A portable and cost-effective microfluidic system for massively parallel single-cell transcriptome profiling

Chuanyu Liu[1,2,10], Tao Wu[1,2,10], Fei Fan[1,2,10], Ya Liu[1,2,10], Liang Wu[1,2,3,10], Michael Junkin[1,4], Zhifeng Wang[1,2], Yeya Yu[1,2,5], Weimao Wang[1,2], Wenbo Wei[1,2], Yue Yuan[1,2,3], Mingyue Wang[1,2,3], Mengnan Cheng[1,2,3], Xiaoyu Wei[1,2,3], Jiangshan Xu[1,2,3], Quan Shi[1,2,6], Shiping Liu[1,2], Ao Chen[1,2], Ou Wang[1,2,6], Ming Ni[1,2], Wenwei Zhang[1,2], Zhouchun Shang[1,2,7], Yiwei Lai[8], Pengcheng Guo[9], Carl Ward[8], Giacomo Volpe[8], Lei Wang[1,2], Huan Zheng[1,2], Yang Liu[1,2,3], Brock A. Peters[1,2,4,7], Jody Beecher[4,7], Yongwei Zhang[1,4], Miguel A. Esteban[1,8], Yong Hou[1,2], Xun Xu[1,2], I-Jane Chen[1,4*] & Longqi Liu[1,2*]

[1]BGI-Shenzhen, Shenzhen 518083, China. [2]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China. [3]BGI Education Center, University of Chinese Academy of Sciences, Shenzhen 518083, China. [4]Complete Genomics Inc., 2904 Orchard Parkway, San Jose, California 95134, USA. [5]BGI College, Zhengzhou University, Zhengzhou 450000, China. [6]Laboratory of Genomics and Molecular Biomedicine, Department of Biology, University of Copenhagen, Copenhagen 2200, Denmark. [7]MGI, BGI-Shenzhen, Shenzhen 518083, China. [8]Laboratory of RNA, Chromatin, and Human Disease, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Science, Guangzhou 510530, China. [9]College of Veterinary Medicine, Jilin University, Changchun 130062, China. [10]These authors contributed equally to this work.

[*]Correspondence should be addressed to IJ.C. (i-janechen@completegenomics.com) or to L.L. (liulongqi@genomics.cn).

Abstract

Single-cell technologies are becoming increasingly widespread and have been revolutionizing our understanding of cell identity, state, diversity and function. However, current platforms can be slow to apply to large-scale studies and resource-limited clinical arenas due to a variety of reasons including cost, infrastructure, sample quality and requirements. Here we report DNBelab C4 (C4), a negative pressure orchestrated, portable and cost-effective device that enables high-throughput single-cell transcriptional profiling. C4 system can efficiently allow discrimination of species-

specific cells at high resolution and dissect tissue heterogeneity in different organs, such as murine lung and cerebral cortex. Finally, we show that the DNBelab C4 system is comparable to existing platforms but has huge benefits in cost and portability and, as such, it will be of great interest for the wider scientific community.

Introduction

The emergence of single-cell sequencing technologies has offered great promises in the interrogation of tissue heterogeneity that could not previously be resolved from measuring the average gene expression of a bulk cell population [1]. During the past 10 years, single-cell genomics technologies have evolved rapidly at scale and power, enabling the profiling of thousands to tens of thousands single-cells per experiment [2]. These efforts have enabled a comprehensive characterization of tissue heterogeneity, developmental trajectories, cellular reprogramming and human diseases at an unprecedented resolution.

Current high-throughput single-cell technologies apply droplet and micro-well based methods, which enable partition and barcoding of single cells inside nano-liter reactors. Micro-particles coated with barcoded oligos are employed to capture mRNA or DNA molecules from each cell [3-5]. These methods, in comparison to low-throughput plate-based methods, are robust in cell type classification since profiling higher number of cells can reduce the negative impact from technical and intrinsic noise [6]. However, the effectiveness of these methods is limited by lengthy approaches and by the use of expensive commercial instruments and reagents. Specifically, clinical samples require to be transported to different laboratories for library preparation because current microfluidic single-cell platforms are not portable. To address these short-comings, we have developed DNBelab C4, a hand-held microfluidic device to perform cell separation before high-throughput single-cell transcriptional profiling. C4 is a vacuum driven and user-friendly droplet system which offers a cost-effective and portable single-cell technology for both basic research and clinical purposes. C4 is also an extensible platform that can be further implemented for a variety of high-throughput omics technologies, such as scATAC-seq and scChIP-seq. Here we show that the DNBelab C4 system is comparable to existing platforms but has benefits in cost and portability.

