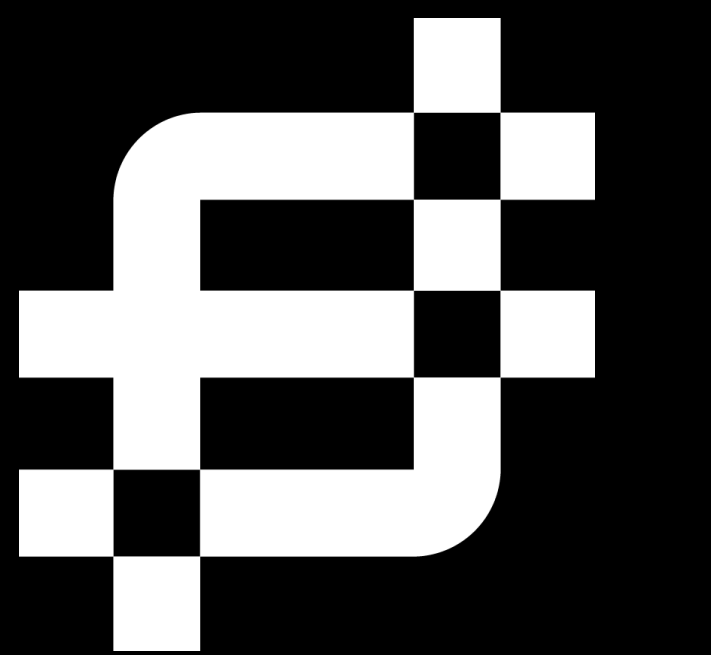


mRNA Sequencing of Single Cells Reveals Vast Heterogeneity in Single-Cell Gene Expression and Provides Insight into the Molecular Mechanisms of Neural Development



J Shuga¹, X Wang¹, AA Leyrat¹, AA Pollen², TJ Nowakowski², JH Lui², N Li¹, L Szpankowski¹, B Fowler¹, P Chen¹, G Sun¹, S Wu¹, B Alvarado¹, J Wang¹, S Weaver¹, MA Unger¹, AR Kriegstein², and JAA West¹

1. Fluidigm Corporation, South San Francisco, CA ,USA
2. Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA, USA
3. Department of Neurology, University of California, San Francisco, San Francisco, CA, USA

