Parallel RNA and DNA analysis after deep sequencing (PRDD-seq) reveals cell type-specific lineage patterns in human brain

August Yue Huang[a,b,c,d,e,f,1], Pengpeng Li[a,b,c,d,e,f,1], Rachel E. Rodin[a,b,c,d,e,f,g,h], Sonia N. Kim[a,b,c,d,e,f,i], Yanmei Dou[j,k], Connor J. Kenny[a,b,c,d,e,f], Shyam K. Akula[a,b,c,d,e,f,g,h], Rebecca D. Hodge[l], Trygve E. Bakken[l], Jeremy A. Miller[l], Ed S. Lein[l], Peter J. Park[j,k], Eunjung Alice Lee[a,b,c,d,e,f], and Christopher A. Walsh[a,b,c,d,e,f,2]

[a]Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA 02115; [b]Howard Hughes Medical Institute, Boston Children's Hospital, Boston, MA 02115; [c]Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA 02115; [d]Department of Neurology, Harvard Medical School, Boston, MA 02115; [e]Department of Pediatrics, Harvard Medical School, Boston, MA 02115; [f]Broad Institute of MIT and Harvard, Cambridge, MA 02141; [g]Program in Neuroscience, Harvard Medical School, Boston, MA 02115; [h]Harvard/MIT MD-PhD Program, Harvard Medical School, Boston, MA 02115; [i]Program in Biological and Biomedical Sciences, Harvard University, Boston, MA 02138; [j]Department of Biomedical Informatics, Harvard Medical School, Boston, MA 02115; [k]Ludwig Center at Harvard, Harvard Medical School, Boston, MA 02115; and [l]Allen Institute for Brain Science, Seattle, WA 98109

This contribution is part of the special series of Inaugural Articles by members of the National Academy of Sciences elected in 2018.

Contributed by Christopher A. Walsh, May 11, 2020 (sent for review April 2, 2020; reviewed by Mary E. Hatten and Pasko Rakic)

Elucidating the lineage relationships among different cell types is key to understanding human brain development. Here we developed parallel RNA and DNA analysis after deep sequencing (PRDD-seq), which combines RNA analysis of neuronal cell types with analysis of nested spontaneous DNA somatic mutations as cell lineage markers, identified from joint analysis of single-cell and bulk DNA sequencing by single-cell MosaicHunter (scMH). PRDD-seq enables simultaneous reconstruction of neuronal cell type, cell lineage, and sequential neuronal formation ("birthdate") in postmortem human cerebral cortex. Analysis of two human brains showed remarkable quantitative details that relate mutation mosaic frequency to clonal patterns, confirming an early divergence of precursors for excitatory and inhibitory neurons, and an "inside-out" layer formation of excitatory neurons as seen in other species. In addition our analysis allows an estimate of excitatory neuron-restricted precursors (about 10) that generate the excitatory neurons within a cortical column. Inhibitory neurons showed complex, subtype-specific patterns of neurogenesis, including some patterns of development conserved relative to mouse, but also some aspects of primate cortical interneuron development not seen in mouse. PRDD-seq can be broadly applied to characterize cell identity and lineage from diverse archival samples with single-cell resolution and in potentially any developmental or disease condition.

PRDD-seq | single-cell MosaicHunter | birthdating | cortical layer | neurodevelopment

Although we have learned a great deal about development of the cerebral cortex from animal models, we have remarkably little direct information about how the human brain, which differs vastly in shape, size, and composition from the brains of nonprimates, forms the neurons of its cerebral cortex (1–4). Recent studies defining the fundamental cell types of the adult and developing human cortex (5–7) form a foundation for understanding how these cell types develop, how the unique aspects of the human cortex come about, and how developmental brain disorders might alter patterns of cell lineage or cell type in human brain. However, whether individual neural progenitor cells (NPCs) in embryonic stages are restricted to produce certain subtypes of neurons, or multipotential to generate all neuronal types, is still an open question even in model animal species, since making this distinction requires simultaneous identification of cell lineage and transcriptional analysis of cell type, which remains a technical challenge (8–12).

Somatic genetic mutations accumulate with each cell division during early development, when spontaneous DNA damage escapes the DNA repair machinery, with single-nucleotide variants (SNVs) being the most common mutation type (13–15). The timing of somatic mutations can be inferred by either the cell fraction that carries each mutation or the cooccurrence status of multiple mutations, in which early mutations should be shared by a large fraction of cells whereas later mutations should be present in nested subpopulations of cells (16). Previous study has shown the ability to use somatic SNVs (sSNVs) as a rich internal lineage map to birthdate the developmental timing of each neuron differentiated from neuronal progenitor cells (14) but has not combined that with direct analysis of the subtypes of neurons, defined by morphology, location, physiology, or RNA transcription pattern.

Significance

Stem cells and progenitors undergo a series of cell divisions to generate the neurons of the brain, and understanding this sequence is critical to studying the mechanisms that control cell division and migration in developing brain. Mutations that occur as cells divide are known as the basis of cancer but have more recently been shown to occur with normal cell divisions, creating a permanent, forensic map of the clonal patterns that define the brain. Here we develop technology to analyze both DNA mutations and RNA gene expression patterns in single cells from human postmortem brain, allowing us to define clonal patterns among different types of human brain neurons, gaining direct insight into how they form.

Author contributions: A.Y.H., P.L., and C.A.W. designed research; A.Y.H. and P.L. performed research; A.Y.H., P.L., R.E.R., S.N.K., C.J.K., R.D.H., T.E.B., J.A.M., and E.S.L. contributed new reagents/analytic tools; A.Y.H., P.L., S.N.K., Y.D., S.K.A., P.J.P., E.A.L., and C.A.W. analyzed data; and A.Y.H., P.L., S.K.A., E.A.L., and C.A.W. wrote the paper.

Reviewers: M.E.H., Rockefeller University; and P.R., Yale University.

Competing interest statement: T.E.B. and P.R. are coauthors on a 2019 article.

Data deposition: Sequencing data were deposited in the National Center for Biotechnology Information Sequence Read Archive with accession nos. SRP041470 and SRP061939. Bulk WGS data are available from the National Institute of Mental Health Data Archive (https://doi.org/10.15154/1503337). Config files of scMH and other scripts about PRDD-seq can be accessed on GitHub at https://github.com/AugustHuang/PRDD-seq.

See Profile on page 13861.

[1]A.Y.H. and P.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: christopher.walsh@childrens.harvard.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2006163117/-/DCSupplemental.

First published June 10, 2020.

Single-cell transcriptomes provide granular information about cell identity (5–7), but they cannot provide lineage maps as it fails to capture most somatic mutations, since somatic mutations occur throughout the genome, most often in intronic or intergenic regions (16, 17). Similarly, DNA sequencing alone fails to provide information about cell identity, and so lineage mapping using only somatic mutations from DNA sequencing is unable to address questions about the lineage of specific cell identities in neurodevelopment. Somatic mutations in mitochondrial DNA have been recently suggested as potential lineage marks as well, but the modest target size of the mitochondrial genome, and the multiple diverse mitochondrial genomes in each cell, represent challenges to the use of mitochondrial mutations as a rich source of stable lineage markers (18).

To address this challenge, we developed parallel RNA and DNA analysis after deep sequencing (PRDD-seq) that identifies sSNVs from single-cell and bulk whole-genome sequencing (WGS) data, with multiplexed detection of sSNVs and multiple RNA marker transcripts from single nuclei. We then benchmarked the performance of the DNA and RNA assays of PRDD-seq against bulk WGS and single-cell RNA sequencing (scRNAseq), respectively. Applying PRDD-seq to two postmortem brains of individuals without neurological disease allowed unprecedented quantitative analysis of cell lineage in the human brain. While revealing the expected patterns of divergence of excitatory and inhibitory lineages and "inside-out" generation of excitatory neurons, our PRDD-seq data also directly suggest complex patterns of interneuron formation in the human brain.

Results

Simultaneous Cell Type and Lineage Analysis of Single Cells by PRDD-Seq. The workflow of PRDD-seq is illustrated in Fig. 1. Single NeuN+ cortical neuronal nuclei from prefrontal cortex (PFC) of postmortem human brain tissue were purified by fluorescence-activated nuclear sorting (FANS) (16) (Fig. 1*A*), and subjected to one-step qRT-PCR with target-specific primers for 1) complementary DNA (cDNA) specific for up to 30 marker genes of major neuronal cell types, and 2) specific genomic DNA (gDNA) loci representing identified somatic mutations (see below) as markers of cell lineage (Fig. 1*B*). Aliquots of the preamplified gDNA and cDNA libraries were analyzed for the presence of specific somatic mutations and transcripts by microfluidic genotyping and gene expression profiling, respectively, using the Fluidigm Biomark system (Fig. 1*C*). The somatic mutations used in PRDD-seq were identified by single-cell MosaicHunter (scMH), described below, a bioinformatic tool to identify lineage-informative sSNVs, jointly considering WGS data from multiple displacement amplification-amplified single cells and matched deep (>200×) WGS from bulk DNA samples collected from the same brain region (Fig. 1*D*).

