

SHORT COURSE 1

Using Single-Cell Genomics to Analyze Neurons, Glia, and Circuits

Organizer: Steven McCarroll, PhD

Short Course 1

Using Single-Cell Genomics to Analyze Neurons, Glia, and Circuits

Organized by Steven McCarroll, PhD



SOCIETY *for*
NEUROSCIENCE

Please cite articles using the model:
[AUTHOR'S LAST NAME, AUTHOR'S FIRST & MIDDLE INITIALS] (2016)
[CHAPTER TITLE] In: Using Single-Cell Genomics to Analyze Neurons, Glia, and Circuits.
(McCarroll S, ed) pp. [xx-xx]. San Diego, CA: Society for Neuroscience.

All articles and their graphics are under the copyright of their respective authors.

Cover graphics and design © 2016 Society for Neuroscience.



SHORT COURSE 1

Using Single-Cell Genomics to Analyze Neurons, Glia, and Circuits

Organized by: Steven McCarroll, PhD

Friday, November 11, 2016

8:30 a.m.–6 p.m.

Location: San Diego Convention Center • Room: 6B • San Diego, CA

TIME	TALK TITLE	SPEAKER
8–8:30 a.m.	Check-in	
8:30–8:40 a.m.	Opening Remarks	Steven McCarroll, PhD • Harvard Medical School
8:40–9:30 a.m.	Massive single-cell RNA-seq analysis to analyze cellular specialization and evolution in the nervous system	Evan Macosko, MD, PhD • Harvard Medical School
9:30–10:20 a.m.	Single-cell analysis of interneurons during brain development	Gordon Fishell, PhD • New York University
10:20–10:50 a.m.	Morning Break	
10:50–11:40 a.m.	Single-cell analysis of somatic mutation and cell lineage in the human nervous system	Christopher Walsh, MD, PhD • Boston Children's Hospital
11:40 a.m.–12:30 p.m.	Combining electrophysiology with single-cell RNA-seq to reveal principles of connectivity	Andreas Tolias, PhD • Baylor College of Medicine
12:30–1:30 p.m.	Lunch	
1:30–2:20 p.m.	A mouse cortical cell taxonomy	Bosiljka Tasic, PhD • Allen Institute for Brain Science
2:20–3:10 p.m.	Single-cell analysis of neuronal differentiation and reprogramming	Alex Pollen, PhD • UC San Francisco
3:10–4 p.m.	Opportunities and directions in single-cell analysis	Steven McCarroll, PhD • Harvard Medical School
4–4:15 p.m.	Afternoon Break	

AFTERNOON BREAKOUT SESSIONS • PARTICIPANTS SELECT DISCUSSION GROUPS AT 4:15 AND 5:15 P.M.

TIME	BREAKOUT SESSIONS	SPEAKERS	ROOM
4:15–5 p.m.	Group 1: Single-cell analyses of development	SC1 faculty	31A
	Group 2: Defining cell types in the nervous system: transcriptomics and beyond	SC1 faculty	31B
	Group 3: Distinguishing cell types from cell states: experimental approaches	SC1 faculty	31C
5–5:15 p.m.	Afternoon Break		
5:15–6 p.m.	Repeat sessions above. Select a second discussion group.		

Table of Contents

Introduction: Scientific Opportunities and Challenges in Single-Cell Analysis <i>Steven McCarroll, PhD</i>	7
Novel Technologies for Single-Cell Resolution Whole-Transcriptome Analysis in CNS Tissue <i>Evan Macosko, MD, PhD</i>	13
Clonally Related Interneurons Are Not Constrained by Functional or Anatomical Boundaries <i>Christian Mayer, PhD, Rachel C. Bandler, MD, PhD, and Gordon Fishell, PhD</i>	21
Genetic Techniques for Cell Lineage Tracing in the Nervous System <i>Kelly Girskis, BA, Mollie Woodworth, PhD, and Christopher Walsh, MD, PhD</i>	31
Correlating Cellular Morphology, Physiology and Gene Expression Using Patch-seq <i>Cathryn R. Cadwell, PhD, and Andreas S. Tolias, PhD</i>	41
Adult Mouse Cortical Cell Taxonomy Revealed by Single-Cell Transcriptomics <i>Bosiljka Tasic, PhD</i>	53
Distinct Molecular Programs Define Human Radial Glia Subtypes During Human Cortical Development <i>Tomasz J. Nowakowski, PhD, and Alex A. Pollen, PhD</i>	65

Introduction: Scientific Opportunities and Challenges in Single-Cell Analysis

Steven A. McCarroll, PhD

Harvard Medical School
Boston, Massachusetts

Broad Institute of MIT and Harvard
Stanley Center for Psychiatric Research
Cambridge, Massachusetts

Background

Individual cells are the basic units with which larger biological systems—circuits, tissues, and entire organisms—are built. Cells in the same tissue or circuit have various biological missions; a cell's missions are reflected in its size, morphology, physiology, and use of its genome. Adjacent cells often use the same genome in dramatically different ways.

Historically, insights about cell types and their specialization were obtained one at a time, as a result of varying combinations of serendipity and painstaking work. The discovery of a cell population with unusual physiological properties might be followed later by the identification of a molecular marker for those cells, and then eventually by insights into these cells' interactions with and connectivity to other cells. Several technological innovations promise to transform the pace of discovery about cell types and their properties—first, by allowing the collection of genome-scale information (e.g., about gene expression or DNA sequence) from individual cells (Tang et al., 2009), and more recently, by allowing genome-scale analyses to be conducted on vast numbers of individual cells at once (Klein et al., 2015; Macosko et al., 2015). The pace of data generation has increased dramatically; the pace of biological insights will, one hopes, begin to increase as well.

Moving From Proofs of Concept to Useful Data Resources to Insights

Emerging fields in genomics often follow a similar trajectory. Early “proof of concept” studies serve to illustrate that new kinds of analysis can be executed. Although the data and analysis methods are often quickly replaced by better approaches, such early results help many readers to expand their sense of the possible.

As experimental approaches begin to stabilize and mature (such that the shelf life of a dataset is longer and its quality more assured), it becomes possible to build data resources that have cumulative value. In human genetics, for example, datasets on human genome variation (alleles and allele frequencies at each site in the genome) are used in thousands of genetic inquiries every day, supporting both genome-scale studies and analyses of individual genes (International HapMap Consortium, 2015; Lek et al., 2016). For single-cell transcriptomics, such resources may increasingly take the form of digital atlases in which the expression profiles of individual

cell types can be looked up (Tasic, et al., 2016). Such resources may come to have great value because they allow routine lookups of genes' expression patterns across cell types. Their immediate results may be more facile, quantitative, and reliable than images collected by laborious slogs involving antibodies of varying qualities, tissues and fixatives with varying properties, and hours of microscopy.

The most rewarding phase can occur as new tools and data resources begin to support scientific insights into molecular and cellular mechanisms, and as broader experimental programs and plans reshape themselves to utilize the opportunities inherent in new kinds of data and new ways of monitoring biological systems.

Approaching Integrated Analysis

For single-cell analysis, a growing scientific opportunity will come from beginning to draw connections among the different ways of characterizing individual cells—to appreciate how morphology, physiology, connectivity, and gene expression are codistributed and interconnected mechanistically. Ideally, the cell atlases of the future will report not only what genes each type of cell expresses but also what shape(s) it assumes, what neurons it connects with, what transmitters it responds to, and what voltage and ionic dynamics it has. Armed with this kind of characterization, we will be able to begin to understand how gene expression, morphology, physiology, and connectivity influence and arise from one another. In an early step in this direction, a recent study related the electrophysiological properties of individual cells to their molecular profiles (Cadwell et al., 2016).

A practical challenge of integrated analysis involves the fact that many kinds of analyses of individual cells (e.g., transcriptomics, fixation for immunohistochemistry) destroy these cells' other properties, leaving little room for subsequent analyses of the same cells. In this Short Course, we will discuss the opportunities that arise from integrating multimodal data types at single-cell resolution and the practical challenges of accomplishing this.

Developing Clearer, More Useful Standards and Metrics

New fields often struggle to clarify their thinking about how to quantify and compare findings and how to distinguish real signals from artifacts. Single-cell analysis of somatic DNA variation, for example, now indicates that rates of somatic retrotransposition are far lower than was reported in

NOTES

some earlier studies (Evrony et al., 2016). Perhaps nowhere has such confusion been more abundant than in single-cell transcriptomics. Today, the thoroughness of single-cell experiments is still often evaluated in terms of “reads per cell”: the ratio of the number of sequencing reads generated to the number of cells analyzed. However, this metric may offer little information about what was learned, because any number of DNA or RNA molecules can be amplified into an arbitrarily large number of copies and then resequenced using an arbitrarily large number of sequencing reads without generating any new information. (Put another way, if a tree falls in the woods, it matters little whether that event is documented by one, 10, or 1000 observers, so long as the fall is recorded reliably and distinguished from that of other trees.)

A similar confusion involves the use of the metric “genes detected per cell.” The number of genes expressed in a cell depends strongly on cell type, and more important, this number is inflated when an analysis is not truly single-cell (e.g., when a cell doublet is assumed to be a single cell). This problem appears to have inflated estimates in early single-cell studies. Significant advances, such as the use of unique molecular indicators (UMIs) (Kivioja et al., 2011), which affix a particular molecular barcode to each cDNA and allow digital counting with correction for amplification effects, are increasingly enabling true estimates of transcript ascertainment. To return to our “tree falling” analogy, UMIs make it possible to recognize when many observers reporting a “tree falling” are in fact all talking about the same tree. Not surprisingly, the figures yielded by UMI-informed analyses—typically quantified as transcripts per cell (trees) rather than reads per cell (observers)—are also far more modest. Still, UMIs have offered a significant step forward in clarity, even if the resulting estimates have less “bling.” A goal of this Short Course will be to try to clarify such terms and help scientists to design, evaluate, and think about experiments in quantitative ways.

Scaling Up Computational Approaches

Most single-cell experimental approaches in use today produce novel kinds of datasets for which computational methods are still in their infancy. For example, methods for collecting gene-expression information from tens of thousands of individual cells have created a new scientific opportunity to

infer cell types and cell states (including rare ones) in “unsupervised” ways that are not constrained by earlier theories, categories, or lists of markers. This opportunity needs to be met increasingly by new analytical approaches. Many computational approaches that were developed for early, small single-cell RNA-seq datasets do not scale up successfully to, or do not realize the opportunities inherent in, the far-larger datasets that are being generated. Thus, an important direction will be to develop algorithms that can recognize patterns and structure in vast multidimensional datasets and then present these patterns in ways that lead to biological insights. This exciting emerging area will benefit from close multidisciplinary collaborations among scientists who have expertise in computer science, biology, statistics, and mathematics.

Seizing the Opportunity Ahead

The functions of tissues and organs derive from interactions and collaborations among specialized individual cells. Elucidating how tissue and circuit functions encompass the actions of specialized cells expressing distinct genes and molecular complexes, with varying proximity and connectivity, is one of the great scientific challenges of our time. Aspiring to such understanding is also increasingly within our grasp.

References

- Cadwell CR, Palasantza A, Jiang X, Berens P, Deng Q, Yilmaz M, Reimer J, Shen S, Bethge M, Tolias KF, Sandberg R, Tolias AS (2016) Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat Biotechnol* 34:199–203.
- Evrony GD, Lee E, Park PJ, Walsh CA (2016) Resolving rates of mutation in the brain using single-neuron genomics. *Elife* 5. pii: e12966.
- International HapMap Consortium (2015) A haplotype map of the human genome. *Nature* 437:1299–1320.
- Kivioja T, Vähärautio A, Karlsson K, Bonke M, Enge M, Linnarsson S, Taipale J. (2011) Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* 9:72–74.
- Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161:1187–1201.

- Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, et al. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536:285–291.
- Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214.
- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6:377–382.
- Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, et al. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 19:335–346.

NOTES

Novel Technologies for Single-Cell Resolution Whole-Transcriptome Analysis in CNS Tissue

Evan Macosko, MD, PhD

Stanley Center for Psychiatric Research
Broad Institute of MIT and Harvard
Harvard Medical School
Cambridge, Massachusetts

Introduction

Individual cells are the building blocks of tissues, organs, and organisms. Each tissue contains cells of many types, and cells of each type can switch among biological states. Especially in the mammalian brain, our knowledge of cellular diversity is incomplete. In particular, the extent of cell-type complexity in the brain remains unknown and is widely debated (Luo et al., 2008; Petilla Interneuron Nomenclature Group et al., 2008). Many important but rare cell populations likely remain undiscovered, potentially limiting our understanding of physiological function. In addition, the overall landscape of transcriptional variation, even among abundant cell types, is mostly undescribed.

A major determinant of each cell type's function is its transcriptional program. Consequently, ascertainment of sufficient numbers of single-cell gene expression profiles may enable a comprehensive taxonomy of cell populations across the mammalian nervous system. Although two molecular techniques for isolating and amplifying small amounts of mRNA were developed some time ago—T7 amplification (Eberwine et al., 1992) and SMART (switching mechanism at 5' end of RNA template) technology (Matz et al., 1999; Zhu et al., 2001)—it was the advent of high-throughput next-generation sequencing technologies, coupled with these amplification techniques, that has made the analysis of meaningful numbers of single-cell gene expression profiles possible. Together with improved techniques for isolating individual cells, barcoding their transcriptional contents, and miniaturizing amplification volumes, single-cell gene expression profiling has moved rapidly from an era in which only a handful of profiles could be gleaned in a major study, to one in which the routine ascertainment of tens of thousands of profiles in a single experiment is now possible.

This chapter is divided into three sections, describing (1) the various technological innovations that made this recent transformation possible; (2) the important technical parameters for assessing the quality of data produced by these techniques; and (3) a discussion of biological applications of single-cell gene expression analysis and future technological directions.

Single-Cell mRNA-seq: From Handfuls to Thousands of Cell Profiles

Amplifying and interrogating small quantities of mRNA

Gene expression analysis at the level of individual cells began soon after the advent of techniques for amplifying minute quantities of mRNA. In 1992,

Eberwine and colleagues used T7 amplification to prepare cDNA libraries from individually hand-picked hippocampal cells (Eberwine et al., 1992). T7 amplification works by reverse transcription of an mRNA pool using an oligo dT primer fused to a T7 RNA polymerase promoter sequence. After second-strand synthesis, the double-stranded cDNA is used as the template for *in vitro* transcription amplification by T7 RNA polymerase. The resulting RNA amplicons are reverse transcribed in bulk to yield an amplified cDNA library. By repeating this process twice, Eberwine's group was able to achieve an amplification factor of $\sim 10^6$. Sometime later, an alternative approach was developed that uses the template-switching capability of MMLV (Moloney murine leukemia virus) reverse transcriptase (known as SMART) to amplify small quantities of cDNA by PCR (Matz et al., 1999). This approach is the basis of the suite of RNA amplification products manufactured and sold by Clontech Laboratories (Mountain View, CA). Initially, the single-cell cDNA libraries produced by these amplification schemes were interrogated by hybridization (Northern blot and microarray analysis). Today, however, the improved throughput, precision, and accuracy of next-generation sequencing have made mRNA sequencing (mRNA-seq) the near-universal choice for measuring the concentration of individual RNA species.

The most common single-cell RNA-seq protocols currently in use continue to feature either T7 or SMART amplification to generate cDNA libraries. The two amplification schemes have different advantages: T7 amplification, because it is linear, is generally believed to produce more even amplification of a diverse cDNA library, while SMART is somewhat less technically demanding.

Approaches to isolating individual cells

A major impediment to high-throughput examination of single-cell profiles is the technical difficulty associated with isolating individual cells. Hand-picking cells (the traditional approach) allows for visual confirmation of cell capture and morphological screening for a desired cell population, but is inherently very time-consuming. Flow cytometry sorting of individual cells into microtiter plates (Jaitin et al., 2014; Tasic et al., 2016) provides a significant improvement in scale and can be combined with fluorescent staining to screen for subsets of cells of interest. Microfluidic techniques have also been developed to isolate cells. Traditional valve-based microfluidic devices capture cells within individual chambers and process the isolated mRNA in parallel

NOTES

(White et al., 2011). Two commercially available products from Fluidigm (South San Francisco, CA) and WaferGen Bio-systems (Fremont, CA) enable several hundred cells to be captured and processed at once. In contrast, microfluidic droplet-generation devices can disperse tens of thousands of precisely sized (“monodisperse”) picoliter-scale or nanoliter-scale droplets per minute (Umbanhowar, 2000; Thorsen et al., 2001). By critically diluting a cell suspension to a concentration far lower than one cell per droplet, individual cells can be isolated in extremely high throughput in these emulsions (tens of thousands per hour).

Massive molecular barcoding

Following technical improvements in the ease and throughput of cell isolation, particularly by droplet microfluidics, the major obstacle to routine, massively multiplexed single-cell mRNA-seq became the cost and time required to prepare individual libraries from so many cells in individual microtiter reactions. If the mRNA content of individual cells could be barcoded at the start of processing, then all subsequent molecular amplification and library preparation steps could be performed in a single bulk reaction, dramatically simplifying the process. Recently, two barcoding approaches were developed that address this problem (Klein et al., 2015; Macosko et al., 2015). In each, a collection of microparticles (beads) is generated, each of which harbors a large number of barcoded oligo dT primers on its surface; the barcode is the same across all the primers on the surface of any one bead but differs from the barcodes on all other beads. In the first method, Drop-seq, barcode diversity is generated through a modified form of chemical oligonucleotide synthesis, in which beads are repeatedly split and pooled to achieve millions of unique sequences (Fig. 1). The second method, inDrop, uses an enzymatic approach to combinatorially stitch together two sets of barcoded oligos, resulting in a pool of beads with hundreds of thousands of individual barcodes. Both methods are able to collectively barcode and process thousands of cells in a single experiment.

Technical assessments of single-cell RNA-seq data

To glean meaningful biological signals from any technology, it is vital to have technical measurements that assess the strengths and limitations of the data. Single-cell RNA-seq (scRNA-seq) technologies should be evaluated by several criteria: (1) the amount of RNA that is captured; (2) the specificity of the signal (how truly “single-cell” the profile is); and (3) how consistent the resulting profile is across individual technical replicates.

RNA capture efficiency

The most common method for estimating the proportion of sampled transcripts is to process a synthetic library of RNAs (known as the External RNA Controls Consortium [ERCC] “spike-in” controls) and compute the fraction of these RNAs that are reported by sequencing. In general, these analyses have produced estimates of between 2% and 12% capture efficiency across different technological platforms (Grun et al., 2014; Klein et al., 2015; Macosko et al., 2015). One study explained the majority of the loss by inefficiency in the mRNA hybridization step (Macosko et al., 2015); it remains unknown whether this step is also the bottleneck for other methods.

A typical mammalian cell contains 5–10 pg of total RNA (Tang et al., 2011), of which 1%–10% is polyadenylated, mature mRNA. This corresponds to ~100,000–500,000 unique mRNA molecules, distributed across thousands of individual genes. This means that, at a capture efficiency of 10%, many minimally expressed genes will go undetected in a given cell. High-throughput single-cell technologies like Drop-seq and inDrop can address this problem by repeatedly sampling cells of the same type to accrue observations of these low-copy genes.

Doublet rates and purity

One mode of failure in any single-cell method involves cells that stick together or happen to otherwise be co-isolated for library preparation. To measure doublet rates, two groups recently sequenced mixtures of cells derived from two species and calculated organism purity rates of individual cell barcodes. For droplet-based approaches (i.e., inDrop and Drop-seq), the doublet rate could be adjusted to arbitrarily low levels by reducing the cell concentration. Although doublet rates can be higher in other systems (e.g., Fluidigm C1), many of these doublets can be identified up front by fluorescence microscopy of the capture chambers (Fluidigm, 2016). Species-mixing experiments enable a careful quantification of single-cell purity across libraries. In Drop-seq, impurity was strongly related to the concentration at which cell suspensions were loaded: organism purity ranged from 98.8% at 12.5 cells/ μ l to 90.4% at 100 cells/ μ l.

Technical reliability

Replication across experimental sessions enables the construction of cumulatively more powerful datasets for detecting subtle biological signals. Technical variation can arise from day-to-day differences in cell preparation, molecular processing and sequencing,

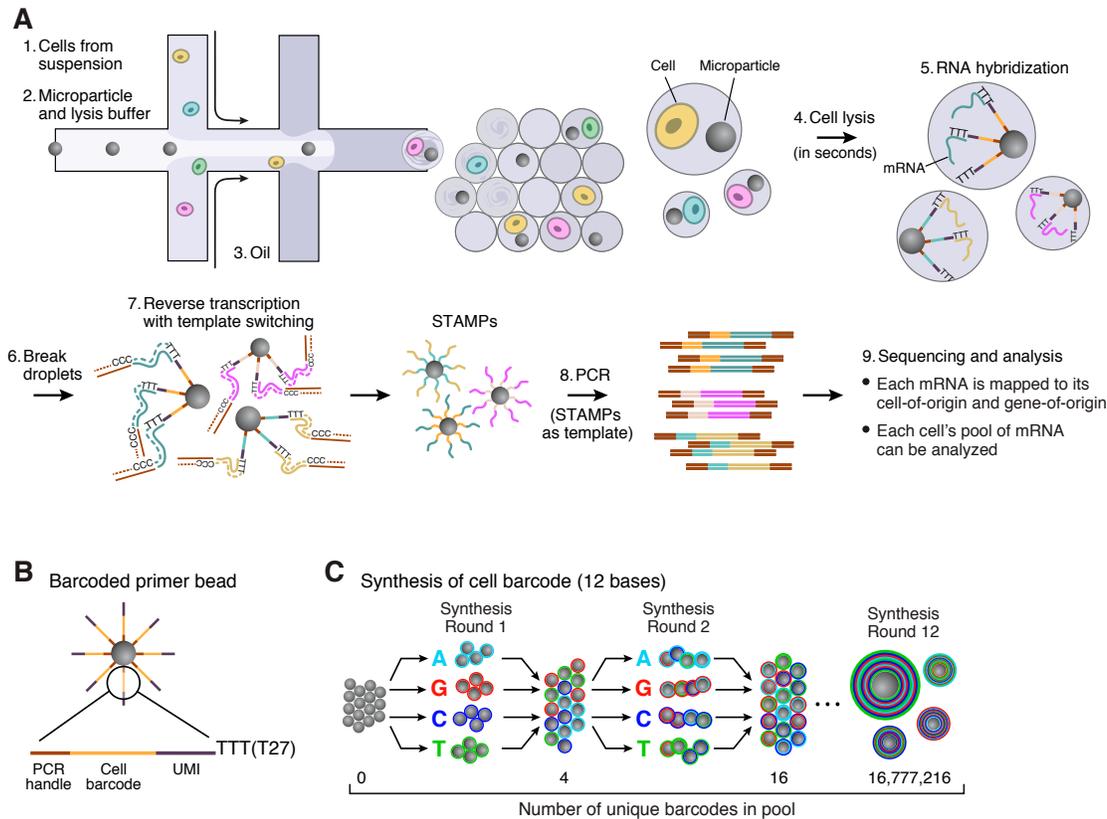


Figure 1. Drop-seq: molecular barcoding of cellular transcriptomes using droplet microfluidics. **A**, Schematic of single-cell mRNA-seq library preparation using Drop-seq. A custom-designed microfluidic device joins two aqueous flows before their compartmentalization into discrete droplets. One flow contains cells, and the other flow contains barcoded primer beads suspended in a lysis buffer. Immediately after droplet formation, the cell is lysed and releases its mRNAs, which then hybridize to the primers on the microparticle surface. The droplets are broken up by adding a reagent to destabilize the oil–water interface, and the microparticles are collected and washed. The mRNAs are then reverse transcribed in bulk, forming STAMPs (single-cell transcriptomes attached to microparticles), and template switching is used to introduce a PCR handle downstream of the synthesized cDNA (Zhu et al., 2001). **B**, Sequence of primers on the microparticle. The primers on all beads contain a common sequence (“PCR handle”) to enable PCR amplification after STAMP formation. Each microparticle contains $>10^8$ individual primers that share the same “cell barcode” (**C**) but have different unique molecular identifiers (UMIs), enabling mRNA transcripts to be digitally counted. A 30 bp oligo dT sequence is present at the end of all primer sequences for capture of mRNAs. **C**, Split-and-pool synthesis of the cell barcode. To generate the cell barcode, the pool of microparticles is repeatedly split into four equally sized oligonucleotide synthesis reactions, to which one of the four DNA bases is added, and then pooled together after each cycle, in a total of 12 split-pool cycles. The barcode synthesized on any individual bead reflects that bead’s unique path through the series of synthesis reactions. The result is a pool of microparticles, each possessing one of 4^{12} (16,777,216) possible sequences on its entire complement of primers. Reprinted with permission from Macosko EZ et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161:1203, 1205; their Figs. 1B, C, and 2A. Copyright 2015, Elsevier.

or peculiarities specific to particular systems. We clustered $\sim 45,000$ Drop-seq–derived single-cell profiles from dissociated mouse retinas over the course of seven experimental sessions. The resulting 39 clusters contained cells from each session, suggesting that the technical variation in gene expression was relatively small even compared with the differences between highly similar cell subtypes. New high-throughput technologies should provide large enough datasets to permit more-rigorous

computational analyses in which portions of the data are withheld (e.g., *k*-fold cross-validation).

Biological applications and technological improvements

Already, studies using scRNA-seq have transformed our understanding of cellular diversity in many mammalian CNS tissues, including the spinal cord (Usoskin et al., 2015), cortex (Zeisel et al., 2015;

NOTES

Tasic et al., 2016), and retina (Macosko et al., 2015). In addition, biologists are quickly recognizing the plethora of scientific opportunities enabled by ascertaining transcriptional variation in individual cells, beyond performing initial taxonomic analyses of tissues. For example, genome-scale genetic studies are identifying large numbers of genes in which genetic variation contributes to disease risk. Finding the cellular sites and biological activities of so many genes is an important but challenging goal. High-throughput single-cell transcriptomics could localize the expression of all risk genes to specific cell types, and in conjunction with genetic perturbations, help to systematically relate each gene to (1) the cell types most affected by loss or perturbation of those genes and (2) the alterations in cell state elicited by such perturbations. Such approaches could help cross the daunting gap from gene discoveries to insights about pathophysiology.

scRNA-seq (possibly coupled to additional manipulations) could be used to generate an information-rich, multidimensional readout of the influence of many kinds of perturbations—such as small molecules, genetic mutations (natural or engineered), pathogens, or other stimuli—on many kinds of cells. When studying the effects of a mutation, for example, scRNA-seq could illuminate pleiotropies by revealing the ways in which the same mutation differentially impacts distinct cell types. Single-cell expression analysis could also be used to characterize the heterogeneous responses of diverse cell populations to a drug or metabolite, or combinations thereof.