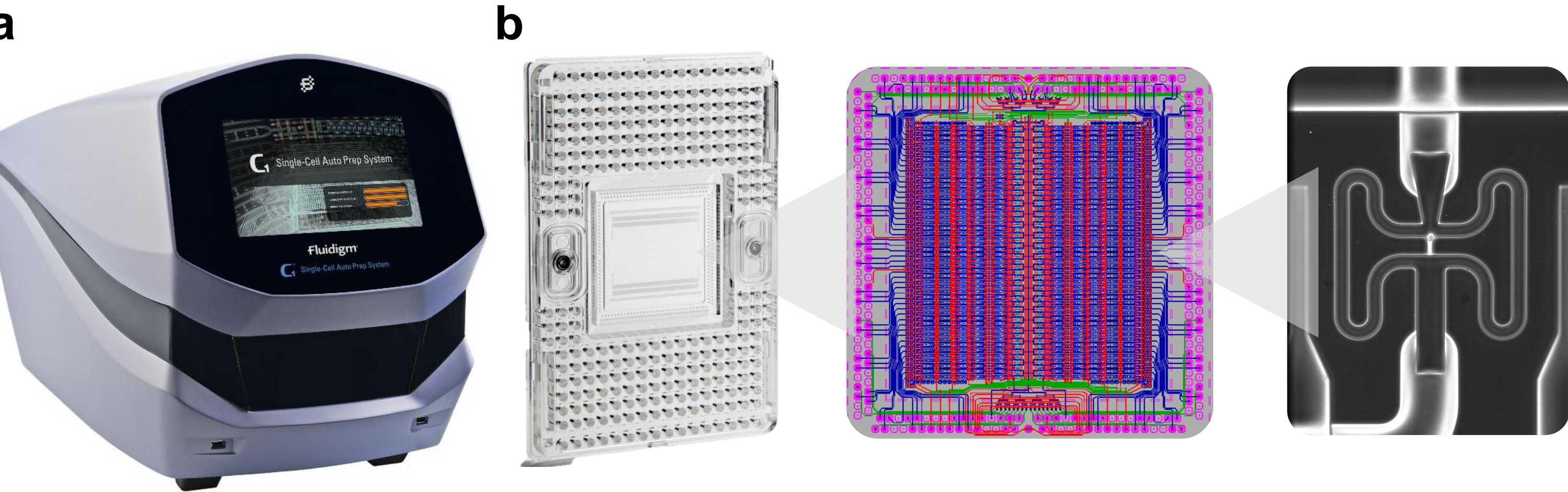
Introduction

The ability to profile gene expression of single cells is essential to advance our understanding of the molecular complexity of living cells, tissues, and organisms. Until recently, high-throughput gene expression profiling methods could not be applied to single cells, and most analyses required enrichment steps for the cell types of interest. By analyzing the transcriptomes of 301 cells from 10 distinct populations, we show that low-coverage (~2.7 x 10⁵ reads per cell) single-cell sequencing is sufficient for unbiased classification of cell identity, for discovery of novel candidate biomarkers, and for detection of candidate effector genes of activated signaling pathways. In a single sample of human fetal cortex, low-coverage sequencing detected radial glia, newborn neurons, and maturing neurons. Our strategy establishes an efficient method for unbiased analysis and comparison of cell populations from heterogeneous tissue by single-cell capture and low-coverage sequencing of many cells.

Validation Dataset

To validate and characterize the method we used External RNA Controls Consortium (ERCC) exogenous spike-in mRNA sequences at loads of ~2.8 x 10⁴ total ERCC spikes per reaction along with single K562 cells (n = 46). We then evaluated detection rates, gene expression levels and variance in the single-cell mRNA seq data. We observed a strong correlation between mean expression levels as determined by mRNA seq and the known ERCC spike-in amounts (Pearson R > 0.96, log-scale). In addition, we used these spike data to assess the sensitivity and technical noise in the single-cell mRNA seq dataset. We found that 5/7 spikes loaded at ~15.9 copies per reaction were detected in all 46 single-cell libraries, but that the other 2/7 spikes loaded at this level (ERCC-00025 and ERCC-00051) were not detected in 5/46 and 2/46 single-cell libraries, respectively. All 26 spikes loaded at higher input amounts, including 5/5 spikes loaded at ~31.8 copies per reaction, were detected in 100% of the single-cell libraries. Seven spikes were loaded at a level of 0.99 (~1) copies per reaction chamber. By comparing the observed detection rate of these seven spikes in our dataset (~33%) to the theoretical fraction of no-copy chambers for one molecule spikes (~63%) we estimate the overall efficiency of detection at ~52%.

C₁TM Single-Cell Auto Prep System and C₁TM Single-Cell Auto Prep Integrated Fluidic Circuit



C₁TM Single-Cell Auto Prep System and C₁TM Single-Cell Auto Prep Integrated Fluidic Circuit (IFC).