We first created a map of neuronal cell types by analyzing >25,000 single neuronal nuclei—FANS-sorted based on NeuN immunoreactivity—by scRNAseq from two different datasets, to create a cell type landscape onto which PRDD-seq analyzed neurons could be located. We performed 10× Genomics scRNAseq of 10,967 NeuN+ nuclei from the same PFC region of one of the brains from which DNA mutations were identified (Fig. 1*E*). The t-distributed Stochastic Neighbor Embedding (t-SNE) analysis of this dataset defined 21 transcriptionally distinct cell clusters, including 8 excitatory neuron clusters that further clustered into upper, middle, and lower layers, and 13 inhibitory neuron clusters that could be further classified into SST+, PV+, VIP+, and LAMP5+ subtypes (Fig. 1*F* and *SI Appendix*, Fig. S1) (5, 7). A recently published Switching Mechanism at 5′ End of RNA Template Sequencing (SMART-seq) dataset of 15,928 single neuronal nuclei from human middle temporal gyrus (MTG) (5), sorted by NeuN immunoreactivity following microdissection of cerebral cortical layers, provided additional direct information about layer location of neuronal types (Fig. 1*G* and *SI Appendix*, Fig. S2) and so was used for cell type mapping in parallel. PFC and MTG share relatively generic cerebral cortical architecture as "association" cortex, and clustering analysis of the two datasets (Fig. 1*H*) shows that they identified similar major cell types, with cells clustering by cell type rather than by platform, although the SMART-seq dataset from MTG defined finer subdivisions of cell type, as expected, because of its larger sample size and deeper sequence depth.

We jointly analyzed single PRDD-seq cells and scRNAseq cells and mapped each PRDD-seq cell onto the t-SNE maps of scRNAseq based on gene expression similarity (Fig. 1*I*; see *Materials and Methods*). The cell type and cortical layer information of each PRDD-seq cell was then imputed based on its assigned cluster in scRNAseq datasets. Finally, the combination of genotype and gene expression information of PRDD-seq cells allowed lineage and birthdate analysis of particular cell types (Fig. 1*J*), as well as analysis of cell type differentiation of particular lineages (Fig. 1*K*).

Discovery of Lineage-Informative sSNVs from Bulk Brain and Single-Neuron WGS Data. The resolution of lineage reconstruction is dependent on having a comprehensive list of somatic mutations identified from the specific brain under analysis. Whereas deep WGS (e.g., 200× to 250× coverage) of "bulk" DNA, isolated from tissue, efficiently identifies sSNVs present in 4% or more cells (19), it is insensitive to detecting later-occurring sSNVs that mark late cell lineage events. On the other hand, WGS of DNA amplified from single neuronal nuclei (16) identifies later-occurring sSNVs but is limited by cost and subject to artifacts during single-cell amplification. Therefore, we developed scMH, which incorporates a Bayesian graphic model (20, 21) that integrates analysis of bulk WGS and single-cell WGS data to distinguish somatic mutations from germline mutations and technical artifacts (Fig. 2*A*; see *Materials and Methods*). The scMH first calculates the likelihood and mosaic fraction of candidate sSNVs from a bulk DNA sample, and then applies these values as the priors to genotype each candidate SNV across every single cell being analyzed. The shared presence of a given sSNV in bulk DNA and one or more single cells serves as validation of the sSNV. To expand the utility of scMH when a matched bulk sample is unavailable, we further designed a "bulk-free" mode that can utilize a "synthetic" bulk WGS dataset, generated by in silico merging of the many WGS datasets of multiple single cells obtained from the same donor. We benchmarked scMH using 45× single-cell WGS of 24 neurons—22 of which were sequenced in previous studies (16, 17) —as well as ∼200× bulk WGS of PFC (both from the brain of the same individual, UMB1465, who died at age 17 with no neurological diagnosis) against existing single-cell sSNV callers including Monovar (22), SCcaller (23), LiRA (24), and Conbase (25). Sensitivity and false discovery rate (FDR) were estimated based on experimentally validated mutations and clade annotations identified previously (16). With either PFC bulk or synthetic bulk, scMH outperformed the other tools and achieved ∼70% sensitivity to detect lineage-informative mutations with <5% FDR; combining both the default and "bulk-free" modes improved detection sensitivity to 93% without increasing the FDR, suggesting that the "bulk-free" mode of scMH can detect sSNVs that are present in multiple single cells but may be undetectable in the bulk 200× WGS samples because of the low mosaic fraction of these late mutations (Fig. 2*B*).

Applying scMH to data from brains of three normal individuals (UMB1465, UMB4638, and UMB4643) (16, 17) identified and validated 42, 19, and 22 sSNVs, respectively (Fig. 2 *C–E* and *SI Appendix*, Table S1), with an overall validation rate of 74.8%

