a high-throughput single-cell molecular characterization. Together, these advances will move us closer to the goal of providing a complete catalog of cell types in the nervous system.

The classification of neural cell types will likely be complemented by other single-cell '-omics' approaches, including genomics, epigenomics and proteomics[125–130]. Further development of single-cell proteomic approaches[127] would be very valuable, as mRNA presence is not always coupled with protein expression[77]. In the case of genomics, single-cell sequencing has revealed that, contrary to the general conception, the genomes of most cells in an organism harbor somatic mutations[126,130]. These mutations accumulate sporadically starting in the zygote and continue throughout the life of the organism, endowing every cell with a 'genetic tattoo'. This phenomenon has been exploited to reconstitute the lineage tree of human cortical cell populations[126]. Similarly, genomic mutations could be induced by targeted genome editing[131] to extract lineage information of neural cell types. In this regard, it has recently been reported that both genome and

transcriptome can be simultaneously obtained for single cells[132,133]. This opens up the possibility that lineage trees and transcriptome-based cell classification could be obtained simultaneously, and could help to elucidate the developmental processes underlying the generation of the immense cellular diversity in the nervous system.

Single-cell transcriptomics is poised to advance neuroscience, even beyond its utility in neuronal classification in healthy adult tissue. Single-cell RNA-seq can be performed in developing neural tissues, various genetic mutants, disease models or in response to pharmacological treatments. In such situations, monitoring gene expression with single-cell resolution presents a more accurate picture than the evaluation of gene expression changes from bulk tissues, as variation in the expression of a given gene could otherwise be occluded by bulk measurements[13]. Single-cell analysis also allows one to decipher gene regulatory networks through coexpression analyses. For instance, Macosko *et al.* identified many genes that were not previously connected to the cell cycle by virtue of their coexpression with known cell cycle genes in individual cells[25]. Recent studies have also used single-cell profiling to examine the process of developmental refinement of odorant receptor expression[134], the establishment of cortical connectivity[135,136] or define cellular diversity in the human brain[53,137–140]. Single-cell transcriptomics is therefore poised to revolutionize the study of many neuronal phenomena, which have too long been hindered by the immense cellular complexity of the brain.

Conclusion

We have reviewed the merits and challenges of single-cell transcriptomic approaches for classification of neural cell types in various parts of the nervous system. Moving forward, it is essential that cell types defined by transcriptomics be evaluated for other characteristics, such as electrophysiological properties and connectivity. For this to be feasible, methods to derive different types of information from the same cell need to be employed (for example, Patch-seq). In addition, new genetic or viral tools that will enable specific access to unique transcriptomically defined cell types need to be developed[141–143]. These tools, in conjunction with platforms to visualize or manipulate neuronal activity, will help to delineate the connections and functions of each neuronal type. Thus, it will be possible to determine the function of each neuronal type, one at a time, just as conditional knockout technology has allowed us to decipher gene functions, one gene at a time. This one-neuron-type-at-a-time approach can eventually be partnered with complementing technologies that allow recordings or imaging of ensembles of neurons[144,145], toward understanding emerging properties of complex neural networks.