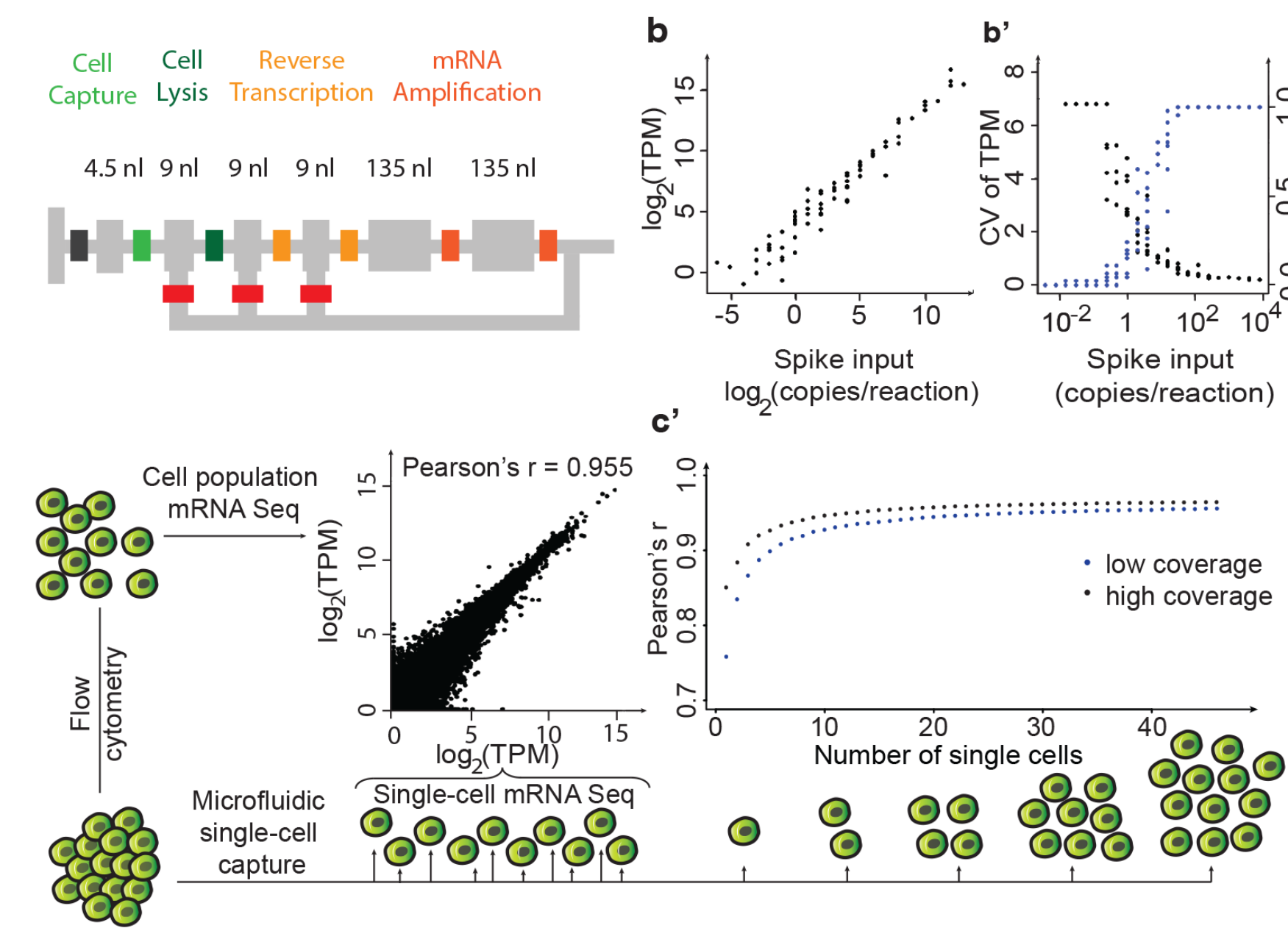
(a) The C₁TM System is used for pneumatic and thermal control of the C₁TM IFC. (b) The C₁TM IFC is detailed. **Left Panel:** the complete IFC with carrier; reagents and cells are loaded into dedicated carrier wells and reaction products are exported into other dedicated carrier wells. **Middle Panel:** CAD diagram of the C₁TM IFC. Connections between the microfluidic lines in the polydimethylsiloxane (PDMS) and the carrier are shown in pink, control lines are shown in red, fluidic lines for preparatory reactions are shown in blue, and lines connecting control lines are shown in green. **Right Panel:** A single cell captured in one of the 96 capture sites.