Results

A hand-held and cost-effective microfluidic system for single-cell profiling

We have previously developed a hand-held, power-free microfluidic device that can stably generate mono-dispersion sub-nanoliter size droplets [7]. To further apply this system to single-cell profiling, we have optimized both the structure of microfluidic chip and parameters of negative pressure to ensure high compability with single-cell RNA sequecing (scRNA-seq) related reagents. To this end, we developed C4 system (Fig. 1a), which is a low cost device composed of a syringe, a microfluidic chip, and a station. C4 is a fully manual system and does not have any electronic parts. The syringe is connected to the chip via connection tubings and placed on the station. After all reagents and samples are loaded in the inlet reservoirs, the plunger of the syringe is pulled to generate negative pressure that drives the reagents flowing into the fluidic passage to form droplets. This single source negative pressure ensures the flows of all reagents occur simultaneously without any fluctuation or lag in flow rates. The air volume in the syringe before and after plunger is pulled determines the degree of vacuum, which follows ideal gas laws. The station holds the plunger in place after it is pulled so the vacuum is uniformly maintained. To evaluate the uniformity of droplets size, more than 1.5 million droplets each run were generated independently from 4 users using 14 chips; we observed an average size of 55.9 μm with a standard deviation of 2.4 μm (Fig 1b, n=14,000), revealing high reproducibility and stability of the negative pressure-driven system in generating large numbers of droplets.

Massively parallel single-cell RNA-seq using C4 system

We next developed an entire workflow for scRNA-seq using C4 system (Fig 1a). Cells, functionalized beads and lysis buffer were encapsulated into emulsion droplets, where each cell was lysed and mRNA transcripts were captured by bead-linked single-strand oligonucleotides consisting of sequencing adaptor, cell barcode, unique molecular identifier (UMI) and oligo-dT. Emulsion droplets were then transferred into a membrane filter (pore size: 8 μm), where emulsion was broken and filtered by either negative pressure or centrifugation. Reverse transcription of mRNA transcripts was conducted on pooled beads, resulting in cDNA molecules containing both cell barcode and UMIs. Subsequently, the cDNA molecules were amplified and sheared for short-read based sequencing library preparation.

To assess the technical performance of C4 system in single-cell transcriptional sequencing, we profiled cells from a mixture of 50% human (HEK293T) and 50% mouse (NIH3T3) cells at the concentration of 1,000 cells/ul. A total of 548 million quality-filtered reads were obtained and then assigned to each cell barcode and were mapped to reference genomes. 61% of the reads mapped to exonic regions (Supplementary Figure 1a), revealing high library quality. One library contained an estimated 2,897 cells based on the distribution of the number of UMIs per barcode (Fig. 1c). A discrimitation between mouse and human UMI revealed only a small percentage (roughly 2%) of cells that contain high fraction of both human and mouse reads (Fig. 1d). In addition, an average of 43,719 and 39,767 UMI-tagged transcripts (assigned to 6,618 and 5,847 genes) were detected in human and mouse cells, respectively (Fig. 1e). These data demonstrate that C4 system shows minimal collision rate and high gene capture efficiency. We performed an independent cross-species experiment and observed similar results (Supplementary Figure 1b-d) and strong correlations between replicates (Pearson correlation = 0.998, Supplementary Figure 2), thus strenghening the effectiveness of our approach.

Unsupervised taxonomy of cellular states

To examine the ability of C4 system to resolve cell populations in complex primary tissues, we isolated nuclei, an extensively adopted type of input sample [8-9], from mouse lung tissue to generate scRNA-seq libraries. After sequencing and read alignment, we collected a total of 3,241 cells based on the UMI distribution (data not shown). Unsupervised clustering identified 9 distinct subpopulations, suggesting high heterogeneity within this tissue (Fig. 2a). Differential gene expression analysis identified genes that were specifically expressed in each subpopulation (Fig. 2b), resulting in the characterization of cell types including epithelial cells such as alveolar type 1 (AT1, expressing *Ager*) and type 2 (AT2, expressing *Sftpb*) cells, endothelial cells (*Pecam1*), fibroblast cells (*Igfbp5* and *Fn1*), and immune cells (*Ptprc*) (Fig. 2a, Supplementary Figure 3). Interestingly, 4 sub-types of cells can be further identified within immune cells including 3 types of macrophages and one type of B lymphocytes, while endothelial cell population can also be further sub-classified into those of vascular and lymphocyte of origin, respectively (Fig. 2a). These data are consistent with previous single-cell analysis of lung tissue from newborn mice [10], suggesting

that C4 system could be faithfully applied into dissecting complex cell populations at high resolution.