ACKNOWLEDGMENTS

We would like to thank L. Laboissonniere, R. Chowdhury, N. Skene, V. Menon and G. Caronia-Brown for comments on the manuscript. J.-F.P. is supported by FRSQ and CIHR. R.A. is supported by the Paul Ruby Foundation, NARSAD, NIH R21NS092034-01A1 and R01 NS096240-01. B.T. is supported by the Allen Institute for Brain Science. J.H.-L. is supported by Swedish Research Council (VR), StratNeuro, Hjärnfonden, and EU FP7/Marie Curie Actions.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Arendt, D. The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* **9**, 868–882 (2008).
2. Nelson, S.B., Sugino, K. & Hempel, C.M. The problem of neuronal cell types: a physiological genomics approach. *Trends Neurosci.* **29**, 339–345 (2006).
3. Luo, L., Callaway, E.M. & Svoboda, K. Genetic dissection of neural circuits. *Neuron* **57**, 634–660 (2008).
4. Bota, M. & Swanson, L.W. The neuron classification problem. *Brain Res. Rev.* **56**, 79–88 (2007).
5. Rowe, M.H. & Stone, J. Naming of neurones. Classification and naming of cat retinal ganglion cells. *Brain Behav. Evol.* **14**, 185–216 (1977).
6. Masland, R.H. Neuronal cell types. *Curr. Biol.* **14**, R497–R500 (2004).
7. Fishell, G. & Heintz, N. The neuron identity problem: form meets function. *Neuron* **80**, 602–612 (2013).
8. Holmberg, J. & Perlmann, T. Maintaining differentiated cellular identity. *Nat. Rev. Genet.* **13**, 429–439 (2012).
9. Deneris, E.S. & Hobert, O. Maintenance of postmitotic neuronal cell identity. *Nat. Neurosci.* **17**, 899–907 (2014).
10. McKenna, N.J. & O'Malley, B.W. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108**, 465–474 (2002).
11. Lein, E.S. *et al.* Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**, 168–176 (2007).
12. Gong, S. *et al.* A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**, 917–925 (2003).
13. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res.* **25**, 1491–1498 (2015).
14. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
15. Brady, G., Barbara, M. & Iscove, N.N. Representative *in vitro* cDNA amplification from individual hemopoietic cells and colonies. *Methods Mol. Cell. Biol.* **2**, 17–25 (1990).
16. Eberwine, J. *et al.* Analysis of gene expression in single live neurons. *Proc. Natl. Acad. Sci. USA* **89**, 3010–3014 (1992).
17. Dulac, C. & Axel, R. A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83**, 195–206 (1995).
18. Tietjen, I. *et al.* Single-cell transcriptional analysis of neuronal progenitors. *Neuron* **38**, 161–175 (2003).
19. Klein, C.A. *et al.* Combined transcriptome and genome analysis of single micrometastatic cells. *Nat. Biotechnol.* **20**, 387–392 (2002).
20. Trimarchi, J.M. *et al.* Molecular heterogeneity of developing retinal ganglion and amacrine cells revealed through single cell gene expression profiling. *J. Comp. Neurol.* **502**, 1047–1065 (2007).
21. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* **14**, 618–630 (2013).
22. Okaty, B.W., Sugino, K. & Nelson, S.B. Cell type-specific transcriptomics in the brain. *J. Neurosci.* **31**, 6939–6943 (2011).
23. Handley, A., Schauer, T., Ladurner, A.G. & Margulies, C.E. Designing cell-type-specific genome-wide experiments. *Mol. Cell* **58**, 621–631 (2015).
24. Usoskin, D. *et al.* Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat. Neurosci.* **18**, 145–153 (2015).
25. Macosko, E.Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
26. Klein, A.M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
27. Hindson, B.J. *et al.* High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal. Chem.* **83**, 8604–8610 (2011).
28. Jaitin, D.A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
29. Shalek, A.K. *et al.* Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* **510**, 363–369 (2014).
30. Rossier, J. *et al.* Cortical fast-spiking parvalbumin interneurons enwrapped in the perineuronal net express the metallopeptidases Adamts8, Adamts15 and Neprilysin. *Mol. Psychiatry* **20**, 154–161 (2015).
31. Subkhankulova, T., Yano, K., Robinson, H.P.C. & Livesey, F.J. Grouping and classifying electrophysiologically-defined classes of neocortical neurons by single cell, whole-genome expression profiling. *Front. Mol. Neurosci.* **3**, 10 (2010).
32. Fuzik, J. *et al.* Integration of electrophysiological recordings with single-cell RNA-seq data identifies neuronal subtypes. *Nat. Biotechnol.* **34**, 175–183 (2016).
33. Cadwell, C.R. *et al.* Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat. Biotechnol.* **34**, 199–203 (2016).
34. Lovatt, D., Bell, T. & Eberwine, J. Single-neuron isolation for RNA analysis using pipette capture and laser capture microdissection. *Cold Spring Harb. Procs* **2015**, pdb.prot072439 (2015).
35. Okaty, B.W., Sugino, K. & Nelson, S.B. A quantitative comparison of cell-type-specific microarray gene expression profiling methods in the mouse brain. *PLoS One* **6**, e16493–e16493 (2011).
36. Lovatt, D. *et al.* Transcriptome *in vivo* analysis (TIVA) of spatially defined single cells in live tissue. *Nat. Methods* **11**, 190–196 (2014).
37. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
38. Grün, D. & van Oudenaarden, A. Design and analysis of single-cell sequencing experiments. *Cell* **163**, 799–810 (2015).
39. Kolodziejczyk, A.A., Kim, J.K., Svensson, V., Marioni, J.C. & Teichmann, S.A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
40. Ramsköld, D. *et al.* Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
41. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).

42. Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5′ end sequencing. *Nat. Protoc.* **7**, 813–828 (2012).
43. Hashimshony, T., Wagner, F., Sher, N. & Yanai, I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* **2**, 666–673 (2012).
44. Sasagawa, Y. *et al.* Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **14**, R31 (2013).
45. Islam, S. *et al.* Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* **11**, 163–166 (2014).
46. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011).
47. Grün, D., Kester, L. & van Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11**, 637–640 (2014).
48. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
49. Baker, S.C. *et al.* The External RNA Controls Consortium: a progress report. *Nat. Methods* **2**, 731–734 (2005).
50. Zeisel, A. *et al.* Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).
51. Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y. & Tyagi, S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* **4**, e309 (2006).
52. Li, C.-L. *et al.* Somatosensory neuron types identified by high-coverage single-cell RNA-sequencing and functional heterogeneity. *Cell Res.* **26**, 83–102 (2016).
53. Pollen, A.A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058 (2014).
54. Heimberg, G., Bhatnagar, R., El-Samad, H. & Thomson, M. Low dimensionality in gene expression data enables the accurate extraction of transcriptional programs from shallow sequencing. *Cell Syst.* **2**, 239–250 (2016).
55. Spaethling, J.M. *et al.* Serotonergic neuron regulation informed by *in vivo* single-cell transcriptomics. *FASEB J.* **28**, 771–780 (2014).
56. Okaty, B.W. *et al.* Multi-scale molecular deconstruction of the serotonin neuron system. *Neuron* **88**, 774–791 (2015).
57. Bacher, R. & Kendziorski, C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* **17**, 63 (2016).
58. Wu, A.R. *et al.* Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11**, 41–46 (2014).
59. Stegle, O., Teichmann, S.A. & Marioni, J.C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
60. Brennecke, P. *et al.* Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).
61. Fan, J. *et al.* Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13**, 241–244 (2016).
62. Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
63. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, 17 (2005).
64. Pierson, E. & Yau, C. ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. *Genome Biol.* **16**, 241 (2015).
65. Tsafrir, D. *et al.* Sorting points into neighborhoods (SPIN): data analysis and visualization by ordering distance matrices. *Bioinformatics* **21**, 2301–2308 (2005).
66. Buettner, F. *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160 (2015).
67. Grün, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255 (2015).
68. Xu, C. & Su, Z. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics* **31**, 1974–1980 (2015).
69. McDavid, A. *et al.* Data exploration, quality control and testing in single-cell qPCR-based gene expression experiments. *Bioinformatics* **29**, 461–467 (2013).
70. Cangelosi, R. & Goriely, A. Component retention in principal component analysis with application to cDNA microarray data. *Biol. Direct* **2**, 2 (2007).
71. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
72. Kharchenko, P.V., Silberstein, L. & Scadden, D.T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11**, 740–742 (2014).
73. Vallejos, C.A., Richardson, S. & Marioni, J.C. Beyond comparisons of means: understanding changes in gene expression at the single-cell level. *Genome Biol.* **17**, 70 (2016).
74. Diaz, A. *et al.* SCell: integrated analysis of single-cell RNA-seq data. *Bioinformatics* **32**, 2219–2220 (2016).
75. Guo, M., Wang, H., Potter, S.S., Whitsett, J.A. & Xu, Y. SINCERA: a pipeline for single-cell RNA-Seq profiling analysis. *PLOS Comput. Biol.* **11**, e1004575 (2015).
76. Spurgeon, S.L., Jones, R.C. & Ramakrishnan, R. High throughput gene expression measurement with real time PCR in a microfluidic dynamic array. *PLoS One* **3**, e1662–e1662 (2008).
77. Vogel, C. & Marcotte, E.M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012).
78. Xue, Z. *et al.* Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013).
79. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90–94 (2011).
80. Tang, F. *et al.* Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* **6**, 468–478 (2010).
81. Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209–1222 (2012).
82. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**, 371–375 (2014).