Cell Types, Sequencing Depths, and Alignment Rates

| Cell Type | Abbreviation | C ₁ TM IFC | Origin | Usage | Cell type | Number of Cells | Low coverage | | | | High coverage | | | | Read depth ratio (HiSeq/MiSeq*) | |
|----------------------------------|--------------|--|-----------|------------|------------------|-----------------|-----------------|-----------------------------|-------------------------------|------------------------|-----------------|-----------------------------|-------------------------------|------------------------|---------------------------------|--------|
| | | | | | | | MiSeq* run type | MiSeq* total reads (B1-B42) | MiSeq* % mapped to the genome | MiSeq* % mapped to RNA | HiSeq* run type | HiSeq* total reads (B1-B42) | HiSeq* % mapped to the genome | HiSeq* % mapped to RNA | | |
| CHL-2338 | 2338 | Medium (10–17 µm) C ₁ TM IFC for mRNA Seq | Cell Line | mean study | epithelial | 22 | PE 30 | 7,500,000 | 1.66E+04 | 88.28% | 1.55% | PE 100 | 1,498,407 | 3,54E+06 | 56.48% | 2.07% |
| CHL-2339 | 2339 | Medium (10–17 µm) C ₁ TM IFC for mRNA Seq | Cell Line | mean study | lymphoblastoid | 17 | PE 30 | 7,610,000 | 1.28E+04 | 87.25% | 1.70% | PE 100 | 1,610,007 | 3,81E+06 | 44.44% | 3.03% |
| PC92 (rat 1) | PC92 | Large (12–25 µm) C ₁ TM IFC for mRNA Seq | Cell Line | mean study | epithelial | 42 | PE 100 | 7,715,000 | 7.89E+04 | 86.05% | 1.17% | PE 100 | 3,35E+06 | 5,55E+06 | 81.82% | 1.00% |
| BJ | BJ | Large (12–25 µm) C ₁ TM IFC for mRNA Seq | Cell Line | mean study | dermal | 37 | PE 25 | 2,02E+05 | 1.17E+05 | 83.26% | 8.31% | PE 100 | 3,54E+06 | 2,02E+06 | 53.23% | 11.32% |
| K562 | K460 | Small (5–10 µm) C ₁ TM IFC for mRNA Seq (46-Seq) | Cell Line | mean study | myeloid | 64 | PE 100 (9) | 2,04E+05 | 5.66E+04 | 42.57% | 5.99% | PE 100 | 3,38E+06 | 2,13E+06 | 47.04% | 9.51% |
| hiPSC | hiPSC | Medium (10–17 µm) C ₁ TM IFC for mRNA Seq | Cell Line | mean study | undifferentiated | 34 | PE 30 | 3,54E+05 | 5.25E+04 | 86.04% | 0.64% | PE 100 | 3,35E+06 | 3,34E+06 | 88.88% | 1.33% |
| lateral ventricle | Kera | Large (12–25 µm) C ₁ TM IFC for mRNA Seq | Primary | mean study | epithelial | 40 | PE 25 | 2,49E+05 | 1.05E+05 | 90.02% | 0.94% | PE 100 | 3,45E+06 | 2,81E+06 | 70.15% | 8.22% |
| total cortex (GW21-2) | GW21-2 | Medium (10–17 µm) C ₁ TM IFC for mRNA Seq* | Primary | mean study | neural | 16 | PE 100 | 3,55E+05 | 7.61E+04 | 80.01% | 0.61% | PE 100 | 3,39E+07 | 2,81E+06 | 45.46% | 1.00% |
| total cortex (GW21-3) | GW21-3 | Medium (10–17 µm) C ₁ TM IFC for mRNA Seq* | Primary | mean study | neural | 8 | PE 100 | 7,55E+05 | 1.32E+05 | 54.74% | 2.37% | PE 100 | 3,48E+07 | 3,98E+06 | 44.92% | 1.40% |
| total cortex (GW21-4) | GW21-4 | Medium (10–17 µm) C ₁ TM IFC for mRNA Seq* | Primary | mean study | neural | 15 | PE 30 | 2,02E+05 | 3.87E+04 | 88.08% | 1.03% | PE 100 | 3,35E+07 | 3,98E+06 | 44.92% | 1.40% |
| total cortex (GW21-5) | GW21-5 | Small (5–10 µm) C ₁ TM IFC for mRNA Seq | Primary | mean study | neural | 35 | PE 100 | 3,39E+05 | 1.31E+05 | 88.77% | 1.09% | PE 100 | 3,45E+07 | 3,85E+06 | 89.12% | 9.40% |
| all single cells (mean study) | | Small (5–10 µm) C ₁ TM IFC for mRNA Seq | na | mean study | na | 301 | na | 2,70E+05 | 1.46E+05 | 86.55% | 4.77% | na | 8,92E+06 | 5,86E+06 | 66.16% | 10.80% |
| K562 (rat 1) | PC92 | Medium (10–17 µm) C ₁ TM IFC for mRNA Seq | Cell Line | validation | epithelial | 46 | PE 100 | 3,70E+05 | 5.62E+04 | 73.34% | 3.04% | PE 125 | 3,71E+06 | 2,79E+06 | 39.47% | 1.32% |
| mean reaction time (on chip MFC) | | Small (5–10 µm) C ₁ TM IFC for mRNA Seq | na | validation | na | 1 | PE 100 | 2,36E+05 | na | 0.00% | na | PE 125 | 3,08E+06 | na | 0.00% | na |
| 1000 population (on chip MFC) | | na | Cell Line | validation | epithelial | 1 | PE 100 | 4,28E+04 | na | 0.00% | na | PE 125 | 3,91E+06 | na | 45.38% | na |