Comparative analysis of C4 and existing platforms

Two recent benchmarking studies by Ding *et al.* and Mereu *et al*. [11, 12] have systematically and comprehensively compared existing single-cell sequencing techniques, providing invaluable resource for users to make informed choices. To compare the technical performance between C4 and these single-cell platforms, we firstly analyzed the data from HEK293T/NIH3T3 mixture cells in replicates and the data generated from the same cell types in the study by Ding *et al*. [11]. We observed a multiplet rate of 2-2.5% in C4 system, which is quite comparable and acceptable in comparison with other high-througput platforms (Fig. 3a). To evaluate the gene detection sensitivity, we first performed down-sampling analysis and found that a median number of over 6,000 genes was identified at 100K reads per cell, which outperforms other platforms (Supplementary Figure 4). Supporting this, we further calculated the number of UMIs and genes of all single cells from each platform. As expected, the plate-based method, Smart-seq2 and CEL-Seq2 showed the highest capture sensitivity (Fig.3b, c). Interestingly, when comparing with high-throughput methods, C4 detected over 6,000 genes in both mouse and human cell lines, which is much higher than all other platforms (Fig. 3b, c), possibly owing this to the higher number of capture oligos (theoretical number: 10^7) on our functionalized micro-particles.

One of the key biological information obtained from single-cell RNA-seq is the identification and recovery of different cell types from heterogeneus populations. In their report, Ding *et al*. [11] compared the ability of different platforms to resolve cellular heterogeneity in mouse cortex and human PBMCs. To further compare this facet of C4 with current platforms, we profiled a total of 6,473 single cells from the nuclei of mouse cortex, which was disassociated and extracted according to the protocol by Ding *et al*. [11]. As expected, we observed slightly higher number of genes in C4 system (Fig. 4a) than all other high-throughput platforms. We next performed unsupervised clustering for the nuclei data based on gene expression matrices, resulting in 9 subpopulations. Marker gene based annotation revealed recovery of known cell types in mouse cortex in all methods, e.g., excitatory or inhibitory neurons,

astrocytes and microglia (Fig. 4b). We further calculated the percentage of each subpopulation and observed successful capture of all major cell types (Fig. 4c). C4 and other droplet-based methods can recover all cell types, whereas the combinatorial idexing based method, sci-RNA-seq, can not detect cell types such as pericytes, suggesting potential bias across different strategies. Taken together, these results strongly indicated that C4 presents high sensitivity in gene detection and comparable recovery ability with most of the current platforms.

Discussion

In summary, we have developed a portable, affordable and user friendly microfludic system that enables scalable single-cell transcriptomics profiling. The portability feature of C4 system is critical since many samples should be freshly prepared and immediately loaded to ensure high data quality. C4 is applicable to different tissues and sample types including viable cells and nuclei, which is important since only nuclei can be isolated from many archived samples including frozen tissues or dissociated cells. In addition, we have systematically compared C4 and other 7 single cell platforms, demonstrating that C4 shows high sensitivity on mRNA molecule detection and robust ability for recovery of cellular heterogeneity in different tissues, such as cortex. Apart from good technical performance, C4 system is of extremely low-cost, making it possible for every life science lab to do single-cell studies without purchasing expensive instruments.

One limitation for current high throughput single-cell platforms (such as droplet, microwell or combinatorial indexing based methods) is that only 3’ or 5’ end exonic read can be measured on short-read based sequencing platforms. The tag information can only be used for expression quantification but does not allow transcriptomic analyses such as alternative splicing and structural variation in a high-throughput manner [1]. However, full-length cDNA library can be obtained from these platforms including C4 system. One potential approach to achieve full-length cDNA sequencing is to combine long read sequencing techniques such as long fragment read (LFR) technology [13], which has been widely used for DNA sequencing. Moreover, C4 could be an extendable platform for efficient profiling of other omics layers, such as chromatin accessibility or simultaneous profiling of multi-omics at single-cell resolution, with specific modifications to oligo on the micro-particles. Overall, C4 is a promising

candidate single-cell system for high-throughput multimodal study of single cells, paving the way to a thorough assessment of cellular heterogeneity in a variety of basic research and clinical applications.