Enormous opportunities exist to improve approaches to single-cell gene expression analysis. First, the extension of existing methods to the analysis of frozen and/or fixed tissue could help relate functional genomic variation to transcriptional variation in specific cell types and provide novel hypotheses for how specific cell types are altered in disease states whose pathogenesis remain mysterious. Second, tissue dissociation before cell processing introduces artifactual signals (as the dissociated cells begin to die) and does not maintain spatial relationships among analyzed cells. Thus, multiple new technologies, including highly multiplexed *in situ* hybridization techniques (Chen et al., 2015; Coskun and Cai, 2016) and approaches to sequencing mRNA directly from tissue slices (Lee et al., 2014; Ståhl et al., 2016) could ultimately make it possible to perform single-cell profiling without tissue dissociation. Finally, the coupling of scRNA-seq with other cellular readouts,

including single-cell epigenetic measurements and DNA sequencing, could someday provide fundamental insights into transcriptional regulation in specialized cell populations.

The functional implications of a gene's expression are a product not just of a gene's intrinsic properties but also of the entire cell-level context in which a gene is expressed. The routine facile, large-scale measurement of single-cell gene expression profiles with new technologies should enable the abundant and routine discovery of such relationships across biology.

Acknowledgments

Parts of this chapter were excerpted with permission from Macosko EZ et al. (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214. Copyright 2015, Elsevier.

References

- Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X (2015) RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348:aaa6090.
- Coskun AF, Cai L (2016) Dense transcript profiling in single cells by image correlation decoding. *Nat Methods* 13:657–660.
- Eberwine J, Yeh H, Miyashiro K, Cao Y, Nair S, Finnell R, Zettel M, Coleman P (1992) Analysis of gene expression in single live neurons. *Proc Natl Acad Sci USA* 89:3010–3014.
- Fluidigm (2016) Double rate and detection on the C1 IFCs. White Paper PN 101-2711 A1. https://mailbox.zimbra.upenn.edu/home/ngsc@zimbra.upenn.edu/Briefcase/PublicFiles/Equipment/Fluidigm/C1/C1_Doublet_wp_101-2711A1_20160106.pdf
- Grun D, Kester L, van Oudenaarden A (2014) Validation of noise models for single-cell transcriptomics. *Nat Methods* 11:637–640.
- Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343:776–779.
- Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz D.A., Kirschner M.W. (2015) Droplet barcoding for single cell transcriptomics and its application to embryonic stem cells. *Cell* 161:1187–1201.

- Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SS, Li C, Amamoto R, Peters DT, Turczyk BM, Marblestone AH, Inverso SA, Bernard A, Mali P, Rios X, Aach J, Church GM (2014) Highly multiplexed subcellular RNA sequencing *in situ*. *Science* 343:1360–1363.
- Luo L, Callaway EM, Svoboda K (2008) Genetic dissection of neural circuits. *Neuron* 57:634–660.
- Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214.
- Matz M, Shagin D, Bogdanova E, Britanova O, Lukyanov S, Diatchenko L, Chenchik A (1999) Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res* 27:1558–1560.
- Petilla Interneuron Nomenclature Group (PING), Ascoli GA, Alonso-Nanclares L, Anderson SA, Barrionuevo G, Benavides-Piccione R, Burkhalter A, Buzsáki G, Cauli B, Defelipe J, Fairén A, Feldmeyer D, Fishell G, Fregnac Y, Freund TF, Gardner D, Gardner EP, Goldberg JH, Helmstaedter M, Hestrin S, et al. (2008) Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat Rev Neurosci* 9:557–568.
- Ståhl PL, Salmén F, Vickovic S, Lundmark A, Navarro JF, Magnusson J, Giacomello S, Asp M, Westholm JO, Huss M, Mollbrink A, Linnarsson S, Codeluppi S, Borg Å, Pontén F, Costea PI, Sahlén P, Mulder J, Bergmann O, Lundeberg J, Frisén J (2016) Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353:78–82.
- Tang F, Lao K, Surani MA (2011) Development and applications of single-cell transcriptome analysis. *Nat Methods* 8:S6–S11.
- Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, et al. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 19:335–346.
- Thorsen T, Roberts RW, Arnold FH, Quake SR (2001) Dynamic pattern formation in a vesicle-generating microfluidic device. *Phys Rev Lett* 86:4163–4166.
- Umbanhowar PB, Prasad V, Weitz DA (2000) Monodisperse emulsion generation via drop break off in a coflowing stream. *Langmuir* 16:347–351.
- Usoskin D, Furlan A, Islam S, Abdo H, Lonnerberg P, Lou D, Hjerling-Leffler J, Haeggstrom J, Kharchenko O, Kharchenko PV, Linnarsson S, Ernfors P (2015) Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci* 18:145–153.
- White AK, VanInsberghe M, Petrivi OI, Hamidi M, Sikorski D, Marra MA, Piret J, Aparicio S, Hansen CL (2011) High-throughput microfluidic single-cell RT-qPCR. *Proc Natl Acad Sci USA* 108:13999–14004.
- Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347:1138–1142.
- Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *Biotechniques* 30:892–897.

Clonally Related Interneurons Are Not Constrained by Functional or Anatomical Boundaries

Christian Mayer, PhD, Rachel C. Bandler, MD, PhD,
and Gordon Fishell, PhD

Neuroscience Institute
NYU Langone Medical Center
New York, New York

Introduction

In 2015, two papers were published in *Neuron* (Harwell et al., 2015, and Mayer et al., 2015) that jointly argued that interneuron lineages were dispersed across functional and structural boundaries. These conclusions were challenged by the laboratory of Songhai Shi (Sultan et al., 2016), and this Short Course chapter presents our response. In it, we discuss ongoing single-cell approaches that combine whole-genome analysis and lineage to take the next step toward understanding the possible links among interneuron lineage, cell type, and position within the brain.

During development, excitatory principal neurons and inhibitory interneurons assemble within the mammalian cortex and integrate into common circuits. However, a fundamental question in developmental neuroscience remains whether clonally related interneurons, like excitatory neurons, maintain a coherent relationship with their siblings while populating specific cortical areas and the local columnar architecture therein. Our laboratory and a copublished article (Harwell et al., 2015; Mayer et al., 2015) independently took advantage of a lineage fate mapping method devised by the Cepko Lab. With this method, a replication-defective retroviral library that contains a highly diverse set of DNA barcodes can be used to tag dividing progenitor cells during embryonic development, thereby permitting the unambiguous determination of lineage relationships across individual cells in the adult. Both studies reported that interneurons derived from a single progenitor lineage within the forebrain disperse widely across both functional and anatomical structures. As outlined in their upcoming article in *Neuron*, the laboratory of Dr. Shi (Sultan et al., 2016) further analyzed our datasets and concluded that clonally related interneurons are not “randomly dispersed,” and we agree with this conclusion. In fact, we never claimed that interneuron clones “randomly disperse” either within or across brain structures. Rather, we reported a finding consistent with Sultan et al. (2016) that ~30% of clones spanned more than one brain structure, providing clear cases in which progenitor lineage is not predictive of an interneuron’s ultimate anatomical or functional fate. In addition, we found that the spatial distribution of clones is similar among progenitors regardless of whether they share a lineal relationship. Based on our findings, we conclude that the integration of interneurons into functional cortical areas is unlikely to be constrained by lineage.

The mammalian cortex is subdivided into areas devoted to vision, sensation, audition, and other functions. Each area can be further divided physiologically into smaller units or functional columns. Excitatory and inhibitory neurons (the two main cell types of the cortex and hippocampus) have very distinct embryonic origins (Anderson et al., 1997) and have segregated into separate lineages by the time the primary prosencephalon has developed into the secondary prosencephalon (Rubenstein et al., 1998). Excitatory cells are derived from the dorsal telencephalon or pallium. Consecutive rounds of asymmetric cell division produce lineage-related sister excitatory neurons that migrate short distances toward the pia and into the overlaying developing cortical plate. After migration, spatially organized vertical clusters of excitatory sibling neurons (referred to as “clonal units”) form functional columnar microcircuits in the neocortex (Noctor et al., 2001; Li et al., 2012). In contrast, inhibitory cells derive entirely from the ventral telencephalon or subpallium (Marin and Rubenstein, 2001; Fishell and Rudy, 2011), most prominently from the medial and caudal ganglionic eminences (MGE and CGE, respectively), and migrate over large distances to integrate into the developing cortex, hippocampus, or other subcortical forebrain structures.

Conflicting Results from Four Recent Studies Examining Interneuron Lineages

Despite the technical difficulties associated with fate mapping interneuron lineages resulting from their complex migration patterns, four recent studies (Brown et al., 2011; Ciceri et al., 2013; Harwell et al., 2015; Mayer et al., 2015) have endeavored to explore whether clonally related interneurons are selectively positioned within cortical units, similar to what is observed in excitatory neurons. If clonally related interneurons were confined to discrete anatomical brain units (e.g., columns of the cortex), this would support the idea that cell lineage is dictating the integration of interneurons into functional cortical networks.

All four groups agreed that before migration, the majority of interneurons are generated from symmetric and asymmetric divisions of MGE progenitor cells, leading to radially aligned interneuron precursors being symmetrically aligned in proximity to each other (Brown et al., 2011; Ciceri et al., 2013; Harwell et al., 2015; Mayer et al., 2015). Postmitotic interneurons

NOTES

reach their final positions within the cortex through long-range tangential migration that requires them to travel 100 times farther than excitatory pyramidal neurons to reach the cortical plate.

However, the four studies drew different conclusions about how lineage contributes to the final location of interneurons after long-range migration. Brown et al. (2011) and Ciceri et al. (2013) described clonal clusters in the cortex that were sufficiently compact to suggest that they were confined by functional boundaries. Specifically, Brown et al. suggested that presumptive clones were aligned into functional columns, very similar to their excitatory counterparts (Brown et al., 2011; Yu et al., 2012), raising the possibility of a lineage-dependent functional matching in the organization of inhibitory and excitatory neurons (Brown et al., 2011). Ciceri et al. (2013) did not detect such radial clusters but rather described exclusively laminar clusters. In contrast, Mayer et al. (2015) and Harwell et al. (2015) both concluded that clonally related interneurons can disperse across anatomical and functional boundaries within the forebrain and are not restricted to narrow cortical columns or lamina. Notably, Mayer et al. (2015) and Harwell et al. (2015) agreed that sibling interneurons reside in a volume that far exceeds functional cortical units, such as the whisker barrels (Bruno et al., 2003) of the somatosensory cortex (the average distance between pairs of sibling neurons was >2 mm in Mayer et al., 2015). These data imply that the integration of interneurons into functional units is unlikely to be determined by lineage.

Can Cluster Analysis Be Used to Determine Lineal Relationships Between Interneurons?

A common feature of all studies considered above is that interneuron progenitors in the MGE of mouse embryos were labeled through infection using very similar fluorescently tagged retroviruses. What then explains the disparate results reported in these four investigations? Discrepancies almost certainly arose from the different methods used to assess and define interneuron clonality. Mayer et al. (2015) and Harwell et al. (2015) used a replication-defective retroviral library containing a highly diverse set of DNA barcodes, an approach pioneered by Walsh and Cepko in the early 1990s (Walsh and Cepko, 1993), to determine lineal relations between labeled interneurons. Recovering the barcodes from the mature progeny of infected progenitor cells enabled Mayer et al. and Harwell et al. to unambiguously determine the lineal relationship between clones regardless of their geometric distribution within the

brain. In contrast, Brown et al. (2011) and Ciceri et al. (2013) used a combination of approaches, including (1) time-lapse imaging (before migration), (2) mixing of red and green retroviruses, and (3) presumptive clonal labeling with low-titer retrovirus injections followed by the use of geometric criteria to infer lineal relationships among retrovirally labeled neurons. For the following reasons, we believe that none of the aforementioned methods used by Brown et al. (2011) and Ciceri et al. (2013) reliably indicated lineal relationships among interneurons.

First, whereas in principle time-lapse imaging could be used to determine lineal relationships, this approach is impractical, given both the distances involved and the protracted time over which interneurons migrate from their birth to their settling position. Second, the use of red and green retroviruses is confounded by technical difficulties that, when addressed by Ciceri et al. (2013), revealed that assigned clusters of interneurons are polyclonal in nature. In brief, they reported that when retroviruses encoding green fluorescent protein (GFP) and mCherry were mixed before ultracentrifugation, “most clusters were likely to include cells from a different progenitor (i.e., a different fluorescent protein), even at very limiting dilutions.” The authors concluded that “this strongly suggested that lineage relationships are not exclusive determinants of interneuron clustering.” Third, whereas low-titer retroviral injections can in principle be used to determine lineal relationships, in practice this proves untenable. If one could reliably label a single progenitor with a single injection, it would of course be possible to trace interneurons in the forebrain, even if individual siblings pursued drastically different migration paths. However, the labeling of a single progenitor cell cannot be guaranteed using current technology. Retroviral labeling of multiple progenitor cells unavoidably results in both lumping errors (clustered cells that are not clonal) and splitting errors (dispersed cells that are clonal but are not recognized as such), particularly if cells undergo complex migration.

Brown et al. (2011) attempted to minimize lumping and splitting errors by using low-titer retroviral injections “to label dividing progenitor cells in the ventricular zone...at clonal density.” Given this claim, we were surprised when we looked at the raw data provided by Dr. Shi (Figs. 1A, B) to see that individual brains showed >500 labeled cells—far exceeding what our analysis indicated would allow for “clonal labeling.” These data, we believe, preclude the assignment of lineage using the geometric criteria used by Brown et al. (2011) and Ciceri et al. (2013). More specifically, to assign lineage after interneuron

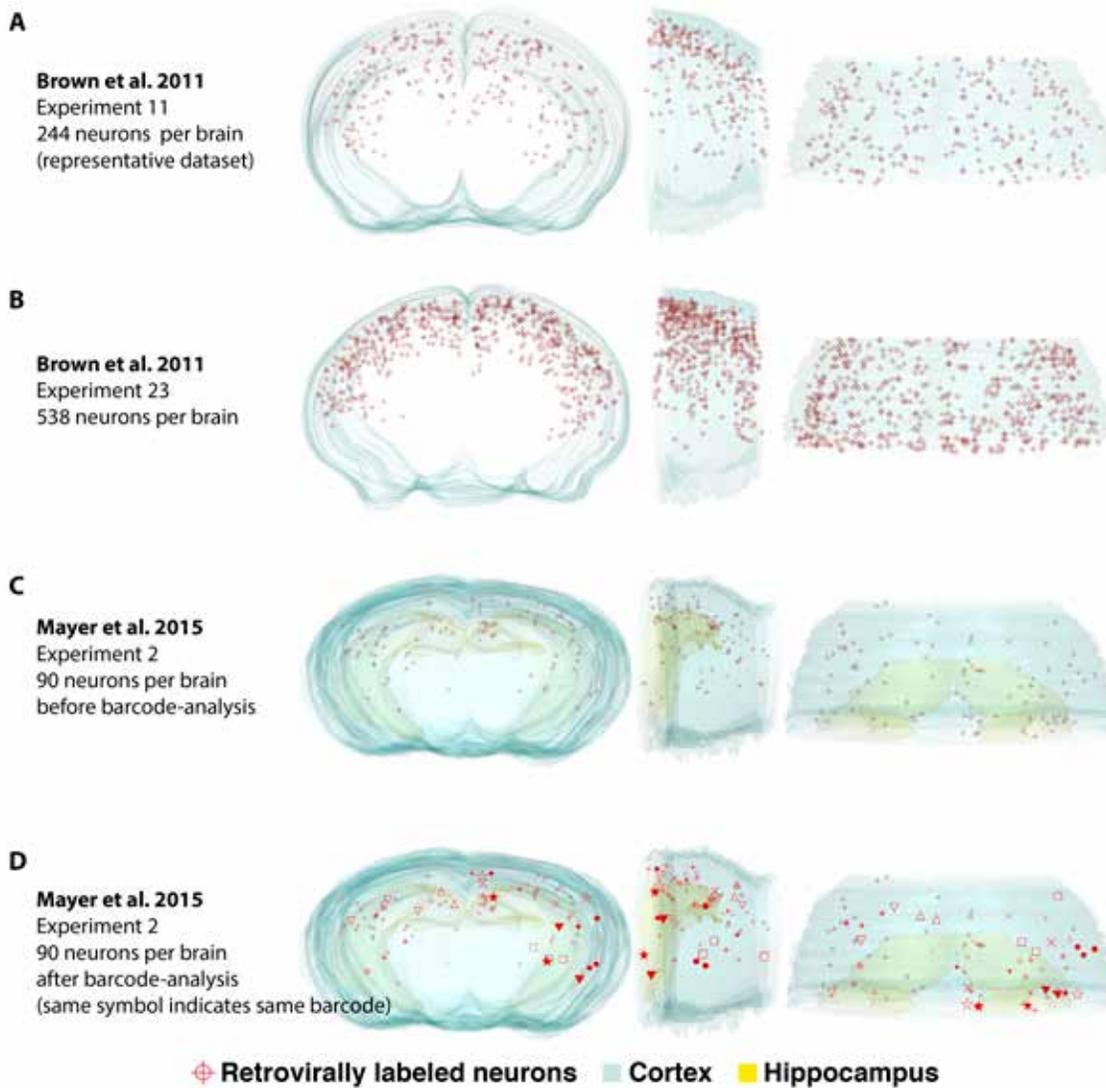


Figure 1. Comparison of the distributions of retrovirally infected interneurons in Mayer et al. (2015) and Brown et al. (2011). **A, B,** Two experimental datasets from Brown et al. are shown. Three-dimensional reconstructions of the distribution of cortical interneurons in a postnatal $Nkx2.1^{Cre/+};R26^{LSL-TVAiLacZ/+}$ mouse infected with retroviruses expressing enhanced green fluorescence protein (EGFP). Datasets in Brown et al. contained ≤ 538 data points per brain. To predict clonal relations of inhibitory interneurons, Brown et al. applied spatial parameters based on the observed distributions of excitatory neuron clusters (not shown). **C, D,** Three-dimensional reconstructions of a representative dataset reproduced from Mayer et al. (2015), illustrating the distribution of cortical interneurons in a postnatal $Nkx2.1^{Cre/+};R26^{LSL-TVAiLacZ/+}$ mouse that was infected with a retroviral library. The same dataset is shown, before **C** and after **D** determination of clonal relations based on retroviral barcodes. The dark red symbols (stars, circles, or triangles) represent single-cell clones (i.e., neurons harboring a barcode that occurred only once in the dataset); light red symbols represent multicell clones, whereby symbols with the same shape indicate the location of sister interneurons (i.e., neurons with the same barcode).

labeling and migration, both Brown et al. and Ciceri et al. compared the distance from each interneuron to its closest neighbor (nearest neighbor distance [NND]) with a randomly computer-simulated dataset to test whether the labeled interneurons were clustered. Ciceri et al. then calculated the number of clusters in the experiment using a threshold distance value that maximized the difference between the number

of clusters observed in the experimental dataset and the mean number of clusters in 100 simulated populations of randomly distributed neurons. Brown et al. used spatial parameters that picked up excitatory neuron clusters to predict clonally related inhibitory interneuron clusters.

Because these methods require that any “clonal”

NOTES

group of cells be constrained to a specified geometric area, as a matter of principle, these methods cannot be used to study dispersed clones that reside in different forebrain structures or distant locations within the neocortex. In addition to missing clonal dispersion across areas, our findings (as well as those of Harwell et al., 2015) demonstrated that the use of such geometric criteria also failed to predict clonality of interneurons within the cortex. When local clusters are deemed to be clonal clusters, lumping errors are a major confounding factor, particularly for datasets with a large number of total neurons (e.g., those used in Brown et al., 2011; Figs. 1A, B). This is because as the number of labeled neurons in a dataset increases, the chance that a nonclonally related cell will be found nearby clonally related cells also increases. Sultan et al. (2016) recognized this point, as they stated, “a clone forming a local cluster does not preclude the presence of nearby non-clonally related interneurons.... The more data points, the shorter the distance in general between them. Therefore, it is crucial to take into consideration the total number of data points in each dataset.” Even with much lower rates of infections per brain (Fig. 1C), Mayer et al. (2015) and Harwell et al. (2015) reported a large number of interneurons that were nearest neighbors but not clonally related (i.e., they had different DNA barcodes, indicating that they originated from different progenitors) (see dendrogram analysis in Mayer et al., 2015). In their recently published article in *Neuron*, Harwell et al. provided an additional detailed analysis, showing that the spatial parameters used in Brown et al. (2011) to cluster interneurons had failed to identify lineal boundaries in either our dataset or their own.

Complications Arising from the Analysis by Sultan et al.

As outlined in their upcoming article in *Neuron*, the lab of Dr. Shi (Sultan et al., 2016) further analyzed our datasets and concluded that clonally related interneurons in our datasets were not randomly dispersed. Their study implied that this contradicted our findings, attributing to us conclusions to which we do not subscribe. We hold that the real discrepancy between our conclusions and those of Sultan et al. is semantic, coming down to how we precisely define a cluster. “Clusters,” per definition, are a group of things that occur close together. In the cases of Brown et al. (2011) and Sultan et al. (2016), clusters were determined geometrically, as groups of cells that occur closer to each other than predicted in a random distribution (random computer-simulated cells). We completely agree that retrovirally labeled cohorts of interneurons appear clustered when compared with

a randomly distributed (computer-simulated) group of data points, but given the biological constraints placed on interneuron development, this should come as no surprise. For example, it is known that interneurons’ ultimate location in the brain is heavily influenced by several factors: (1) their position and time of birth (Miyoshi et al., 2007), (2) prescribed paths of migration (Tanaka et al., 2006; Marin, 2013), and (3) stereotyped radial migration from the marginal and subventricular zones to the cortical plate (Miyoshi and Fishell, 2011). All these factors indicate that although the dispersion of interneurons is perhaps stochastic, it is also tightly regulated, and therefore a random dispersion model will be grossly inaccurate.

Are Interneuron Clones Preferentially Clustered?

Similar to the analysis done in Mayer et al. (2015), but for cortical clones only, we further examined whether the average distance between pairs of lineage-related interneurons is preferentially reduced compared with unrelated interneurons. We found that the results for average distance between pairs of neurons is not influenced by the total number of data points in individual datasets (unlike, e.g., the NND; Fig. 2A), thus providing a robust measure for comparing clonally related and unrelated cells. Notably, both lineage-related and lineage-unrelated interneurons were labeled at the same time and with the same method, ensuring that they shared similar birthdates and migratory trajectories. The “intraclonal distance” was calculated as the average distance between pairs of clonally related interneurons, and the “interclonal distance” was calculated as the average distance between pairs of unrelated cells within one hemisphere (Fig. 2B). “Pairs of unrelated cells” included the distance between all possible pairs of interneurons with different barcodes: (1) individual members of “multicell clones” with different barcodes, (2) “single-cell clones,” and (3) individual members of “multicell clones” and “single-cell clones.” Significantly, the average distance between 40 pairs of clonally related interneurons in the cortex of P16 mice (average distance [AD] = 2134 ± 213 , SEM) was not statistically different from 926 pairs of clonally unrelated interneurons (AD = 2145 ± 34 , SEM; $p > 0.9$, Kruskal–Wallis test, multiple comparison; $p = 0.6$, Mann–Whitney nonparametric t test) (Fig. 2C). When we broke down the analysis by dataset (i.e., for each retrovirally infected brain), we did not detect a statistical difference despite the low numbers of clonally related pairs in each analysis ($p > 0.1$ in all three datasets, Mann–Whitney nonparametric t test) (Fig. 2D). Taken together, our results indicate

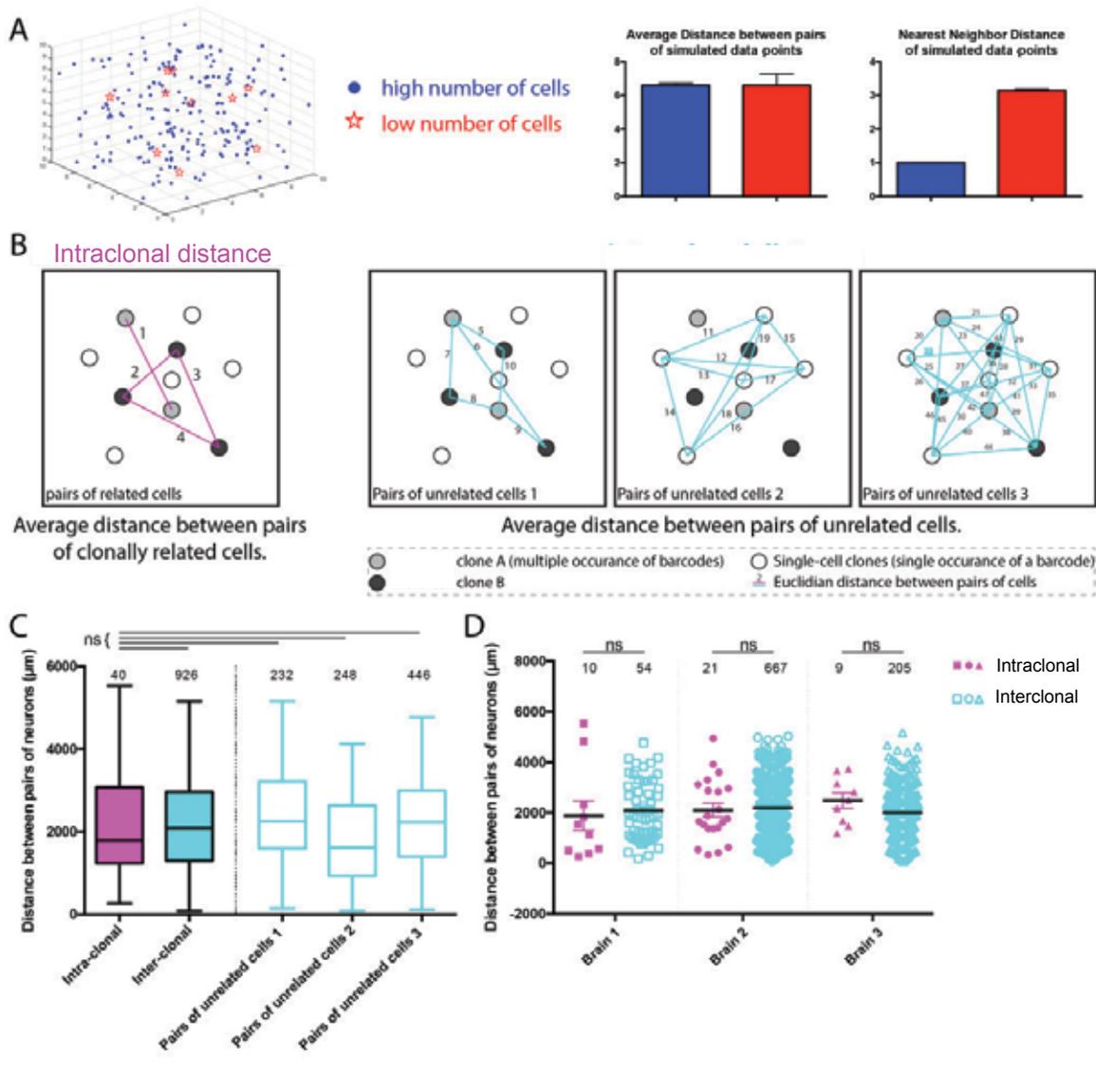


Figure 2. Interneuron clones within the cortex in Mayer et al. (2015) are not spatially segregated when compared with a biologically appropriate control group. **A**, The NND decreases as the number of cells per dataset increases. Notably, the AD between pairs of neurons is not influenced by the total number of data points in individual datasets. To illustrate this principle, NNDs and ADs were calculated for simulated datasets containing a high number of cells (200; blue dots) and a low number of cells (10; red five-pointed stars) in a given volume; $N = 100$ simulations; **B**, Schematic illustration showing an analysis similar to that done in Mayer et al. (2015), except that in the present case, only included cortical intraclonal and interclonal distances were calculated for interneurons. The intraclonal distance was calculated as the average distance between pairs of clonally related interneurons. The interclonal distance represents the sum of distances between (1) individual members of “multicell clones” with different barcodes, (2) “single-cell clones,” and (3) individual members of “multicell clones” and “single-cell clones.” **C**, Box-and-whiskers plot of the intraclonal and interclonal distance. Whiskers indicate minimum-to-maximum values. All three datasets from Mayer et al. were included in this analysis. The interclonal distance represents the sum of the three pairwise comparisons between (1) multicellular but unrelated clones, (2) single-cell clones, and (3) individual members of multicellular and single-cell clones (**B**). No significant difference in separation was observed when comparing intraclonal and interclonal distances (Kruskal–Wallis test, multiple comparison; Mann–Whitney nonparametric t test). The number above the boxes indicates the number (n) of interneuron pairs. **D**, Scatter plot of intraclonal and interclonal distances by brain (1–3). No significant difference in separation was observed when intraclonal and interclonal distances were compared (Kruskal–Wallis test, multiple comparison; Mann–Whitney nonparametric t test); the number above the boxes indicates the number (n) of interneuron pairs.