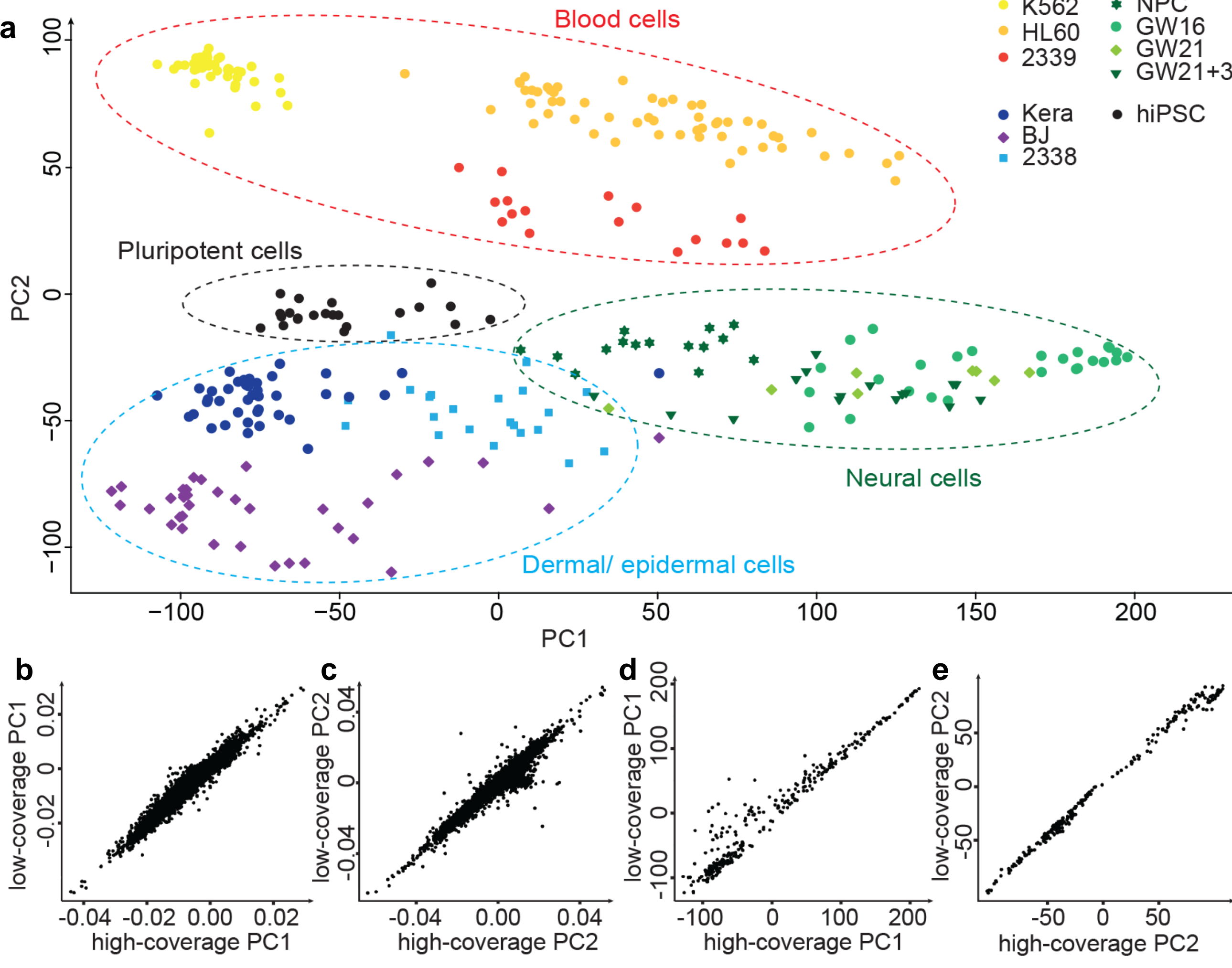
* Note: the Medium (10–17 µm) C₁TM IFC was used for these cells because the optimal Small (5–10 µm) C₁TM IFC was not available at the time.