Methods

Methods will be released when this manuscript is officially published.

References

1. Chen X, Teichmann SA and Meyer KB. From tissues to cell types and back: single-cell gene expression analysis of tissue architecture. Annual Review of Biomedical Data Science. 2018;1:29-51.
2. Svensson V, Vento-Tormo R and Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. Nature protocols. 2018;13 4:599.
3. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell. 2015;161 5:1202-14.
4. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161 5:1187-201.
5. Gierahn TM, Wadsworth II MH, Hughes TK, Bryson BD, Butler A, Satija R, et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nature methods. 2017;14 4:395.
6. Luecken MD and Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Molecular systems biology. 2019;15 6.
7. Chen I-J, Wu T and Hu S. A hand-held, power-free microfluidic device for monodisperse droplet generation. MethodsX. 2018;5:984-90.
8. Habib N, Avraham-Davidi I, Basu A, Burks T, Shekhar K, Hofree M, et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nature methods. 2017;14 10:955.
9. Bakken TE, Hodge RD, Miller JA, Yao Z, Nguyen TN, Aevermann B, et al. Single-nucleus and single-cell transcriptomes compared in matched cortical cell types. PloS one. 2018;13 12:e0209648.
10. Guo M, Du Y, Gokey JJ, Ray S, Bell SM, Adam M, et al. Single cell RNA analysis identifies cellular heterogeneity and adaptive responses of the lung at birth. Nature communications. 2019;10 1:37.
11. Ding J, Adiconis X, Simmons SK, Kowalczyk MS, Hession CC, Marjanovic ND, et al. Systematic comparative analysis of single cell RNA-sequencing methods. BioRxiv. 2019:632216.
12. Mereu E, Lafzi A, Moutinho C, Ziegenhain C, MacCarthy DJ, Alvarez A, et al. Benchmarking Single-Cell RNA Sequencing Protocols for Cell Atlas Projects. BioRxiv. 2019:630087.
13. Wang O, Chin R, Cheng X, Wu MKY, Mao Q, Tang J, et al. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. Genome research. 2019;29 5:798-808.

Acknowledgements

We would like to acknowledge the ongoing contributions and support of all BGI-Shenzhen and Complete Genomics employees. This work was supported by the National Natural Science Foundation of China (No. 31900466 and No. 31900582), the Science, Technology and Innovation Commission of Shenzhen Municipality (No. JCYJ20170303151334808 and No. JSGG20170412153009953) and the Strategic Priority Research Program of the Chinese Academy of Sciences (No. XDA16030502).

Author contributions

L.L. and IJ.C. conceived the idea. C.L. and L.L. designed the experiments. IJ.C., T.W. and M.J. designed the microfluidic chip and instrument. C.L. and Z.W. designed the reagents. F.F., O.W., B.A.P. and W.W. designed and optimized the micro-particles. Y.L., L.W. and W.Wei designed the filter. L.Wu. and L.L. analyzed the data. C.L., Z.W. and Y.Y. conducted experiments. Y.Yuan., M.W., M.C., J.X., Y.Lai, P.G. and H.Z. assisted with the experiments. Q.S., X.W., S.L. and Y.Liu assisted with the data analyses. L.L. and IJ.C. wrote the manuscript. M.A.E., C.W. and G.V. revised the manuscript. X.X., Y.H., Y.Z., J.B., A.C., Z.S., M.N., W.Z. and M.A.E. provided helpful comments on this study. All authors reviewed and approved the final manuscript.

Competing interests

Employees of BGI and Complete Genomics have stock holdings in BGI.

