Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex

Alex A Pollen1,2,4, Tomasz J Nowakowski1,2,4, Joe Shuga3,4, Xiaohui Wang3,4, Anne A Leyrat3, Jan H Lui1,2, Nianzhen Li3, Lukasz Szpankowski3, Brian Fowler3, Peilin Chen3, Naveen Ramalingam3, Gang Sun3, Myo Thu3, Michael Norris3, Ronald Lebofsky3, Dominique Toppani3, Darnell W Kemp II3, Michael Wong3, Barry Clerkson3, Brittnee N Jones3, Shiquan Wu3, Lawrence Knutsson3, Beatriz Alvarado3, Jing Wang3, Lesley S Weaver3, Andrew P May3, Robert C Jones3, Marc A Unger3, Arnold R Kriegstein1,2 & Jay A A West3

Large-scale surveys of single-cell gene expression have the potential to reveal rare cell populations and lineage relationships but require efficient methods for cell capture and mRNA sequencing1–4. Although cellular barcoding strategies allow parallel sequencing of single cells at ultralow depths5, the limitations of shallow sequencing have not been investigated directly. By capturing 301 single cells from 11 populations using microfluidics and analyzing single-cell transcriptomes across downsampled sequencing depths, we demonstrate that shallow single-cell mRNA sequencing (~50,000 reads per cell) is sufficient for unbiased cell-type classification and biomarker identification. In the developing cortex, we identify diverse cell types, including multiple progenitor and neuronal subtypes, and we identify EGR1 and FOS as previously unreported candidate targets of Notch signaling in human but not mouse radial glia. Our strategy establishes an efficient method for unbiased analysis and comparison of cell populations from heterogeneous tissue by microfluidic single-cell capture and low-coverage sequencing of many cells.

To routinely capture single cells, we designed the C1 Single-Cell Auto Prep System (Fig. 1a). The microfluidic system performs reverse transcription and cDNA amplification in nanoliter reaction volumes (Fig. 1b,c), which increase the effective concentration of reactants and may improve the accuracy of mRNA sequencing (mRNA-seq)6. We sequenced libraries from single cells at high coverage (~8.9 × 10^6 reads per cell) and used the results as a reference to explore the consequences of reduced sequencing depth. To explore the current practical limits of low-coverage sequencing, we pooled dozens of barcoded single-cell libraries in single MiSeq System runs (Illumina; ~2.7 × 10^5 reads per cell) and downsampled high-coverage results to ultra-low depths. We prepared sequencing libraries after cDNA amplification with the SMARTer Ultra Low RNA Kit for Illumina Sequencing (Clontech) and the Nextera XT kit (Illumina). Genomic alignment rates and other quality metrics were similar across libraries, whereas empty negative control wells showed no appreciable sequence alignment (<1%) (Supplementary Table 1).

We assessed the accuracy, detection rates and variance of RNA-level estimates from low-coverage sequencing of single-cell libraries by comparing the results with known quantities of spike-in RNA transcripts7 and with high-coverage sequencing of the same libraries. The levels of RNA spikes determined by low-coverage mRNA sequencing correlated strongly with the known input quantities (r = 0.968; Fig. 1d). For inputs above 32 copies, all spikes could be detected in all samples with minimal variance (Fig. 1e)8,9. In a representative cell, the majority of genes detected by high-coverage sequencing were also detected by low-coverage sequencing (Fig. 1f,g). Of the genes detected by high- but not low-coverage sequencing, the vast majority (98%) were not expressed at high levels (transcripts per million (TPM) > 100), and most (63%) were expressed at low levels (1 < TPM < 10; Supplementary Fig. 1). Across 301 cells from a range of sources, the average correlation between the estimates of single-cell gene expression from low- and high-coverage sequencing was 0.91 (Fig. 1h,i and Supplementary Fig. 2). However, for transcripts with low expression levels (1 < TPM < 10), the correlation dropped to 0.25, demonstrating a limitation in quantifying low-abundance transcripts in individual cells using shallow sequencing. Despite this limitation, combining low-coverage results from as few as ten individual K562 cells accurately reflected results from a pooled population of K562 cells captured by flow cytometry (r > 0.92) (Fig. 2a,b). We conclude that single-cell capture and low-coverage sequencing can be used to profile the gene expression of individual cells and that combined results reflect the properties of a given cell population.

To examine whether low-coverage sequencing can distinguish between cell types, we first compared cells from sources that are expected to show robust differences in gene expression: pluripotent, skin, blood and neural cells (Supplementary Tables 1 and 2).