Overview and Validation



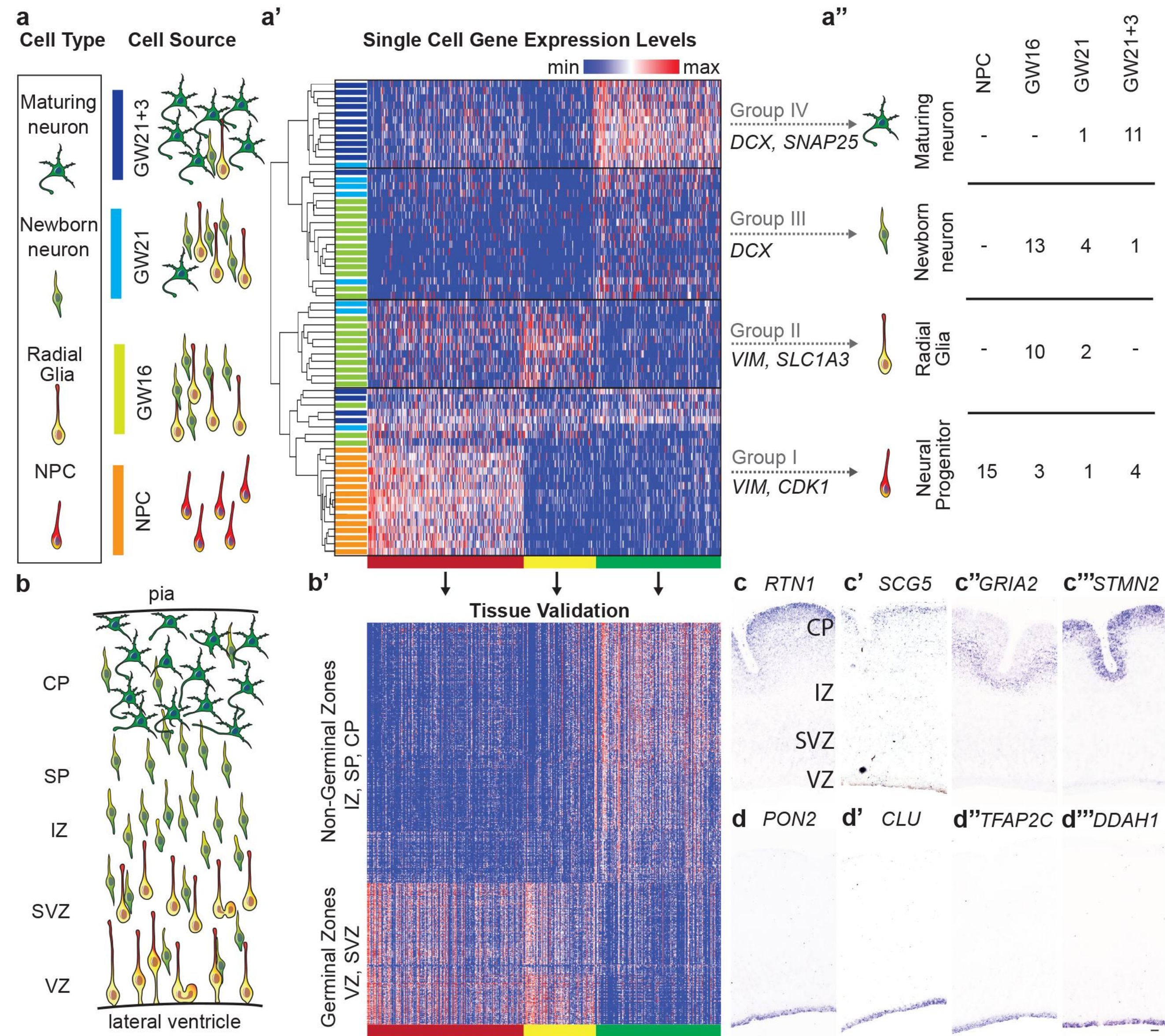
Overview and Validation (a) Schematic outline of the C₁TM Single-Cell Auto Prep Integrated Fluidic Circuit (IFC). Processing of individual cell samples takes place in nanoliter reaction volumes. (b) Transcript per million (TPM) values versus input spike copy number across 46 K562 cells sequenced at 1.7 x 10⁵ read pairs per cell with ERCC RNA Spike-In Mix 1 at 2.8 x 10⁴ copies/reaction (r = 0.968). (b') The fraction of positive reactions (ERCC transcripts detected above one TPM) in single cells and the coefficient of variation for ERCC levels are both plotted versus the spike input amounts. (c) The average TPM values from single-cell sequencing of 46 K562 cells correlate strongly with TPM values from a population of 100 K562 cells isolated by flow cytometry. (c') The correlation between individual K562 cells and the population improves with diminishing returns as additional single-cell results are combined.