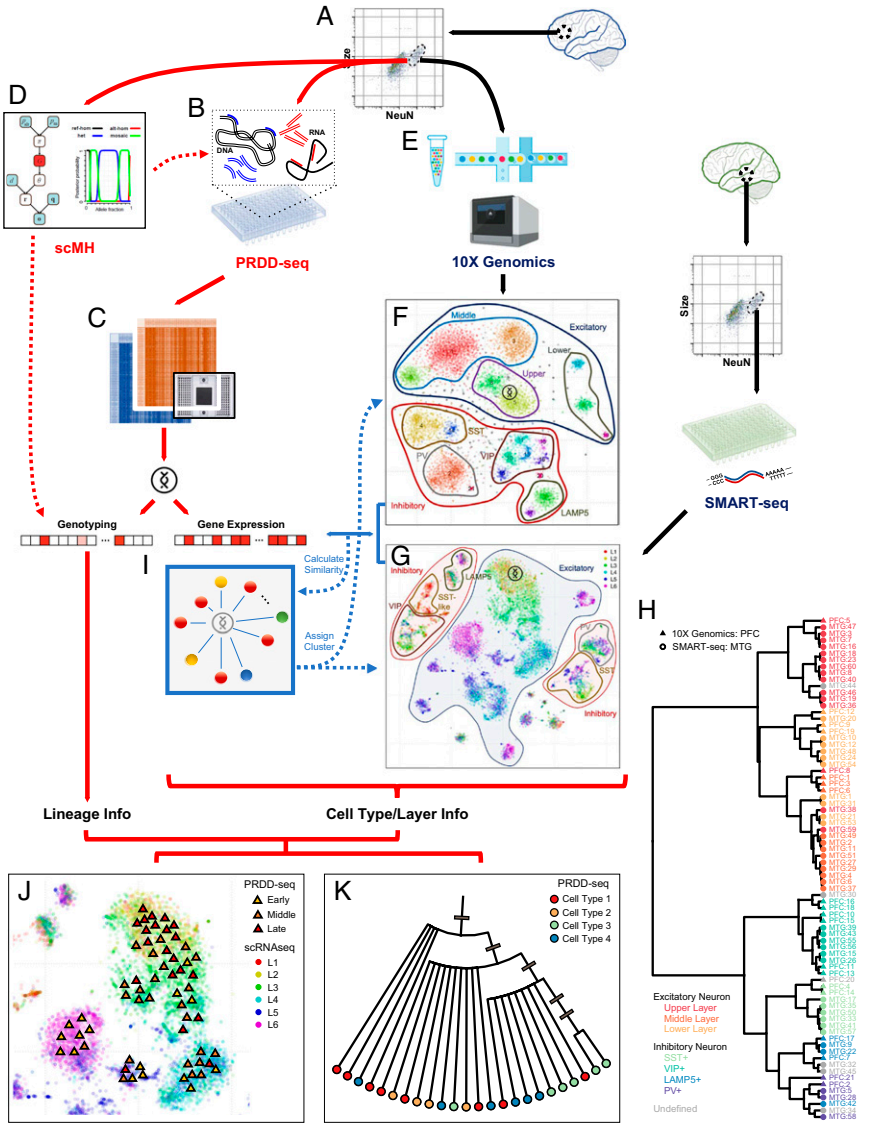

Fig. 1. PRDD-seq enables simultaneous assessment of cell identity and lineage in single cells. (*A*) Neuronal nuclei from postmortem human brain were based on NeuN+ immunoreactivity. (*B*) Target-specific one-step qRT-PCR amplification of cDNA and gDNA fragments of interest. (*C*) The scMH coanalysis of single-cell and bulk deep sequencing data to identify lineage-informative sSNVs. (*D*) Multiplex analysis of the amplified cDNA and gDNA fragments to genotype the sSNVs and profile 30 cell type-specific markers of gene expression. (*E*) The 10× Genomics scRNAseq was performed on NeuN+ nuclei isolated from the same PFC region. (*F*) The 21 cell clusters were identified based on 10× Genomics gene expression data, and then divided into upper, middle, and lower layers of excitatory neurons and four subtypes of inhibitory neurons. (*G*) A second scRNAseq dataset (5) performed on nuclei isolated from the MTG region of another postmortem healthy human brain was also analyzed, where layer information was identified based on layer microdissection. Cell types were identified based on gene expression data. (*H*) Transcriptional clustering revealed similar single-cell expression profiles between 10× Genomics PFC and SMART-seq MTG scRNAseq datasets. Cell clusters were color coded to denote different cell type annotation, and clusters derived from 10× Genomics PFC (triangle) and SMART-seq MTG (circle), in general, clustered by cell type but not by platform. (*I*) Each PRDD-seq cell was mapped to the t-SNE maps by the cosine similarity of gene expression to scRNAseq cells, and then assigned cell type and dissected layer accordingly by majority voting of 25 nearest neighbors. (*J* and *K*) A combination of genotype and gene expression information of PRDD-seq cells allowed (*J*) lineage and birthdate analysis of particular cell types/layers and (*K*) cell type differentiation analysis of particular lineage reconstructed by somatic mutations. Colored triangles in *I* indicate PRDD-seq cells. Gray bars in *K* indicate occurrences of somatic mutations, whereas all cells in one corresponding subclade share the same somatic mutation.

determined by Sanger sequencing of independently sorted neurons from the same brain region. The number and validation rate of lineage-informative sSNVs detected by scMH dramatically increased from previous studies (16, 17). The sSNVs identified from all three brains showed an enrichment in C > T mutations, especially in CpG sites (*SI Appendix*, Fig. S3), a pattern observed in other studies of embryonic mutations and cancer mutations (13, 26), since such C > T mutations appear to be caused by cytosine deamination that is replicated into a fixed SNV before it can be repaired (27). Unsupervised clustering analysis grouped the 24 sequenced neurons from UMB1465 into six different clades; no cells harbored mutations of multiple clades, suggesting the high accuracy of scMH for single-cell genotyping of sSNVs (Fig. 2*C*). In clades C and E, we observed neurons that shared early mutations but harbored different sets of later mutations, suggesting that they were derived from different branches of the same clades (Fig. 2*C*). Clustering of 10 and 9 sequenced neurons from UMB4638 and UMB4643—respectively by their sSNVs—demonstrated similar nested patterns forming three primary clades for each individual, and also showed evidence for branches

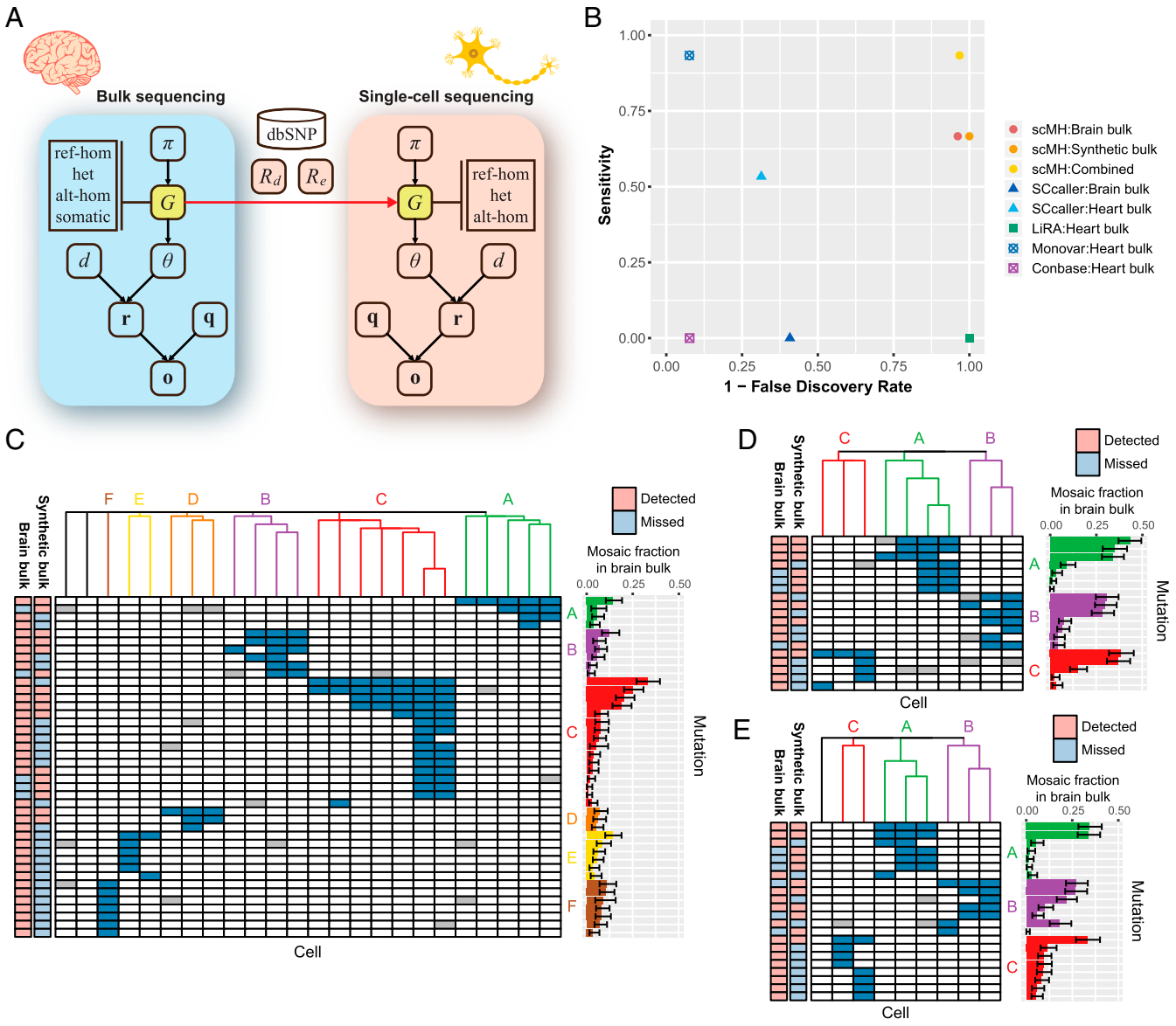

Fig. 2. The scMH identifies lineage-informative sSNVs from the joint analysis of bulk brain and single neurons. (*A*) Overview of the extended Bayesian model of scMH to use bulk sequencing data to facilitate sSNV calling in single cells. *G* denotes the genotype state, π denotes the prior probabilities of genotype, and *d*, **o**, and **q** denote the depth, observed bases, and their base qualities, respectively, in bulk (*Left*) or single-cell (*Right*) sequencing data. (*B*) Specificity and precision of identifying sSNVs using scMH and other published callers. The scMH outcompeted other callers in both precision and sensitivity. (*C–E*) Validated lineage-informative sSNVs identified by scMH in (*C*) UMB1465, (*D*) UMB4638, and (*E*) UMB4643. Heatmaps demonstrate the genotyping status of sSNVs; dark blue and white squares denote the presence or absence of sSNVs in a given cell, whereas gray squares denote unknown genotype due to locus dropout in single-cell WGS. Bar graphs show the mosaic fraction of each sSNV in WGS of bulk brain sample. Clade E in *C* and clade C in *E* represent likely branching clades where early shared mutations are present, while later sSNVs mark two branches with distinct mutations. Error bars reflect 95% CIs.

of these clades (Fig. 2 *D* and *E*). The mosaic fraction of each sSNV in "bulk" DNA (Fig. 2 *C–E*) was used as an additional indicator of the sequence in which sSNV occurred, since early sSNVs tend to be found in many single cells, as well as at higher mosaic fraction in bulk DNA, whereas later mutations appear in fewer cells and lower mosaic fraction in bulk DNA. These two findings correlated very strongly.