NOTES

that in general, clonally related cells are not located closer to each other than a biologically similar group of nonrelated interneurons.

Dendrogram Analysis Has Limited Utility in Determining Clonal Clusters

As mentioned above, lumping errors and splitting errors cannot be avoided if clonal clusters are defined geometrically because these methods implicitly assume that neighboring cells are clonally related. The separation between cells that are clustered versus not clustered is strongly influenced by the total number of data points in the dendrogram. In Mayer et al. (2015), we performed a dendrogram analysis to illustrate that this is an inherent problem when using geometrical methods. In brief, we grouped GFP-labeled neurons, regardless of lineage, by their proximity and displayed the results in dendrograms (Mayer et al., 2015). We then labeled the neurons according to their lineage relationship (i.e., barcode identity). Despite the fact that we labeled a relatively small number of neurons (e.g., much lower than in Brown et al., 2011; Fig. 2) in our dataset, only 52% of clones (12 out of 23) formed closest nearest neighbors (lowest hierarchical branch in the dendrogram; Mayer et al., 2015). In addition, a number of the clones that were closest nearest neighbors on the dendrogram had at least one “split” sibling on a far branch on the dendrogram. Sultan et al. (2016) reached a very similar result: they also found that 52% of clones (14/27) were closest nearest neighbors. It is critical to note that our results preclude the use of dendrogram analysis to determine the lineage relationships between neurons, and it was never our intention to use it for that purpose.

We would like to address the “error corrections” made by Sultan et al. (2016) when they reanalyzed the dendrogram analysis presented in Mayer et al. (2015). In particular, Sultan et al. stated that we failed to add clone #32 to our dendrogram. However, calling this an “error correction” is inaccurate because we deliberately excluded this clone from our analysis. Clone #32 was located within the olfactory bulb, and the dendrogram analysis in Mayer et al. “focused on cortical, hippocampal, and striatal clones only.” In another such “error correction,” Sultan et al. noted that clone #12 contained three cells in the cortex and three cells in the hippocampus, requiring them to “add all six clones to the dendrogram.” This statement implies that we incorrectly excluded all these cells from our dendrogram, which again is inaccurate. We deliberately divided clones that

crossed anatomical boundaries for analysis within brain structures.

The Use of Euclidian Distance Measurements

Sultan et al. (2016) noted that the use of Euclidian distances in structures such as the cortex is problematic, as it is clear that in many if not most cases, migration along straight lines (e.g., in cases where such trajectories would cross ventricles or sulci) is not biologically tenable. Nonetheless, all distances between pairs of neurons described in Mayer et al. (2015) as well as Sultan et al. (2016) and Brown et al. (2011) were calculated as Euclidian distances. Given the impossibility of determining more realistic trajectories, this approach is at least systematic, and by its nature chronically underestimates the real distances between neurons. This only strengthens our conclusion that clonal dispersion does not respect functional boundaries, as properly corrected measurements of the distance between clones would only be larger rather than smaller.

Interneuron Clones Can Span Multiple Brain Structures

Although the results of Mayer et al. (2015) and Harwell et al. (2015) demonstrated that interneuron clones are not obliged to populate particular anatomical structures, this does not rule out the possibility that they are predetermined to occupy particular brain regions. Sultan et al. (2016) discussed this point as follows:

1. Should lineage relationship have no influence on interneuron distribution, the relatively total interneuron output to different forebrain structures and the small clone size dictate that virtually all clones must be located in the cortex, the cortex and hippocampus, or the cortex and striatum. Interestingly, a significant fraction of clones was observed to be restricted to the hippocampus or striatum (Mayer et al., 2015), suggesting that some MGE/PoA [medial ganglionic eminence/preoptic area] progenitors specifically produce interneurons destined for these two brain structures.
2. While it is evident that the majority (~66% in Mayer et al. and 80% in Harwell et al.) of clones are located within one brain structure, i.e. the cortex, some are dispersed in more than one brain structure. However, this clonal dispersion largely occurs between the cortex and hippocampus, the two highly related forebrain structures emerging

side-by-side in the dorsal telencephalon. The same tangential migration routes are responsible for interneuron distribution in the cortex and hippocampus (Ayala et al., 2007; Marin and Rubenstein, 2001, 2003). In comparison, only a small fraction (~12.5% in Mayer et al. and 20% in Harwell et al.) of clones is dispersed between developmentally unrelated brain structures such as the cortex and striatum, or globus pallidus, or olfactory bulb.

Although these statements are factual, understanding their implications requires a more nuanced analysis. Both the absolute size of the cortex, hippocampus, and striatum as well as the density of interneurons within these structures differ dramatically. For example, 20% of the cells within the cortex and hippocampus are interneurons (Fishell and Rudy, 2011), whereas the percentage of interneurons within the striatum is only 3% (Marin et al., 2000; Tepper et al., 2010). These facts demonstrate that even if interneurons were randomly distributed to different structures, probabilistically, they would be preferentially found in the cortex. That said, we reiterate that we do not believe that the distribution of interneurons is random. But what rules then underlie the distribution of discrete interneuron lineages? Our results definitively indicate that if interneuron lineages do have a covert logic as to how they populate different structures, clearly the rules of allocation are not as simple as an interneuron lineage being earmarked for cortex or hippocampus per se. Further examination of interneuron lineages will be required to address whether there is a degree of predetermination in the positioning of sibling neurons derived from a common lineage.

Final Remarks

We have shown here and in previous work (Mayer et al., 2015) that clonally related interneurons are no more closely clustered than nonlinearly related interneurons (proximally generated brethren). These findings, of course, neither should nor do end the debate as to whether lineage contributes to the development, subtype differentiation, or connectivity of interneurons. Our results were limited by the fact that the lineages we assembled were only partially reconstructed, so we can say nothing regarding the fate of those sibling cells that we failed to recover. In addition, we know startling little about the phenotypic identity and nothing about the connectivity of clonally related siblings, both of which would be fascinating to explore. We would, however, implore any further examination of lineage to confine itself to methods

that provide a high degree of confidence about the lineage relationships of cells designated as clones.

Acknowledgments

We are grateful to Xavier Jaglin, Robert Machold, and Timothy Petros for comments on this manuscript. Research in the Fishell Laboratory is supported by the National Institutes of Health (grants NS081297, MH095147, and P01NS074972) and the Simons Foundation.

This paper was excerpted and modified with permission from a previously published article: Mayer, C; Bandler, RC; Fishell, G (2016) Lineage is a poor predictor of interneuron positioning within the forebrain. *Neuron* 92:45 – 51, copyright 2016, Elsevier.

References

- Anderson SA, Eisenstat DD, Shi L, Rubenstein JL (1997) Interneuron migration from basal forebrain to neocortex: dependence on *Dlx* genes. *Science* 278:474–476.
- Ayala R, Shu T, Tsai LH (2007) Trekking across the brain: the journey of neuronal migration. *Cell* 128:29–43.
- Brown KN, Chen S, Han Z, Lu C-H, Tan X, Zhang X-J, Ding L, Lopez-Cruz A, Saur D, Anderson SA, Huang K, Shi S-H (2011) Clonal production and organization of inhibitory interneurons in the neocortex. *Science* 334:480–486.
- Bruno RM, Khatri V, Land PW, Simons DJ (2003) Thalamocortical angular tuning domains within individual barrels of rat somatosensory cortex. *J Neurosci* 23:9565–9574.
- Ciceri G, Dehorter N, Sols I, Huang ZJ, Maravall M, Marin O (2013) Lineage-specific laminar organization of cortical GABAergic interneurons. *Nat Neurosci* 16:1199–1210.
- Fishell G, Rudy B (2011) Mechanisms of inhibition within the telencephalon: “where the wild things are.” *Annu Rev Neurosci* 34:535–567.
- Harwell CC, Fuentealba LC, Gonzalez-Cerrillo A, Parker PRL, Gertz CC, Mazzola E, Garcia MT, Alvarez-Buylla A, Cepko CL, Kriegstein AR (2015) Wide dispersion and diversity of clonally related inhibitory interneurons. *Neuron* 87:999–1007.
- Li Y, Lu H, Cheng P-L, Ge S, Huuatai X, Shi S-H, Dan Y (2012) Clonally related visual cortical neurons show similar stimulus feature selectivity. *Nature* 486:118–121.

NOTES

- Marin O (2013) Cellular and molecular mechanisms controlling the migration of neocortical interneurons. *Eur J Neurosci* 38:2019–2029.
- Marin O, Anderson SA, Rubenstein JL (2000) Origin and molecular specification of striatal interneurons. *J Neurosci* 20:6063–6076.
- Marin O, Rubenstein JLR (2001) A long, remarkable journey: tangential migration in the telencephalon. *Nat Rev Neurosci* 2:780–790.
- Marin O, Rubenstein JL (2003) Cell migration in the forebrain. *Annu Rev Neurosci* 26:441–483.
- Mayer C, Jaglin XH, Cobbs LV, Bandler RC, Streicher C, Cepko CL, Hippenmeyer S, Fishell G (2015) Clonally related forebrain interneurons disperse broadly across both functional areas and structural boundaries. *Neuron* 87:989–998.
- Miyoshi G, Butt SJB, Takebayashi H, Fishell G (2007) Physiologically distinct temporal cohorts of cortical interneurons arise from telencephalic Olig2-expressing precursors. *J Neurosci* 27:7786–7798.
- Miyoshi G, Fishell G (2011) GABAergic interneuron lineages selectively sort into specific cortical layers during early postnatal development. *Cereb Cortex* 21:845–852.
- Noctor SC, Flint AC, Weissman TA, Dammerman RS, Kriegstein AR (2001) Neurons derived from radial glial cells establish radial units in neocortex. *Nature* 409:714–720.
- Rubenstein JL, Shimamura K, Martinez S, Puelles L (1998) Regionalization of the prosencephalic neural plate. *Annu Rev Neurosci* 21:445–477.
- Sultan KT, Shi W, Shi SH (2016) Clonally related GABAergic interneurons do not randomly disperse but frequently form local clusters in the forebrain. *Neuron*: in press.
- Tanaka DH, Maekawa K, Yanagawa Y, Obata K, Murakami F (2006) Multidirectional and multizonal tangential migration of GABAergic interneurons in the developing cerebral cortex. *Development* 133:2167–2176.
- Tepper JM, Tecuapetla F, Koos T, Ibáñez-Sandoval O (2010) Heterogeneity and diversity of striatal GABAergic interneurons. *Front Neuroanat* 4:150. doi:10.3389/fnana.2010.00150
- Walsh C, Cepko CL (1993) Clonal dispersion in proliferative layers of developing cerebral cortex. *Nature* 362:632–635.
- Yu Y-C, He S, Chen S, Fu Y, Brown KN, Yao X-H, Ma J, Gao KP, Sosinsky GE, Huang K, Shi S-H (2012) Preferential electrical coupling regulates neocortical lineage-dependent microcircuit assembly. *Nature* 486:113–117.

Genetic Techniques for Cell Lineage Tracing in the Nervous System

Kelly Girsakis, BA,¹ Mollie Woodworth, PhD,¹
and Christopher Walsh, MD, PhD^{1,2,3}

¹Division of Genetics and Genomics
Manton Center for Orphan Disease Research and
Howard Hughes Medical Institute
Boston Children's Hospital
Boston, Massachusetts

²Departments of Pediatrics and Neurology
Harvard Medical School
Boston, Massachusetts

³Broad Institute of MIT and Harvard
Cambridge, Massachusetts

Introduction

Deriving lineage relationships between cells in a developing organism, or between an early dividing cell of unknown potential and its descendants, has been a long-standing interest in developmental biology. In recent years, many new methods have emerged to enable cell lineage tracing with increasing resolution, leading to substantial biological insights. In model organisms, novel cellular labels, such as barcoded retroviral libraries (Gerrits et al., 2010) and a rainbow of available fluorescent proteins (Cai et al., 2013), have increased the number of founder cells that can be uniquely labeled and traced. Unlike most early cellular tracers, labels inserted into the genome can permanently mark lineages in a variety of experimental organisms without being diluted by cell division, and these modifications are facilitated by genome editing technologies (Hsu et al., 2014). In addition, recent advances in sequencing have enabled naturally occurring somatic mosaic mutations to be used as lineage marks in both cancerous tissue (Navin et al., 2011; Wang et al., 2014) and normal tissue (Behjati et al., 2014; Lodato et al., 2015), illuminating a future in which lineage tracing moves from experimental organisms into humans.

Prospective Methods of Lineage Tracing

A classic genetic approach to cell lineage analysis is performed by labeling a single founder cell and tracing its progeny over time. This prospective method has been used since biological dyes mapped the fate of cells within chicken and mouse embryos in early observational studies (Beddington, 1981; Serbedzija et al., 1989), and continues to be used in current lineage tracing experiments. Whereas early developmental studies hoped to achieve clonal labeling by microinjecting small amounts of dye into an area of interest, advancements in genetic tools for prospective lineage tracing now allow for far greater cell and tissue specificity, recombinase-based intersectional analyses, and single-cell resolution.

Sparse retroviral labeling for lineage tracing

Since the advent of recombinant DNA technology in the late 1980s, retroviral libraries containing reporter transgenes such as β -galactosidase (β -gal) and green fluorescent protein (GFP) have been used for cell labeling and lineage tracing in vertebrate animal models (Turner and Cepko, 1987; Frank and Sanes, 1991). Retroviral vector-mediated gene transfer allows viruses to introduce recombinant DNA into a host cell's genome. The integrated exogenous DNA is then inherited by all descendants of the

infected cell. The DNA encodes a histochemical or fluorescent protein that can be easily assayed to label cells of a “clone” and elucidate cell fate choices within that clone. Histological and morphological analyses of the progeny of virally infected cells allow for *post hoc* fate mapping within a clonally related cell population.

Sparse retroviral infection has also been used in live cell imaging of progenitors and their progeny in organotypic slice culture. Mouse, ferret, chimpanzee, and human progenitors have all been analyzed using time-lapse imaging. Individual progenitors labeled by fluorescent reporter genes are visualized using confocal microscopy for multiple cellular divisions. At the end of the imaging experiment, immunohistochemistry and cellular morphology can then be used to analyze cell fate within the imaged clone (Noctor et al., 2001, 2004; Brown et al., 2011; Gertz et al., 2014; Dehay et al., 2015). Although *ex vivo* organotypic culturing conditions closely mimic the *in vivo* cellular environment, such experiments typically can be performed for only a few days at most, and so cannot typically relate clonal relationship to adult structure.

Sparse retroviral labeling requires that clonality be inferred based solely on proximity of cells expressing a reporter gene. Early studies in the cerebral cortex soon showed that sibling cells dispersed widely from one another in some clones (Walsh and Cepko, 1988). To analyze such widespread clones, the first retroviral libraries were developed, encoding the *lacZ* gene as a reporter, but also short DNA fragments to act as barcode tags (Walsh and Cepko, 1992). Clonal relationships were then directly revealed by PCR amplification of the integrated barcode tags from cells dissected from tissue sections, rather than being inferred based on proximity alone. Cells derived from a common progenitor share the same DNA tag at the vector integration site regardless of their patterns of migration, whereas clonally unrelated cells harbor different barcodes. The first library of a hundred tags soon expanded to a thousand tags (Walsh and Cepko, 1993; Reid et al., 1995) and then to essentially unlimited complexity using random oligonucleotide barcodes of identical size but distinct sequence (Golden et al., 1995; Fuentealba et al., 2015).

Advancements in transgenic animal lines have also extended the applications of retroviral genetic tagging and fate mapping. Cell-type specificity can now be achieved with transgenic mouse lines expressing virus receptors under the control of a cell-type-specific promoter (Harwell et al., 2015; Mayer et al., 2015). Only dividing cells that contain the virus receptor can

NOTES

be infected and express the reporter gene or barcode, allowing for more-precise viral targeting. Barcode tags can then be recovered using fluorescence-activated cell sorting (FACS) by the fluorescent reporter transgene, or laser capture microdissection (LCM) techniques that can preserve cellular position within the infected tissue for future reconstruction and analysis.

Although retroviral library labeling is useful for determining lineage relationships both *in vivo* and, it has some considerations and limitations: (1) only cells with the capacity to divide will propagate the barcode to progeny, (2) retroviral vectors typically spontaneously silence, so many retrovirally transfected cells are no longer histochemically labeled even though their DNA can be detected in the tissue, and (3) barcode tag recovery from single cells can prove challenging (Mayer et al., 2015). To circumvent the possibility of spontaneous retroviral silencing, new studies have been combining retroviral library labeling with RNA-sequencing (RNA-seq) technology. These studies not only recover barcodes to trace clonal lineage relationships but can also elucidate cell type using transcriptomics in sparse or heterogeneous cell populations (Lu et al., 2011). This valuable advance allows for the overlay of phenotypic cell identity with genetic lineage information for a more comprehensive view of clonal relationships.

Plasmid transfection labeling for lineage tracing

In addition to viral infection, reporter transgenes for cell labeling and fate mapping can be introduced into cells using DNA plasmid transfection. Lipofection, a common lipid-based system, has been used to transfect the developing *Xenopus* retina and to trace retinal cell fate *in vivo* (Holt et al., 1990). Electroporation, an alternative nonviral delivery method, has been used to deliver reporter transgenes encoding fluorescent proteins to trace cells both *in vitro* and in various vertebrate animal models (Fukuchi-Shimogori and Grove, 2001; Emerson and Cepko, 2011). Reporter gene plasmids can be injected into the developing brain's ventricles and introduced into neural progenitors lining the ventricular wall by electrical pulses. A reporter transgene, such as *GFP*, is then carried episomally by the progenitor cell and passed on to subsequent daughter cells. Unlike retroviral labeling, however, plasmid DNA is not integrated into the progenitor's genome and becomes diluted or inactivated in progeny after serial cellular divisions. Plasmid electroporation techniques, therefore, are transient and fail to label the entire lineage (LoTurco et al., 2009).

A solution to plasmid loss or inactivation is a DNA transposon system, which stably integrates the reporter transgene into the progenitor's genome. Transposon systems include Mos1, Tol2, Sleeping Beauty (SB), and piggyBac (PB), which all use a dual-plasmid approach with a "cut-and-paste" mechanism (Wu et al., 2006; VandenDriessche et al., 2009; Yoshida et al., 2010). The typical transposon system includes a donor plasmid containing the reporter transgene of interest and a helper plasmid that expresses the transposase. The donor plasmid includes terminal repeats flanking the transgene, which allows for genomic integration by the transposase. The transgene is then propagated to all progeny within the lineage, but the transposase (like any episomal plasmid) will be diluted over cellular divisions. Donor and helper plasmids can be driven by different promoters, allowing for cell-type specificity and genetic intersectional analyses. Compared with the other transposon systems, PB has a more precise cut-and-paste mechanism, higher transposition efficiency, and a larger cargo capacity (Chen and LoTurco, 2012). These attributes have made the PB transposon system particularly popular. In addition, PB transposase can be co-electroporated with multiple fluorescent reporter constructs, each driven by a cell-type-specific promoter. In this experimental design, multiple lineages can be examined in a single animal (Siddiqi et al., 2014). PiggyBac has been successfully used in multiple mammalian cell lines and in combination with *in utero* electroporation (IUE) to trace and manipulate cell lineages in animal models (Ding et al., 2005; Wilson et al., 2007; Woltjen et al., 2009; Siddiqi et al., 2014).

The piggyBac transposon plasmid system allows remarkable flexibility and cell-type specificity, but as with any random genomic insertion event, the precise location or number of transposition occurrences introduces a risk of confounded results due to mutagenesis. Transposition of the reporter transgene may cause endogenous genes at or near the insertion site to become unintentionally dysregulated. One study, however, found no evidence of mutagenesis by transposon insertion in cells labeled with the PB IUE method (Chen and LoTurco, 2012). Transposase plasmid systems are a remarkable tool for transgenesis and cell lineage tracing in both classically genetically modifiable animal models, such as mice, and otherwise non-genetically tractable animals, such as the ferret.

Genetic recombination for lineage tracing

Cell lineage tracing by genetic recombination leverages the expression of recombinase enzymes in a cell-specific or tissue-specific manner to activate expression of a conditional reporter gene. Two genetically encoded, site-specific recombination systems include Cre-loxP and FLP-FRT. In the Cre-loxP system, mice are engineered to express Cre recombinase under the control of a chosen promoter, limiting Cre expression to a specific tissue or cell type (Orban et al., 1992). These mice are then crossed with a second mouse line in which a reporter transgene, such as *lacZ* or *GFP*, is preceded by a loxP-flanked transcriptional stop cassette. In cells expressing Cre recombinase, the STOP sequence is excised, and the reporter transgene is expressed. Temporal control of recombination can be gained by using an inducible Cre system, which selectively activates Cre under promoters that are also active at undesired time points such as embryogenesis. In an inducible system, Cre recombinase is fused to the human estrogen or progesterone receptor and activated only with the presence of an anti-estrogen such as tamoxifen or an anti-progestin, respectively. A pulse-chase strategy with an inducible Cre system can be used to determine lineage relationships. Leakiness, however, is a common problem of inducible Cre systems; nonetheless, these inducible systems have been used for lineage tracing in many adult tissues.

Dual or multicolor reporter lines have become increasingly popular for tracing cell lineage relationships. Mosaic analysis with double markers (MADM) uses a Cre-loxP system to express GFP and red fluorescent protein in cell populations of interest (Zong et al., 2005). Before recombination, no reporter transgene is expressed, but after Cre recombinase is activated, one or both transgenes are reconstituted. Green, red, or double-labeled yellow cells are generated depending on the recombination and the chromosomal segregation type. MADM can be used with cell-type-specific and inducible Cre systems to provide single-cell resolution and to more precisely examine progenitor division patterns (Zong et al., 2005; Hippenmeyer et al., 2010; Bonaguidi et al., 2011; Mayer et al., 2015). Multicolor lineage tracing is also possible with recent mouse reporter lines, including Brainbow and Confetti (Livet et al., 2007; Snippert et al., 2010). The Brainbow mouse lines harness stochastic Cre-mediated recombination using incompatible loxP sites to drive combinatorial expression of fluorescent reporter transgenes. The Brainbow mouse can label individual cells with ≤ 90 distinguishable colors by stochastic expression of several fluorescent reporter

transgenes. Cells expressing a particular color share a common lineage. A modified line, the Confetti mouse, ubiquitously expresses Cre from the ROSA26 locus and has been used to trace individual stem cell lineages in the mouse intestinal crypt (Snippert et al., 2010). Owing to the expression of a multitude of unique colors, costaining with antibodies to determine protein expression within Brainbow or Confetti mice is nearly impossible. Endogenous fluorescence of the reporter genes, however, can be used for imaging clones. Advancements in microscopy, such as the two-photon microscope, continue to make these lines an attractive choice for *in vivo* cell lineage tracing.