Low-Coverage Single-Cell mRNA Seq is Sufficient to Detect Genes Contributing to Cell Identity



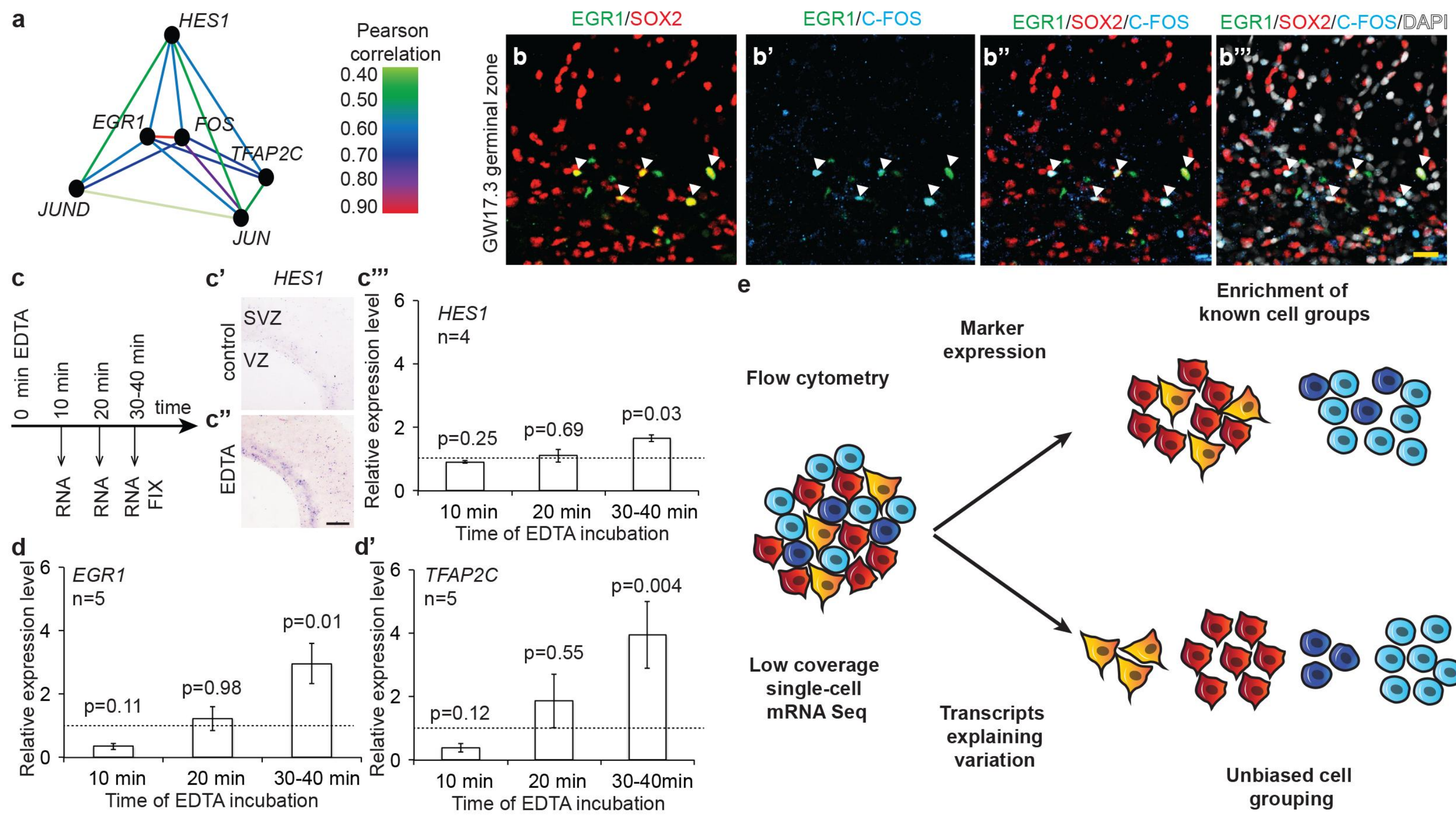
Low-Coverage Single-Cell mRNA Seq is Sufficient to Detect Genes Contributing to Cell Identity (a) Distinct groups of cells corresponding to pluripotent, blood, skin, and neural cells can be identified by PCA of 301 cells sequenced at low coverage. (b–c) Results from low- and high-coverage datasets correlate strongly for the eigenvectors defining PC1 (b) and PC2 (c). (d–e) Similarly, sample scores derived from low- and high-coverage data and calculated using eigenvectors from high-coverage data correlate strongly across all 301 cells for PC1 (d) and PC2 (e).