Figure legends

Figure 1. Massively parallel single-cell transcriptome profiling using DNBelab C4 system. (**a**) Schematic representation of the workflow for scRNA-seq using DNBelab C4 system. (**b**) Upper panel: an image of droplets generated by C4 system. Bottom panel: distribution of the size of droplets generated by different users and from different chips. (**c**) An image showing droplets contaning single bead, single cell or both of them. (**d**) Number of UMIs captured in each cell barcode. The cell barcodes are ordered by the UMIs number. (**e**) Scatter plot of human and mouse UMI counts detected in a mixture of 293T and NIH3T3 cells. Cell barcodes containing human reads are colored in red and termed ‘Human’; cell barcodes with mouse reads are colored in green and termed ‘Mouse’; and cell barcodes with significant human reads are coloured in light grey and termed ‘Multiplet’. The percentage of multiplets detected

in indicated in the panel (2%). (**f**) Left panel: number of UMIs per cell for the human and mouse cells. Right panel: number of genes per cell for the human and mouse cells.

Figure 2. DNBelab C4 can recover diverse cell types from heterogeneous samples. (**a**) Unsupervised clustering identified 14 distinct subpopulations in mouse lung, which were then assigned to cell types based on gene expression, including 4 immune subtypes and 2 endothelial subtypes. (**b**) Heatmap showing differential gene expression analysis result that specifically expressed in each subpopulation.

Figure 3. Comparative analysis of DNBelab C4 and existing platforms. (**a**) Assessment of multiplet frequency. Cells were ordered based on the number of UMIs. For a given x value, the plot shows the percent of the top x cells that are multiplets. (**b**) Violin plots showing the number of UMIs per cell in each method. (**c**) Violin plot showing the number of genes per cell in each method. For (**b**) and (**c**), the top panel shows the results for human cells, and the bottom panel shows the results for mouse cells. Boxplots denote the medians (labelled on the right) and the interquartile ranges (IQRs).

Figure 4. Comparative analysis of C4 and existing systems in the recovery of different cell types within mouse brain cortex. (**a**) Left panel: Violin plot showing the number of UMIs per cell from mouse cortex scRNA-seq data. Right panel: Violin plot showing the number of genes per cell in data generated by each method. (**b**) Cell clustering and cell type annotation for data generated by each method. (**c**) Proportion of the indicated cell types detected from the data by each method.

Supplementary Figure 1. Quality metrics of C4 generated scRNA-seq. (**a**) Percentages of reads mapped to the indicated regions of the human genome. (**b**) Number of UMIs captured in each cell barcode. The cell barcodes are ordered by the UMIs number. (**c**) Scatter plot of human and mouse UMI counts detected in a mixture of 293T and NIH3T3 cells. The percentage of multiplets detected in indicated in the panel (2.6%). (**d**) Left panel: number of UMIs per cell for the human and mouse cells. Right panel: number of genes per cell for the human and mouse cells.

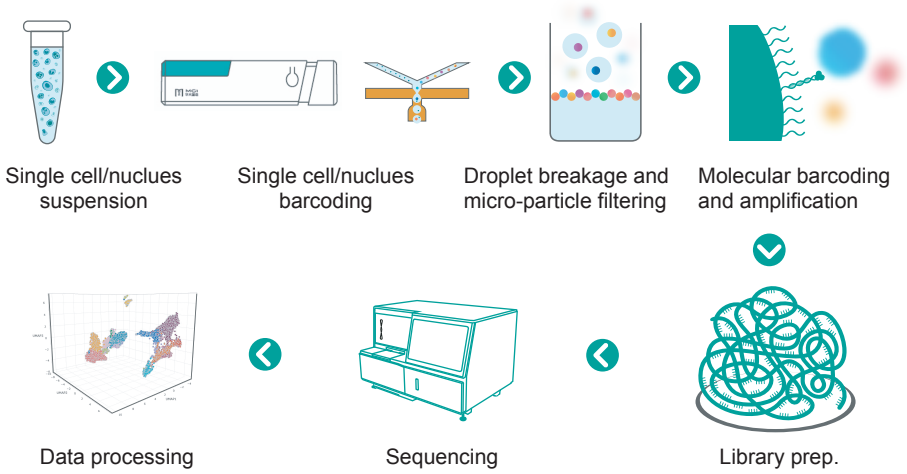
Supplementary Figure 2. Correlation of pseudo-bulk expression profiles. The Pearson correlation coefficients between pseudo-bulk profiles of (**a**) human and (**b**) mouse cells generated from DNBelab C4 system. For each replicate, we compared the log2 transcripts per 10K (log2 [TP10K+1]) for pseudo-bulk from scRNA-seq on a gene by gene basis. The color of each box in the plane is calculated from the log count of gene.