Recent methodological advances in prospective lineage tracing

Innovations in both microfluidic platforms and genome editing strategies have also recently been used to prospectively trace cell lineage. Microfluidic technologies allow for capture and culture of single progenitor cells and up to five generations of their progeny on a single chip. *In vitro* time-lapse imaging for both division kinetics and identification of lineage relationships can be coupled with on-chip immunohistochemistry to assess cell fate within the captured clones. Clones can also be retrieved after culturing for single-cell transcriptomics with known lineage relationships. Kimmerling et al. (2015) used this microfluidic trap array technology, paired with single-cell RNA-seq, to look at both interclonal and intracolon variability in activated CD8⁺ T cells; they demonstrated that lineage-dependent transcriptional profiles corresponded to functional cellular phenotypes. This study was the first to link single-cell transcriptomics with cell lineage history (Kimmerling et al., 2015).

Recently, CRISPR/Cas9 genome editing technology (CRISPR signifies clustered regularly interspaced short palindromic repeats; Cas9 is a class of RNA-guided endonucleases) has been applied to trace and synthetically reconstruct cell lineage relationships in complex, multicellular organisms. McKenna et al. developed genome editing of synthetic target arrays for lineage tracing (GESTALT), a highly multiplexed method that uses barcodes composed of multiple CRISPR/Cas9 target sites (McKenna et al., 2016). These barcodes progressively and stably accumulate unique mutations over cellular divisions and can be recovered by targeted sequencing. Cell lineage relationships are determined based on the pattern of shared mutations among analyzed cells. While prospective in the sense that the barcode is introduced at the start of the experiment, the GESTALT method also parallels retrospective,

NOTES

NOTES

somatic-mutation-based tracing, discussed below. The incrementally edited barcodes from thousands of cells were used in large-scale reconstructions of multiple cell lineages within cell culture and zebrafish. Although precise anatomical position and cell type of each assayed cell cannot be determined with this method, this published study and others in progress demonstrate the potential for cumulative and combinatorial barcode editing in prospective lineage tracing of whole organisms (Junker et al., 2016; Kalhor et al., 2016; McKenna et al., 2016). Advances during the past 30 years, since the advent of genetic barcoding and recombinase-based transgenic animals, have allowed prospective cell lineage tracing experiments to not only uncover clonal relationships at the single-cell level but also map cell fate choices in a wide variety of cells, tissues, and model organisms.

Retrospective Methods of Lineage Tracing

It has only recently become possible to harness naturally occurring mutations to retrospectively infer cell lineage information, owing to advances in genome sequencing. Like prospective lineage tracers in model organisms, somatic mutations indelibly mark the progeny of the dividing cell in which they occurred, and the cells bearing these naturally occurring lineage marks can be analyzed later to reconstruct the genealogy of organs and cell types (Salipante et al., 2010). To use naturally occurring somatic mutations for lineage tracing, it is first necessary to discover mutations shared among multiple cells of that individual; however, somatic mutations are difficult to identify by sequencing a mixed population of cells at conventional depths, as they are low-frequency by nature. Nonetheless, the declining cost of deep next-generation genome sequencing and the advent of single-cell genome sequencing have made it possible to discover rare mutations that mark minority lineages within a larger cellular population (Shapiro et al., 2013). These variants—from the least frequently somatically mutated to the most—include retrotransposons, copy number variants, single-nucleotide variants (SNVs), and microsatellites. The different rates at which these variants occur in somatic tissues allow lineage tracing experiments to be conducted at different levels of granularity according to the types of variants, tissue, and disease state selected. Single-cell genome sequencing promises to revolutionize lineage tracing in humans, although potential technical artifacts and complications must be considered when planning a single-cell genome sequencing experiment. Critically, whole-genome sequencing

currently requires considerably more DNA than the 6 pg present in a single cell, necessitating a presequencing genome amplification step that may introduce errors (Grün and van Oudenaarden, 2015; Gawad et al., 2016).

Somatic mutations for lineage tracing in normal tissue

Endogenous retroelements, principally including long interspersed nuclear element-1 (LINE-1 or L1), compose much of the human genome; L1 elements alone make up nearly one-fifth of the genome (Ostertag and Kazazian, 2001). Some of these L1 elements retain the ability to mobilize in humans and can insert into a new genomic location during somatic cell division (Muotri et al., 2005). This mobilization has raised substantial interest in their potential contribution to somatic diversity, especially within complex tissues like the brain (Erwin et al., 2014). Estimates of L1 mobilization frequency derived by sorting single neurons, amplifying the whole genome, and analyzing L1 retrotransposition at a single-cell level (Evrony et al., 2012) suggest there are fewer than one somatic insertion per neuronal genome on average (Evrony et al., 2012). A single-neuron whole-genome sequencing study confirmed the low rate of L1 retrotransposition events but also illustrated the striking spatial distribution patterns of clonal retrotransposition events, providing strong proof of principle for the use of spontaneous somatic L1 events for lineage tracing (Evrony et al., 2015).

SNVs are a significant source of evolutionary and disease-causing mutations, yet they can also occur very frequently in noncoding portions of the genome without having functional effects on somatic cells. Somatic SNVs represent a rich source of lineage-marking mutations because they are both abundant and frequently functionally neutral. Indeed, work in mouse stomach, intestine, and prostate (Behjati et al., 2014), mouse brain (Hazen et al., 2016), and human brain (Lodato et al., 2015) suggests that somatic SNVs can be identified from single cells or clones and used to reconstruct developmental lineages; in one study, 9 of 16 sequenced neurons, and 136 of 226 total neurons from the same area of cortex, could be placed in a lineage tree with four independent clades that diverged before gastrulation. One clade contained a nested set of 11 somatic mutations, which were progressively regionally restricted across the brain and present in progressively decreasing frequency in bulk tissue (Lodato et al., 2015). These results suggest that analysis of such nested mutations might enable the analysis of the progressively narrower lineage trees characterizing the developing embryo.

The most frequently mutated somatic loci are microsatellites (Ellegren, 2004). Because of the instability of microsatellite repeats, analysis of all microsatellite locations in the genome is predicted to be capable of reconstructing the entire cell lineage tree of an organism (Frumkin et al., 2005), using methods adapted from organism-level phylogenetic analysis (Salipante et al., 2010). Like microsatellites, the polyadenylated tracts following somatic L1 retrotransposition events are subject to frequent polymerase slippage, and therefore, lineages defined by a somatic L1 retrotransposition event can be further delineated by analyzing poly-A tail polymorphisms (Evrony et al., 2015).

Perspective

When designing a lineage tracing experiment, it is important to consider the strengths and weaknesses of prospective and retrospective approaches. For prospective lineage tracing, there must be genetic access to the population in question, whether by a regionally directed method such as viral injection or electroporation, or by population-specific marker lines or promoters. Because prospective lineage tracing depends on labeling and follow-up analysis, its use is restricted to experimental organisms and cell-culture systems. Alternately, retrospective lineage tracing can investigate lineage directly in human tissue, allowing unprecedented access to lineage information relevant to human development and disease. Currently, retrospective lineage tracing relies heavily on sequencing, frequently of single cells, and is therefore lower-throughput and more expensive than most prospective methods. Although emerging prospective lineage systems are engineering revolutionary ways to investigate lineage in model organisms, it will always be necessary to retrospectively map lineage in naturally occurring tissues without engineered lineage marks.

No longer limited to tracing a small number of cells with serially diluted dyes, biologists can now access a variety of methods for tracing lineage forward from the application of a genetic label. Additionally, recent advances in sequencing—particularly genome sequencing of single cells—allow lineage tracing to be performed retrospectively, reconstructing lineage decisions that occurred well before sequencing. A hundred years after the first investigations of cell lineage, developmental biologists have built a tremendously enriched genetics toolkit for examining the developmental fate decisions that construct a whole organism.

References

- Beddington SP (1981) An autoradiographic analysis of the potency of embryonic ectoderm in the 8th day postimplantation mouse embryo. *J Embryol Exp Morphol* 64:87–104.
- Behjati S, Huch M, van Boxtel R, Karthaus W, Wedge DC, Tamuri AU, Martincorena I, Petljak M, Alexandrov LB, Gundem G, Tarpey PS, Roerink S, Blokker J, Maddison M, Mudie L, Robinson B, Nik-Zainal S, Campbell P, Goldman N, van de Wetering M, et al. (2014) Genome sequencing of normal cells reveals developmental lineages and mutational processes. *Nature* 513:422–425.
- Bonaguidi MA, Wheeler MA, Shapiro JS, Stadel RP, Sun GJ, Ming G-L, Song H (2011) clonal analysis reveals self-renewing and multipotent adult neural stem cell characteristics. *Cell* 145:1142–1155.
- Brown KN, Chen S, Han Z, Lu CH, Tan X, Zhang XJ, Ding L, Lopez-Cruz A, Saur D, Anderson SA, Huang K, Shi SH (2011) Clonal production and organization of inhibitory interneurons in the neocortex. *Science* 334:480–486.
- Cai D, Cohen KB, Luo T, Lichtman JW, Sanes JR (2013) Improved tools for the Brainbow toolbox. *Nat Methods* 10:540–547.
- Chen F, LoTurco J (2012) A method for stable transgenesis of radial glia lineage in rat neocortex by piggyBac mediated transposition. *J Neurosci Methods* 207:172–180.
- Dehay C, Kennedy H, Kosik KS (2015) The outer subventricular zone and primate-specific cortical complexification. *Neuron* 85:683–694.
- Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* 122:473–483.
- Ellegren H (2004) Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5:435–445.
- Emerson MM, Cepko CL (2011) Identification of a retina-specific *Otx2* enhancer element active in immature developing photoreceptors. *Dev Biol* 360:241–255.
- Erwin JA, Marchetto MC, Gage FH (2014) Mobile DNA elements in the generation of diversity and complexity in the brain. *Nat Rev Neurosci* 15:497–506.

NOTES

- Evrony GD, Cai X, Lee E, Hills LB, Elhosary PC, Lehmann HS, Parker JJ, Atabay KD, Gilmore EC, Poduri A, Park PJ, Walsh CA (2012) Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* 151:483–496.
- Evrony GD, Lee E, Mehta BK, Benjamini Y, Johnson RM, Cai X, Yang L, Haseley P, Lehmann HS, Park PJ, Walsh CA (2015) Cell lineage analysis in human brain using endogenous retroelements. *Neuron* 85:49–59.
- Frank E, Sanes JR (1991) Lineage of neurons and glia in chick dorsal root ganglia: analysis *in vivo* with a recombinant retrovirus. *Development* 111:895–908.
- Frumkin D, Wasserstrom A, Kaplan S, Feige U, Shapiro E (2005) Genomic variability within an organism exposes its cell lineage tree. *PLoS Comput Biol* 1:e50.
- Fuentealba LC, Rompani SB, Parraguez JI, Obernier K, Romero R, Cepko CL, Alvarez-Buylla A (2015) Embryonic origin of postnatal neural stem cells. *Cell* 161:1644–1655.
- Fukuchi-Shimogori T, Grove EA (2001) Neocortex patterning by the secreted signaling molecule FGF8. *Science* 294:1071–1074.
- Gawad C, Koh W, Quake SR (2016) Single-cell genome sequencing: current state of the science. *Nat Rev Genet* 17:175–188.
- Gerrits A, Dykstra B, Kalmykova OJ, Klauke K, Verovskaya E, Broekhuis MJC, de Haan G, Bystriykh LV (2010) Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* 115:2610–2618.
- Gertz CC, Lui JH, LaMonica BE, Wang X, Kriegstein AR (2014) Diverse behaviors of outer radial glia in developing ferret and human cortex. *J Neurosci* 34:2559–2570.
- Golden JA, Fields-Berry SC, Cepko CL (1995) Construction and characterization of a highly complex retroviral library for lineage analysis. *Proc Natl Acad Sci USA* 92:5704–5708.
- Grün D, van Oudenaarden A (2015) Design and analysis of single-cell sequencing experiments. *Cell* 163:799–810.
- Harwell CC, Fuentealba LC, Gonzalez-Cerrillo A, Parker PRL, Gertz CC, Mazzola E, Garcia MT, Alvarez-Buylla A, Cepko CL, Kriegstein AR (2015) Wide dispersion and diversity of clonally related inhibitory interneurons. *Neuron* 87:999–1007.
- Hazen JL, Faust GG, Rodriguez AR, Ferguson WC, Shumilina S, Clark RA, Boland MJ, Martin G, Chubukov P, Tsunemoto RK, Torkamani A, Kupriyanov S, Hall IM, Baldwin KK (2016) The complete genome sequences, unique mutational spectra, and developmental potency of adult neurons revealed by cloning. *Neuron* 89:1223–1236.
- Hippenmeyer S, Youn YH, Moon HM, Miyamichi K, Zong H, Wynshaw-Boris A, Luo L (2010) Genetic mosaic dissection of *Lis1* and *Ndel1* in neuronal migration. *Neuron* 68:695–709.
- Holt CE, Garlick N, Cornel E (1990) Lipofection of cDNAs in the embryonic vertebrate central nervous system. *Neuron* 4:203–214.
- Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157:1262–1278.
- Junker JP, Spanjaard B, Peterson-Maduro J, Alemany A, Hu B, Florescu M, van Oudenaarden A (2016) Massively parallel whole-organism lineage tracing using CRISPR/Cas9 induced genetic scars. *bioRxiv preprint posted online June 1, 2016*. doi: <http://dx.doi.org/10.1101/056499>.
- Kalhor R, Mali P, Church GM (2016) Rapidly evolving homing CRISPR barcodes. *bioRxiv preprint posted online May 27, 2016*. doi: <http://dx.doi.org/10.1101/055863>.
- Kimmerling RJ, Szeto GL, Li JW, Genshaft AS, Kazer SW, Payer KR, de Riba Borrajo J, Blainey PC, Irvine DJ, Shalek AK, Manalis SR (2015) A microfluidic platform enabling single-cell RNA-seq of multigenerational lineages. *Nat Commun* 7:10220.
- Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* 450:56–62.
- Lodato MA, Woodworth MB, Lee S, Evrony GD, Mehta BK, Karger A, Lee S, Chittenden TW, D’Gama AM, Cai X, Luquette LJ, Lee E, Park PJ, Walsh CA (2015) Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 350:94–98.
- LoTurco J, Manent JB, Sidiqi F (2009) New and improved tools for *in utero* electroporation studies of developing cerebral cortex. *Cereb Cortex* 19:i120–i125.

- Lu R, Neff NF, Quake SR, Weissman IL (2011) Tracking single hematopoietic stem cells *in vivo* using high-throughput sequencing in conjunction with viral genetic barcoding. *Nat Biotechnol* 29:928–933.
- Mayer C, Jaglin XH, Cobbs LV, Bandler RC, Streicher C, Cepko CL, Hippenmeyer S, Fishell G (2015) Clonally related forebrain interneurons disperse broadly across both functional areas and structural boundaries. *Neuron* 87:989–998.
- McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J (2016) Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 353:aaf7907.
- Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV, Gage FH (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435:903–910.
- Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, Muthuswamy L, Krasnitz A, McCombie WR, Hicks J, Wigler M (2011) Tumour evolution inferred by single-cell sequencing. *Nature* 472:90–94.
- Noctor SC, Flint AC, Weissman TA, Dammerman RS, Kriegstein AR (2001) Neurons derived from radial glial cells establish radial units in neocortex. *Nature* 409:714–720.
- Noctor SC, Martínez-Cerdeño V, Ivic L, Kriegstein AR (2004) Cortical neurons arise in symmetric and asymmetric division zones and migrate through specific phases. *Nat Neurosci* 7:136–144.
- Orban PC, Chui D, Marth JD (1992) Tissue- and site-specific DNA recombination in transgenic mice. *Proc Natl Acad Sci USA* 89:6861–6865.
- Ostertag EM, Kazazian HH (2001) Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 35:501–538.
- Reid CB, Liang I, Walsh C (1995) Systematic widespread clonal organization in cerebral cortex. *Neuron* 15:299–310.
- Salipante SJ, Kas A, McMonagle E, Horwitz MS (2010) Phylogenetic analysis of developmental and postnatal mouse cell lineages. *Evol Dev* 12:84–94.
- Serbedzija GN, Bronner-Fraser M, Fraser SE (1989) A vital dye analysis of the timing and pathways of avian trunk neural crest cell migration. *Development* 106:809–816.
- Shapiro E, Biezuner T, Linnarsson S (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* 14:618–630.
- Siddiqi F, Chen F, Aron AW, Fiondella CG, Patel K, LoTurco JJ (2014) Fate mapping by piggyBac transposase reveals that neocortical GLAST⁺ progenitors generate more astrocytes than Nestin⁺ progenitors in rat neocortex. *Cereb Cortex* 24:508–520.
- Snippert HJ, van der Flier LG, Sato T, van Es JH, van den Born M, Kroon-Veenboer C, Barker N, Klein AM, van Rheenen J, Simons BD, Clevers H (2010) Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* 143:134–144.
- Turner DL, Cepko CL (1987) A common progenitor for neurons and glia persists in rat retina late in development. *Nature* 328:131–136.
- VandenDriessche T, Ivics Z, Izsvak Z, Chuah MKL (2009) Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. *Blood* 114:1461–1468.
- Walsh C, Cepko CL (1988) Clonally related cortical cells show several migration patterns. *Science* 241:1342–1345.
- Walsh C, Cepko CL (1992) Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. *Science* 255:434–440.
- Walsh C, Cepko CL (1993) Clonal dispersion in proliferative layers of developing cerebral cortex. *Nature* 362:632–635.
- Wang Y, Waters J, Leung ML, Unruh A, Roh W, Shi X, Chen K, Scheet P, Vattathil S, Liang H, Multani A, Zhang H, Zhao R, Michor F, Meric-Bernstam F, Navin NE (2014) Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature* 512:155–160.
- Wilson MH, Coates CJ, George AL (2007) PiggyBac transposon-mediated gene transfer in human cells. *Molecular Therapy* 15:139–145.
- Woltjen K, Michael IP, Mohseni P, Desai R, Mileikovsky M, Hämmäläinen R, Cowling R, Wang W, Liu P, Gertsenstein M, Kaji K, Sung H-K, Nagy A (2009) piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* 458:766–770.

NOTES

Wu SC-Y, Meir Y-JJ, Coates CJ, Handler AM, Pelczar P, Moisyadi S, Kaminski JM (2006) piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proc Natl Acad Sci USA* 103:15008–15013.

Yoshida A, Yamaguchi Y, Nonomura K, Kawakami K, Takahashi Y, Miura M (2010) Simultaneous expression of different transgenes in neurons and glia by combining in utero electroporation with the Tol2 transposon-mediated gene transfer system. *Genes Cells*:1–12.

Zong H, Espinosa JS, Su HH, Muzumdar MD, Luo L (2005) Mosaic analysis with double markers in mice. *Cell* 121:479–492.

Correlating Cellular Morphology, Physiology, and Gene Expression Using Patch-seq

Cathryn R. Cadwell, PhD, and Andreas S. Tolias, PhD

Department of Neuroscience
Baylor College of Medicine
Houston, Texas

Introduction

More than a century ago, Ramon y Cajal and others speculated that even the most complex functions of the human brain—perception, memory, and decision-making—might eventually be understood at the level of neuronal cell types and their connections (Cajal et al., 2002). Since that time, it has become increasingly clear that different brain regions contain distinct molecularly specified neuronal cell types with characteristic morphological and electrophysiological properties. Furthermore, these different kinds of neurons are arranged in stereotypical circuits that are essential to the functions that each brain area performs. True understanding of the workings of the normal and pathological brain will require identification of all the constituent cell types, mapping their interconnections, and determining their function *in vivo*.

Approaches to Cell-Type Classification

For decades, the gold standard for classification of neuronal cell types has been their complex and diverse morphology (Cajal et al., 2002; Burkhalter, 2008; Petilla Interneuron Nomenclature Group

et al., 2008). In particular, axonal geometry and projection patterns have been the most informative morphological features for predicting how a neuron is integrated into the local circuit (Burkhalter, 2008). To better understand the extensive diversity of cell types in the neocortex and how they are connected into functional circuits, we recently performed a census of morphologically defined neuronal types (primarily GABAergic interneurons) in adult mouse visual cortex layers 1, 2/3, and 5 (L1, L2/3, and L5) using octuple simultaneous, whole-cell patch-clamp recordings, and an improved avidin–biotin–peroxidase staining technique that allowed detailed recovery of axonal and dendritic arbor morphology (Fig. 1) (Jiang et al., 2015). We identified 15 major types of interneurons, each of which has stereotypical electrophysiological properties and morphological features and can be differentiated from all others by cell-type-specific axonal geometry and axonal projection patterns. Notably, each type of neuron has its own unique input–output connectivity profile, connecting with other constituent neuronal types with varying degrees of specificity in postsynaptic targets, laminar location, and synaptic characteristics. Despite specific connection patterns for each cell type, we found that a small number of simple

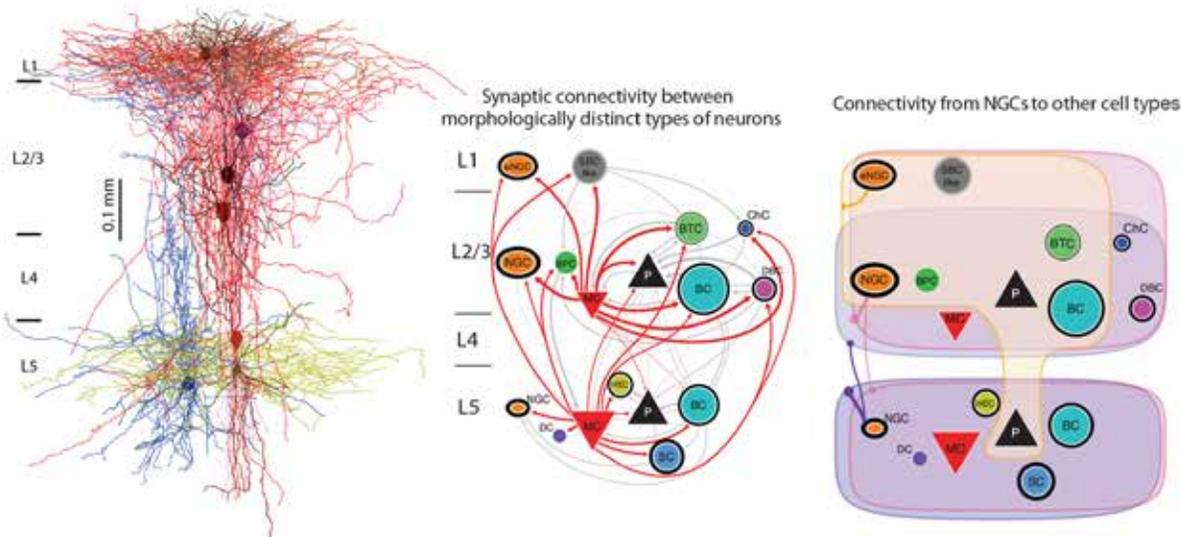


Figure 1. Connectivity among morphologically defined cell types in adult neocortex. Left panel, simultaneous octuple whole-cell recording to study connectivity followed by morphological reconstruction. Scale bar, 0.1 mm. Middle panel, synaptic connectivity among morphologically distinct types of neurons, including pyramidal neurons. Right panel, connectivity from NGCs to other cell types. This connectivity is believed to be nonsynaptic and mediated by volume transmission. B, basket cell; BP, bipolar cell; BT, bitufted cell; Ch, chandelier cell; D, deep projecting cell; DB, double bouquet cell; E, horizontally elongated cell; M, Martinotti cell; NG, neurogliaform cell; P, pyramidal neuron; S, shrub cell. Adapted from Jiang X et al., 2015, Principles of connectivity among morphologically defined cell types in adult neocortex, *Science* 350:aac9462, their Figs. 3A, 6A, and 6B, with permission from AAAS.

NOTES

connectivity motifs are repeated across layers and cell types, defining a canonical cortical microcircuit.

Recent advances in molecular biology, particularly high-throughput single-cell RNA-sequencing (RNA-seq) (Tang et al., 2009; Sandberg, 2014), have begun to reveal the rich genetic programs that give rise to cellular diversity (Fishell and Heintz, 2013). These advances have enabled *de novo* identification of cell types in many tissues, including neuronal subtypes in the retina, neocortex, and hippocampus (Macosko et al., 2015; Zeisel et al., 2015; Tasic et al., 2016). Unfortunately, it has been difficult to reconcile these molecular classification schemes with the classical morphologically defined cell types (Burkhalter, 2008; Petilla Interneuron Nomenclature Group et al., 2008; DeFelipe et al., 2013). Currently available transgenic lines for targeting molecular subclasses of neurons paint a picture of the cortex in broad strokes, with insufficient resolution to distinguish many of the known morphologically defined cell types. For instance, in our study of interneuron subtypes, we recorded from three widely used transgenic lines (targeting parvalbumin [PV]-expressing, somatostatin [SST]-expressing, and vasoactive intestinal peptide [VIP]-expressing interneurons). We found that each molecular class included a number of distinct morphological subtypes, some of which were identified in more than one molecular class, and some of which were not represented in any of the lines (Jiang et al., 2015). Novel molecular markers and techniques to correlate gene expression and morphology at the level of single cells are therefore needed to arrive at a comprehensive cell-type classification scheme that incorporates molecular, morphological, and physiological criteria.

Development of the Patch-seq Protocol

We developed a protocol called Patch-seq that combines whole-cell patch-clamp recordings with high-quality RNA-seq of single neurons, and used L1 of the mouse neocortex as a simple proof of principle to demonstrate the feasibility of this approach to cell-type classification (Cadwell et al., 2016). L1 is known to contain only two main morphological classes of neurons, both of which are inhibitory interneurons, with their own distinct firing patterns and connectivity profiles: elongated neurogliaform cells (eNGCs) and single bouquet cells (SBCs) (Jiang et al., 2013). Using standard electrophysiology techniques in cortical slices, we first generated

a dataset of 72 L1 interneurons, for which we recorded their firing pattern in response to sustained depolarizing current and also reconstructed their detailed morphology using avidin–biotin–peroxidase staining (Figs. 2a, b). Using this as training data, we built an automatic cell-type classifier based on electrophysiological properties that could predict morphological cell class with ~98% accuracy (Figs. 2d, e). In a separate set of experiments, we patched an additional set of 67 L1 interneurons in acute cortical slices using the Patch-seq protocol. This protocol makes use of an optimized mechanical recording approach (tip size, volume inside pipette, etc.) as well as a modified intracellular recording solution to extract and preserve as much full-length mRNA from each cell as possible (see Cadwell et al., 2016, for a detailed protocol). For downstream RNA-seq analysis, we recorded their firing patterns (Fig. 2c) and extracted their cell contents until the cell had visibly shrunken (Fig. 2g). Each neuron from this RNA-seq dataset was assigned to a neuronal class of either eNGC or SBC by blinded expert examination of the firing pattern and using the automated classifier just described. Both classifications were performed independently and led to very similar cell-type labels ($r = 0.91$) (Fig. 2f). In addition, we recorded from 32 L1 interneurons *in vivo* in anesthetized animals and extracted their cell contents for RNA-seq. Large fluctuations in the resting membrane potential, likely resulting from ongoing activity in the local circuit and/or fluctuations in cortical state (Reimer et al., 2014), made it difficult to classify neurons recorded *in vivo* based on their electrophysiological properties. Thus, these cells did not receive a cell-type label. Although we aimed to target L1 interneurons, we occasionally patched an excitatory neuron ($n = 1$ *ex vivo*; $n = 7$ *in vivo*) or astrocyte ($n = 1$ *in vivo*) near the L1/L2 border. Rather than discarding these samples, we proceeded with RNA-seq in the same manner as for the L1 interneurons and used them as additional controls to validate cell-type-specific markers (see below). In addition, each experiment included at least one negative control, in which a recording pipette was inserted into the tissue but no cell was patched. The negative controls were processed in the same manner as the rest of the samples to assess the amount of background contamination during sample collection and amplification.