83. Roeper, J. Dissecting the diversity of midbrain dopamine neurons. *Trends Neurosci.* **36**, 336–342 (2013).
84. Poulin, J.-F. *et al.* Defining midbrain dopaminergic neuron diversity by single-cell gene expression profiling. *Cell Rep.* **9**, 930–943 (2014).
85. Grimm, J., Mueller, A., Hefti, F. & Rosenthal, A. Molecular basis for catecholaminergic neuron diversity. *Proc. Natl. Acad. Sci. USA* **101**, 13891–13896 (2004).
86. Niederkofler, V., Asher, T.E. & Dymecki, S.M. Functional interplay between dopaminergic and serotonergic neuronal systems during development and adulthood. *ACS Chem. Neurosci.* **6**, 1055–1070 (2015).
87. Chiu, I.M. *et al.* Transcriptional profiling at whole population and single cell levels reveals somatosensory neuron molecular diversity. *eLife* **3**, e04660 (2014).
88. Kepecs, A. & Fishell, G. Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014).
89. Ascoli, G.A. *et al.* Petilla Interneuron Nomenclature Group et al. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat. Rev. Neurosci.* **9**, 557–568 (2008).
90. Lodato, S. & Arlotta, P. Generating neuronal diversity in the mammalian cerebral cortex. *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015).
91. Greig, L.C., Woodworth, M.B., Galazo, M.J., Padmanabhan, H. & Macklis, J.D. Molecular logic of neocortical projection neuron specification, development and diversity. *Nat. Rev. Neurosci.* **14**, 755–769 (2013).
92. Masland, R.H. The neuronal organization of the retina. *Neuron* **76**, 266–280 (2012).
93. Cherry, T.J., Trimarchi, J.M., Stadler, M.B. & Cepko, C.L. Development and diversification of retinal amacrine interneurons at single cell resolution. *Proc. Natl. Acad. Sci. USA* **106**, 9495–9500 (2009).
94. Sanes, J.R. & Masland, R.H. The types of retinal ganglion cells: current status and implications for neuronal classification. *Annu. Rev. Neurosci.* **38**, 221–246 (2015).
95. Siegert, S. *et al.* Genetic address book for retinal cell types. *Nat. Neurosci.* **12**, 1197–1204 (2009).
96. Markram, H. *et al.* Reconstruction and simulation of neocortical microcircuitry. *Cell* **163**, 456–492 (2015).
97. O'Leary, T., Sutton, A.C. & Marder, E. Computational models in the age of large datasets. *Curr. Opin. Neurobiol.* **32**, 87–94 (2015).
98. Huang, Z.J. & Zeng, H. Genetic approaches to neural circuits in the mouse. *Annu. Rev. Neurosci.* **36**, 183–215 (2013).
99. Beier, K.T. *et al.* Anterograde or retrograde transsynaptic labeling of CNS neurons with vesicular stomatitis virus vectors. *Proc. Natl. Acad. Sci. USA* **108**, 15414–15419 (2011).
100. Callaway, E.M. & Luo, L. Monosynaptic circuit tracing with glycoprotein-deleted rabies viruses. *J. Neurosci.* **35**, 8979–8985 (2015).
101. Miyamichi, K. *et al.* Cortical representations of olfactory input by trans-synaptic tracing. *Nature* **472**, 191–196 (2011).
102. Zador, A.M. *et al.* Sequencing the connectome. *PLoS Biol.* **10**, e1001411 (2012).
103. Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat. Methods* **6**, 875–881 (2009).
104. Fenno, L., Yizhar, O. & Deisseroth, K. The development and application of optogenetics. *Annu. Rev. Neurosci.* **34**, 389–412 (2011).
105. Sternson, S.M. & Roth, B.L. Chemogenetic tools to interrogate brain functions. *Annu. Rev. Neurosci.* **37**, 387–407 (2014).
106. Awatramani, R., Soriano, P., Rodriguez, C., Mai, J.J. & Dymecki, S.M. Cryptic boundaries in roof plate and choroid plexus identified by intersectional gene activation. *Nat. Genet.* **35**, 70–75 (2003).
107. Dymecki, S.M. & Kim, J.C. Molecular neuroanatomy's "Three Gs": a primer. *Neuron* **54**, 17–34 (2007).
108. Madisen, L. *et al.* Transgenic mice for intersectional targeting of neural sensors and effectors with high specificity and performance. *Neuron* **85**, 942–958 (2015).
109. Fenno, L.E. *et al.* Targeting cells with single vectors using multiple-feature Boolean logic. *Nat. Methods* **11**, 763–772 (2014).
110. Plummer, N.W. *et al.* Expanding the power of recombinase-based labeling to uncover cellular diversity. *Development* **142**, 4385–4393 (2015).
111. Brichta, L. & Greengard, P. Molecular determinants of selective dopaminergic vulnerability in Parkinson's disease: an update. *Front. Neuroanat.* **8**, 152 (2014).
112. Sulzer, D. & Surmeier, D.J. Neuronal vulnerability, pathogenesis, and Parkinson's disease. *Mov. Disord.* **28**, 715–724 (2013).
113. Liu, G. *et al.* Aldehyde dehydrogenase 1 defines and protects a nigrostriatal dopaminergic neuron subpopulation. *J. Clin. Invest.* **124**, 3032–3046 (2014).