Supplementary Figure 3. Expression patterns of the indicated marker genes in each cell type within the lung tissue. UMAP 2D scatter plot of the transcriptome of each cell. The expression value is normalized as log2 transcripts per 10K (log2 [TP10K+1]).

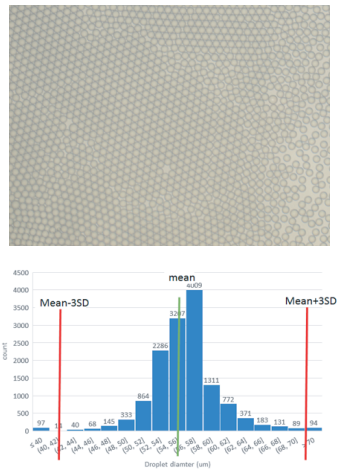
Supplementary Figure 4. Rarefaction curves showing the effect of sequencing depth on the mean number of detected genes per cell.

Figure 1

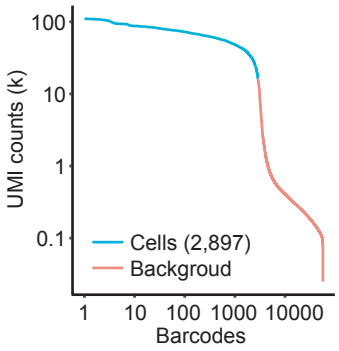
a



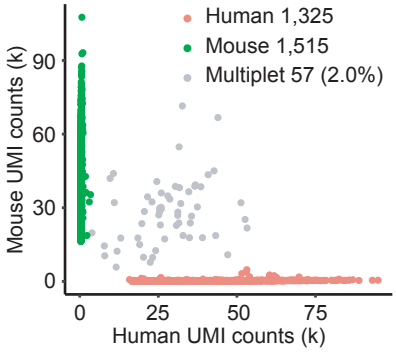
b



c



d



e

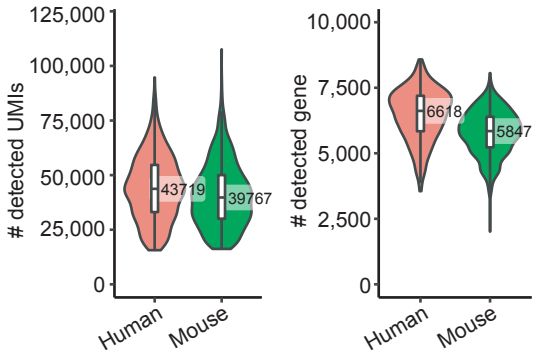
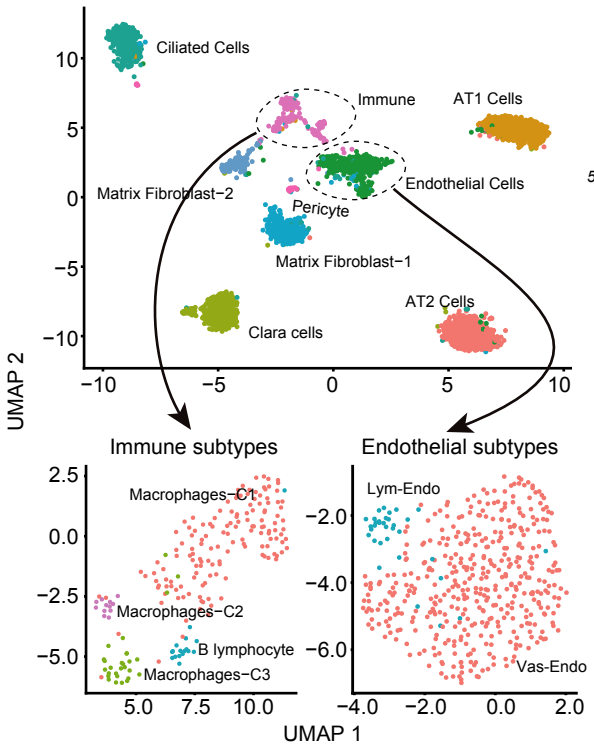