After harvesting the cell contents, single-cell mRNA was converted to cDNA and used to generate sequencing libraries following a protocol similar to Smart-seq2 (Picelli et al., 2013; Cadwell et al., 2016). Libraries with low cDNA yield (<200

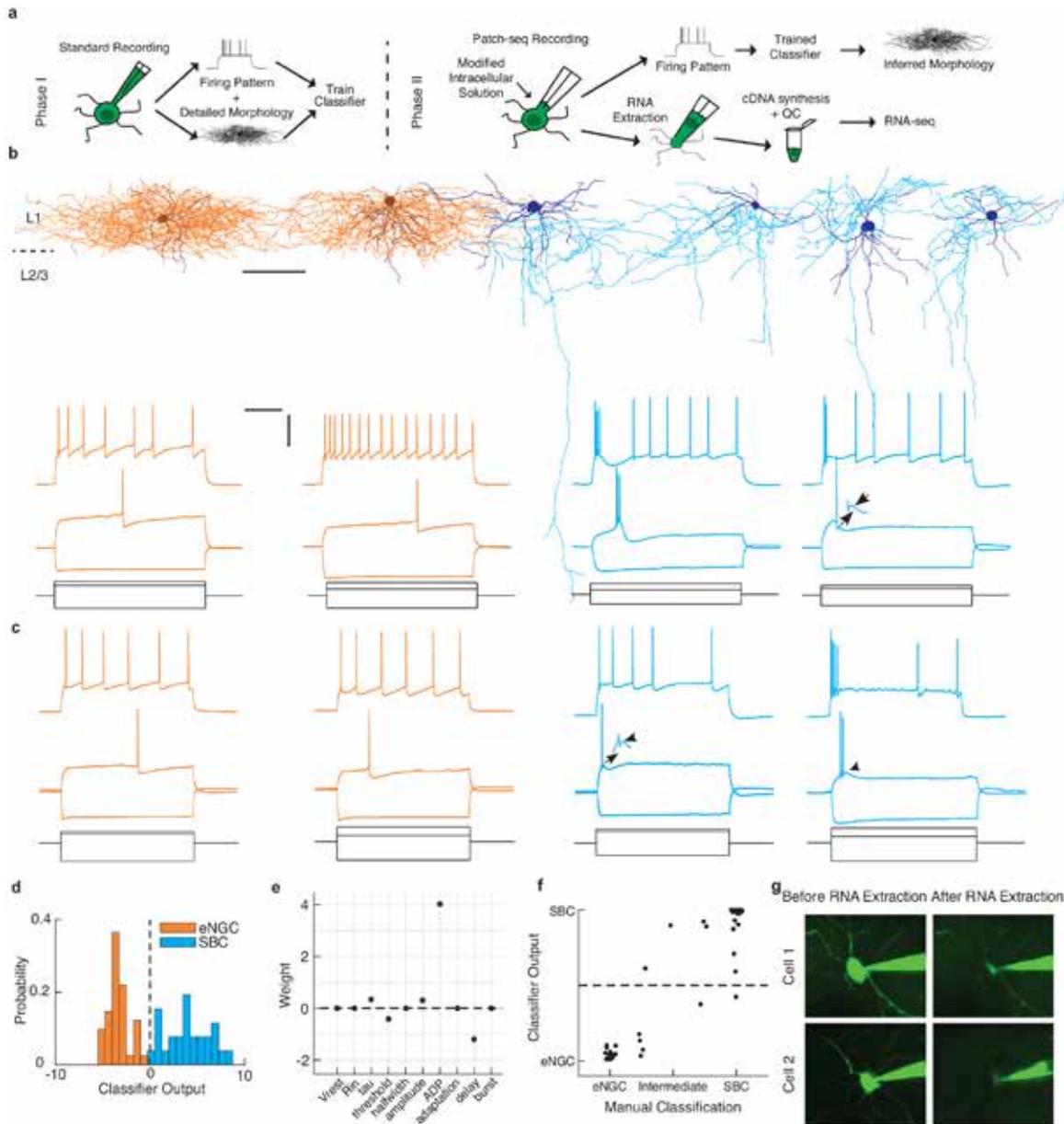


Figure 2. Two morphologically and electrophysiologically distinct neuronal classes in neocortical L1. **a**, Schematic of experimental approach. QC, quality control. **b**, Representative examples of the morphology (top) and firing pattern (bottom) of the two main types of neurons found in L1: eNGCs (orange) and SBCs (cyan). For morphological reconstructions, the darker outline represents the somatodendritic region, and the lighter color is the axonal arbor. Scale bar, 100 μm . For firing patterns, gray lines represent current steps used to elicit the firing patterns shown above. Scale bars, 300 ms (horizontal bar), 40 mV and 500 pA (vertical bar). Arrows denote prominent after-depolarization in SBCs. **c**, Neurons recorded using Patch-seq protocol display similar firing responses as seen using standard electrophysiological techniques, as shown in **b**. **d**, Output of automated cell-type classifier robustly predicts morphological class based on electrophysiological features. **e**, Weights of features used in the automated cell-type classifier. **f**, Results of the automated classifier highly correlate with an independent, blinded expert classification of the electrophysiological properties as "eNGC-like" or "SBC-like"; $r = 0.91$. **g**, Example cells before and after RNA extraction. Reprinted with permission from Cadwell CR et al., 2016, Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq, Nat Biotech 34:199–203, Fig. 1. Copyright 2016, Nature Publishing Group.

NOTES

pg/ μ l) or poor quality suggesting cDNA degradation (<1500 bp mean size) were excluded from further analysis (50/108 cells and 32/32 negative controls). A higher fraction of *in vivo* samples was excluded (31/40) compared with *ex vivo* samples (19/68), likely because of a combination of lower amounts of cDNA obtained as well as increased contamination during *in vivo* sample acquisition (i.e., the pipette must penetrate the dura and traverse more tissue in order to reach the target cell). We sequenced the 58 single-cell libraries that met our inclusion criteria; they corresponded to 48 L1 interneurons patched in slices, 5 L1 interneurons patched *in vivo*, 1 pyramidal neuron patched in slices, 3 pyramidal neurons patched *in vivo*, and 1 astrocyte patched *in vivo*. Analyses of the sequenced libraries revealed that, on average, 65% of reads mapped uniquely to the mouse genome, and 60% of those mapped within exons. As expected, the pyramidal neuron and astrocyte samples showed clear differences in gene expression compared with the L1 interneurons (Fig. 3a) (Cadwell et al., 2016), consistent with known cell-type-specific markers (Bignami et al., 1972; Marshak, 1990; Chan et al., 2001; Fremerey et al., 2001). We subsequently focused our analyses on the L1 interneurons, which expressed interneuron markers including *Gad1*, *Reln*, and *Cplx3* (Alcantara et al., 1998; Stuhmer et al., 2002). We detected ~7000 genes per interneuron (Fig. 3b), with an average Spearman correlation of 0.59 and 0.56 between *ex vivo* and *in vivo* cells, respectively (Fig. 3c). This result was on par with those of high-quality cDNA libraries used for molecular cell-type classification in other tissue types (Jaitin et al., 2014; Treutlein et al., 2014) and had a higher detection of genes per cell than a recent study using dissociated neurons (Zeisel et al., 2015).

Correlation of Morphology, Physiology, and Gene Expression Using Patch-seq

In order to explore the interneuron transcriptomes and to resolve the molecular cell classes in an unbiased manner, we performed unsupervised clustering and dimensionality reduction analysis using the 3000 most variable genes. Affinity propagation was used to cluster cells in this high-dimensional gene space (without prespecifying the number of clusters), and we reduced the dimensionality of the data to visualize the resulting clusters using *t*-distributed stochastic neighbor embedding (*t*-SNE). We identified two molecular interneuron clusters (Fig. 3d) (Cadwell et al., 2016) with high correspondence to the eNGC and SBC classification (41/47 cells, 87%) (Figs. 3d, e). Random subsampling of the data demonstrated that the two cell classes could be robustly distinguished

using as few as 31 samples. In addition, we asked whether we could predict cell class based on single-cell gene expression using a regularized generalized linear model (GLM). The classifier performed at ~86% accuracy for predicting cell type (Fig. 3f). Together, these results demonstrate a strong agreement between cell-type assignments based on morphological, electrophysiological, and transcriptional profiles.

Next we asked whether specific physiological properties could also be predicted using single-neuron gene expression data. We trained a sparse, regularized GLM for each of seven quantitative electrophysiological measurements using the single-cell transcriptome data (selecting the most variable 50–250 genes across cells) as input. Three of these measurements (after-hyperpolarization amplitude [AHP], after-depolarization amplitude [ADP], and action potential [AP] amplitude) could be predicted based on differential gene expression, as shown by the correlation between cross-validated predictions and the ground truth for individual neurons (Figs. 3g–i). The remaining variables (membrane time constant, adaptation index, AP width, and resting membrane potential) could not be modeled using gene expression data, suggesting either that variability along these features may reflect factors other than differential gene expression or that a larger dataset is needed to infer these properties from single-cell gene expression.

Transcriptome analyses of cells collected *in vivo* assigned many of them to a specific cell class (Fig. 3e). They also suggested a shift in gene expression compared with cells collected *ex vivo* (Fig. 3e, second *t*-SNE component [tSNE2]) that may reflect an increased stress response in the acute slice preparation (e.g., increased *Fos* expression *ex vivo* compared with *in vivo*). Notably, these results demonstrate that high-quality, whole-transcriptome data can be obtained even from single neurons in intact animals, and that the gene expression profile within a cell class is mostly preserved across *in vivo* and *ex vivo* preparations. Extension of cell-type classification to include dynamic functional properties, such as receptive fields and tuning properties (which can be measured only *in vivo*) may ultimately lead to better understanding of cell types in terms of their role in information processing in the cortex.

Identification of Novel Cell-Type Markers Using Patch-seq

Cell-type-specific transcriptome data can be used to generate improved driver lines for cell-type targeting. As noted earlier, current genetic cell-type-specific

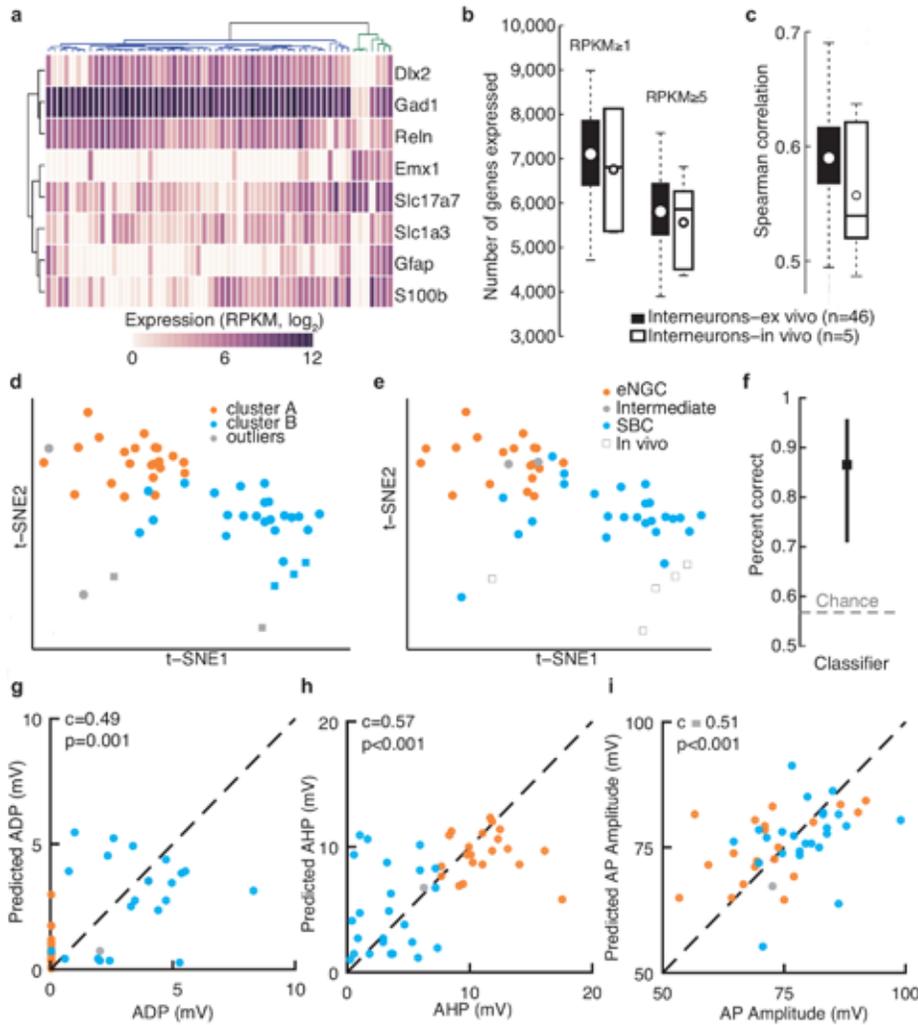


Figure 3. Single-neuron transcriptome profiles predict cell type and electrophysiological properties. **a**, Clustering analysis separates interneurons (blue dendrogram subtree) from other neuronal classes (green dendrogram subtree), includes four pyramidal neurons and one astrocyte based on marker gene expression. Two L1 interneurons clustered with non-interneuron cell types, indicating possible contamination of these samples, and so these two cells were excluded from our analysis of interneuron subtypes. **b**, Number of genes detected per neuron using two different expression thresholds, shown for both *ex vivo* and *in vivo* collection methods. **c**, Pairwise Spearman correlation across all detected genes for *ex vivo* and *in vivo* patched interneurons. **d**, Two-dimensional *t*-SNE representation of gene expression for all L1 interneurons. Cells are colored according to affinity propagation-based clustering in gene space spanned by the 3000 most variable genes before dimensionality reduction. **e**, The same two-dimensional map as in **d**, but with cells color-coded according to expert classification of cell type based on electrophysiological properties. Performance of GLMs using single-neuron gene expression to predict cell type (**f**), ADP (**g**), AHP (**h**), or AP amplitude (**i**). RPKM, reads per kilobase of transcript per million reads. Reprinted with permission from Cadwell CR et al., 2016, Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq, *Nat Biotech* 34:199–203, Fig. 2. Copyright 2016, Nature Publishing Group.

markers often lack sufficient specificity to capture the known diversity of morphological cell classes (Burkhalter, 2008; Petilla Interneuron Nomenclature Group et al., 2008; Jiang et al., 2015). In the case of L1 interneurons, previous studies have suggested that late-spiking eNGCs express Reelin, whereas burst-spiking SBCs express vasoactive intestinal peptide (VIP) (Miyoshi et al., 2010). However, other

studies have shown that Reelin is found in similar proportions of both cell types, and only ~20% of burst-spiking cells express VIP (Ma et al., 2014). We found that neither of these markers was very useful for distinguishing eNGCs from SBCs at the mRNA level (Fig. 4a). This finding calls into question whether single-neuron reverse transcriptase (RT)-PCR and protein-level studies are well suited for predicting

NOTES

which mRNA transcripts are differentially expressed between cell types. Using single-cell differential expression (SCDE) analysis (Kharchenko et al., 2014), we identified several genes that are strongly differentially expressed between the two cell types (Fig. 4b). These genes have the potential to serve as more robust cell-type markers and facilitate future studies on the functional roles of these cell types in the cortical microcircuit.

Patch-seq Provides Insight Into Mechanisms of Synaptic Specificity and Disease Pathophysiology

In the past several decades, we have witnessed a revolution in human genetics that has revealed hundreds of gene mutations that correlate with neuropsychiatric disorders such as autism spectrum

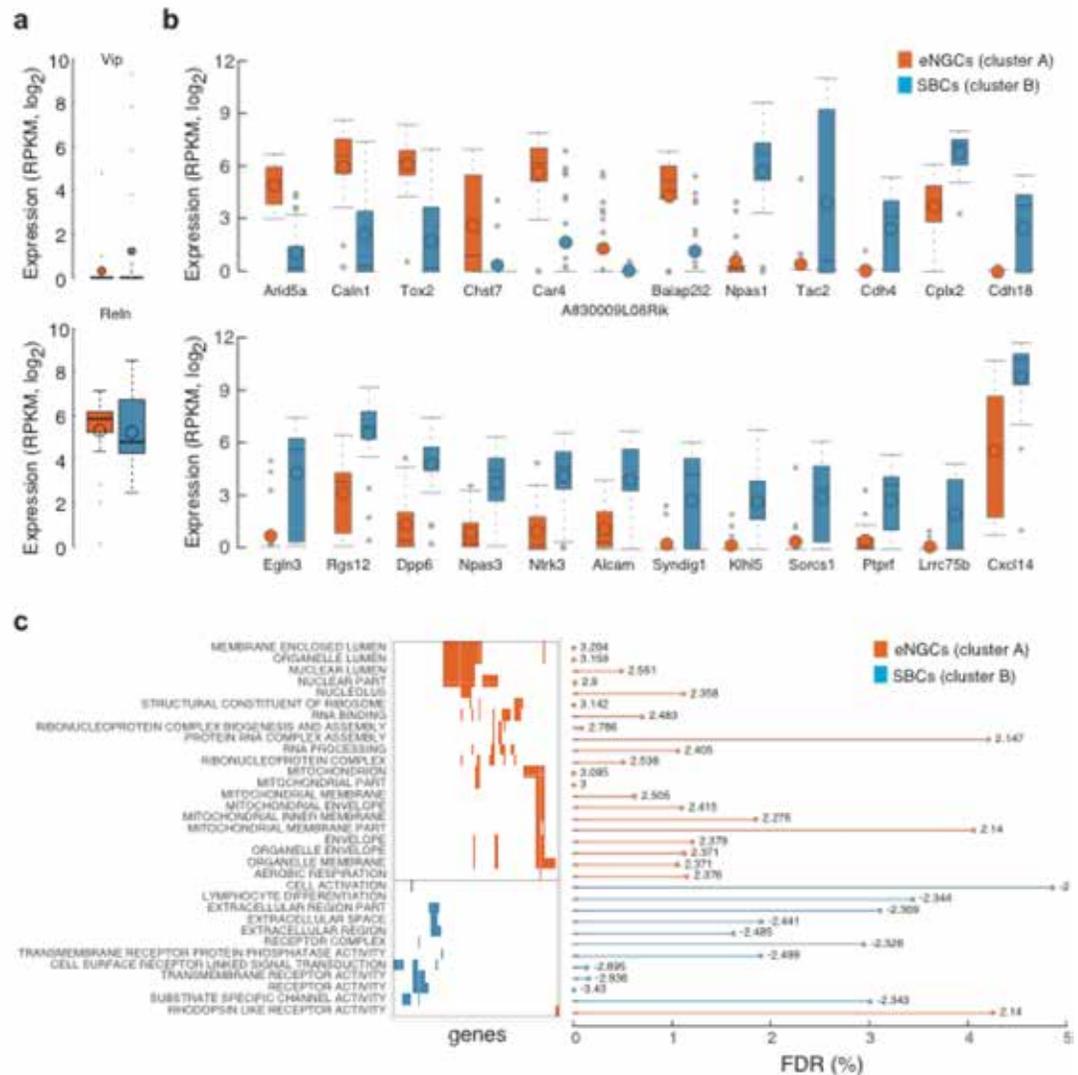


Figure 4. Differential gene expression analysis reveals novel markers for L1 interneuron classes. **a**, Box plots summarize the cell-type expression level of previously proposed marker genes (*Vip* and *Reelin*). **b**, Box plots with expression levels across cell types for novel differentially expressed genes identified between the two affinity propagation clusters. **c**, Significant gene ontology categories from GSEAs on ranked genes from SCDE analysis of SBCs and eNGCs. The gene matrix illustrates gene overlap among categories; the bar plot shows the false discovery rates (FDR), and the numbers indicate normalized enrichment scores per category from GSEA. RPKM, reads per kilobase of transcript per million reads. Reprinted with permission from Cadwell CR et al., 2016, Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq, Nat Biotech 34:199–203, Fig. 3. Copyright 2016, Nature Publishing Group.

disorders, schizophrenia, and depression. Many of these disease-related genes have been linked to synapse formation and function (Spooren et al., 2012; Delorme et al., 2013). However, the expression of disease-associated genes has not been systematically mapped to specific cell types or circuits. Knowing which cell type(s) a disease-associated gene is expressed in is crucial to understanding the disease mechanism and developing novel therapeutic strategies (Siegert et al., 2012). Moreover, having reference transcriptomes for different neuronal cell types will facilitate cell-type engineering through the reprogramming of pluripotent stem cells into specific types of neurons and could lead to more principled treatments for neurological disorders.

In our study of L1 interneurons, gene set enrichment analysis (GSEA) revealed that genes involved in cell–cell signaling (transmembrane and extracellular proteins, receptors, ion channels, and intracellular signaling molecules) were particularly upregulated in SBCs, whereas genes involved in RNA processing and mitochondrial function were upregulated in eNGCs (Fig. 4c). These findings are consistent with previous reports that eNGCs communicate nonspecifically with all cell types using volume transmission, whereas SBCs form highly selective synapses onto particular neuronal types (Olah et al., 2009; Jiang et al., 2013, 2015). In particular, our results predict that increased expression of cell adhesion molecules (including CDH18 [cadherin 18], CDH4, and ALCAM [activated leukocyte cell adhesion molecule]) and synaptic regulatory proteins (such as SYNDIG1 [synapse differentiation inducing 1]) may play an important role in shaping the synaptic specificity of SBCs (Jiang et al., 2013, 2015). Taken together, these results demonstrate that whole-transcriptome profiling of patched neurons is a useful approach to identify novel, unpredicted mechanisms of synaptic specificity.

A number of the differentially expressed genes we identified are also associated with human disease. For example, the genes encoding the transcription factors NPAS1 (neuronal PAS domain protein 1) and NPAS3 are highly expressed in SBCs but not in eNGCs (Fig. 4b). Notably, these proteins have been implicated in autism spectrum disorders (ASD) and schizophrenia and were previously shown to regulate the generation of specific neocortical interneurons (Macintyre et al., 2010; Stanco et al., 2014). SBCs also preferentially express *Dpp6* (dipeptidylpeptidase 6) and *Cplx2* (complexin 2) (Fig. 4b). DPP6 is an auxiliary subunit of the Kv4 family of voltage-gated K⁺ channels implicated in ASD that regulates channel

function and dendrite morphogenesis (Lin et al., 2013), whereas CPLX2 is a presynaptic protein linked to schizophrenia that controls neurotransmitter release and presynaptic differentiation (Brose, 2008). Our observation that four disease genes implicated in neuropsychiatric illness are significantly upregulated in SBCs, combined with previous studies suggesting that SBCs may play an important role in the detection of salient sensory information and the mediation of top-down influences (Jiang et al., 2013), raises the question of whether SBC dysfunction may contribute to the pathophysiology of autism and schizophrenia. The ability to map disease-associated genes onto specific neuronal cell types will lay the foundation for a more principled, circuit-level understanding of neuropsychiatric disorders.

Conclusions and Future Directions

Generating a complete census of neocortical cell types that integrates morphological, electrophysiological, and gene expression data into a cohesive classification scheme presents a tremendous challenge for the field of neuroscience. We have developed a technique to bridge these three distinct modalities, bringing them into a common framework by combining whole-cell patch-clamp recordings and high-quality RNA-sequencing of individual neurons. Using Patch-seq, we demonstrated that cellular morphology, physiology, and gene expression can be integrated at the single-cell level to generate a comprehensive profile of neuronal cell types, using neocortical L1 interneurons as a proof of principle. In addition, we identified several molecular markers that can be used to target these cell types for further study, generate new hypotheses regarding the molecular mechanisms of their synaptic specificity, and link specific cell types to neuropsychiatric illness. Notably, this approach can be used broadly to characterize neuronal cell types in any brain region, in different mouse models of disease, and even in nongenetically tractable organisms such as primates. We hope that the ability to perform unbiased, whole-genome transcriptome analysis and to physiologically characterize individual neurons will help to resolve long-standing questions in the field of neuroscience and initiate entirely new directions of investigation.

Acknowledgments

Portions of this chapter were adapted from Jiang X et al. (2015) Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* 350:aac9462, reprinted with permission from AAAS; and Cadwell et al. (2016) Electrophysiological, transcriptomic and morphologic profiling of single

NOTES

neurons using Patch-seq. *Nat Biotech* 34:199–203. Copyright 2016, reprinted with permission from Nature Publishing Group.