114. Brichta, L. *et al.* Identification of neurodegenerative factors using translatome-regulatory network analysis. *Nat. Neurosci.* **18**, 1325–1333 (2015).
115. Ferguson, J.M. SSRI antidepressant medications: adverse effects and tolerability. *Prim. Care Companion J. Clin. Psychiatry* **3**, 22–27 (2001).
116. Steinbeck, J.A. & Studer, L. Moving stem cells to the clinic: potential and limitations for brain repair. *Neuron* **86**, 187–206 (2015).
117. Sterneckert, J.L., Reinhardt, P. & Schöler, H.R. Investigating human disease using stem cell models. *Nat. Rev. Genet.* **15**, 625–639 (2014).
118. Treutlein, B. *et al.* Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* **534**, 391–395 (2016).
119. Camp, J.G. *et al.* Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. USA* **112**, 15672–15677 (2015).
120. Crosetto, N., Bienko, M. & van Oudenaarden, A. Spatially resolved transcriptomics and beyond. *Nat. Rev. Genet.* **16**, 57–66 (2015).
121. Ke, R. *et al. In situ* sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* **10**, 857–860 (2013).
122. Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090–aaa6090 (2015).
123. Lubeck, E., Coskun, A.F., Zhiyentayev, T., Ahmad, M. & Cai, L. Single-cell *in situ* RNA profiling by sequential hybridization. *Nat. Methods* **11**, 360–361 (2014).
124. Lee, J.-H. *et al.* Highly multiplexed subcellular RNA sequencing *in situ*. *Science* **343**, 1360–1363 (2014).
125. Cusanovich, D.A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910–914 (2015).
126. Lodato, M.A. *et al.* Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).
127. Bendall, S.C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
128. Bheda, P. & Schneider, R. Epigenetics reloaded: the single-cell revolution. *Trends Cell Biol.* **24**, 712–723 (2014).
129. Hu, Y. *et al.* Simultaneous profiling of transcriptome and DNA methylome from a single cell. *Genome Biol.* **17**, 88 (2016).
130. Behjati, S. *et al.* Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* **513**, 422–425 (2014).
131. McKenna, A. *et al.* Whole organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
132. Dey, S.S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol.* **33**, 285–289 (2015).
133. Macaulay, I.C. *et al.* G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12**, 519–522 (2015).
134. Hanchate, N.K. *et al.* Single-cell transcriptomics reveals receptor transformations during olfactory neurogenesis. *Science* **350**, 1251–1255 (2015).
135. Ye, Z. *et al.* Instructing perisomatic inhibition by direct lineage reprogramming of neocortical projection neurons. *Neuron* **88**, 475–483 (2015).
136. Fuccillo, M.V. *et al.* Single-cell mRNA profiling reveals cell-type-specific expression of neurexin isoforms. *Neuron* **87**, 326–340 (2015).
137. Johnson, M.B. *et al.* Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat. Neurosci.* **18**, 637–646 (2015).
138. Pollen, A.A. *et al.* Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015).
139. Thomsen, E.R. *et al.* Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat. Methods* **13**, 87–93 (2016).
140. Darmanis, S. *et al.* A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. USA* **112**, 7285–7290 (2015).
141. Gerfen, C.R., Paletzki, R. & Heintz, N. GENSAT BAC cre-recombinase driver lines to study the functional organization of cerebral cortical and basal ganglia circuits. *Neuron* **80**, 1368–1383 (2013).
142. Harris, J.A. *et al.* Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Front. Neural Circuits* **8**, 76 (2014).
143. Tang, J.C.Y. *et al.* Cell-type-specific manipulation with GFP-dependent Cre recombinase. *Nat. Neurosci.* **18**, 1334–1341 (2015).
144. Buzsáki, G. Large-scale recording of neuronal ensembles. *Nat. Neurosci.* **7**, 446–451 (2004).
145. Rajasethupathy, P. *et al.* Projections from neocortex mediate top-down control of memory retrieval. *Nature* **526**, 653–659 (2015).
146. Ahmed, Z.M. *et al.* Double homozygous waltzer and Ames waltzer mice provide no evidence of retinal degeneration. *Mol. Vis.* **14**, 2227–2236 (2008).
147. Lacar, B. *et al.* Nuclear RNA-seq of single neurons reveals molecular signatures of activation. *Nat. Commun.* **7**, 11022 (2016).
148. Spitzer, N.C. Neurotransmitter switching? No surprise. *Neuron* **86**, 1131–1144 (2015).
149. Landis, S.C. Target regulation of neurotransmitter phenotype. *Trends Neurosci.* **13**, 344–350 (1990).
150. Bikoff, J.B. *et al.* Spinal inhibitory interneuron diversity delineates variant motor microcircuits. *Cell* **165**, 207–219 (2016).