Figure 2

a


Ciliated Cells
Immune
AT1 Cells
Matrix Fibroblast−2
Endothelial Cells
Pericyte
Matrix Fibroblast−1
AT2 Cells
Clara cells
UMAP 2
UMAP 1
Immune subtypes
Endothelial subtypes
Macrophages−C1
Macrophages−C2
B lymphocyte
Macrophages−C3
Lym-Endo
Vas-Endo

b

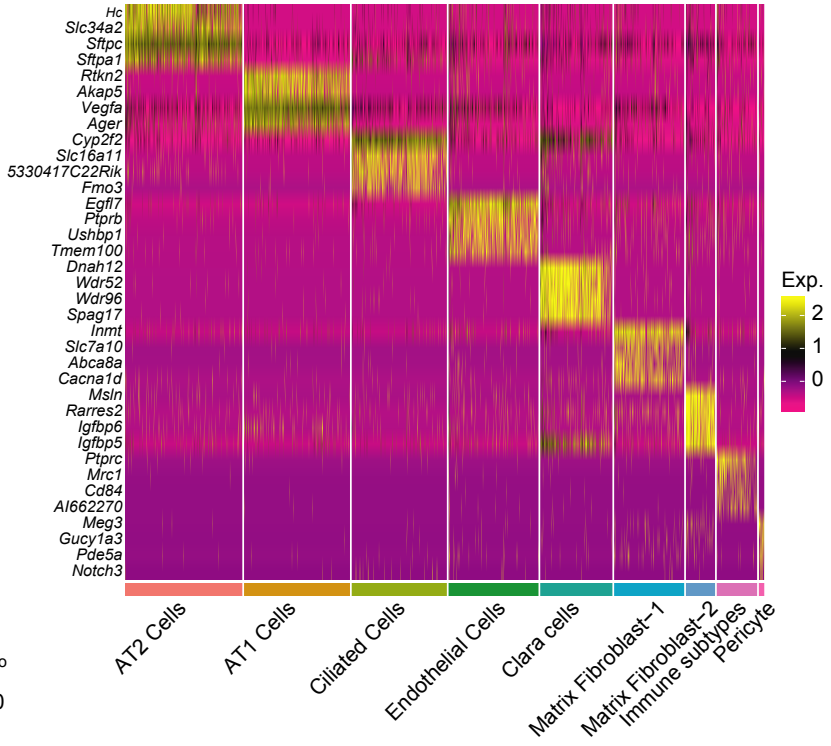

Hc
Slc34a2
Sftpc
Sftpa1
Rtkn2
Akap5
Vegfa
Ager
Cyp2f2
Slc16a11
5330417C22Rik
Fmo3
Egfl7
Ptprb
Ushbp1
Tmem100
Dnah12
Wdr52
Wdr96
Spag17
Inmt
Slc7a10
Abca8a
Cacna1d
Msln
Rarres2
Igfbp6
Igfbp5
Ptprc
Mrc1
Cd84
AI662270
Meg3
Gucy1a3
Pde5a
Notch3
Exp.
AT2 Cells
AT1 Cells
Ciliated Cells
Endothelial Cells
Clara cells
Matrix Fibroblast−1
Matrix Fibroblast−2
Immune subtypes
Pericyte