References

- Alcantara S, Ruiz M, D’Arcangelo G, Ezan F, de Lecea L, Curran T, Sotelo C, Soriano E (1998) Regional and cellular patterns of *reelin* mRNA expression in the forebrain of the developing and adult mouse. *J Neurosci* 18:7779–7799.
- Bignami A, Eng LF, Dahl D, Uyeda CT (1972) Localization of the glial fibrillary acidic protein in astrocytes by immunofluorescence. *Brain Res* 43:429–435.
- Brose N (2008) For better or for worse: complexins regulate SNARE function and vesicle fusion. *Traffic* 9:1403–1413.
- Burkhalter A (2008) Many specialists for suppressing cortical excitation. *Front Neurosci* 2:155–167.
- Cadwell CR, Palasantza A, Jiang X, Berens P, Deng Q, Yilmaz M, Reimer J, Shen S, Bethge M, Tolias KF, Sandberg R, Tolias AS (2016) Electrophysiological, transcriptomic and morphologic profiling of single neurons using Patch-seq. *Nat Biotechnol* 34:199–203.
- Cajal SR (2002) *Texture of the nervous system of man and the vertebrates, Vol III* (Pasik P, Pasik T, eds). New York: Springer.
- Chan CH, Godinho LN, Thomaidou D, Tan SS, Gulisano M, Parnavelas JG (2001) *Emx1* is a marker for pyramidal neurons of the cerebral cortex. *Cereb Cortex* 11:1191–1198.
- DeFelipe J, López-Cruz PL, Benavides-Piccione R, Bielza C, Larrañaga P, Anderson S, Burkhalter A, Cauli B, Fairén A, Feldmeyer D, Fishell G, Fitzpatrick D, Freund TF, González-Burgos G, Hestrin S, Hill S, Hof PR, Huang J, Jones EG, Kawaguchi Y, et al. (2013) New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat Rev Neurosci* 14:202–216.
- Delorme R, Ey E, Toro R, Leboyer M, Gillberg C, Bourgeron T (2013) Progress toward treatments for synaptic defects in autism. *Nat Med* 19:685–694.
- Fishell G, Heintz N (2013) The neuron identity problem: form meets function. *Neuron* 80:602–612.
- Freneau RT Jr, Troyer MD, Pahner I, Nygaard GO, Tran CH, Reimer RJ, Bellocchio EE, Fortin D, Storm-Mathisen J, Edwards RH (2001) The expression of vesicular glutamate transporters defines two classes of excitatory synapse. *Neuron* 31:247–260.
- Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343:776–779.
- Jiang X, Wang G, Lee AJ, Stornetta RL, Zhu JJ (2013) The organization of two new cortical interneuronal circuits. *Nat Neurosci* 16:210–218.
- Jiang X, Shen S, Cadwell CR, Berens P, Sinz F, Ecker AS, Patel S, Tolias AS (2015) Principles of connectivity among morphologically defined cell types in adult neocortex. *Science* 350:aac9462.
- Kharchenko PV, Silberstein L, Scadden DT (2014) Bayesian approach to single-cell differential expression analysis. *Nat Methods* 11:740–742.
- Lin L, Sun W, Throesch B, Kung F, Decoster JT, Berner CJ, Cheney RE, Rudy B, Hoffman DA (2013) *DPP6* regulation of dendritic morphogenesis impacts hippocampal synaptic development. *Nat Commun* 4:2270.
- Ma J, Yao XH, Fu Y, Yu YC (2014) Development of layer I neurons in the mouse neocortex. *Cereb Cortex* 24:2604–2618.
- Macintyre G, Alford T, Xiong L, Rouleau GA, Tibbo PG, Cox DW (2010) Association of *NPAS3* exonic variation with schizophrenia. *Schizophr Res* 120:143–149.
- Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214.
- Marshak DR (1990) S100 beta as a neurotrophic factor. *Prog Brain Res* 86:169–181.
- Miyoshi G, Hjerling-Leffler J, Karayannis T, Sousa VH, Butt SJ, Battiste J, Johnson JE, Machold RP, Fishell G (2010) Genetic fate mapping reveals that the caudal ganglionic eminence produces a large and diverse population of superficial cortical interneurons. *J Neurosci* 30:1582–1594.

- Olah S, Fule M, Komlosi G, Varga C, Baldi R, Barzo P, Tamas G (2009) Regulation of cortical microcircuits by unitary GABA-mediated volume transmission. *Nature* 461:1278–1281.
- 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* 10:1096–1098.
- Petilla Interneuron Nomenclature Group (PING), Ascoli GA, Alonso-Nanclares L, Anderson SA, Barrionuevo G, Benavides-Piccione R, Burkhalter A, Buzsáki G, Cauli B, Defelipe J, Fairén A, Feldmeyer D, Fishell G, Fregnac Y, Freund TE, Gardner D, Gardner EP, Goldberg JH, Helmstaedter M, Hestrin S, et al. (2008) Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat Rev Neurosci* 9:557–568.
- Reimer J, Froudarakis E, Cadwell CR, Yatsenko D, Denfield GH, Tolias AS (2014) Pupil fluctuations track fast switching of cortical states during quiet wakefulness. *Neuron* 84:355–362.
- Sandberg R (2014) Entering the era of single-cell transcriptomics in biology and medicine. *Nat Methods* 11:22–24.
- Siegert S, Cabuy E, Scherf BG, Kohler H, Panda S, Le YZ, Fehling HJ, Gaidatzis D, Stadler MB, Roska B (2012) Transcriptional code and disease map for adult retinal cell types. *Nat Neurosci* 15:487–495, S481–S482.
- Spooren W, Lindemann L, Ghosh A, Santarelli L (2012) Synapse dysfunction in autism: a molecular medicine approach to drug discovery in neurodevelopmental disorders. *Trends Pharmacol Sci* 33:669–684.
- Stanco A, Pla R, Vogt D, Chen Y, Mandal S, Walker J, Hunt RF, Lindtner S, Erdman CA, Pieper AA, Hamilton SP, Xu D, Baraban SC, Rubenstein JL (2014) NPAS1 represses the generation of specific subtypes of cortical interneurons. *Neuron* 84:940–953.
- Stuhmer T, Anderson SA, Ekker M, Rubenstein JL (2002) Ectopic expression of the *Dlx* genes induces glutamic acid decarboxylase and *Dlx* expression. *Development* 129:245–252.
- Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA (2009) mRNA-seq whole-transcriptome analysis of a single cell. *Nat Methods* 6:377–382.
- Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, et al. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 19:335–346.
- Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509:371–375.
- Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347:1138–1142.

Adult Mouse Cortical Cell Taxonomy Revealed by Single-Cell Transcriptomics

Bosiljka Tasic, PhD

Allen Institute for Brain Science
Seattle, Washington

Introduction

In the mammalian brain, the neocortex is essential for sensory, motor, and cognitive behaviors. Although different cortical areas have dedicated roles in information processing, they exhibit a similar layered structure, with each layer harboring distinct neuronal populations (Harris and Shepherd, 2015). In the adult cortex, many types of neurons have been identified by characterizing their molecular, morphological, connectional, physiological, and functional properties (Sugino et al., 2006; Rudy et al., 2011; DeFelipe et al., 2013; Greig et al., 2013; Sorensen et al., 2013). Despite much effort, however, objective classification on the basis of quantitative features has been challenging, and our understanding of the extent of cell-type diversity remains incomplete (Toledo-Rodriguez et al., 2004; DeFelipe et al., 2013; Greig et al., 2013).

Cell types can be preferentially associated with molecular markers that underlie their unique structural, physiological, and functional properties, and these markers have been used for cell classification. Transcriptomic profiling of small cell populations from fine dissections (Belgard et al., 2011; Hawrylycz et al., 2012) on the basis of cell surface (Cahoy et al., 2008; Zhang et al., 2010) or transgenic markers (Sugino et al., 2006; Doyle et al., 2008) has been informative; however, any population-level profiling obscures potential heterogeneity in collected cells. Recently, robust and scalable transcriptomic single-cell profiling has emerged as a powerful approach to characterization and classification of single cells, including neurons (Pollen et al., 2014; Usoskin et al., 2014; Macosko et al., 2015; Zeisel et al., 2015). We used single-cell RNA-sequencing (RNA-seq) to characterize and classify >1600 cells from the primary visual cortex in adult male mice. The annotated dataset and a single-cell gene expression visualization tool are freely accessible via the Allen Brain Atlas data portal (<http://casestudies.brain-map.org/celltax>).

Cell-Type Identification by Single-Cell Transcriptomics

To minimize the potential variability in cell types that results from differences in cortical region, age, and sex, we focused on a single cortical area in adult (8-week-old) male mice. We selected the primary visual cortex (VISp or V1), which receives and transforms visual sensory information and is one of the main models for understanding cortical computation and function (Glickfeld et al., 2014). To access both abundant and rare cell types in VISp, we selected a set of 24 transgenic mouse lines in

which Cre, Dre, or Flp recombinases are expressed in specific subsets of cortical cells (Tasic et al., 2016). To isolate individual cells for transcriptional profiling, we sectioned fresh brains from adult transgenic male mice; microdissected the full cortical depth, combinations of sequential layers, or individual layers (L1, 2/3, 4, 5, and 6) of VISp; and generated single-cell suspensions using a previously published procedure (Sugino et al., 2006; Hempel et al., 2007) with some modifications (Fig. 1a) (Tasic et al., 2016). We developed a robust procedure for isolating individual adult live cells from the suspension by fluorescence-activated cell sorting (FACS); reverse-transcribed and amplified full-length poly(A)-RNA using the SMARTer protocol (SMARTer Ultra Low RNA Kit for Illumina Sequencing, Clontech, Mountain View, CA); converted the cDNA into sequencing libraries by tagmentation (Nextera XT, Illumina, San Diego, CA); and sequenced them using next-generation sequencing (NGS) (Fig. 1a). We established quality control (QC) criteria to monitor the experimental process and data quality (Tasic et al., 2016). Our final QC-qualified dataset contains 1679 cells, with >98% of cells sequenced to a depth of $\geq 5,000,000$ total reads (median, $\sim 8,700,000$; range, $\sim 3,800,000$ – $84,300,000$).

To identify cell types, we developed a classification approach that takes into account all expressed genes and is agnostic as to the origin of cells (Fig. 1b). Briefly, we applied two parallel and iterative approaches for dimensionality reduction and clustering, iterative principal component analysis (PCA), and iterative weighted gene coexpression network analysis (WGCNA); we then validated the cluster membership from each approach using a nondeterministic machine learning method (random forest). The results from these two parallel cluster identification approaches were intersected and subjected to another round of cluster membership validation. This step assessed the consistency of individual cell classification: we refer to the 1424 cells that were consistently classified into the same cluster as “core cells” and refer to the 255 cells that were classified into more than one cluster by the random forest approach as “intermediate cells” (Fig. 1b).

This analysis segregated cells into 49 distinct core clusters (Fig. 1c). On the basis of known markers for major cell classes, we identified 23 GABAergic neuronal clusters (*Snap25*⁺, *Slc17a7*⁻, *Gad1*⁺); 19 glutamatergic neuronal clusters (*Snap25*⁺, *Slc17a7*⁺, *Gad1*⁻); and 7 non-neuronal clusters (*Snap25*⁻, *Slc17a7*⁻, *Gad1*⁻) (Fig. 1c). We assigned location and identity to cell types within VISp on the basis of three complementary lines of evidence: layer-enriching

NOTES

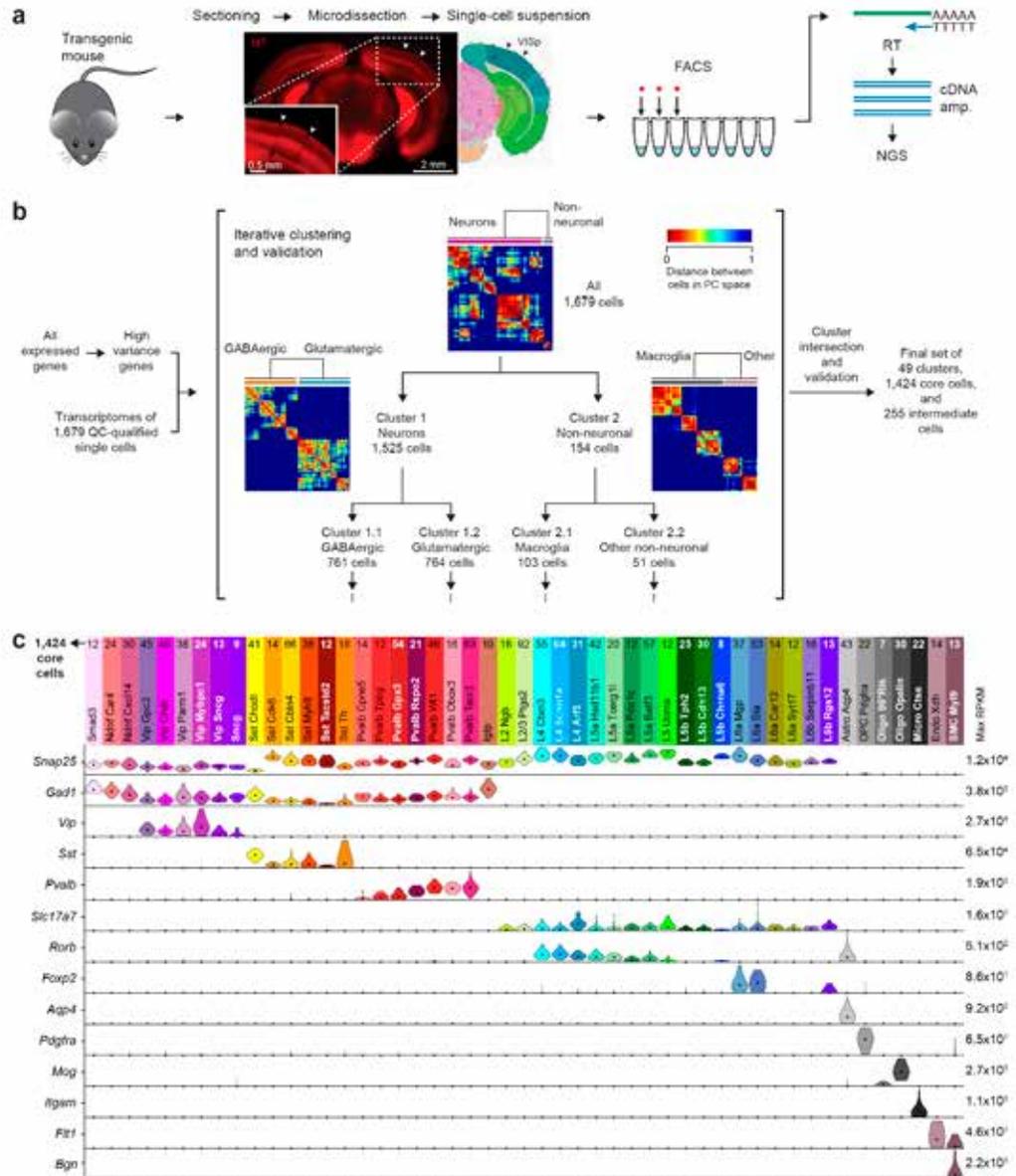


Figure 1. Workflow overview and cell-type identification. **a**, Experimental workflow started with the isolation, sectioning, and microdissection of the primary visual cortex from a transgenic mouse. The tissue samples were converted into a single-cell suspension; single cells were isolated by FACS; poly(A)-RNA from each cell was reverse transcribed (RT); and cDNA was amplified (SMARTer protocol, Clontech), tagged (Nextera XT, Illumina), and sequenced on an NGS platform. **b**, Analysis workflow started with the definition of high-variance genes and iterative clustering based on two different methods—PCA (shown here) and WGCNA—and cluster membership validation using a random forest classifier. Cells that are classified consistently into one cluster are referred to as “core cells” ($N = 1424$), whereas cells that are mapped to more than one cluster are labeled as “intermediate cells” ($N = 255$). After the termination criteria are met, clusters from the two methods are intersected and iteratively validated until all core clusters contain at least four cells. **c**, The final 49 clusters were assigned an identity based on cell location and marker gene expression. Each type is represented by a color bar with the name and number of core cells representing that type. The violin plots represent distribution of mRNA expression on a linear scale, adjusted for each gene (maximum RPKM on the right), for major known marker genes: *Snap25* (pan-neuronal); *Gad1* (pan-GABAergic); *Vip*, *Sst*, and *Pvalb* (GABAergic); *Slc17a7* (pan-glutamatergic); *Rorb* (mostly L4 and L5a); *Foxp2* (L6); *Aqp4* (astrocytes); *Pdgfra* (OPCs); *Mog* (oligodendrocytes); *Itgam* (microglia); *Flt1* (endothelial cells); and *Bgn* (SMCs). RPKM, reads per kilobase per million. Reprinted with permission from Tasic B et al. (2016), Adult mouse cortical cell taxonomy revealed by single cell transcriptomics, Nature Neuroscience 19:335–346, their Fig. 1. Copyright 2016, Nature Publishing Group.

dissections from specific Cre lines, expression of previously reported and/or newly discovered marker genes in our RNA-seq data, and localized expression patterns of marker genes determined using RNA *in situ* hybridization (ISH) (Tasic et al., 2016).

Our single-cell analysis detects most previously known marker genes and identifies many new differentially expressed genes. For a select set of markers, we used single-label and double-label RNA ISH and quantitative reverse transcription PCR (qRT-PCR) to confirm predicted specificity of marker expression or cell location obtained from layer-enriching dissections (Tasic et al., 2016). Our Cre line-based approach also enabled the characterization of these lines' specificity, thereby informing their proper use for labeling and perturbing specific cellular populations (Taniguchi et al., 2011; Olsen et al., 2012; Harris et al., 2014; Huang, 2014). In general, we found that the examined Cre lines mostly label the expected cell types based on promoters and other genetic elements that control Cre recombinase expression in each line; however, all but one Cre line (*Chat-IRES-Cre*) labeled more than one transcriptomic cell type (Tasic et al., 2016).

Cortical Cell Types: Markers and Relationships

To provide an overall view of the transcriptomic cell types that we identified, we integrated our data into constellation diagrams that summarize the identity, select marker genes, and putative location of these types along the pia-to-white-matter axis (Figs. 2a–c). In these diagrams, each transcriptomic cell type is represented by a disk whose surface area corresponds to the number of core cells in our dataset belonging to that type. Intermediate cells are represented by lines connecting the disks; the line thickness is proportional to the number of intermediate cells. We separately present GABAergic, glutamatergic, and non-neuronal constellations because we detected only a single intermediate cell between these major classes. This mode of presentation paints the overall phenotypic landscape of cortical cell types as a combination of continuity and discreteness: the presence of a large number of intermediate cells between a particular pair of core types suggests a phenotypic continuum, whereas a lack of intermediate cells connecting one type to others suggests its more discrete character (Figs. 2a–c). We represent the overall similarity of gene expression between the transcriptomic cell types by hierarchical clustering of groups of their core cells based on all genes expressed above a variance threshold (Fig. 2d). These two views of transcriptomic cell types are complementary: one

shows the extent of intermediate phenotypes, and the other shows the overall similarity in gene expression among cluster cores.

We identified 18 transcriptomic cell types belonging to three previously described major classes of GABAergic cells named after the corresponding markers *Vip* (vasoactive intestinal peptide), *Pvalb* (parvalbumin), and *Sst* (somatostatin) (Gonchar et al., 2008; Xu et al., 2010; Rudy et al., 2011). In a substantial portion of these cells, we detected more than one of these markers; however, our method, which takes into account genome-wide gene expression, usually classified these double-expressing cells into the major type corresponding to the most highly expressed major marker in that cell.

We identified five additional GABAergic types. In accord with a previous report (Pfeffer et al., 2013), we detected *Tnfrif8l3* and *Sema3c* in these types. We named two of them on the basis of a gene for a putative neuropeptide—neuron-derived neurotrophic factor (*Ndnf*)—and we found that they corresponded to neurogliaform cells (Tasic et al., 2016). We refer to the three other types according to markers they express: synuclein gamma (*Sncg*), interferon gamma-induced GTPase (*Igtp*), and SMAD family member 3 (*Smad3*).

Beyond the major types, correspondence of our transcriptomic types to those previously described in the literature was not straightforward and relied on the existence of a Rosetta stone: a shared reagent, feature, or molecular marker with unambiguous translational power. Potential inferences on correspondence to previously proposed types were further complicated by previous studies' use of a variety of animal models, at various ages, focusing on different cortical areas and a few molecular markers (Rudy et al., 2011; DeFelipe et al., 2013).

We found only one *Sst* type (*Sst-Cbln4*) that was prevalent in upper cortical layers, whereas all the other *Sst* types appeared to be enriched in lower layers (Fig. 2a) (Tasic et al., 2016). On the basis of upper-layer enrichment and *Calb2* expression of the *Sst-Cbln4* type, we propose that it likely corresponds to previously characterized *Calb2*-positive Martinotti cells that are enriched in the upper cortical layers (Xu et al., 2006) and are fluorescently labeled in transgenic *GIN* (GABAergic interneuron) mice (Oliva et al., 2000). Our analysis revealed only one additional *Calb2*-positive *Sst* type, which we refer to as *Sst-Chodl*. On the basis of the expression of tachykinin-receptor 1 (*Tacr1*), neuropeptide Y

NOTES

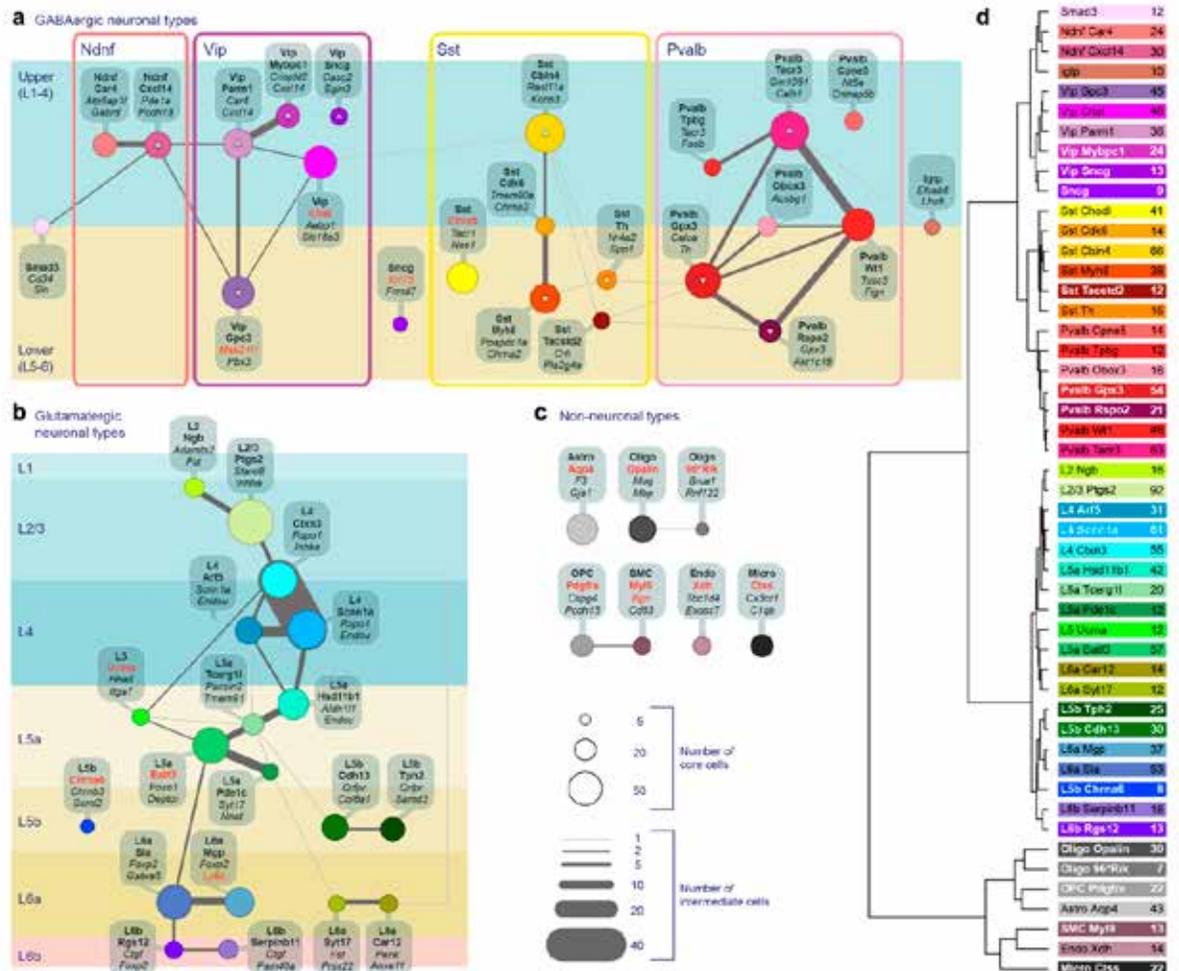


Figure 2. Cell-type summary and relationships. **a–c**, Constellation diagrams showing core and intermediate cells for all cell types. Core cells ($N = 1424$; 664 GABAergic, 609 glutamatergic, and 151 non-neuronal) are represented by colored disks with areas corresponding to the number of core cells for each cluster. Linked tags include cell-type names based on marker genes and layers; unique markers are in red. Intermediate cells ($N = 255$; 97 GABAergic, 155 glutamatergic, and 3 non-neuronal) are represented by lines connecting disks; line thickness corresponds to the number of such cells. **a**, GABAergic types are grouped according to major classes and arranged by their preferential location (enrichment) in upper versus lower cortical layers. Up and down arrows in disks represent statistically significant enrichment determined by layer-enriching dissections. Locations for other clusters are estimates that combine marker gene expression or Cre-line expression based on RNA ISH. The position at the border of upper and lower layers represents lack of evidence for location preference. **b**, Glutamatergic types are arranged according to cortical layer. **c**, Non-neuronal types share few intermediate cells with one another. *96**Rik**, *9630013A20Rik*. **d**, Dendrogram depicting relatedness of the mean gene expression pattern for all cell types based on core cells ($N = 1424$) and genes ($N = 13,878$) with SD for expression >1 across all types. The distance metric is Pearson's correlation coefficient over the genes in the $\log_{10}(\text{RPKM}+1)$ space. The tree was generated using standard hierarchical clustering with average linkage. RPKM, reads per kilobase per million. Reprinted with permission from Tasic B et al. (2016), Adult mouse cortical cell taxonomy revealed by single cell transcriptomics, Nature Neuroscience 19:335–346, their Fig. 4. Copyright 2016, Nature Publishing Group.

(*Npy*), high levels of nitric oxide synthase (*Nos1*), and the absence of *Calb1* (Tasic et al., 2016), we concluded that this type most likely corresponds to *Nos1* type I neurons (Seress et al., 2005). *Nos1* type I neurons are enriched in L5 and 6 (Lee and Jeon, 2005) and are likely long-range projecting (Tomioaka et al., 2005), sleep-active neurons (Gerashchenko et al., 2008).