Received 30 January; accepted 25 June; published online 3 August 2014; doi:10.1038/nbt.2967
Figure 1 Capturing single cells and quantifying mRNA levels using the C1 Single-Cell Auto Prep System. (a) Key functional components of the C1 System are labeled, including the pneumatic components necessary for control of the microfluidic integrated fluidic circuit (IFC) and the thermal components necessary for preparatory chemistry. (b) Left, complete IFC with carrier; reagents and cells are loaded into dedicated carrier wells, and reaction products are exported to other dedicated carrier wells. Middle, diagram of the IFC, with connections between the polydimethylsiloxane microfluidic chip and carrier (pink circles), control lines (red), fluidic lines for preparatory chemistry (blue) and lines connecting the control lines (green) shown. Right, a single cell captured in a 4.5-nl capture site; there are 96 captures sites per IFC. The average single-cell capture rate was 72 ± 5 cells (mean ± s.e.m.) per chip (Supplementary Tables 1 and 2). IFC diagram reprinted with permission from ref. 32. (c) Schematic for a C1 reaction line, with the reaction line colored light gray and the isolation valves shown in varied colors. All reagents are delivered through a common central bus line (a segment of the bus line is shown on the far left). Each reaction begins in the 4.5-nl capture site. Delivery of the lysis reagent expands the reaction to also include the first 9-nl chamber. The reaction is expanded again after delivery of the reverse transcription (RT) reagent to include the second and third 9-nl chambers. The two 135-nl reaction chambers are included to provide the larger volume required for the PCR reagents. After the addition of RT reagent, the contents of the reaction line are pumped in a loop using a bypass line (bottom) for mixing, and the IFC is then incubated at 42°C for RT. Mixing is repeated after the addition of the PCR reagents, and thermal cycling is performed. After preparatory chemistry, each single-cell reaction product exits the chip using a dedicated fluidic path to the carrier (path shown to the right). (d) Sequencing of reaction products from 46 K562 cells at low coverage (1.7 × 10^9 reads per cell) reveals that expression level estimates correlate strongly with known copy numbers of input spikes (Pearson’s r = 0.968) from the External RNA Controls Consortium (ERCC) RNA Spike-In Control Mix 1 (2.8 × 10^4 copies per reaction). (e) The fraction of positive reactions in which ERCC transcripts are detected above 1 TPM in single cells and the coefficient of variation (CV) for ERCC levels are plotted against the spike input amounts. (f–i) Pools of barcoded libraries from 301 cells were sequenced at high coverage by HiSeq and at low coverage by MiSeq. (f) In a representative cell, 4,644 genes were detected above 1 TPM in both data sets. (g) The average number of genes expressed at various levels detected by high coverage sequencing in each cell type (Online Methods). Data are shown as the mean ± s.d. (h) In a representative cell, the expression levels of genes detected in high- and low-coverage data sets were highly correlated (r = 0.91). (i) Histogram of correlation coefficients for all single cells (n = 301). The mean correlation coefficients increased with expression level: 0.25 (1 < TPM < 10, red), 0.66 (10 ≤ TPM < 100, green), 0.93 (TPM > 100, blue) and 0.91 (all genes with TPM > 1).
variation in PC1, PC2 and PC3 separated the cells into four broad groups, with cells from each source contributing to multiple groups (Fig. 3b,c). The group membership of individual cells largely overlapped between the high- and low-coverage sequencing data (Fig. 3d) and could be identified at downsampled depths of between 5,000 and 50,000 reads per cell (Supplementary Figs. 9 and 10). Cells in groups I and II expressed high levels of neuroepithelial markers, including VIM, SOX2 and PAX6. In addition, cells in group I also expressed high levels of the proliferative markers CDK1 and ASPM, whereas cells in group II expressed high levels of the mature radial glial markers SLC1A3 and HES1. In contrast, cells in groups III and IV expressed the pan-neuronal marker DCX, and cells in group IV also expressed many markers of neuronal maturation, including MEF2C, SATB2 and SNAP25 (Supplementary Figs. 10 and 11). Thus, we interpreted the groups to represent dividing neural progenitors (group I), radial glia (group II), newborn neurons (group III) and maturing neurons (group IV). To independently validate our results, we examined the expression of genes distinguishing each group across 599 tissue samples collected from distinct regions of the developing human cortex. Genes defining neural progenitors and radial glia in the single-cell analysis were strongly enriched in germinal zones, whereas genes defining newborn and maturing neurons were strongly enriched outside of germinal zones (Fig. 3e–g and Supplementary Table 5). Similarly, in situ hybridization confirmed that newly identified markers of radial glia, newborn and maturing neurons are expressed in zones where these cell types are abundant (Fig. 3b–o and Supplementary Figs. 12 and 13).

In addition to the four broad groups identified using hierarchical clustering, distinct subgroups also corresponded to other known and potentially new cell types (Supplementary Figs. 10 and 11). For example, cells in group Ib expressed multiple markers of intermediate neural progenitors. Cells in group IIb expressed the canonical markers of inhibitory interneurons GAD1 and DLX genes, as well as previously unreported markers such as PDZRN3 (Fig. 3p), whereas the remaining cells in group III expressed the proneural genes NEUROD1 and NEUROD6. In addition, group III cells expressed UNCS5, a gene that is transiently upregulated in newly generated mouse excitatory neurons and is required for the earliest phases of migration, and other genes such as ROBO2 and NTM (Fig. 3q), whose possible roles in newborn cortical neurons remain to be investigated. Group IV could be further divided into maturing neurons expressing high levels of CAMKV and cells expressing high levels of ADRA2A (Supplementary Fig. 11). Complementary expression patterns of CAMKV and ADRA2A proteins in the cortical plate (Fig. 3r,s) indicate that these finer subgroups may reflect additional heterogeneity within maturing cortical neurons. Although many of the genes explaining variation among single cells related to cell identity, a subset of genes with strong PCA loading were enriched for mitotic markers and have not been studied in radial glia development (Supplementary Fig. 13). The candidate mitotic markers CKS2 and HMGB2 were detected specifically in a subset of human radial glia undergoing cell division at the edge of the lateral ventricle (Fig. 3t–w). Thus, low-coverage sequencing of single cells collected from primary tissue can be used to identify cell types, states and candidate biomarkers.