Lineage and Cell Type Identity of Single Neurons Revealed by PRDD-Seq. To assess the performance of PRDD-seq in capturing lineage and cell type information from single cells, we applied PRDD-seq to 1,710 cortical neurons from UMB1465 PFC, using probes to detect 30 out of 42 validated sSNVs in UMB1465, for which we successfully designed highly specific and sensitive probes (*SI Appendix*, Table S1), along with 30 marker genes whose expression levels distinguish major inhibitory and excitatory neuronal subtypes and cortical layers identified in the scRNAseq datasets (5, 7) (*SI Appendix*, Table S2). Overall, PRDD-seq mapped 1,112/1,710 (65%) cortical neurons from UMB1465 PFC into 20 lineage branches and six major clades (Fig. 3*A*). For each major clade, birthdate-ordered lineage branches were inferred from the nested sSNVs, where earlier derived neurons contained fewer clonal mutations, and neurons generated later harbored additional mutations from subsequent cell divisions (16). The nested nature of sSNVs in clades allows cells to be placed into clades using multiple sSNVs, so that cells whose genomes were subject to allelic dropout—which is not uncommon when single-cell DNA molecules are amplified—could still be placed into

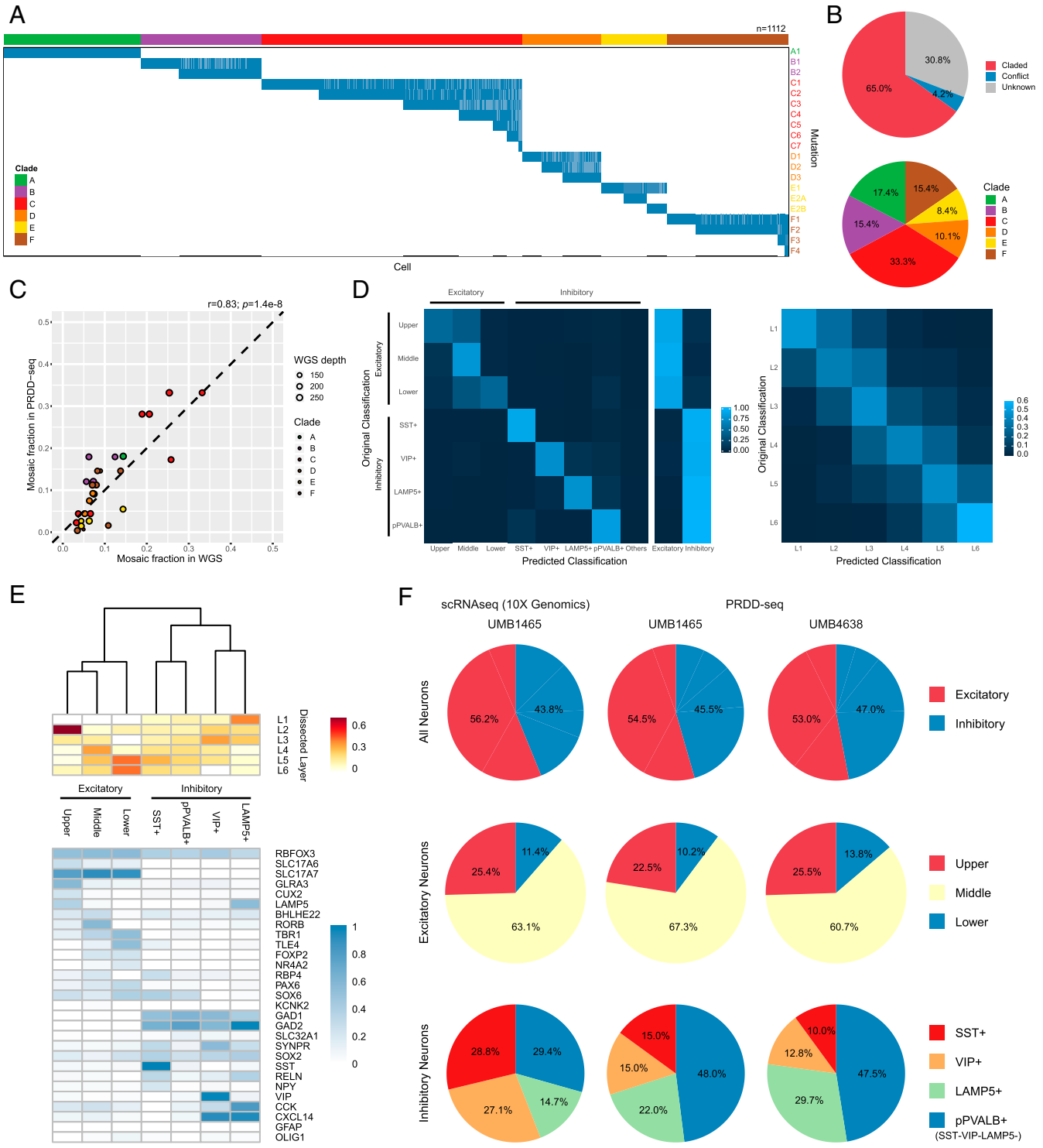

Fig. 3. PRDD-seq profiles single neurons with varied lineage markers and distinct cell type identity. (*A*) Genotyping results of 30 sSNVs (by rows) from 20 lineages across PRDD-seq cells (by columns) from UMB1465. Blue and white squares represent the presence or absence of sSNV, respectively, whereas light blue squares represent the sSNVs that were dropouts in PRDD-seq assay but were inferred by the presence of deeper mutations from the same clade. (*B*, *Lower*) Clade classification of PRDD-seq cells profiled in UMB1465. (*Upper*) PRDD-seq cells which contained sSNVs from multiple or no clades are labeled as "conflict" and "unknown," respectively. (*C*) Correlation of mosaic fractions from WGS and PRDD-seq (calculated as percent of assayed cells carrying a given sSNV) in UMB1465. Both methods showed significantly concordant mosaic fractions (Pearson correlation's $P < 0.001$). (*D*) Accuracy of cell type (*Left*) and cortical layer (*Right*) classification based on the expression profile of 30 marker genes used in PRDD-seq. The scRNAseq cells from each cell type (10× Genomics) and cortical layer (SMART-seq) were randomly sampled and then reassigned to clusters of t-SNE map using 30 marker genes under PRDD-seq mapping strategy. (*E*, *Lower*) Taxonomy of three excitatory layers and four inhibitory subtypes based on average expression of 30 marker genes in PRDD-seq cells. (*Upper*) Relative density of cortical layers for each subgroup is also shown. Here, pPVALB+ denotes PVALB+/SST-VIP- LAMP5- subtype of inhibitory neurons. (*F*) Relative ratio between excitatory and inhibitory neurons (*Upper*) and across different cell types of excitatory neurons (*Middle*) or inhibitory neurons (*Lower*) between PRDD-seq and 10× Genomics scRNAseq.

clades based on other sSNV from the same clade (Fig. 3*A* and *SI Appendix*, Table S1). On the other hand, only 71/1,710 (4.2%) neurons contained sSNVs from multiple clades, suggesting a low rate of false positive amplification or sorting of multiple nuclei into single wells in the DNA assay of PRDD-seq (Fig. 3 *B*, *Upper*). And 527/1,710 (30.8%) neurons showed the absence of any sSNVs from the six clades; these neurons may be from other clades in which we did not discover sSNV markers (Fig. 3 *B*, *Upper*). In PRDD-seq cells, mosaic fractions of sSNVs correlated linearly with the fractions calculated from ∼200× bulk WGS, indicating generally unbiased sSNV detection (Fig. 3 *B*, *Lower* and *C*), and allowing confident inference of the developmental sequence of sSNVs according to the nested pattern.

Among the 1,112 PRDD-seq cells that were successfully claded, we ran the RNA assay of PRDD-seq to measure the expression of 30 marker genes for each cell. Our evaluation using simulation data derived from our own and published scRNAseq datasets (see *Materials and Methods*) suggested that these 30 marker genes were sufficiently informative to infer many aspects of cell type and dissected layer annotation (Fig. 3*D*), with an average accuracy of 84% for cortical layer classification (within ±1-layer difference) and 83% for inhibitory neuron subtype classification. We then utilized expression of these 30 markers to successfully classify 747/1,112 PRDD-seq neurons (67.2%) from UMB1465 into three excitatory subgroups—corresponding to upper, middle, or lower cortical layers—and four inhibitory subgroups: somatostatin positive (SST+), vasoactive intestinal peptide-positive (VIP+), lysosomal associated membrane protein 5-positive (LAMP5+), and putative parvalbumin-positive (putative PVALB+, or pPVALB+), since probes for PVALB were not always directly assayed (Fig. 3*E*). PRDD-seq cells assigned to upper, middle, and lower layers by the 10× PFC scRNAseq dataset were also enriched in L2–L3, L4–L5, and L6 markers, respectively, according to the SMART-seq MTG scRNAseq dataset, indicating the similarity of the cell type compositions between PFC and MTG, the similarity of the results with the two scRNAseq methods, and the robustness of the mapping algorithm (Fig. 3 *E*, *Upper*). Both our 10× scRNAseq dataset and PRDD-seq analysis of UMB1465 and UMB4638 showed higher proportions of inhibitory neurons (43 to 47%) than reported with other methods; however, this ratio was very similar between the three experiments, suggesting that the ratio reflects our particular NeuN+ sorting protocol rather than technical aspects of the cell typing methods (Fig. 3 *F*, *Upper*). We observed remarkably similar layer and subtype distribution between PRDD-seq and scRNAseq cells for excitatory neurons (χ^2 test; Fig. 3 *F*, *Middle*). Among inhibitory neurons, pPVALB+ inhibitory neurons showed a higher proportional representation in PRDD-seq than in scRNAseq, suggesting that a few neurons in this category might reflect amplification failure of the other inhibitory probes (SST, VIP, and LAMP5). In summary, our analysis suggests that PRDD-seq captures the major aspects of cell types, without systematic loss of any given cell type.