Low-Coverage Single-Cell mRNA Seq Distinguishes Diverse Neural Cell Types



Low-Coverage Single-Cell mRNA Seq Distinguishes Diverse Neural Cell Types (a) Schematic of cell types and sources selected to represent stages of neuronal differentiation. Cultured NPCs represent early undifferentiated stages, while primary fetal cortical samples are expected to contain radial glia, newborn, and maturing neurons. (a') Unsupervised hierarchical clustering of 65 single cells across the top 500 PCA genes identified four major groups of cells (I–IV) and unsupervised hierarchical clustering based on the Pearson correlation method identifies three clusters of genes (red, yellow, green). (a'') Table showing number of cells of specific types captured from each source. (b–b') Schematic of the distribution of cell types (b) and heatmap of gene expression values for PCA genes (b') in the developing cortex. Genes belonging to the red and yellow gene clusters show high expression in the ventricular and subventricular zones (VZ, SVZ), while genes belonging to the green cluster show high expression in the intermediate zone, (IZ), subplate (SP), and cortical plate (CP). (c–c'') *In situ* hybridization for representative genes belonging to neuronal (c–c'') and radial glial (d–d'') clusters in GW 14.5 human cortical sections. Scale bar: 100 µm.

Single-Cell mRNA Seq Identifies Candidate Response Genes in Activated Signaling Pathways



Single-Cell mRNA Seq Identifies Candidate Response Genes in Activated Signaling Pathways

(a) Schematic representation of correlation coefficients between gene expression values across single radial glia cells. (b–b'') Immunohistochemical detection of coordinated expression of EGR1 and FOS in a subset of SOX2-positive radial glia in the SVZ of GW 17.3 human fetal cortex. Scale bar: 25µm. (c) To activate Notch signaling, human fetal cortical slices were incubated with EDTA for 10, 20, and 30–40 minutes. (c'–c'') Activation of Notch signaling was confirmed in EDTA-treated slices by detecting higher levels of *HES1* by qRT-PCR (c') and by *in situ* hybridization in control (c'') and experimental slices (c'''). (d–d'') Increased levels of (d) *EGR1* and (d') *TFAP2C* mRNA were detected in EDTA-treated human fetal cortical slices by qRT-PCR. (e) Low-coverage mRNA seq of single cells allows for unbiased detection of cell types. Flow cytometry uses established staining characteristics to enrich for known cell types in heterogeneous samples. In contrast, low-coverage single-cell mRNA seq identifies the major genes explaining variation across single cells allowing for unbiased discovery and further analysis of distinct cell populations and states.

Conclusion

- Low-coverage sequencing of single cells at depths currently possible by the MiSeq® System using barcoded libraries from up to 96 cells is sufficient for unbiased detection of diverse cell types and states in a heterogeneous sample.
- We were able to identify candidate cell type-specific biomarkers for the main cell types found in a mid-gestation human cortex and to identify radial glia in different states of signaling pathway activation using low-coverage mRNA seq.
- Although the observed level of a given transcript in a single cell can vary due to 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 based on shared signatures of gene expression.

References

- Jiang, L. et al. “Synthetic spike-in standards for RNA-seq experiments.” *Genome research* 21 (2011): 1543-1551.
- Brennecke, P. et al. “Accounting for technical noise in single-cell RNA-seq experiments.” *Nature methods* 10 (2013): 1093-1095.
- Wu, A.R. et al. “Quantitative assessment of single-cell RNA-sequencing methods.” *Nature methods* 11 (2014): 41-46.
- Li, B. and Dewey, C.N. “RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.” *BMC bioinformatics* 12 (2011): 323.

Acknowledgement

AAP is supported by a Damon Runyon Cancer Research Foundation postdoctoral fellowship (DRG 2013). LS, NL, and MAU are partially supported by CIRM Tools and Technologies II grant RT2-02052 to MAU, and the generation and maintenance of the hiPS and NPC cells was also supported by RT2-02052 to MAU. We thank Shuyuan Yao and Sherman Ku at Allen Institute for assistance in the NPC induction protocol and Guangwen Wang at Stanford University for providing hiPS cells. We thank the staff at the San Francisco General Hospital for providing access to donated fetal tissue. This research was also supported by NINDS awards R01NS075998 and R01NS072630 to ARK.



Fluidigm Corporation

7000 Shoreline Court, Suite 100 • South San Francisco, CA 94080 Toll-free: 1.866.FLUIDLINE www.fluidigm.com/singlecellgenomics
© Fluidigm Corporation. All rights reserved. Fluidigm, the Fluidigm logo, C₁, BioMark HD, and Dynamic Array, are trademarks or registered trademarks of Fluidigm Corporation