Transcription of immediate early genes has been studied extensively in activated neurons, but the strong PCA loading scores of EGR1 and FOS suggest that their expression may also reflect important aspects of cellular diversity in the developing cortex (Supplementary Fig. 13 and Supplementary Table 5). Indeed, in situ hybridization revealed mosaic expression of EGR1 and FOS in the ventricular zone (Fig. 4a–d). The levels of EGR1 and FOS were highly correlated across single radial glial cells, and the corresponding proteins were coexpressed in a subset of radial glia, suggesting that these genes could be transcribed in response to the same signaling pathway (Fig. 4e,h and Supplementary Fig. 14). Multiple signaling pathways, including FGF, Notch and Wnt, orchestrate radial glia development, but asynchronous activation of these signaling pathways in neighboring cells makes identifying downstream effector genes challenging. Coordinated patterns of pathway activation in other tissues have facilitated the identification of candidate downstream effector genes,
but these target genes often depend on cellular context and vary across species\(^\text{16,18}\) (Fig. 4i–k).

To determine which signaling pathway might be responsible for the coordinated activation of immediate early genes in human radial glia, we examined the correlation of EGR1 and FOS mRNA levels with the levels of canonical signaling pathway effectors that have been established by studies of other developmental processes. Across single cells, EGR1 and FOS mRNA correlated more strongly with the FGF effectors DUSP1, DUSP4, SPRY2 or SPRY4 or the WNT effectors AXIN2 or MYC (Fig. 4i and Supplementary Fig. 15). To examine whether activation of Notch signaling induces changes in EGR1 and FOS expression, we activated Notch signaling in cultured human cortical slices by removing extracellular calcium\(^\text{15}\). Incubation of primary human cortical slices with EDTA induced a rapid (30–40 min) increase in the levels of HES1, as well as EGR1, FOS and another highly correlated gene, TFFAP2C (Fig. 4m–p). In other stem cell contexts, EGR1 and FOS have a role in quiescence and retention in the stem cell niche\(^\text{20–22}\), but the role of these genes as candidate Notch targets in radial glia remains to be examined. Notably, EGR1 and C-FOS were rarely detected in mouse or ferret radial glia (Fig. 4e–g and Supplementary Fig. 14), indicating that these factors could contribute to differences in radial glia development across species, which include a markedly longer G1 phase and increased proliferative capacity in human radial glia\(^\text{23}\).

Identifying gene expression profiles of cells of the same type or state has numerous applications in modern biology. Here we demonstrate that ultra–low coverage single-cell analysis can be more generally applied to identify cells in different states of signaling pathway activation and candidate downstream target genes.