Early Divergence of Progenitors for Excitatory and Inhibitory Neurons. The simultaneous analysis of lineage and gene expression from the same neurons enabled us to study the change of cell type contribution during early neurogenesis. Using PRDD-seq, we profiled >2,700 neurons from two brains, UMB1465 and UMB4638, and successfully captured both lineage and cell type information from 747 and 480 neurons, respectively. In both UMB4638 and UMB1465, all lineage clades showed early sSNVs in both excitatory and inhibitory neurons, reflecting mutations occurring during early embryogenesis before the divergence of these cell types, whereas late SNVs show progressive restriction to one or the other cell type (Fig. 4 *A* and *B*). Among the six major clades in UMB1465, clade C contained seven nested branches with mosaic fractions diminishing from 0.33 to 0.0067 (Fig. 3*A* and *SI Appendix*, Table S1), with an increasing percentage of excitatory neurons containing mutations C1 to C5, and only excitatory neurons containing mutations C6 to C7 (Fig. 4*A*), while clade F showed similar progressive restriction. Similarly, both clades A and B in UMB4638 showed nested mutations that became progressively limited to excitatory neurons (Fig. 4*B*). Interestingly, the excitatory neurons appeared exclusively in branches with mosaic fraction below ∼0.04 (Fig. 4 *A* and *B* and *SI Appendix*, Table S1), corresponding to a progenitor giving rise to about 4% of the total cells in that cortical sample. Considering that ∼40% of cortical cells are excitatory neurons, with the remainder being glial cells or inhibitory neurons (28, 29), this observation suggests that 10 or more excitatory NPCs generate excitatory neurons in a given cortical area, or "column"; the fact that six to seven (including a branched clade) excitatory precursors are explicitly marked by nonoverlapping clades, and account for 60 to 70% of excitatory neurons in our sample, independently supports this estimate. On the other hand, two clades (clades A and B) from UMB1465 are statistically enriched for inhibitory neurons (two-sided one-proportion Z test's $P < 0.05$), with the percentage of inhibitory neurons increasing from B1 to B2 (Fig. 4*A*). These results show that at least some human NPCs demonstrate restricted cell type output and that excitatory and inhibitory neurons are generated from distinct progenitor regions, supporting the model first established in mice (30–32). In fact, the ganglionic eminence between humans and nonhumans are quite conserved (33, 34), which is consistent with our findings.

"Inside-Out" Order of Cortical Layer Formation for Excitatory Neurons. Further subtyping of excitatory neurons using laminar markers revealed layer-specific patterns of excitatory neuron neurogenesis. For example, in UMB1465, the percentage of lower layer neurons carrying a mutation decreased from mutations C1 to C4, and no deep-layer neurons were detected carrying C5 to C7, with the percentage of upper layer neurons increasing correspondingly from C1 to C7 (Pearson correlation's $P = 2.9 \times 10^{-3}$; Fig. 4 *C*, *Upper*). To gain more precise layer identities of PRDD-seq cells, we mapped them to the SMART-seq MTG scRNAseq dataset obtained after layer microdissection using the same methods as earlier (5), which generated similar "birthdate" patterns in clade C, with early lineage sSNVs present in all layers, and later sSNVs restricted to middle and upper layers (Pearson correlation's $P = 1.4 \times 10^{-3}$; Fig. 4 *C*, *Lower*). A similar trend was also observed in clades A and B in UMB4638. Mapping PRDD-seq cells of UMB4638 to both 10× PFC and SMART-seq MTG scRNAseq datasets showed that cells with later lineage markers were restricted to middle and upper layers (Fig. 4*D*). These results together directly indicate that human cortical excitatory neurons are formed in "inside-out" sequence after preplate cells are born, similar to mouse and nonhuman primates (35–37). Furthermore, this suggests that neurons in lower cortical layers begin becoming postmitotic relatively quickly after progenitors are specialized for excitatory neuron production.

Diverse Spatiotemporal Patterns of Development of Inhibitory Neuron Subtypes. Mapping PRDD-seq cells onto two different scRNAseq datasets also allowed analysis of cortical inhibitory neurons, which originate from multiple developmentally transient structures of the ventral telencephalon, including the medial, lateral, and caudal ganglionic eminences (MGE, LGE, and CGE), and migrate into dorsal cortex (30, 38). However, the highly dispersed nature of inhibitory neuron clones observed in animal models (39–41) suggests that sSNVs in the inhibitory lineage are likely to be present at exceedingly low allele frequencies in bulk DNA and tiny fractions of single cells, so that only sSNVs occurring relatively early in development have been analyzed so far. Inhibitory neurons derived from MGE and CGE can be distinguished by