The *Pvalb* types are highly interconnected in the constellation diagrams (Fig. 2a). Using layer-enriching dissections, we found that some types were preferentially present in upper (*Pvalb-Tpbg*, *Pvalb-Tacr3*, *Pvalb-Cpne5*) or lower (*Pvalb-Gpx3* and *Pvalb-Rspo2*) layers (Tasic et al., 2016). To relate our transcriptomic types to previously described *Pvalb* types, we isolated cells from the upper layers of the *Nkx2.1-CreERT2* line, which, when induced with tamoxifen perinatally, labels a subset of neocortical interneurons, including chandelier cells (Taniguchi et al., 2013). Our analysis classified cells from this line in all three upper layer-enriched *Pvalb* types (Tasic et al., 2016). We suggest that *Pvalb-Cpne5* corresponds to chandelier cells for several reasons: it was most transcriptionally distinct among *Pvalb* types, it was enriched in upper layers, and it did not express *Etv1* (also known as *Er81*), as previously shown for chandelier cells (Dehorter et al., 2015).

The *Vip* major class can be divided into several transcriptomic cell types, all of which appeared to be enriched in upper cortical layers, except the *Vip-Gpc3* type (Fig. 2a). In accord with previous reports (von Engelhardt et al., 2007; Gonchar et al., 2008), our *Vip-Chat* transcriptomic type was located in upper cortical layers and displayed unique expression of choline acetyltransferase (*Chat*) in *Vip*-positive cells. These cells have been reported to either express (von Engelhardt et al., 2007) or not express *Calb2* at the protein level (Gonchar et al., 2008); we found that they robustly expressed *Calb2* mRNA.

For glutamatergic cells, we identified six major classes of transcriptomic types—L2/3, L4, L5a, L5b, L6a, and L6b—on the basis of the layer-specific expression of marker genes and layer-enriching dissections (Fig. 2b); this is consistent with many previous studies (Lein et al., 2007; Molyneaux et al., 2007; Greig et al., 2013; Sorensen et al., 2013). We discovered subdivisions among all of these layer-specific major types. In L2/3, we identified two major types, one of which (L2-*Ngb*) appeared to be located more superficially based on marker gene expression (for example, *Ngb*, *Fst*, *Syt17*, and *Cdh13*). In L4, we identified three types (L4-*Ctxn3*, L4-*Scnn1a*, and L4-*Arf5*) with

high gene expression similarity (Fig. 2d) and a large number of intermediate cells (Fig. 2b). We identified eight different transcriptomic types in L5. Four of these types expressed the L5a marker *Deptor* (L5a-*Hsd11b1*, L5a-*Tcerg11*, L5a-*Batf3*, and L5a-*Pde1c*), whereas three expressed the L5b marker *Bcl6* (L5b-*Cdh13*, L5b-*Tph2*, and L5b-*Chrna6*). One of these L5b types (L5b-*Chrna6*), together with the L5-*Ucma* type, appeared most distinct among L5 types, on the basis of both gene expression and the small number of intermediate cells between them and other L5 types (Fig. 2b). We identified six transcriptomic cell types in L6: four L6a types and two L6b types. Among L6a types, two highly related types (L6a-*Sla* and L6a-*Mgp*) expressed the marker *Foxp2* (Molyneaux et al., 2007; Zeng et al., 2012; Sorensen et al., 2013) and were derived primarily from the *Ntsr1-Cre* line, whereas the other two (L6a-*Syt17* and L6a-*Car12*) did not express *Foxp2* and were isolated as tdT⁻ cells from L6 of the same Cre line. For the latter two types, we discovered several new markers that can be used to identify them (*Car12*, *Prss22*, *Syt17*, and *Penk*). The two L6b types (L6b-*Serp1b11* and L6b-*Rgs12*) expressed the known L6b marker *Ctgf* (Molyneaux et al., 2007; Zeng et al., 2012; Sorensen et al., 2013) and several other previously identified L6b markers (e.g., *Trh*, *Tnmd*, and *Mup5*) (Sorensen et al., 2013).

Despite the neuronal focus of this study, our sampling strategy captured enough cells to identify the major non-neuronal classes. We found seven non-neuronal types: astrocytes, microglia, oligodendrocyte precursor cells (OPCs), two types of oligodendrocytes, endothelial cells, and smooth muscle cells (SMCs). In accord with previous population-level studies (Cahoy et al., 2008; Zhang et al., 2014), these types could be distinguished by many combinatorial and unique markers (Figs. 1c and 2c).

Discussion and Outlook

The adult mouse visual cortex contains ~1,000,000 cells, of which approximately half are neurons (Herculano-Houzel et al., 2013) that can be divided into glutamatergic (80%) and GABAergic cells (20%) (DeFelipe, 2002). Our description of the 49 transcriptomic cortical cell types includes all the major types reported in the literature, some additional new types, as well as subdivisions among the major types. Our approach also provides an experimental and computational workflow to systematically catalog cell types in any region of the mouse brain and relate them to the tools used to examine those cell types (Cre lines and viruses). The discovery of new marker genes enables the generation of new specific Cre lines

NOTES

and provides guidance for intersectional transgenic strategies (Tasic et al., 2016) to enable specific access to cortical cell types that do not express unique marker genes.

Our method relies on dissociation and FACS isolation of single cells, thereby exposing them to stress that might lead to changes in gene expression. However, in our dataset, the majority of marker genes showed excellent correspondence to RNA ISH data from the Allen Brain Atlas (Lein et al., 2007) (~72% of 228 examined genes), suggesting that our procedure did not markedly alter the transcriptional signatures of cell types. Most of the other examined transcripts in this set, which appeared to be very specific markers based on RNA-seq and qRT-PCR (for example, *Chodl*), were not detected by the Allen Brain Atlas in VISp. This discrepancy is probably a consequence of low sensitivity for a subset of ISH probes.

To classify cells based on their transcriptomes, we used two iterative clustering methods and one machine learning-based validation method. The latter assessed the robustness of cluster membership for each cell and suggested the existence of cells with intermediate transcriptomic phenotypes. Previous studies either excluded intermediate cells explicitly (Macosko et al., 2015) or allowed cells to have only a single identity (Usoskin et al., 2014; Pollen et al., 2015; Zeisel et al., 2015). We chose to develop a data analysis approach that accommodates these intermediate cells, as they may be a reflection of actual phenotypic continua. However, as in any approach, both biological and technical aspects contributed to our datasets. For example, similarly to a previous single-cell transcriptomic study (Zeisel et al., 2015), we estimate that we detected only ~23% of mRNA molecules present in a cell (Tasic et al., 2016). Use of a highly efficient transcriptomic method that sampled the cells in their native environment and in proportion to their abundance would provide a more complete and accurate description of the transcriptomic cell-type landscapes. Inclusion of additional cells, even with the current method, is likely to segregate some of the types we defined here into additional subtypes. This subdivision is already apparent in our dataset, as we observed more subtypes if we decreased the threshold for the minimal number of core cells required to define a type (Tasic et al., 2016). In contrast, additional cell sampling may also reveal previously undetected intermediate cells that would define new continua between discrete types. Finally, although we attempted to cover all major

types by choosing a variety of Cre lines, including pan-glutamatergic and pan-GABAergic lines, it is still possible we did not sample some rare types.

We used substantially deeper sequencing per cell than several other studies did (Jaitin et al., 2014; Pollen et al., 2014; Macosko et al., 2015). One of the main advantages of low-depth sequencing is reduction of experimental cost. However, we note that when we downsampled our data from full depth to 1,000,000 or 100,000 mapped reads per cell, we lost the power to detect many types (Tasic et al., 2016). Thus, when subsampling to 100,000 reads, we found only 35 types instead of 49. This decrease in resolution could be compensated for by sampling many more cells, but the appropriate balance between sequencing depth and cell number depends on a variety of factors, including the selected RNA-seq method, informative transcript abundance, tissue and cell-type abundance/accessibility, and desired resolution between cell types.

Our results suggest many new directions for further investigation. At the forefront is the question of the correspondence and potential causal relationships between transcriptomic signatures and specific morphological, physiological, and functional properties. Are certain transcriptomic differences representative of cell state or activity, rather than cell type? In fact, is there a clear distinction between the state and the type? For example, recent evidence suggests that Pvalb basket cells acquire specific firing properties in an activity-dependent manner that may result in a continuum of basket-cell phenotypes (Dehorter et al., 2015), perhaps mirroring the large numbers of intermediate cells that we found for upper-layer *Etw1(Et81)*-positive Pvalb cells (Fig. 2a). Although these questions await further studies, our approach provides an overview of adult cell types in a well-defined cortical area based on a highly multidimensional dataset and is an essential step toward understanding the most complex animal organ—the mammalian brain.

Acknowledgments

This paper was excerpted with permission, with minor modifications, from Tasic B et al. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics, *Nature Neuroscience* 19:335–346, copyright 2016, Nature Publishing Group. The author would like to thank the Allen Institute for Brain Science founders, Paul G. Allen and Jody Allen, for their vision, encouragement, and support.

References

- Belgard TG, Marques AC, Oliver PL, Abaan HO, Sirey TM, Hoerder-Suabedissen A, Garcia-Moreno F, Molnar Z, Margulies EH, Ponting CP (2011) A transcriptomic atlas of mouse neocortical layers. *Neuron* 71:605–616.
- Cahoy JD, Emery B, Kaushal A, Foo LC, Zamanian JL, Christopherson KS, Xing Y, Lubischer JL, Krieg PA, Krupenko SA, Thompson WJ, Barres B (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J Neurosci* 28:264–278.
- DeFelipe J (2002) Cortical interneurons: from Cajal to 2001. *Prog Brain Res* 136:215–238.
- DeFelipe J, López-Cruz PL, Benavides-Piccione R, Bielza C, Larrañaga P, Anderson S, Burkhalter A, Cauli B, Fairén A, Feldmeyer D, Fishell G, Fitzpatrick D, Freund TF, González-Burgos G, Hestrin S, Hill S, Hof PR, Huang J, Jones EG, Kawaguchi Y, et al. (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. *Nat Rev Neurosci* 14:202–216.
- Dehorter N, Ciceri G, Bartolini G, Lim L, del Pino I, Marin O (2015) Tuning of fast-spiking interneuron properties by an activity-dependent transcriptional switch. *Science* 349:1216–1220.
- Doyle JP, Dougherty JD, Heiman M, Schmidt EF, Stevens TR, Ma G, Bupp S, Shrestha P, Shah RD, Doughty ML, Gong S, Greengard P, Heintz N (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell* 135:749–762.
- Gerashchenko D, Wisor JP, Burns D, Reh RK, Shiromani PJ, Sakurai T, de la Iglesia HO, Kilduff TS (2008) Identification of a population of sleep-active cerebral cortex neurons. *Proc Natl Acad Sci USA* 105:10227–10232.
- Glickfeld LL, Reid RC, Andermann ML (2014) A mouse model of higher visual cortical function. *Curr Opin Neurobiol* 24:28–33.
- Gonchar Y, Wang Q, Burkhalter AH (2008) Multiple distinct subtypes of GABAergic neurons in mouse visual cortex identified by triple immunostaining. *Front Neuroanat* 1:3.
- Greig LC, Woodworth MB, Galazo MJ, Padmanabhan H, Macklis JD (2013) Molecular logic of neocortical projection neuron specification, development and diversity. *Nat Rev Neurosci* 14:755–769.
- Harris JA, Hirokawa KE, Sorensen SA, Gu H, Mills M, Ng LL, Bohn P, Mortrud M, Ouellette B, Kidney J, Smith KA, Dang C, Sunkin S, Bernard A, Oh SW, Madisen L, Zeng H (2014). Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Front Neural Circuits* 8:76.
- Harris KD, Shepherd GM (2015) The neocortical circuit: themes and variations. *Nat Neurosci* 18:170–181.
- Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, van de Lagemaat LN, Smith KA, Ebbert A, Riley ZL, Abajian C, Beckmann CF, Bernard A, Bertagnolli D, Boe AF, Cartagena PM, Chakravarty MM, Chapin M, Chong J, Dalley RA, et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 489:391–399.
- Hempel CM, Sugino K, Nelson SB (2007) A manual method for the purification of fluorescently labeled neurons from the mammalian brain. *Nat Protoc* 2:2924–2929.
- Herculano-Houzel S, Watson C, Paxinos G (2013) Distribution of neurons in functional areas of the mouse cerebral cortex reveals quantitatively different cortical zones. *Front Neuroanat* 7:35.
- Huang ZJ (2014) Toward a genetic dissection of cortical circuits in the mouse. *Neuron* 83:1284–1302.
- Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343:776–779.
- Lee JE, Jeon CJ (2005) Immunocytochemical localization of nitric oxide synthase-containing neurons in mouse and rabbit visual cortex and colocalization with calcium-binding proteins. *Mol Cells* 19:408–417.
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ, Chen L, Chen L, Chen TM, Chin MC, Chong J, Crook BE, Czaplinska A, Dang CN, Datta S, Dee NR, et al. (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–176.

NOTES

- Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214.
- Molyneaux BJ, Arlotta P, Menezes JR, Macklis JD (2007) Neuronal subtype specification in the cerebral cortex. *Nat Rev Neurosci* 8:427–437.
- Oliva AA Jr, Jiang M, Lam T, Smith KL, Swann JW (2000) Novel hippocampal interneuronal subtypes identified using transgenic mice that express green fluorescent protein in GABAergic interneurons. *J Neurosci* 20:3354–3368.
- Olsen SR, Bortone DS, Adesnik H, Scanziani M (2012) Gain control by layer six in cortical circuits of vision. *Nature* 483:47–52.
- Pfeffer CK, Xue M, He M, Huang ZJ, Scanziani M (2013) Inhibition of inhibition in visual cortex: the logic of connections between molecularly distinct interneurons. *Nat Neurosci* 16:1068–1076.
- Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, Ramalingam N, Sun G, Thu M, Norris M, Lebofsky R, Toppani D, Kemp DW 2nd, Wong M, Clerkson B, Jones BN, et al. (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol* 32:1053–1058.
- Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, Shuga J, Liu SJ, Oldham MC, Diaz A, Lim DA, Leyrat AA, West JA, Kriegstein AR (2015) Molecular identity of human outer radial glia during cortical development. *Cell* 163:55–67.
- Rudy B, Fishell G, Lee S, Hjerling-Leffler J (2011) Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev Neurobiol* 71:45–61.
- Seress L, Abraham H, Hajnal A, Lin H, Totterdell S (2005) NOS-positive local circuit neurons are exclusively axo-dendritic cells both in the neo- and archi-cortex of the rat brain. *Brain Res* 1056:183–190.
- Sorensen SA, Bernard A, Menon V, Royall JJ, Glatfelder KJ, Desta T, Hirokawa K, Mortrud M, Miller JA, Zeng H, Hohmann JG, Jones AR, Lein ES (2013) Correlated gene expression and target specificity demonstrate excitatory projection neuron diversity. *Cereb Cortex* 25:433–449.
- Sugino K, Hempel CM, Miller MN, Hattox AM, Shapiro P, Wu C, Huang ZJ, Nelson SB (2006) Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nat Neurosci* 9:99–107.
- Taniguchi H, He M, Wu P, Kim S, Paik R, Sugino K, Kvitsiani D, Fu Y, Lu J, Lin Y, Miyoshi G, Shima Y, Fishell G, Nelson SB, Huang ZJ (2011) A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* 71:995–1013.
- Taniguchi H, Lu J, Huang ZJ (2013) The spatial and temporal origin of chandelier cells in mouse neocortex. *Science* 339:70–74.
- Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, et al. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 19:335–346.
- Toledo-Rodriguez M, Blumenfeld B, Wu C, Luo J, Attali B, Goodman P, Markram H (2004) Correlation maps allow neuronal electrical properties to be predicted from single-cell gene expression profiles in rat neocortex. *Cereb Cortex* 14:1310–1327.
- Tomioka R, Okamoto K, Furuta T, Fujiyama F, Iwasato T, Yanagawa Y, Obata K, Kaneko T, Tamamaki N (2005) Demonstration of long-range GABAergic connections distributed throughout the mouse neocortex. *Eur J Neurosci* 21:1587–1600.
- Usooskin D, Furlan A, Islam S, Abdo H, Lonnerberg P, Lou D, Hjerling-Leffler J, Haeggstrom J, Kharchenko O, Kharchenko PV, Linnarsson S, Ernfors P (2014) Unbiased classification of sensory neuron types by large-scale single-cell RNA sequencing. *Nat Neurosci* 18:145–153.
- von Engelhardt J, Eliava M, Meyer AH, Rozov A, Monyer H (2007) Functional characterization of intrinsic cholinergic interneurons in the cortex. *J Neurosci* 27:5633–5642.
- Xu X, Roby KD, Callaway EM (2006) Mouse cortical inhibitory neuron type that coexpresses somatostatin and calretinin. *J Comp Neurol* 499:144–160.
- Xu X, Roby KD, Callaway EM (2010) Immunochemical characterization of inhibitory mouse cortical neurons: three chemically distinct classes of inhibitory cells. *J Comp Neurol* 518:389–404.

- Zeisel A, Muñoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Juréus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347:1138–1142.
- Zeng H, Shen EH, Hohmann JG, Oh SW, Bernard A, Royall JJ, Glattfelder KJ, Sunkin SM, Morris JA, Guillozet-Bongaarts AL, Smith KA, Ebbert AJ, Swanson B, Kuan L, Page DT, Overly CC, Lein ES, Hawrylycz MJ, Hof PR, Hyde TM, et al. (2012) Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* 149:483–496.
- Zhang C, Frias MA, Mele A, Ruggiu M, Eom T, Marney CB, Wang H, Licatalosi DD, Fak JJ, Darnell RB (2010) Integrative modeling defines the Nova splicing-regulatory network and its combinatorial controls. *Science* 329:439–443.
- Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O’Keeffe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N, Deng S, Liddelow SA, Zhang C, Daneman R, Maniatis T, Barres BA, Wu JQ (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci* 34:11929–11947.

NOTES

Distinct Molecular Programs Define Human Radial Glia Subtypes During Human Cortical Development

Tomasz J. Nowakowski, PhD, and Alex A. Pollen, PhD

Department of Neurology
Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research
University of California, San Francisco
San Francisco, California

Introduction

The cell and the genome represent two of the fundamental units in biology. The evolution of multicellular organisms relied on the ability of individual cells to adopt distinct morphologies, physiological properties, and functional states by utilizing the same genome in different ways through diverse transcriptional regulation mechanisms. However, until recently, technical limitations prevented large-scale analysis of gene expression signatures in individual cells. Instead, unbiased sampling of the transcriptome required combining thousands or millions of cells. The average landscape of gene expression from a mixture of cells may obscure the distinct molecular features of diverse cell types or states. Recent advances in microfluidics and molecular biology now enable routine transcriptional profiling from single cells using high-throughput sequencing. These advances support powerful strategies for decoding the mutations, genes, and pathways that distinguish diverse cell lineages and cell types. Here we discuss how we have applied single-cell RNA sequencing (scRNA-seq) to studying the development of the human nervous system.

Astonishing Diversity

More than 100 years ago, Ramon y Cajal appreciated the astonishing diversity of cells in the nervous system, using silver stains to visualize single cells (Cajal, 2002). Cajal likened himself to an entomologist searching for colorful butterflies, “whose beating of wings may one day reveal to us the secrets of the mind” (Garcia-Lopez, 2012). We now know that the human cerebral cortex contains more than 16 billion neurons and even more non-neuronal cells (Herculano-Houzel, 2009) distributed across dozens of cortical areas, and that a single region of mouse cortex may contain more than 40 cell types (Petilla Interneuron Nomenclature Group et al., 2008; Zeisel et al., 2015; Tasic et al., 2016) distributed across six layers. These diverse neuronal cell types emerge from an initially homogenous neuroepithelium during embryonic development.

In human development, the primordial forebrain emerges at around gestational week 7 (GW7) after sequential waves of induction and patterning of the neural tube. Neuroepithelial stem cells undergo symmetric expansive divisions as the neural tube grows before giving rise to the founder neural stem cell population of radial glia along the ventricular zone (VZ) at ~GW11. These ventricular radial glia (vRG) can divide asymmetrically to generate intermediate progenitor cells (IPCs) that will subsequently give

rise to postmitotic neurons (Fig. 1A). In early development, radial glia in the dorsal forebrain generate glutamatergic neurons that migrate to deep cortical layers and generally project to subcortical regions, whereas neurons born at later stages migrate to upper cortical layers and tend to project to neurons within the cortex. In contrast, inhibitory cortical interneurons are generated by radial glia in the ventral forebrain and migrate tangentially toward the cortical plate (Fig. 1B).

The human neocortex also contains a second population of radial glia cells concentrated in the outer subventricular zone (OSVZ) that emerge around GW14, and are rarely observed in mouse. These outer radial glia (oRG) cells differ from vRG cells with respect to position, morphology, and dynamic cell behavior during cell division (Hansen et al., 2010) (Fig. 1C). In the VZ, vRG cells are bipolar and possess apical processes that directly contact the lateral ventricle and transduce signals from the CSF. These signals are critical for their survival, proliferation, and neurogenic capacity (Lehtinen et al., 2011). In synchrony with the cell cycle, cell bodies of vRG cells undergo interkinetic nuclear migration (INM) (Fig. 1C). In contrast, oRG cells have unipolar morphology, lack apical junctions, and undergo a distinct dynamic behavior—mitotic somal translocation (MST)—that directly precedes cytokinesis (Hansen et al., 2010) (Fig. 1C). Thus, vRG and oRG cells reside in distinct niches defined by differences in anatomical location, provision of growth factors, and their ability to be further distinguished based on morphology and cell behavior (Fietz et al., 2010). Although oRG cells are able to generate the majority of cortical neurons (Smart et al., 2002), the molecular features sustaining neural stem cell properties of oRG cells in the OSVZ niche are largely unknown, and the long-term proliferative capacity of these cells has not been examined.

Applying Single-Cell RNA Sequencing to Developmental Heterogeneity

We sought to identify genes and pathways distinguishing vRG and oRG cells during cortical development (Pollen et al., 2015). Previous studies attempted to find markers for oRG cells by comparing gene expression between microdissected samples (Fietz et al., 2012; Miller et al., 2014) or between cell populations expressing particular surface proteins (Florio et al., 2015; Johnson et al., 2015). However, inferring the oRG signature in heterogeneous tissue

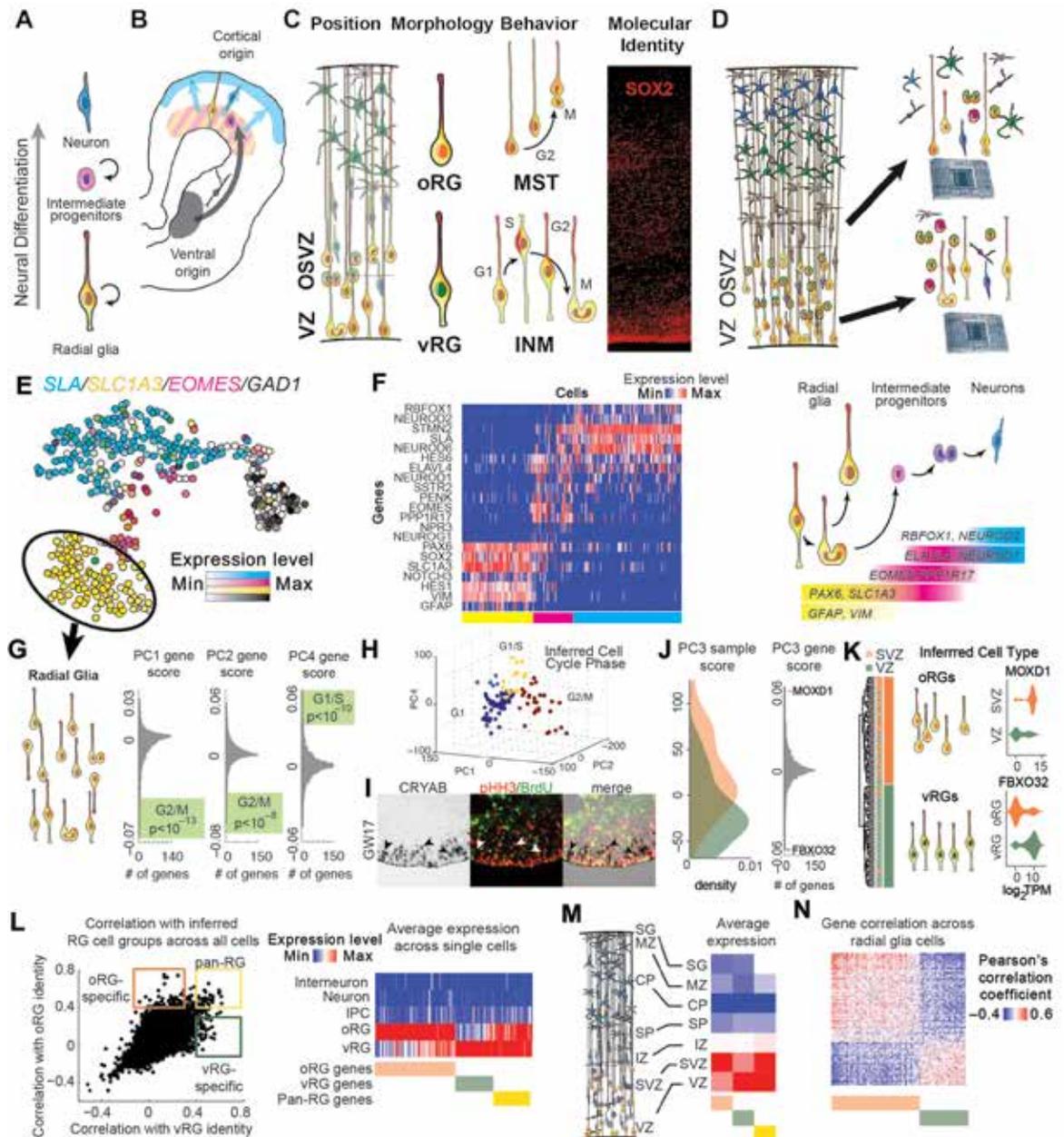


Figure 1. Single-cell mRNA-seq reveals molecular heterogeneity among cells in the developing human brain. **A, B**, Schematics show major modes of neurogenesis of cortical neurons. **C**, Comparison of the closely related radial glia subpopulations. **D**, Strategy for unbiased sampling of single cells from the germinal zone. **E**, Unbiased clustering of 393 single cells highlights major cell populations distinguished by nonoverlapping patterns of marker gene expression. **F**, Overlapping patterns of radial glia and neuronal marker gene expression highlight IPCs as the intermediate step during neurogenesis. **G, H**, Iterative analysis of the molecular variation within classically defined radial glia reveals cell cycle as the dominant source of transcriptional variation. **I**, CRYAB is a candidate G1 marker in vRG cells expressed in cells that do not express M-phase marker pHH3 or incorporate S-phase-specific BrdU tracer. **J, K**, Stem cell niche contributed to transcriptional variation within radial glia and highlights candidate vRG and oRG marker genes. TPM, transcripts per million. **L–N**, Ideal vector correlation analysis identifies radial glia subtype-specific genes. CP, cortical plate; IZ, intermediate zone; MZ, marginal zone; SG, subplate granular layer; SP, subplate. Adapted with permission from Pollen AA et al. (2015) Molecular identity of human outer radial glia during cortical development, *Cell* 163:55–67, their Figs. 1, 2, and 3. Copyright 2015, Elsevier.

proved challenging. We chose to examine gene expression in single cells captured from the VZ and OSVZ without additional enrichment steps for specific cell types. We then isolated radial glia from other cell types *in silico* by analyzing thousands of genes that vary across cell types (Zeisel et al., 2015) and examined the major sources of variation among radial glia. In contrast to this approach of capturing diverse cell types without enrichment, another elegant study used flow cytometry to enrich for cells in the G1 phase of the cell cycle that expressed radial glia markers; the study then analyzed single-cell gene expression specifically among radial glia to identify a similar set of oRG marker genes (Thomsen et al., 2016).