Figure 3 Low-coverage single-cell mRNA sequencing distinguishes diverse neural cell types and identifies biomarkers in heterogeneous tissue. (a) Schematic of cell types and sources selected to represent stages of neuronal differentiation. Cultured neural progenitor cells (NPC) represent early undifferentiated stages, and primary cortical samples are expected to contain radial glia, as well as newborn and maturing neurons. (b) Hierarchical clustering of 65 single cells across 500 genes with the strongest PC1–PC3 loading scores identifies four major groups of cells (I–IV), and k-means clustering identifies three clusters of genes (red, yellow and green). (c) Major groups can be interpreted on the basis of the expression of known genes. The table shows the number of cells of specific types captured from each source. (d) Cell classification on the basis of low-coverage data largely overlaps with classification based on high-coverage data. (e) Schematic of the distribution of cell types in the developing cortex at midgestation. VZ, ventricular zone; SVZ, subventricular zone; IZ, intermediate zone; SP, subplate; CP, cortical plate. (f) Heat map of gene expression values for PCA genes (columns) in 599 regions of the developing cortex\(^\text{11}\) (rows). (g) Genes belonging to the red cluster (\(n = 218\)) and yellow cluster (\(n = 98\)) are enriched in VZ and SVZ, whereas genes belonging to the green cluster (\(n = 176\)) are enriched in IZ, SP and CP; \(P\) values were calculated using Wilcoxon signed-rank test. Error bars, s.d. \(h\)–\(o\) Immunostaining for candidate cell division markers in the progenitor gene cluster. (h) Heat map of gene expression values for candidate newborn neuron markers PDZRN3 and CKS2 in GW16.5 human cortex (red) showing CKS2 (red) and HMGB2 (yellow) expression in radial glia undergoing mitosis at the edge of the ventricular surface revealed by immunoreactivity for phosphorilated (Ser82) vimentin (pVIM; u, w). Scale bars, 100 \(\mu m\) (h–o); 200 \(\mu m\) (p); 100 \(\mu m\) (q–s); 25 \(\mu m\) (t–w).
cortex, including radial glia in different stages of cell cycle progression and signaling pathway activation, and newly generated neurons in the earliest phases of migration. Few specific markers exist for the purification of these distinct cell states and transient developmental intermediates using flow cytometry. In contrast to flow cytometry, low-coverage single-cell sequencing detects thousands of abundant transcripts that can be analyzed to group cells according to cell type or state (Fig. 4q). Although the observed level of a given transcript in a single cell can vary because of transcriptional bursts and technical noise associated with low quantities of input RNA, the simultaneous profiling of multiple differentially expressed transcripts enables unbiased discovery of cell groups on the basis of shared signatures of gene expression. By grouping cells of a given identity, the bulk transcriptome for that population may then be accurately reconstructed. We anticipate that the unbiased classification of cells by efficient low-coverage single-cell sequencing will be applied to large-scale surveys of primary tissue samples to identify cell type–specific biomarkers, compare gene expression in cells of a given type across samples and reconstruct developmental lineages of related cell types.
METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Relevant data sets, including raw reads, have been deposited in the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/Traces/sra/) under accession number SRP041736.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank M. Bershetein (University of California, San Francisco (UCSF)), E. Di Lullo (UCSF), C. Gertz (UCSF), C. McLean (23andMe), F. Ng (Fluidigm Corporation), J. Liu (UCSF), C. Lowe (Stanford University), M. Oldham (UCSF), A. Diaz (UCSF) and H. Retallick (UCSF) for reading the manuscript and for useful suggestions. S. Wang and Y. Wang for technical assistance, K. Davies (University of Oxford) for providing the antibody to syncoilin, A. Simese (Institute of Genetics and Biophysics, Adriano Buzzati-Traverso) for providing the Em2x probe, S. Yao and S. Ku (Allen Institute for Brain Research) for assistance in the neural progenitor induction protocol, G. Wang (Stanford University) for providing human pluripotent stem cells (hiPSCs) and the staff at the San Francisco General Hospital for providing access to donated fetal tissue. L.S., N.L. and M.A.U. are partially supported by California Institute for Regenerative Medicine Tools and Technologies II grant RT2-02052 to M.A.U., and the generation and maintenance of the hiPSCs and neural progenitor cells was also supported by RT2-02052 to M.A.U. A.A.P. is supported by a Damon Runyon Cancer Research Foundation postdoctoral fellowship (DRG-2166-13). This research was also supported by National Institute of Neurological Disorders and Stroke awards R01NS075998 and R01NS072630 to A.R.K.

AUTHOR CONTRIBUTIONS

A.A.P., A.A.L., J.A.A.W., J.S., T.J.N. and X.W. conceived and designed the study. A.A.P., A.A.L., B.A., J.H.L., J.S., I.W., L.K., I.S., N.L., P.C. and T.J.N. performed experiments. A.A.P., A.A.L., J.S., S.W., T.J.N. and X.W. analyzed the data. A.A.L., A.P.M., B.C., B.F., B.N.J., D.W.K., D.T., G.S., J.S., J.A.A.W., J.W., M.T., M.N., M.W., N.R., P.C., R.L., I.S.W. and X.W. were involved in system development. A.P.M., A.R.K., J.A.A.W., M.A.U. and R.C.J. supervised the project, helped with design and interpretation, and provided laboratory space and financial support. A.A.P., J.S. and T.J.N. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Shalek, A.K. et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498, 236–240 (2013).
2. Shapiro, E., Biezuner, T. & Linnarsson, S. Single-cell sequencing-based technologies will revolutionize whole-organism science. Nat. Rev. Genet. 14, 618–630 (2013).
3. Kawaguchi, A. et al. Single-cell gene profiling defines proliferative progenitor subclasses in mammalian neurogenesis. Development 135, 3113–3124 (2008).
4. Treutlein, B. et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 509, 371–375 (2014).
5. Jain, D. et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science 343, 776–779 (2014).
6. Wu, A.R. et al. Quantitative assessment of single-cell RNA-sequencing methods. Genome Res. 21, 1543–1551 (2011).
7. Jiang, L. et al. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 21, 1106–1117 (2011).
8. Ren, B. et al. Graph theoretical analysis of single-cell RNA sequencing data. Genome Biol. 14, R10 (2013).
9. Kriegstein, A., Noctor, S. & Martinez-Cerdeno, V. Patterns of neural stem and progenitor cell division may underlie evolutionary cortical expansion. Nat. Rev. Neurosci. 7, 883–890 (2006).
10. Ross, M.E. & Walsh, C.A. Human brain malformations and their lessons for neuronal migration. Annu. Rev. Neurosci. 24, 1041–1070 (2001).
11. Miller, J.A. et al. Transcriptomic landscape of the prenatal human brain. Nature 508, 199–206 (2014).
12. Miyoshi, G. & Fishell, G. Dynamic FoxG1 expression coordinates the integration of multipolar pyramidal neuron precursors into the cortical plate. Neuron 74, 1045–1058 (2012).
13. Taricci, G. et al. EGR1 and the ERK-ERF axis drive mammary cell migration in response to EGF. FASEB J. 26, 1582–1592 (2012).
14. Biddle, I. et al. Molecular profiling of inflammatory breast cancer: identification of a poor-prognosis gene expression signature. Clin. Cancer Res. 10, 6789–6795 (2004).
15. Fischer, A.J., Scott, M.A., Ritchey, E.R. & Sherwood, P. Mitogen-activated protein kinase-signaling regulates the ability of Muller glia to proliferate and protect retinal neurons against excitotoxicity. Glia 57, 1538–1552 (2009).
16. Krol, A.J. et al. Evolutionary plasticity of segmentation clock networks. Development 138, 2783–2792 (2011).
17. Shimojo, H., Ohtsuka, T. & Kageyama, R. Oscillations in notch signaling regulate maintenance of neural progenitors. Neuron 58, 52–64 (2008).
18. Hansson, M.L. et al. MAMM1 acts cooperatively with EGR1 to activate EGR1-regulated promoters: implications for nephrogenesis and the development of renal cancer. PLoS ONE 7, e46001 (2012).
19. Harper, B.E. et al. Transcriptional dynamics elicited by a short pulse of notch activation involves feed-forward regulation by E(spl)Hes genes. PLoS Genet. 9, e1003162 (2013).
20. Min, I.M. et al. The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. Cell Stem Cell 2, 380–391 (2008).
21. Okada, S., Fukuda, T., Inada, K. & Tokuhisa, T. Prolonged expression of c-fos suppresses cell cycle entry of dormant hematopoietic stem cells. Blood 93, 816–825 (1999).
22. Bonnert, T.P. et al. Molecular characterization of adult mouse subventricular zone progenitor cells during the onset of differentiation. Eur. J. Neurosci. 24, 661–675 (2006).
23. Kornack, D.R. & Rakic, P. Changes in cell-cycle kinetics during the development and evolution of primate neocortex. Proc. Natl. Acad. Sci. USA 95, 1242–1246 (1998).
24. Islam, S. et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Nat. Commun. 3, 1093 (2012).
25. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343, 193–196 (2014).
26. Sims, D., Sudbery, I., Hott, N.E., Heger, A. & Ponting, C.P. Sequencing depth and coverage key considerations in genomic analyses. Nat. Rev. Genet. 15, 121–132 (2014).
27. Faddah, D.A. et al. Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse ESCs. Cell Stem Cell 13, 23–29 (2013).
28. Islam, S. et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods 11, 163–166 (2014).
29. Dalerba, P. et al. Single-cell dissection of transcriptional heterogeneity in human colon tumors. Nat. Biotechnol. 29, 1120–1127 (2011).
30. Ramsköld, D. et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat. Biotechnol. 30, 777–782 (2012).
31. Guo, G. et al. Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev. Cell 18, 675–685 (2010).
32. Shalek, A.K. et al. Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature 509, 363–369 (2014).
ONLINE METHODS