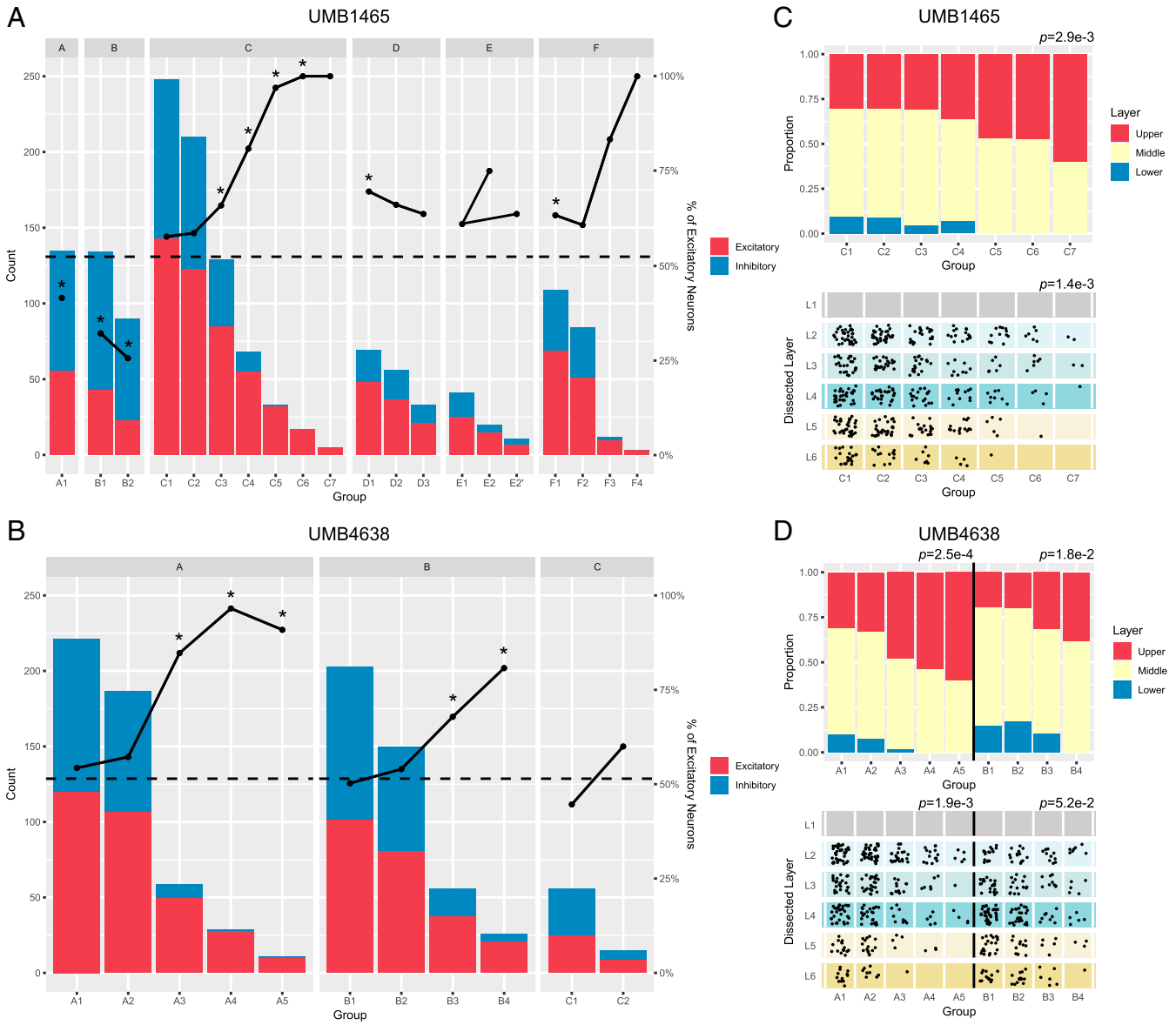

Fig. 4. PRDD-seq reveals distinct developmental sequence of excitatory neurons in different cortical layers. (*A* and *B*) The total number (bar plot) and ratio (dot plot) of excitatory and inhibitory neurons in different lineage clades defined by one or more sSNVs in (*A*) UMB1465 and (*B*) UMB4638. Percentage of excitatory neurons increased in later lineage time points in clades C and F in UMB1465 and clades A and B in UMB4638. In Clade E of UMB1465, E1 branches into two subclades E2A and E2B. Dashed line denotes average excitatory neuron percentage. Asterisk denotes significantly different excitatory–inhibitory ratio from the average (two-sided one-proportion *Z* test's $P < 0.05$). In clades C and F from UMB1465, and clades A and B from UMB4638, later mutations become progressively limited to excitatory neurons. (*C* and *D*) Layer distributions of excitatory neurons in representative excitatory lineages in (*C*) UMB1465 and (*D*) UMB4638. Layers are determined by mapping PRDD-seq cells onto (*Upper*) human PFC scRNAseq or (*Lower*) human MTG scRNAseq based on the expression profile similarity of marker genes. In all three illustrated clades, the percentage of upper layer neurons increased while that of lower layer neurons decreased in cells containing sSNVs present at lower mosaic fraction. *P* value was calculated by Pearson correlation with ordinal variables.

expression of specific markers (5, 6), and PRDD-seq analysis showed that interneurons with diverse marker genes were generated over the same developmental window (Fig. 5 *A* and *B*). The analyzed sSNVs were shared by multiple inhibitory subtypes, with hints that late marks might be more limited to cell types, but no differences that reached statistical significance (FDR-adjusted χ^2 test's $P > 0.05$). Previous studies cataloging interneurons in mouse and human have suggested that MGE-derived inhibitory neuron subtypes (SST+ and PVALB+) are enriched in infragranular cortical layers, while CGE-derived interneuron subtypes (LAMP5/PAX6+, VIP+) tend to occupy upper cortical layers preferentially (5, 42, 43), and thus our mapping of PRDD-seq cells onto scRNAseq reflected these patterns. Birthdating analyses in mice and nonhuman primates have reached contradictory conclusions about whether inhibitory neurons follow inside-out patterns of generation similar to excitatory neurons (44, 45), although recent analyses in mice suggest that previous contradictions may reflect the convolution of multiple patterns of generation that may be subtype specific (46). We found that MGE-derived pPVALB+ subtype neurons, enriched in layers IV to VI, showed, if anything, a trend for the latest-generated neurons to show markers of deeper layers (Fig. 5 *C* and *D*). SST+ neurons, widely distributed in layers II to VI, similarly did not show an inside-out pattern detectable with the mutations and cells analyzed (Fig. 5 *C* and *D*). We robustly detected SST+ neurons with expression of layer I markers in human PFC (SST-like subclass) (Fig. 5 *C* and *D*),

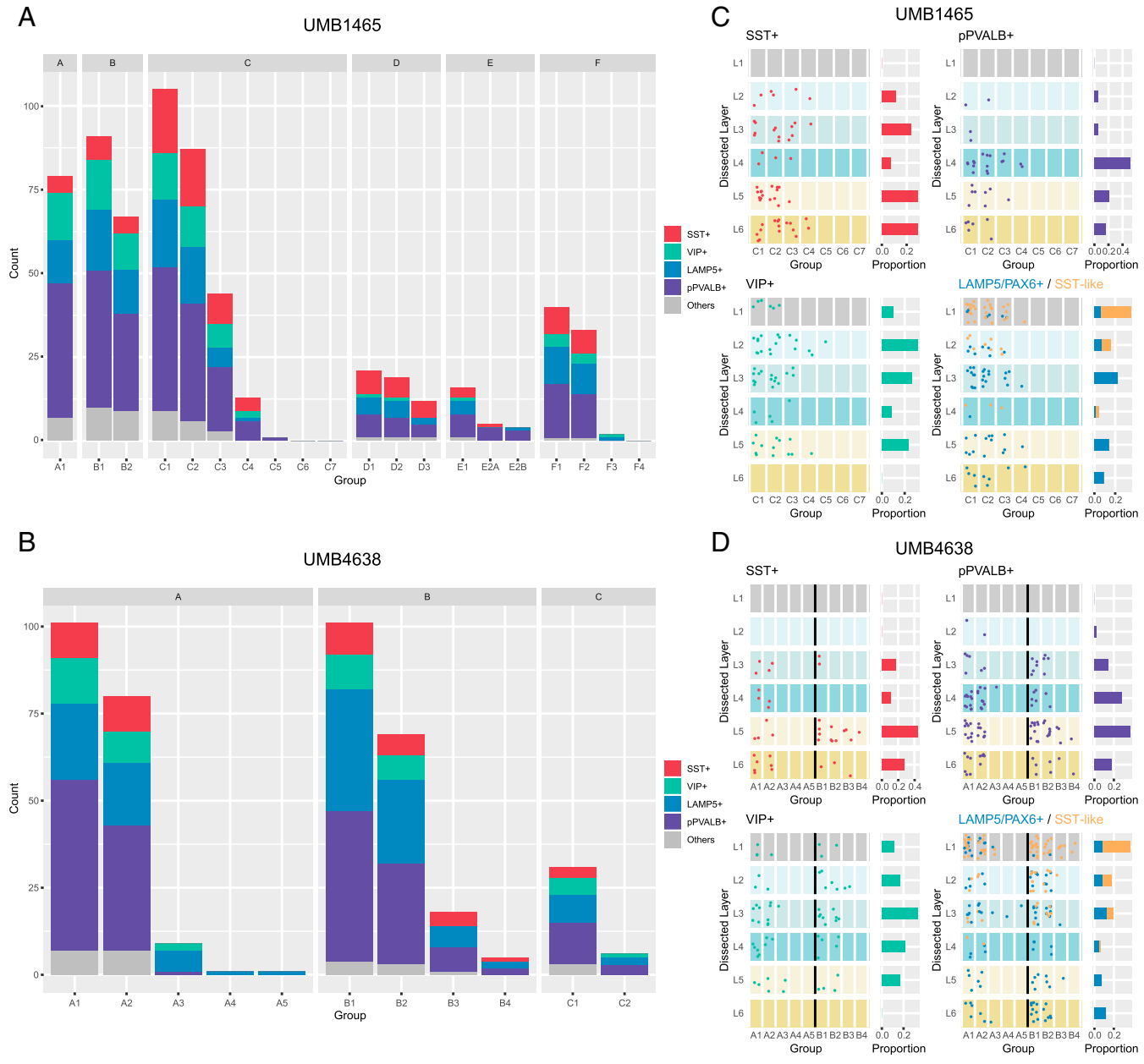

Fig. 5. PRDD-seq reveals heterogeneous developmental process for inhibitory neurons. (*A* and *B*) Distribution of different subtypes of inhibitory neurons in different lineages in (*A*) UMB1465 and (*B*) UMB4638. Major subtypes of inhibitory neurons are widely distributed in different lineages. (*C* and *D*) Layer distributions of inhibitory subtypes in representative lineages in (*C*) UMB1465 and (*D*) UMB4638. Bar graphs show the proportion of each subtype of neurons in different layers. MGE-derived (SST+ and pPVALB+) and CGE-derived (VIP+, LAMP5/PAX6+, and SST-like) interneurons showed similar mutation profiles, suggesting that the groups are produced simultaneously. The pPVALB+ subtype neurons were enriched in layers IV to VI, while MGE-derived SST+ interneurons showed a similar laminar distribution as pPVALB+ interneurons, with no clear evidence of an "inside-out" birth dating pattern. CGE-derived interneurons were broadly distributed across cortical layers, with SST-like cells heavily favoring supragranular layers; LAMP5+, including SST-like cells, were enriched for later lineage marks, suggesting they may be produced later in development than other subtypes.

consistent with observations in MTG (5, 47) and in mice, where such layer I SST+ expressing cells are rare but present (43, 47). These upper layer, CGE-derived SST-like cells are a subclass of LAMP5+ interneurons that are more transcriptionally related to VIP neurons than MGE-derived SST+ interneurons, although they lack VIP expression (5, 47). Our data further confirm that LAMP5+ interneurons express markers suggesting broad laminar location, but also did not reveal a simple inside-out progression of formation (5). Interestingly, we observed a substantial proportion of LAMP5+ inhibitory neurons, particularly the SST-like class, labeled by later mutations, indicating that this subtype may be generated later during development than other inhibitory cell types (Fig. 5 *C* and *D*). Overall, our findings suggest little evidence of the inside-out patterns of neurogenesis demonstrated by excitatory neurons, but also show that detailed analysis of interneurons will likely require deep datasets of sSNV occurring at late stages of interneuron development, and higher-throughput methods of analysis.