To capture single cells, we used the Fluidigm C1 system (Fluidigm, South San Francisco, CA) and analyzed single cells from paired regions of VZ and OSVZ microdissected from three samples (Fig. 1D). After sequencing, we further analyzed 393 nonoutlier cells based on the number of genes detected, a metric that largely overlapped with several outlier removal approaches. To determine cell-type identity, we performed principal component (PC) analysis and used expectation-maximization clustering to group cells based on their position in PC space. Major clusters of cells were visualized using *t*-distributed stochastic neighbor embedding. Based on the expression of known marker genes, we interpreted groups to represent cells along the cortical excitatory lineage and inhibitory interneurons (Fig. 1E).

Among the excitatory lineage cells (Fig. 1F), we identified groups of cells that robustly expressed markers of human radial glia (*SLC1A3*, *PAX6*, *SOX2*, *HES1*, and *GLI3*; yellow bar, Fig. 1F) and groups of cells that expressed markers of postmitotic neurons (*RBFOX1*, *NEUROD2*, and *STMN2*). In addition, we identified groups of cells that retained a reduced level of some radial glia markers but also expressed early neuronal markers such as *STMN2* and *NEUROD6*. These cells were defined by a gene expression module that included known and novel markers for intermediate progenitor cells, including *EOMES* (*TBR2*), *ELAVL4*, *NEUROG1*, *NEUROD1*, *NEUROD4*, *PPP1R17*, and *PENK* (magenta bar, Fig. 1F) (Hevner et al., 2006; Kawaguchi et al., 2008). This analysis demonstrated that scRNA-seq successfully recovers the major sources of molecular variation expected to distinguish cells at different stages of projection neuron differentiation.

Major Sources of Transcriptional Variation Among Radial Glia Relate to Cell Cycle and Stem Cell Niche

We next analyzed variation in gene expression across 107 cells that robustly expressed canonical markers of radial glia but not markers of other major cell populations (Fig. 1G). Given the high proliferative capacity of neural progenitors, we anticipated that cell cycle would be a major source of transcriptional variation across single cells we profiled. Indeed, genes involved in cell-cycle regulation, mitosis, and DNA replication explained most variation along PC1, PC2, and PC4. Clustering radial glia based on variation along these axes revealed cell groups representing G1, G1/S checkpoint, and G2/M checkpoint (Fig. 1H), and we confirmed that select markers were specific for stages of the radial glia cell cycle (Fig. 1I). Thus, differentiation and cell cycle are major sources of transcriptional heterogeneity among cells in the germinal zone, and single-cell analysis reveals novel molecular features of these states.

Given the distinct morphologies and mitotic behaviors of vRG and oRG cells, we hypothesized that niche occupancy would also contribute to variation among radial glia. Indeed, we found that the spatial source of radial glia was significantly associated with the position of cells along PC3 (Fig. 1J). By clustering radial glia based on the 1% of genes most strongly loading PC3, we identified two transcriptionally distinct groups: one almost purely composed of cells from the VZ, which we interpreted as vRG cells, and another composed of cells from both the VZ and the subventricular zone (SVZ), which we interpreted as oRG cells (Fig. 1K).

Predicted Markers Relate to Position, Morphology, and Behavior of oRG Cells

To relate these distinct transcriptional states to the stem cell niches of the developing neocortex, we first searched for genes likely to distinguish predicted radial glia subtypes. We measured the specificity of genes by their correlation with an ideal marker gene uniformly expressed in only one putative radial glia subpopulation across all 393 cells. We identified 67 candidate marker genes strongly correlated with the oRG population, 33 candidate genes strongly correlated with the vRG population, and 31 genes strongly correlated with both radial glia populations (Fig. 1L, orange, green, and yellow boxes, respectively). In support of these

NOTES

predictions, we observed that candidate vRG markers showed higher expression in the VZ, whereas candidate oRG markers showed higher expression in the SVZ across human cortical tissue samples profiled by the Allen Brain Institute using microarray technology (Fig. 1M) (Miller et al., 2014). In addition, the expression levels of predicted oRG and vRG markers were inversely correlated across radial glia cells (Fig. 1N).

To further investigate candidate marker genes, we performed *in situ* hybridization in primary tissue samples. We found that expression of vRG candidates *CRYAB*, *PDGFD*, *TAGLN2*, *FBXO32*, and *PALLD* was strongest in the VZ, while expression of oRG candidates *HOPX*, *PTPRZ1*, *TNC*, *FAM107A*, and *MOXD1* was strongest in the OSVZ (Fig. 2A). We confirmed specificity by co-immunolabeling stained tissue samples with antibodies against the classical radial glia marker *SOX2* (SRY [sex-determining region-Y]-box2) and intermediate progenitor marker *EOMES* (eomesodermin) (Fig. 2B). In contrast to the radial-glia-specific expression of these transcripts, expression of the predicted novel marker of IPCs, *PPP1R17*, correlated strongly with classical marker *EOMES* (Fig. 2C). Immunostaining for *HOPX* (*HOP* homeobox), *PTPRZ1* (protein phosphatase zeta-1), and *TNC* (tenascin C) proteins revealed their expression in cells with basal fibers that lacked *EOMES* expression, linking this molecular identity to the typical morphology of oRG cells (Fig. 2D). To next relate this molecular identity to distinctive oRG behaviors, we performed time-lapse imaging of organotypic cortical slices (between GW15 and GW19.5) infected with green fluorescent protein (GFP)-expressing adenovirus, and then examined the expression of the most specific oRG marker, *HOPX* (representative example shown in Fig. 2E). We observed that cells undergoing mitotic somal translocation behavior of oRG cells can generate *SOX2/HOPX* double-positive daughter cells with long basal processes characteristic throughout neurogenesis. Together, these results link the molecular identity determined from scRNA-seq to the anatomical location, morphology, and behavior of oRG cells.

Beyond simply marking oRG cells, the genes we identified belong to common pathways that suggest mechanisms by which human oRG cells actively maintain the OSVZ as a neural stem cell niche (Fig. 2F). Many of these genes promote growth factor signaling, including *TNC*, *PTPRZ1*, *ITGB5*, *SDC3*, *HS6ST1*, *IL6ST*, and *LIFR* (Wiese et al., 2012). For example, *TNC* potentiates fibroblast growth

factor (FGF) signaling to support the maturation of neural stem cells (Garcion et al., 2004), whereas integrin signaling along the basal fiber promotes radial glia identity (Fietz et al., 2010). Interestingly, *TNC* contains epidermal growth factor (EGF)-like repeats and multiple binding domains for *PTPRZ1*, syndecans, integrins, and other cell-surface receptors (von Holst, 2008). Thus, *TNC* expression in oRG cells is able to couple key protein networks regulating growth factor signaling, migration, and self-renewal in the massively expanded human OSVZ (Fig. 2G). In addition, *LIFR/STAT3* signaling is known to maintain radial glia neural stem cell identity (Bonaguidi et al., 2005), and we found that p-Y705-*STAT3* signaling is necessary for normal cell-cycle progression in oRG cells but is surprisingly absent in vRG cells. Finally, we directly examined the neural stem cell properties of oRG cells using single-cell clonal lineage analysis. We found that single oRG cells could generate clones of nearly 1000 daughter cells of neuron and glial cell types, highlighting the remarkable proliferative capacity of human oRG cells compared with mouse radial glia, which typically generate only 10–100 daughter cells throughout the neurogenic period (Qian et al., 2000; Gao et al., 2014).

Conclusion

Our study identified neuronal differentiation, cell-cycle progression, and anatomical position as major sources of transcriptional variation across single cells sampled from germinal niches of the developing human cortex. Using *in situ* hybridization and immunostaining, we connected gene expression signatures predicted by scRNA-seq to the position, morphology, and dynamic behavior of cells in tissue. Together, this multimodal characterization establishes an integrative identity for oRG cells. These neural stem cells are characterized by the expression of markers that also appear in astrocytes, but not in vRG cells, including *HOPX*, *TNC*, and *ITGB5*, as well as pan-radial glia markers such as *VIM*, *HES1*, and *ATP1A2*; the presence of a basal, but not apical fiber; mitotic-somal translocation behavior; and extensive proliferative and neurogenic capacity. The oRG subtype is most abundant in the OSVZ stem cell niche, for which it was named, but also resides in the inner SVZ and VZ, and the transcriptional state first emerges in the VZ during early cortical neurogenesis. The oRG marker genes potentially enable the construction of molecular tools for selectively visualizing, manipulating, or purifying oRG cells in tissue and for evaluating the identity of human cortical progenitor cells generated from pluripotent stem cells (Qian et al., 2016). In

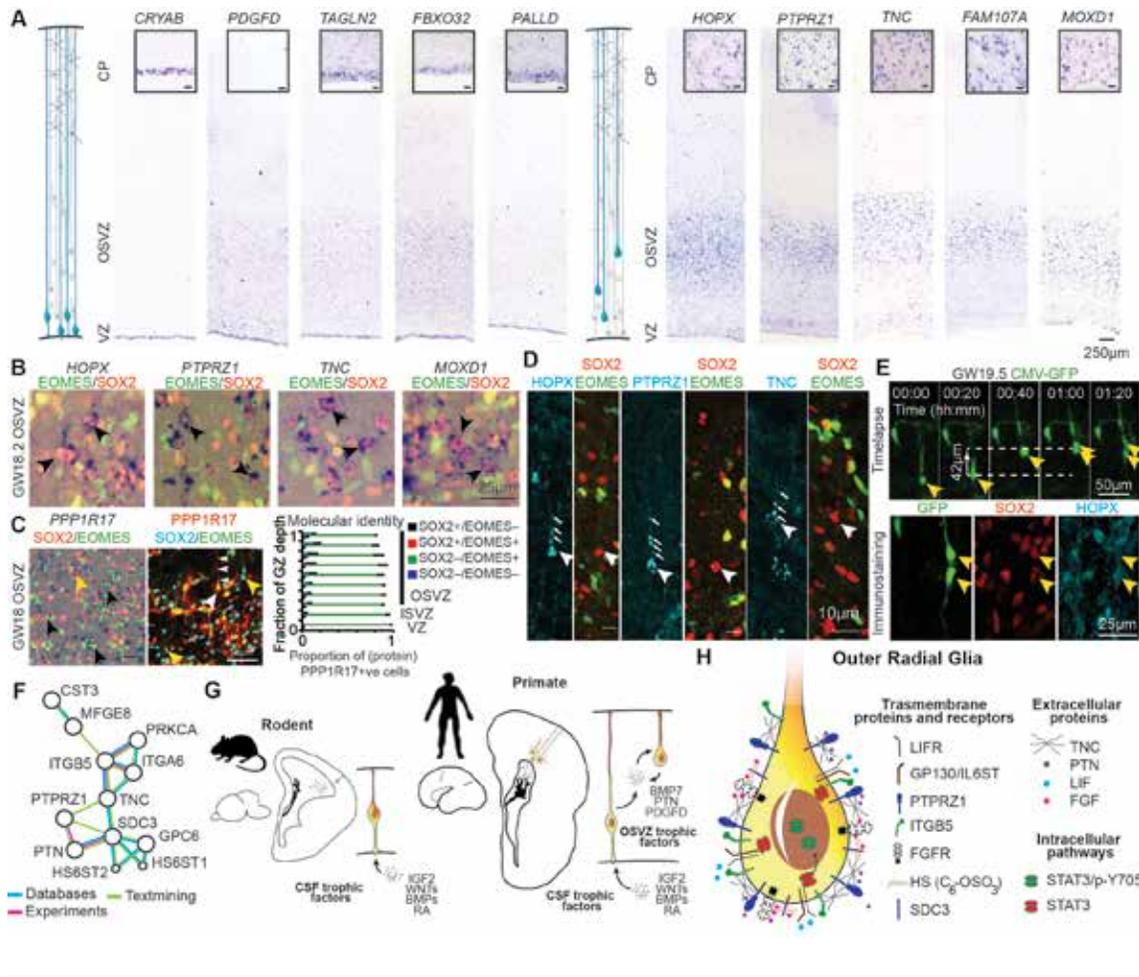


Figure 2. oRG-specific genes relate to functional properties. **A–C,** *In situ* hybridization validation of key vRG, oRG, and IPC transcripts combined with immunohistochemical detection of classical protein markers. Scale bars: **A,** 250 μ m; **B,** 25 μ m; **C,** 50 μ m. **D,** **E,** Candidate oRG cell markers related to morphology and dynamic cell behavior characteristic for oRG cells. Scale bars: **D,** 10 μ m; **E,** 50 and 25 μ m. **F,** Many of the candidate oRG cell markers relate to known functional properties. **G,** Schematic highlighting of oRG-specific local production of growth factors may contribute to neural stem cell niche maintenance in the expanded human OSVZ. **H,** scRNA-seq-based characterization of oRG-enriched transcripts related to signaling pathway activation. GZ, germinal zone. Adapted with permission from Pollen AA et al. (2015) Molecular identity of human outer radial glia during cortical development, *Cell* 163:55–67, their Figs. 3, 4, and 7. Copyright 2015, Elsevier.

addition, these genes may provide insights into the cell types affected in neurodevelopmental disorders and infectious disease (Nowakowski et al., 2016).

Sequencing single-cell mRNA while retaining cell position information provides a general method for identifying distinct subpopulations whose molecular identity possibly relates to microenvironment and functional properties. Here, we explored variation in radial glia gene expression while considering stem cell niche as a covariate. Our results reveal novel molecular features of neural stem cell populations previously distinguished only by cell behavior,

morphology, and position. Together with recent findings (Fietz et al., 2012; Lui et al., 2014), these results highlight three mechanisms that may maintain the “stemness” of the expanded oRG population in the OSVZ stem cell niche: local production of trophic factors such as PTN (pleiotrophin) and BMP7 (bone morphogenetic protein-7) by radial glia, expression of extracellular matrix proteins that potentiate growth factor signaling, and activation of the LIFR/p-STAT3 signaling pathway (Figs. 2G, H). Because the oRG population is thought to be responsible for the majority of human cortical neurogenesis, and OSVZ size correlates with the evolutionary expansion of

NOTES

the brain, future studies could investigate the role of these genes in neurodevelopmental disorders and cortical evolution.

Future Directions

Cell capture is no longer the bottleneck for surveying cell diversity in heterogeneous tissue. New technologies allow massively parallel single-cell capture (Klein et al., 2015; Macosko et al., 2015), and the low dimensionality of gene expression data permits exploration of cells in heterogeneous tissue at extremely low sequencing depths (Jaitin et al., 2014; Pollen et al., 2014; Heimberg et al., 2016). Meanwhile, advances in data analysis and interpretation have resulted in new methods for clustering cell types, thereby predicting patterns of sequential gene expression during signaling pathway activation and lineage progression (Trapnell, 2015). In addition, new techniques have enabled analysis of additional molecular features, including genome sequence and chromatin state (Buenrostro et al., 2015).

One promising area of future work is to use sequencing to also survey cellular phenotypes in a similarly high-throughput manner. In our study, we combined gene expression data with separate low-throughput studies of spatial position, morphology, cell behavior, and developmental lineage potential. Ideally, we could measure these cellular phenotypes and community properties alongside gene expression. *In situ* sequencing (Lee et al., 2014) and other techniques for spatial transcriptomics and proteomics have the potential to help capture the spatial enrichments and neighborhoods of cell types. Similarly, studies using barcoded viruses that cross synapses may be able to label the connections between cell types (Pollock et al., 2014; Kechschull et al., 2016). Finally, the activity, behavior, and lineage relationships between cells may be recorded either through direct imaging of isolated cells prior to capture or through genome modification in response to processes such as cell division or activity (McKenna et al., 2016; Shipman et al., 2016). Together, these emerging technologies may someday enable integrated high-throughput analysis of single-cell gene expression patterns combined with single-cell phenotypes, as well as the connections and interactions between cells in nervous system tissue.

Acknowledgments

This chapter was adapted with permission from Pollen AA et al. (2015) Molecular identity of human outer radial glia during cortical development, *Cell* 163:55–67. Copyright 2015, Elsevier.

References

- Bonaguidi MA, McGuire T, Hu M, Kan L, Samanta J, Kessler JA (2005) LIF and BMP signaling generate separate and discrete types of GFAP-expressing cells. *Development* 132:5503–5514.
- Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ (2015) Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523:486–490.
- Cajal SR (2002) Texture of the nervous system of man and the vertebrates, Vol III (Pasik P, Pasik T, eds). New York: Springer.
- Fietz SA, Kelava I, Vogt J, Wilsch-Brauninger M, Stenzel D, Fish JL, Corbeil D, Riehn A, Distler W, Nitsch R, Huttner WB (2010) OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat Neurosci* 13:690–699.
- Fietz SA, Lachmann R, Brandl H, Kircher M, Samusik N, Schroder R, Lakshmanaperumal N, Henry I, Vogt J, Riehn A, Distler W, Nitsch R, Enard W, Paabo S, Huttner WB (2012) Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proc Natl Acad Sci USA* 109:11836–11841.
- Florio M, Albert M, Taverna E, Namba T, Brandl H, Lewitus E, Haffner C, Sykes A, Wong FK, Peters J, Guhr E, Klemroth S, Pruffer K, Kelso J, Naumann R, Nusslein I, Dahl A, Lachmann R, Paabo S, Huttner WB (2015) Human-specific gene *ARHGAP11B* promotes basal progenitor amplification and neocortex expansion. *Science* 347:1465–1470.
- Gao P, Postiglione MP, Krieger TG, Hernandez L, Wang C, Han Z, Streicher C, Pappusheva E, Insolera R, Chugh K, Kodish O, Huang K, Simons BD, Luo L, Hippenmeyer S, Shi SH (2014) Deterministic progenitor behavior and unitary production of neurons in the neocortex. *Cell* 159:775–788.

- Garcia-Lopez P (2012) Sculpting the brain. *Front Hum Neurosci* 6:5.
- Garcion E, Halilagic A, Faissner A, ffrench-Constant C (2004) Generation of an environmental niche for neural stem cell development by the extracellular matrix molecule tenascin C. *Development* 131:3423–3432.
- Hansen DV, Lui JH, Parker PR, Kriegstein AR (2010) Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* 464:554–561.
- Heimberg G, Bhatnagar R, El-Samad H, Thomson M (2016) Low dimensionality in gene expression data enables the accurate extraction of transcriptional programs from shallow sequencing. *Cell Syst* 2:239–250.
- Herculano-Houzel S (2009) The human brain in numbers: a linearly scaled-up primate brain. *Front Hum Neurosci* 3:31.
- Hevner RF, Hodge RD, Daza RA, Englund C (2006) Transcription factors in glutamatergic neurogenesis: conserved programs in neocortex, cerebellum, and adult hippocampus. *Neurosci Res* 55:223–233.
- Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343:776–779.
- Johnson MB, Wang PP, Atabay KD, Murphy EA, Doan RN, Hecht JL, Walsh CA (2015) Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat Neurosci* 18:637–646.
- Kawaguchi A, Ikawa T, Kasukawa T, Ueda HR, Kurimoto K, Saitou M, Matsuzaki F (2008) Single-cell gene profiling defines differential progenitor subclasses in mammalian neurogenesis. *Development* 135:3113–3124.
- Kebschull JM, Garcia da Silva P, Reid AP, Peikon ID, Albeanu DF, Zador AM (2016) High-throughput mapping of single-neuron projections by sequencing of barcoded RNA. *Neuron* 91:975–987.
- Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161:1187–1201.
- Lee JH, Daugharthy ER, Scheiman J, Kalhor R, Yang JL, Ferrante TC, Terry R, Jeanty SS, Li C, Amamoto R, Peters DT, Turczyk BM, Marblestone AH, Inverso SA, Bernard A, Mali P, Rios X, Aach J, Church GM (2014) Highly multiplexed subcellular RNA sequencing *in situ*. *Science* 343:1360–1363.
- Lehtinen MK, Zappaterra MW, Chen X, Yang YJ, Hill AD, Lun M, Maynard T, Gonzalez D, Kim S, Ye P, D'Ercole AJ, Wong ET, LaMantia AS, Walsh CA (2011) The cerebrospinal fluid provides a proliferative niche for neural progenitor cells. *Neuron* 69:893–905.
- Lui JH, Nowakowski TJ, Pollen AA, Javaherian A, Kriegstein AR, Oldham MC (2014) Radial glia require PDGFD-PDGFR β signalling in human but not mouse neocortex. *Nature* 515:264–268.
- Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161:1202–1214.
- McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J (2016) Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* 353:aaf7907.
- Miller JA, Ding SL, Sunkin SM, Smith KA, Ng L, Szafer A, Ebbert A, Riley ZL, Royall JJ, Aiona K, Arnold JM, Bennet C, Bertagnolli D, Brouner K, Butler S, Caldejon S, Carey A, Cuhaciyan C, Dalley RA, Dee N, et al. (2014) Transcriptional landscape of the prenatal human brain. *Nature* 508:199–206.
- Nowakowski TJ, Pollen AA, Di Lullo E, Sandoval-Espinosa C, Bershteyn M, Kriegstein AR (2016) Expression analysis highlights AXL as a candidate Zika virus entry receptor in neural stem cells. *Cell Stem Cell* 18:591–596.
- Petilla Interneuron Nomenclature Group (PING), Ascoli GA, Alonso-Nanclares L, Anderson SA, Barrionuevo G, Benavides-Piccione R, Burkhalter A, Buzsáki G, Cauli B, Defelipe J, Fairén A, Feldmeyer D, Fishell G, Fregnac Y, Freund TF, Gardner D, Gardner EP, Goldberg JH, Helmstaedter M, Hestrin S, et al. (2008) Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nat Rev Neurosci* 9:557–568.

NOTES

- Pollen AA, Nowakowski TJ, Shuga J, Wang X, Leyrat AA, Lui JH, Li N, Szpankowski L, Fowler B, Chen P, Ramalingam N, Sun G, Thu M, Norris M, Lebofsky R, Toppani D, Kemp DW 2nd, Wong M, Clerkson B, Jones BN, et al. (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol* 32:1053–1058.
- Pollen AA, Nowakowski TJ, Chen J, Retallack H, Sandoval-Espinosa C, Nicholas CR, Shuga J, Liu SJ, Oldham MC, Diaz A, Lim DA, Leyrat AA, West JA, Kriegstein AR (2015) Molecular identity of human outer radial glia during cortical development. *Cell* 163:55–67.
- Pollock JD, Wu DY, Satterlee JS (2014) Molecular neuroanatomy: a generation of progress. *Trends Neurosci* 37:106–123.
- Qian X, Shen Q, Goderie SK, He W, Capela A, Davis AA, Temple S (2000) Timing of CNS cell generation: a programmed sequence of neuron and glial cell production from isolated murine cortical stem cells. *Neuron* 28:69–80.
- Qian X, Nguyen HN, Song MM, Hadiono C, Ogden SC, Hammack C, Yao B, Hamersky GR, Jacob F, Zhong C, Yoon KJ, Jeang W, Lin L, Li Y, Thakor J, Berg DA, Zhang C, Kang E, Chickering M, Nauen D, et al. (2016) Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* 165:1238–1254.
- Shipman SL, Nivala J, Macklis JD, Church GM (2016) Molecular recordings by directed CRISPR spacer acquisition. *Science* 353:aaf1175.
- Smart IH, Dehay C, Giroud P, Berland M, Kennedy H (2002) Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. *Cereb Cortex* 12:37–53.
- Tasic B, Menon V, Nguyen TN, Kim TK, Jarsky T, Yao Z, Levi B, Gray LT, Sorensen SA, Dolbeare T, Bertagnolli D, Goldy J, Shapovalova N, Parry S, Lee C, Smith K, Bernard A, Madisen L, Sunkin SM, Hawrylycz M, et al. (2016) Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci* 19:335–346.
- Thomsen ER, Mich JK, Yao Z, Hodge RD, Doyle AM, Jang S, Shehata SI, Nelson AM, Shapovalova NV, Levi BP, Ramanathan S (2016) Fixed single-cell transcriptomic characterization of human radial glial diversity. *Nat Methods* 13:87–93.
- Trapnell C (2015) Defining cell types and states with single-cell genomics. *Genome Res* 25:1491–1498.
- von Holst A (2008) Tenascin C in stem cell niches: redundant, permissive or instructive? *Cells Tissues Organs* 188:170–177.
- Wiese S, Karus M, Faissner A (2012) Astrocytes as a source for extracellular matrix molecules and cytokines. *Front Pharmacol* 3:120.
- Zeisel A, Munoz-Manchado AB, Codeluppi S, Lonnerberg P, La Manno G, Jureus A, Marques S, Munguba H, He L, Betsholtz C, Rolny C, Castelo-Branco G, Hjerling-Leffler J, Linnarsson S (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347:1138–1142.