Origin of cell lines and tissue samples. hiPSCs were originally derived from neonatal male human foreskin B fibroblasts by G. Wang at the Department of Genetics at Stanford University using Sendai virus from Life Technologies (A16517). Cultured undifferentiated hiPSCs were maintained in Essential 8 Medium (Life Technologies). StainAlive Tra-1–60 antibody (DyLight 488) staining (Stemgent) was used to confirm an undifferentiated state. After a dissociation with StemPro Accutase Cell Dissociation Reagent (LifeTechnologies), single cells were plated onto Matrigel-coated plates at 2.5 × 10^5 cells per cm^2. NPC differentiation was induced using DMEM and F-12 medium (GIBCO DMEM and Nutrient Mixture F-12 from Life Technologies) supplemented with B27 (without vitamin A; Life Technologies, 12587010), N-2 (Life Technologies), 0.1 mM nonessential amino acids (Sigma), 0.5% bovine serum albumin (Sigma), 1 mM β-mercaptoethanol (BME; Sigma), 50 mM LDN-193189 (Stemgent), 5 µM SB431542 (Stemgent) and 1 µM Stмолекуюle Cyclopamine (Stemgent). After 12 d in culture, >90% of cells were immuno-positive for Pax6.

ATCC PCS-200-010 cell (foreskin keratinocytes, abbreviated Kera) culture was maintained in the dermal cell basal medium ATCC PCS-200-040 supplemented with the keratinoocyte growth kit (ATCC PCS-200-040). ATCC CRL-3338 cells (derived from a primary stage IIA, grade 3 invasive ductal carcinoma with no lymph node metastases, abbreviated CRL-2338) were cultured in the complete growth medium RPMI-1640 (ATCC 30-2001) supplemented with 10% fetal bovine serum (FBS; GIBCO 16000-077). ATCC CRL-3339 cells (Epstein-Barr virus–transformed B lymphoblasts, abbreviated CRL-2339) were cultured in the growth medium RPMI-1640 (ATCC 30-2001) supplemented with 10% FBS. ATCC CCL-240 cells (promyelocytic peripheral blood leukocytes obtained by leukopheresis from a patient with acute promyelocytic leukemia, abbreviated HL60) were cultured in Iscove’s modified Dulbecco’s medium (IMDM) (ATCC 30-2005) supplemented with 20% FBS. ATCC CCL-243 cells (lymphoblastic cells isolated from the pleural effusion of a patient with chronic myelogenous leukemia in terminal blast crises, abbreviated K562) were cultured in IMDM (ATCC 30-2005) supplemented with 10% FBS (Life Technologies, 16000-077). Stentgen BJ human fibroblasts were cultured in DMEM and F-12 (Life Technologies) supplemented with 10% FBS. All cultures were passaged using 0.05% trypsin supplemented with 0.02% EDTA or using 1× TrypLE Select (Life Technologies). For systems verification tests of fusions were passaged using 0.05% trypsin supplemented with 0.02% EDTA or using 0.05% trypsin supplemented with 0.02% EDTA or using 1× TrypLE Select (Life Technologies).