Discussion

We have developed scMH and PRDD-seq that allowed us simultaneous analysis of cell lineage and transcriptional cell type in human brain—and, potentially, any mammalian brain—through

improved identification of sSNVs in deep bulk and single-cell sequencing data. Our analysis of a single cortical area (PFC) in two individual brains revealed some conserved patterns of cell lineage compared to nonhumans, including that inhibitory and excitatory neurons diverge early in humans, and that excitatory neurons form following a similar "inside-out" order as seen in the animal models. However, PRDD-seq also provides a quantitative estimate in any species of the number of progenitor cells (~10) that generate the excitatory neurons in a given cortical area. Furthermore, PRDD-seq also provided some direct insight into inhibitory neuron development in humans, supporting parallel development of different subtypes of inhibitory neurons, with spatial and temporal associations specific only to some subtypes. Our data show that, as methods improve to capture sSNVs present in small numbers of cells, the natural occurrence of sSNVs with each cell division (13, 14, 17) is likely sufficient to provide a very rich map of cell lineage patterns in any given postmortem human brain.

The human cerebral cortex has been thought to contain ~80% excitatory glutamatergic neurons and 20% GABAergic interneurons (48), although recent scRNAseq studies have reported a somewhat lower ratio of about 70% excitatory neurons (*SI Appendix*, Table S3) (5, 49, 50). Although our PRDD-seq analysis showed 661 excitatory versus 566 inhibitory PRDD-seq cells in total for UMB1465 and UMB4638, which represents 54% excitatory neurons (*SI Appendix*, Table S3), this higher proportion of inhibitory neurons seems to reflect either aspects of the tissue (which was stored for long periods frozen) or our NeuN+ sorting method, since similar ratios are seen in 10× scRNAseq from the one brain analyzed (*SI Appendix*, Table S3). On the other hand, PRDD-seq cells are studied as containing at least one sSNV identified from scMH using a small number of deeply sequenced neuronal nuclei isolated from the same region, and so do not represent an unbiased sampling of the human brain region. Nonetheless, the fact that we can assign 60 to 70% of all excitatory neurons to clades in UMB1465, and that neurons with identified SNVs represent most major neuronal types in scRNAseq (Fig. 3*E*), suggests that our sampling has captured the majority of the lineage of the cortical patch, although rare lineages are likely to be missed without much deeper sequencing. Moreover, the presence of six to seven explicitly marked clades, and the ability to correlate the allele frequency of a sSNV to the excitatory restriction of the cells carrying that sSNV, allows two independent quantitative assessments of how many progenitors (~10) contribute to the neurons of the patch of cortex from which neurons were isolated, illustrating the remarkable quantitative potential of this approach.

Since occasional dropout of DNA marks and RNA markers in PRDD-seq is unavoidable, limited by the quality of isolated nuclei, we emphasize that our results are most robust when analyzing cells positive for both. The quality of postmortem brain tissues can influence the integrity of both gDNA and messenger RNA (mRNA). Regarding DNA, since no whole-genome amplification is performed prior to targeted preamplification, only a single molecular copy of each allele is available for genotyping of each sSNV, so occasional dropout is inevitable. However, our lineage strategy is based not only on the presence of clade-specific sSNVs but also on the absence of many sSNVs from other clades (Fig. 3*A*), so the chance for misassigning cells should be relatively small. Nevertheless, mapping our sSNVs onto our scRNAseq dataset suggests that lineage marks are present in the major neuronal subtypes, although rare neuronal types are likely to be missed given our modest sample size. Regarding RNA, single nuclei from postmortem human brain contains only a small amount of mRNAs. Fluidigm Biomark assays are microfluidics-based qPCR assays that are sensitive to subtle changes of the input or environment. As a result, we observed a 30.4% dropout rate of DNA markers and similar level of dropout of RNA marker dropout. However, since PRDD-seq analyses excluded these dropout events, and were completely based on the relative cell type proportions across different stages within one lineage, we have no reason to think that the dropouts are systematic with respect to cell type, with one exception: the relatively larger proportion of pPVALB+ neurons in PRDD-seq than scRNAseq, likely reflecting the failure of some probes for SST, VIP, and LAMP5. Better and richer probe sets are likely to be able to resolve this in the future.

There are limitations to our analysis, since we are analyzing a small sample of the vast size of the human brain, and PRDD-seq is relatively low throughput and expensive, so our initial analysis only can make conclusions about relatively common cell types. The present analysis is somewhat limited in the analysis of late mutations present in 1% of cells, especially interneurons, since it is challenging to detect those mutations with great sensitivity, but will await single-cell studies on subtypes of neurons in the future. On the other hand, the combined analysis of sSNVs and cell types is archival and progressive. The vast size of the human brain means that each subsequent round of DNA sequencing—whether of bulk tissue or of single or pooled cells—adds to the total depth of sequence data, and provides progressively richer information about late sSNVs. Indeed, the likely dispersed nature of inhibitory clones suggests that analyzing one cortical region could provide sequence data useful in the analysis of a completely different cortical region for these cell types.

Overall, PRDD-seq has many advantages even beyond the quantitative analysis of lineages and mosaic fractions that we begin to illustrate here. Since the method uses sSNVs as lineage marks, it is inherently genomic and so not only allows correlation of normal developmental patterns but would immediately capture alterations to lineage patterns caused by function-altering germline or somatic mutations. In addition, since sSNVs serve as in vivo cellular markers for drawing a developmental lineage map without any transgenic manipulation as demonstrated in this study, the method promises to be applicable, in principle, to any species or human disease condition for which postmortem brain is available.

Materials and Methods

Methodological details of human tissues WGS, estimation of cell-specific dropout rate and error rate, sSNV calling and performance comparison, validation of sSNVs, generation of simulated single-cell WGS data, design and selection of Taqman genotyping and gene expression probes, 10× Genomics preparation and sequencing, scRNAseq analysis, joint analysis of PRDD-seq and scRNAseq cells, and quantification and statistical analysis are described in *SI Appendix, Supplementary Materials and Methods*.

Framework of scMH. The overall framework of scMH is illustrated in Fig. 2*A*. The sSNV candidates were first called from the bulk sequencing data using a Bayesian graphical model (20, 21), in which the likelihoods of somatic mutation and three genotypes of inherited mutation were calculated with the consideration of binomial sampling variation and base-calling errors (Fig. 2 *A*, *Left*). The presence or absence of somatic mutation in each single cell was then inferred by adapting the likelihood and allele fraction of somatic mutation estimated from bulk sample as prior probability, after controlling cell-specific allele dropout rate (*d*) and error rate (*e*) (Fig. 2 *A*, *Right*). Detailed information about the Bayesian model and error filters of scMH is provided in *SI Appendix, Supplementary Materials and Methods*.

PRDD-Seq. Single nuclei from postmortem brain samples were isolated using FANS for NeuN as described previously (51). Isolated single neuronal nuclei were directly sorted into CellsDirect One-Step qRT-PCR (Thermo Fisher Scientific) preamplification buffers containing 0.14× Taqman gene expression assays and single nucleotide polymorphisms genotyping assays. Preamplification of all cDNA and gDNA amplicons were performed directly after the FANS sorting. Following preamplification, samples were diluted 10-fold and loaded onto 96.96 genotyping or 192.24 gene expression dynamic assay integrated fluidic circuits for standard amplification per manufacturer's instructions (Biomark, Fluidigm). Genotype and gene expression were

further determined by Biomark machine and analyzed by Biomark/EP1 software (Fluidigm).

Data and Code Availability. Single-cell WGS data were deposited in the National Center for Biotechnology Information Sequence Read Archive with accession numbers SRP041470 and SRP061939. Bulk WGS data are available from the National Institute of Mental Health Data Archive (https://doi.org/10.15154/1503337). MosaicHunter is publicly available at http://mosaichunter.cbi.pku.edu.cn/. Config files of scMH and other scripts about PRDD-seq can be accessed at https://github.com/AugustHuang/PRDD-seq.