For experiments in which exogenous spike-in controls were used, the spikes were added to the lysis mix at a 20,000-fold dilution. The PCR thermal protocol was adapted from a recent publication that optimized template-switching chemistry for single-cell mRNA-seq and is outlined in the C1 Single-Cell Auto Prep System protocol. For the population control experiment, we used reagent formulations and workflows exactly as described in the SMARTer Ultra Low RNA Kit user manual (634833, 1 kit for 10 C1s, iFCSs), except that the thermal protocol followed the recommendations outlined in the C1 Single-Cell Auto Prep System user guide (PN 100-7168).

For the population control experiment, we sorted 100K K562 cells into 3.5 µl of Clontech Reaction Buffer containing exogenous spike-in controls using a BD FACSAria III. The 20,000-fold–diluted ERCC spike-in controls were further diluted (9:3,500) in Clontech Reaction Buffer such that an equal mass (rather than an equal concentration) of the spikes was included in the population control reaction. After the sort, cells were frozen at –80 °C overnight before continuing the SMARTer Ultra Low RNA Kit protocol according to the manufacturer’s recommendations.

The cDNA reaction products were quantified using the Quanti-IT PicoGreen dsDNA (double-stranded DNA) Assay Kit (Life Technologies) and high-sensitivity DNA chips (Agilent) and were then diluted to a final concentration of 0.15–0.30 ng/µl using C1 Harvest Reagent. The diluted cDNA reaction products were then converted into mRNA-seq libraries using the Nextera XT DNA Sample Preparation Kit (Illumina, FC-131-1096 and FC-131-1002, 1 kit used for 4 C1, iFCSs and 384 samples) following the manufacturer’s instructions with minor modifications. Specifically, reactions were run at one-quarter of the recommended volume, the tagmentation step was extended to 10 min, and the extension time during the PCR step was increased from 30 to 60 s. After the PCR step, samples were pooled, cleaned twice with 0.9× Agencourt AMPure XP SPRI beads (Beckman Coulter), eluted in Tris + EDTA buffer and quantified using a high-sensitivity DNA chip (Agilent). For high-cover-sequencing, libraries from a subset of captured cells from each source were pooled to reach a target of ten million aligned reads per cell.

Processing the mRNA sequencing data. An index for RNA-Seq by expectation maximization (RSEM) was generated on the basis of the hg19 RefSeq transcriptome downloaded from the UCSC Genome Browser database (23,637 total genes). Read data were aligned directly to this index using RSEM/bowtie2.9,35. FASTQ files from high-cover-sequencing data were downsampled to 11 sequencing depths (100; 500; 1,000; 5,000; 10,000; 50,000; 100,000; 150,000; 200,000; 250,000; and 300,000 reads) using a Python script to randomly select reads, and downsampled results were also aligned to the same index using RSEM/bowtie. Quantification of gene expression levels in TPM for all genes in all samples was performed using RSEM v1.2.4 (ref. 29). Genomic mappings were performed with TopHat v2.0.4 (ref. 36), and the resulting alignments were used to calculate genomic mapping percentages. Raw sequencing read data were aligned directly to the human RNA sequences using Bowtie v2.0, and the percentage of reads aligned to RNA was then calculated as reads aligned to these sequences divided by the total reads. Linear expression data were imported into the Fluidigm SINGUlar Analysis Toolset 2.0 (the R scripts and user guide can be found at http://www.fluidigm.com) and converted into log-space. Transcripts with TPM values less than 1 were dropped from further analysis before log transformation. To identify outlier cells from each chip, we considered a set of genes detected in at least half of the samples,
and samples with median expression values below the 15th percentile for these genes were removed using the identifyOutliers function in the SINGuLAR package. No additional normalization was performed between individual samples. Sequencing results obtained from capture sites with no detectable calcein AM staining that were not flagged as sequencing outliers were retained in the data set. Capture sites containing multiple live cells on the basis of calcein AM staining or brightfield microscopy were removed from further analysis. To assess technical variation during library preparation, we split cDNA from a single cell (GW21+3,1) into two samples, and two independent libraries were prepared with the Nextera XT DNA Sample Preparation Kit (Illumina). The correlation between log2 TPM expression values for technical replicates (0.993) was greater than that between any pair of distinct cells.

**PCA and clustering.** PCA was performed in the Fluidigm SINGuLAR Analysis Toolset 2.0 R package, which calls the princomp R package (http://stat.ethz.ch/R-manual/R-patched/library/stats/html/princomp.html). The 500 top-ranked PCA genes were selected on the basis of the maximum absolute value of each gene loading score in the first three eigenvectors (PC1, PC2 and PC3). To compare sample scores between downsampling-low-coverage data sets with high-coverage mRNA-seq data sets (Fig. 2 and Supplementary Fig. 9), the eigenvectors derived from the high-coverage data were applied to the low-coverage data using the applyPCA function in the SINGuLAR package. Hierarchical clustering of the top 500 PCA genes across 301 cells was also performed in the Fluidigm SINGuLAR package. Genes are clustered on the basis of Pearson correlation. Samples are clustered on the basis of a Euclidian distance matrix with complete linkage. Significance of the cluster assignment shown in Supplementary Figures 9 and 10 was tested using Piculet	extsuperscript{37}, which employs a multiple bootstrap resampling algorithm to calculate the approximately unbiased (AU) probability values for cluster distinctions. We performed the clustering for 50,000 bootstraps.

**Comparison of low- and high-coverage gene expression data.** Pearson’s correlation coefficients were calculated using the log-transformed TPM values for genes shared in both the low- and high-coverage data sets (TPM\textsubscript{low} > 1 and TPM\textsubscript{high} > 1) and also separately for all genes shown in Supplementary Figure 2. In addition, gene transcripts were binned, on the basis of the high-coverage data, into low-expression (1 < TPM\textsubscript{high} < 10), medium-expression (10 < TPM\textsubscript{high} < 100) and high-expression (100 < TPM\textsubscript{high}) bins, and Pearson’s correlation coefficients were again calculated for each of these sub-sets. We assessed the number of dropouts (TPM < 1) that were excluded from the correlation analysis by counting the number of genes that were detected only in the low-coverage data (1 < TPM\textsubscript{low} and TPM\textsubscript{high} < 1) and the number of genes that were detected only in the high-coverage data (1 < TPM\textsubscript{high} and TPM\textsubscript{low} = 1).

**Validation using K562 cells and population mRNA-seq data including spikes.** An additional validation data set was generated using K562 cells with exogenous spike-in controls (Life Technologies, 4456740) delivered in the Fluidigm SINGuLAR package. Exogenous spike-in controls delivered in the Fluidigm SINGuLAR package. The same approach was used to examine the expression of candidate cell type-specific biomarkers shown in Supplementary Figure 11, but samples from the marginal zone and subpial granular layer were also included in the analysis, and results across samples were further averaged for distinct laminae.

**Immunohistochemistry and in situ hybridization.** Timed-pregnant Swiss Webster mice were obtained from Simonsen Laboratories and maintained according to protocols approved by the UCSD Institutional Animal Care and Use Committee. Pregnant dams were deeply anesthetized with inhaled isoflurane and euthanized by cervical dislocation, and two litters were collected. Embryos were decapitated, and dissected brains were fixed in 4% paraformaldehyde overnight. A timed-pregnant ferret (Marshall BioResources) was maintained according to protocols approved by the UCSD Institutional Animal Care and Use Committee. The E35 pregnant dam was deeply anesthetized with ketamine before the administration of inhaled isoflurane. Ovariectomy for fetus collection was performed for embryo collection. Embryos were perfused transcardially with cold PBS and 4% paraformaldehyde. Dissected brains were fixed in 4% paraformaldehyde overnight.

For immunohistochemistry and in situ hybridization, human fetal cortical samples were fixed overnight in 4% paraformaldehyde, cryoprotected in 30% sucrose and embedded in a 1:1 mixture of 30% sucrose and optimal cutting temperature compound (Thermo Scientific). Thin 20-μm cryosections were collected on superfrust slides (VWR) using a Leica CM3050S cryostat. For immunohistochemistry, heat-induced antigen retrieval was performed in 10 mM sodium citrate buffer, pH 9.4. For antibodies to CAMK5, NTS and SYN, we did not perform antigen retrieval. Primary antibodies to ADR2A (1:100, Thermofisher Scientific, PAI-048), CAMK5 (1:100, Novus Biologicals, NB1P1-68097), EGR1 (1:50, Cell Signaling, 14535), C-FOS (1:100, Santa Cruz, SC-8047), CTIP2 (1:500, Abcam, ab18465), NTSM (1:100, R&D Systems, AF1235), phosphorylated vimentin (Ser82) (1:500, MBL International, D095-3), SATB2 (1:500, Sigma, SC81376), SOX2 (1:200, Santa Cruz, SC17320) and SYN, isoform 1 (1:100, a kind gift from K. Davies, University of Oxford) were diluted in blocking buffer containing 10% donkey serum, 0.5% Triton X-100 and 0.02% gelatin. Binding was revealed using an appropriate Alexa Fluor 488 (A21206), Alexa Fluor 546 (A11056) or Alexa Fluor 647 (A31471) fluorophore-conjugated secondary antibody (Life Technologies). Cell nuclei were counterstained using 4′,6-diamidino-2-phenylindole (Life Technologies). Images were collected using a Leica TCS SP5 X confocal microscope and processed using ImageJ or Imaris (Bitplane).