ACKNOWLEDGMENTS. We thank R. Mattieu, K. Brownstein, J. Li, Flow Cytometry Facility in Boston Children's Hospital, Boston Children's Hospital Intellectual and Developmental Disabilities Research Center Molecular Genetics Core Facility, and the Research Computing group at Harvard Medical School for assistance. We thank G. Fishell, C. Harwell, and F. Vaccarino for comments on the manuscript. Human tissue was obtained from the NIH NeuroBioBank at the University of Maryland and Autism BrainNet, and we thank the donors and their families for their invaluable donations for the advancement of science. P.L. is a Howard Hughes Medical Institute–Helen Hay Whitney Foundation Fellow. R.E.R. is supported by the Stuart H. Q. and Victoria Quan Fellowship in Neurobiology and by the Harvard/MIT MD-PhD program (Grant T32GM007753). S.N.K is supported by the Stuart H. Q. and Victoria Quan Fellowship in Neurobiology. E.A.L is supported by grants from the National Institute on Aging (Grant K01AG051791), Suh Kyungbae Foundation, and the Allen Discovery Center program through The Paul G. Allen Frontiers Group. C.A.W. is supported by the Manton Center for Orphan Disease Research, the Allen Discovery Center program through The Paul G. Allen Frontiers Group, Grant R01NS032457 from the National Institute of Neurological Disorders and Stroke, and Grant U01MH106883 from the National Institute of Mental Health. C.A.W. is an Investigator of the Howard Hughes Medical Institute.

1. P. Rakic, Evolution of the neocortex: A perspective from developmental biology. *Nat. Rev. Neurosci.* **10**, 724–735 (2009).
2. D. H. Geschwind, P. Rakic, Cortical evolution: Judge the brain by its cover. *Neuron* **80**, 633–647 (2013).
3. M. Heide, K. R. Long, W. B. Huttner, Novel gene function and regulation in neocortex expansion. *Curr. Opin. Cell Biol.* **49**, 22–30 (2017).
4. C. S. Raju *et al.*, Secretagogin is expressed by developing neocortical GABAergic neurons in humans but not mice and increases neurite arbor size and complexity. *Cereb. Cortex* **28**, 1946–1958 (2018).
5. R. D. Hodge *et al.*, Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019).
6. S. Zhong *et al.*, A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).
7. B. B. Lake *et al.*, Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018).
8. A. McKenna *et al.*, Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science* **353**, aaf7907 (2016).
9. K. L. Frieda *et al.*, Synthetic recording and in situ readout of lineage information in single cells. *Nature* **541**, 107–111 (2017).
10. B. Raj *et al.*, Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. *Nat. Biotechnol.* **36**, 442–450 (2018).
11. B. Spanjaard *et al.*, Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. *Nat. Biotechnol.* **36**, 469–473 (2018).
12. A. Rodriguez-Meira *et al.*, Unravelling intratumoral heterogeneity through high-sensitivity single-cell mutational analysis and parallel RNA sequencing. *Mol. Cell* **73**, 1292–1305.e8 (2019).
13. Y. S. Ju *et al.*, Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* **543**, 714–718 (2017).
14. T. Bae *et al.*, Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* **359**, 550–555 (2018).
15. S. De, Somatic mosaicism in healthy human tissues. *Trends Genet.* **27**, 217–223 (2011).
16. M. A. Lodato *et al.*, Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* **350**, 94–98 (2015).
17. M. A. Lodato *et al.*, Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* **359**, 555–559 (2018).
18. L. S. Ludwig *et al.*, Lineage tracing in humans enabled by mitochondrial mutations and single-cell genomics. *Cell* **176**, 1325–1339.e22 (2019).
19. Y. Dou *et al.*, Accurate detection of mosaic variants in sequencing data without matched controls. *Nat. Biotechnol.* **38**, 314–319 (2020).
20. A. Y. Huang *et al.*, MosaicHunter: Accurate detection of postzygotic single-nucleotide mosaicism through next-generation sequencing of unpaired, trio, and paired samples. *Nucleic Acids Res.* **45**, e76 (2017).
21. A. Y. Huang *et al.*, Postzygotic single-nucleotide mosaicisms in whole-genome sequences of clinically unremarkable individuals. *Cell Res.* **24**, 1311–1327 (2014).
22. H. Zafar, Y. Wang, L. Nakhleh, N. Navin, K. Chen, Monovar: Single-nucleotide variant detection in single cells. *Nat. Methods* **13**, 505–507 (2016).
23. X. Dong *et al.*, Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat. Methods* **14**, 491–493 (2017).
24. C. L. Bohrson *et al.*, Linked-read analysis identifies mutations in single-cell DNA-sequencing data. *Nat. Genet.* **51**, 749–754 (2019).
25. J. Hård *et al.*, Conbase: A software for unsupervised discovery of clonal somatic mutations in single cells through read phasing. *Genome Biol.* **20**, 68 (2019).
26. A. Y. Huang *et al.*, Distinctive types of postzygotic single-nucleotide mosaicisms in healthy individuals revealed by genome-wide profiling of multiple organs. *PLoS Genet.* **14**, e1007395 (2018).
27. T. Helleday, S. Eshtad, S. Nik-Zainal, Mechanisms underlying mutational signatures in human cancers. *Nat. Rev. Genet.* **15**, 585–598 (2014).
28. C. S. von Bartheld, J. Bahney, S. Herculano-Houzel, The search for true numbers of neurons and glial cells in the human brain: A review of 150 years of cell counting. *J. Comp. Neurol.* **524**, 3865–3895 (2016).
29. H. Markram *et al.*, Interneurons of the neocortical inhibitory system. *Nat. Rev. Neurosci.* **5**, 793–807 (2004).
30. S. A. Anderson, D. D. Eisenstat, L. Shi, J. L. Rubenstein, Interneuron migration from basal forebrain to neocortex: Dependence on Dlx genes. *Science* **278**, 474–476 (1997).
31. G. Fishell, C. A. Mason, M. E. Hatten, Dispersion of neural progenitors within the germinal zones of the forebrain. *Nature* **362**, 636–638 (1993).
32. S. Anderson, M. Mione, K. Yun, J. L. Rubenstein, Differential origins of neocortical projection and local circuit neurons: Role of Dlx genes in neocortical interneuronogenesis. *Cereb. Cortex* **9**, 646–654 (1999).
33. T. Ma *et al.*, Subcortical origins of human and monkey neocortical interneurons. *Nat. Neurosci.* **16**, 1588–1597 (2013).
34. D. V. Hansen *et al.*, Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat. Neurosci.* **16**, 1576–1587 (2013).
35. P. Rakic, Neurons in rhesus monkey visual cortex: Systematic relation between time of origin and eventual disposition. *Science* **183**, 425–427 (1974).
36. J. B. Angevine, Jr, R. L. Sidman, Autoradiographic study of cell migration during histogenesis of cerebral cortex in the mouse. *Nature* **192**, 766–768 (1961).
37. P. Gao *et al.*, Deterministic progenitor behavior and unitary production of neurons in the neocortex. *Cell* **159**, 775–788 (2014).
38. C. Mayer *et al.*, Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018).
39. M. Turrero García, E. Mazzola, C. C. Harwell, Lineage relationships do not drive MGE/PoA-derived interneuron clustering in the brain. *Neuron* **92**, 52–58 (2016).
40. K. N. Brown *et al.*, Clonal production and organization of inhibitory interneurons in the neocortex. *Science* **334**, 480–486 (2011).
41. C. B. Reid, S. F. Tavazoie, C. A. Walsh, Clonal dispersion and evidence for asymmetric cell division in ferret cortex. *Development* **124**, 2441–2450 (1997).
42. M. J. Nigro, Y. Hashikawa-Yamasaki, B. Rudy, Diversity and connectivity of layer 5 somatostatin-expressing interneurons in the mouse barrel cortex. *J. Neurosci.* **38**, 1622–1633 (2018).
43. B. Rudy, G. Fishell, S. Lee, J. Hjerling-Leffler, Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011).
44. E. S. Ang Jr, T. F. Haydar, V. Gluncic, P. Rakic, Four-dimensional migratory coordinates of GABAergic interneurons in the developing mouse cortex. *J. Neurosci.* **23**, 5805–5815 (2003).
45. V. V. Rymar, A. F. Sadikot, Laminar fate of cortical GABAergic interneurons is dependent on both birthdate and phenotype. *J. Comp. Neurol.* **501**, 369–380 (2007).
46. S. M. Kelly, R. Raudales, M. Moissidis, G. Kim, Z. J. Huang, Multipotent radial glia progenitors and fate-restricted intermediate progenitors sequentially generate diverse cortical interneuron types. bioRxiv:10.1101/735019 (15 August 2019).
47. E. Boldog *et al.*, Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nat. Neurosci.* **21**, 1185–1195 (2018).
48. C. P. Wonders, S. A. Anderson, The origin and specification of cortical interneurons. *Nat. Rev. Neurosci.* **7**, 687–696 (2006).
49. N. Habib *et al.*, Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).
50. B. B. Lake *et al.*, Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016).
51. G. D. Evrony *et al.*, Cell lineage analysis in human brain using endogenous retroelements. *Neuron* **85**, 49–59 (2015).