Probes complementary to the target human mRNA used for RNA in situ hybridization were generated specifically for this study, except for EMX2,
which was generated against the mouse sequence and generously provided by A. Simeone (Institute of Genetics and Biophysics, Adriano Buzzati-Traverso). To generate RNA in situ probes, we extracted total RNA from primary human cortical samples aged GW14–GW21 using the RNeasy RNA extraction kit (Qiagen) and reverse transcribed it with the Superscript III First Strand Synthesis System with random hexamers (Life Technologies). Primers specific to the target genes of interest were designed using Primer3 and amplified by PCR using Phusion proofreading DNA polymerase (Thermo Scientific). Specific genes were amplified using the following primers: ANXA2: forward primer, CCA GGA GCT GCA GGA AAT TA, reverse primer, TGT TAG CTG GAA GCA TGG TG (it should be noted that the target ANXA2 mRNA sequence is indistinguishable from a related retrotransposed pseudogene, ANXAP2, and our probe would not distinguish between transcripts from these loci); Clorofl: forward primer, TCC AAG AAG AAG CCG CCT CA, reverse primer, CAG GTA CAG TGG GCT TCC TG; Cks2: forward primer, GCC CTG TCG TTT CAT CTT CT, reverse primer, GCA CTT AAG AGA AAA ACT GAC TGG; CLu: forward primer, CCG AGG CCT CAC TCC TTC TT, reverse primer, GTA TTC CTG CAG CCG GCT TCT CT; DDAH1: forward primer, CCC CTA ACG CTC CCG AAG, reverse primer, TAG CCG TGG TCA CTC ATC TG; Egr1: forward primer, CTG CAC GCT TCT CAG TGT TC, reverse primer, CAT GTC CCT CAC AAT TGC AC; Fos: forward primer, AGC AGT GAG CCT GCT CCT AC, reverse primer, CAG GAA AGG TGG CAA ACT AGG; Gria2: forward primer, TTT GCT GAA TC; Hmg18: forward primer, GCC ATT TTT CAA ACC CTC TTC, reverse primer, CAC CTT GGG CCT TT; Hmg18: forward primer, GGT ACC CAG ACC CTC CAA AA; Hnmt: forward primer, TTT CAG ACT GCC CCG GCA GCT GA; Hear: forward primer, TAT GGA CAG CAT ATT AA; Hes1: forward primer, TTT AGC ACT CCT TCC GGT TG, reverse primer, AAA CAC CTT AGC CGC CTC TG; Hmgb2: forward primer, GCC TT TTT CAA ACC CTC TTC, reverse primer, CAC CTT TGG GAG GCT CCA AAC; Mmp24: forward primer, CTT CAG ACT GCC CCG GCA GCT GA; Nfat: forward primer, TTT CTC GAC CAC CCA CCT AC, reverse primer, AGG AGC ACC TGA TGA TAC GG; Pdzrn3: forward primer, AGC AAC GAG TCT TTC ATT TCG, reverse primer, GCT CTC CGC TCT TTG TTT TC; Pn2: forward primer, CCG AAG GTA TCT GGG GAA AT; reverse primer, TGG ATC CCA TTT GCT GAA TC; Rtn1: forward primer, CCC CTC CCT CCA GTA CCA TA, reverse primer, TGA ATC CAT TAG GAA CTA CAG AGA AA; Scgg5: forward primer, GGT ACC CAG ACC CTC CAA AT, reverse primer, CCA AGG GCT GGA TGA ACT AC; Sparc: forward primer, CTT CAG ACT GCC CCG AGA, reverse primer, CAG CCG GTC CTC ATT CTC AT; Srgap3: forward primer, CCG AGA AGA TGT TCC CCA AC, reverse primer, CCG AGT TAC TAT GGG CCT TT; Stmn2: forward primer, AAT GGA TCA TGC GAT TTC AGG, reverse primer, GCC AAA GCA CAT TTG TAG CA; Tgaln3: forward primer, GGG CTT GAT TGA CAC AGG AG, reverse primer, GAA CTG GGA GAT TGG CCA TCA; Tfpap2c: forward primer, GAC TAC TGT CGA AGG AG; reverse primer, CCA GTA CAG CAT TTG AAG GG, reverse primer, CCT CCA GGA GCT CCA AAC TG; Tpap2c: forward primer, TGA TCT GGT CCT GGG GAA AT, reverse primer, TGG ATC CCA TTT GCT GAA TC; Zeb4: forward primer, CCA AAA TGG TCA CAA AGG AG; reverse primer, AGC AGT TAC TAT GGG CCT TT. Quantification and comparisons of gene expression levels were performed using the ΔΔCt method, and statistical analyses of differences between control and EDTA-treated samples were performed using paired two-tailed Student’s t-test.

Quantitative RT-PCR. RNA extraction and cDNA synthesis were described as above, and quantitative RT-PCR was performed using the Quantitect SYBR Green PCR Mix (Qiagen) in a Roche LightCycler 480 II. The following primer pairs were used in this study to detect specific mRNAs, blinded to the treatment status of each sample: Gapdh; forward primer, GAG TCA ACG GAT TAG TTG GTC GT, reverse primer, TTT ATT TTG GGA GGA TCT CG; Actb; forward primer, GGA CTG GTA TTG. Quantification and comparisons of gene expression levels were performed using the ΔΔCt method, and statistical analyses of differences between control and EDTA-treated samples were performed using paired two-tailed Student’s t-test.

33. Fan, J.B. et al. Highly parallel genome-wide expression analysis of single mammalian cells. PLoS ONE 7, e30794 (2012).