Figure 3

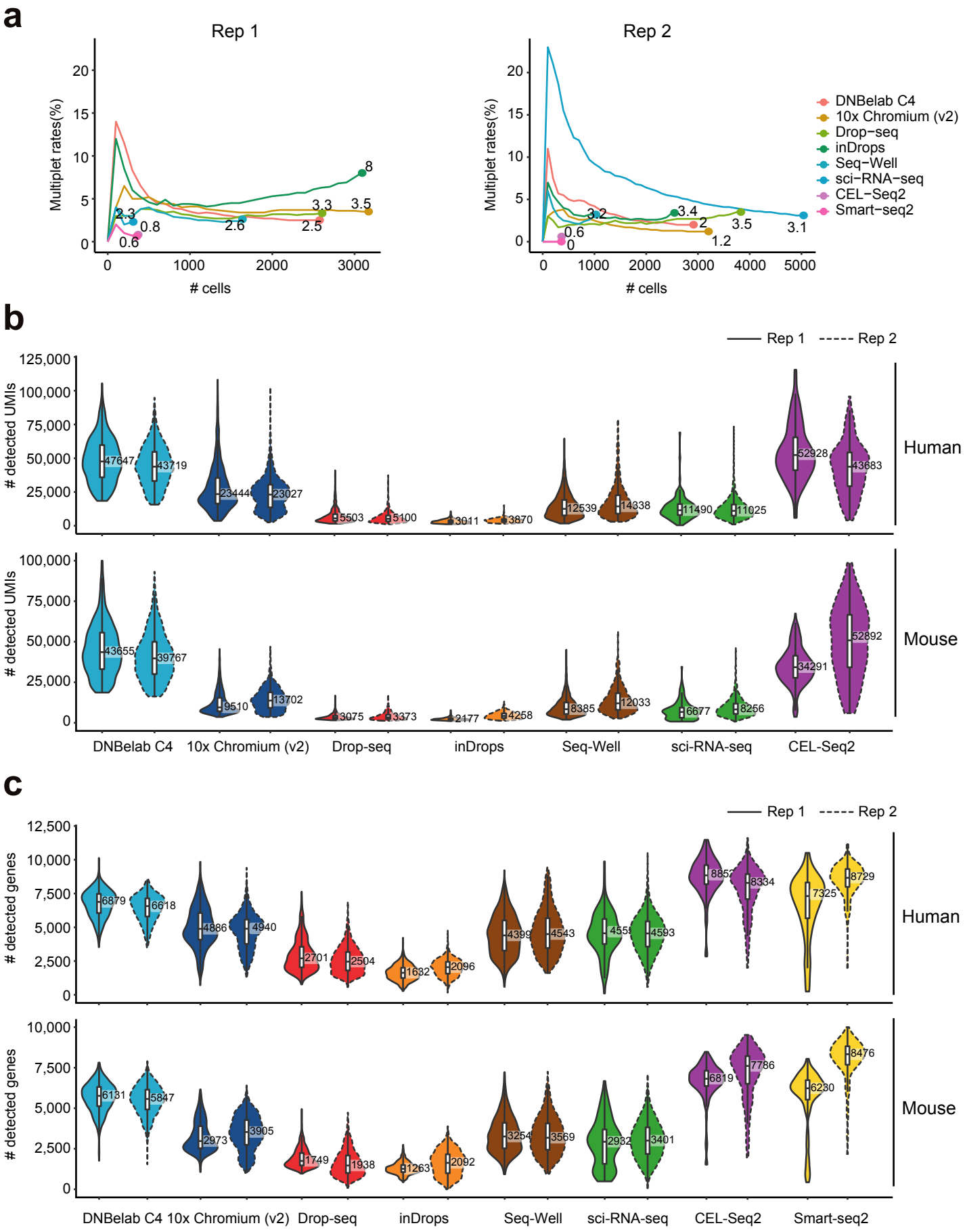
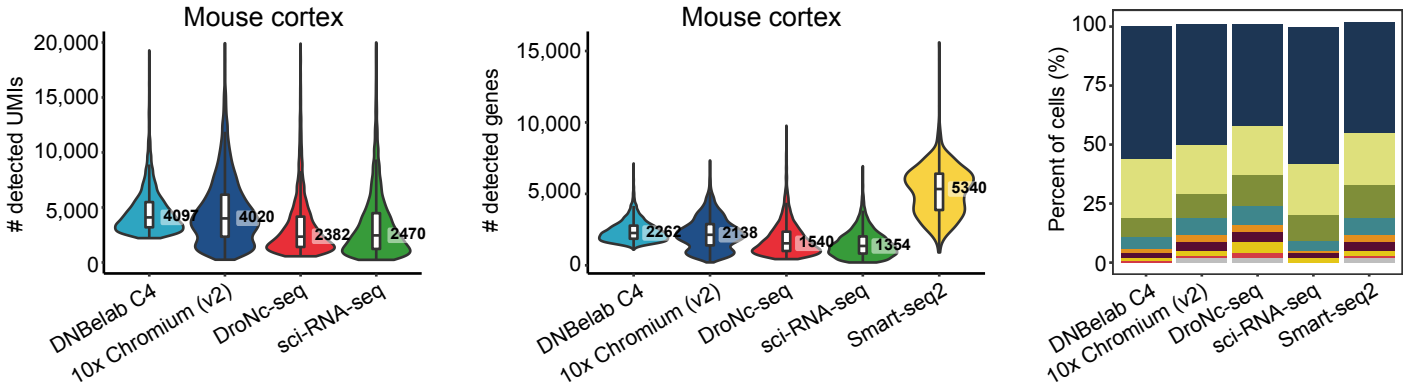
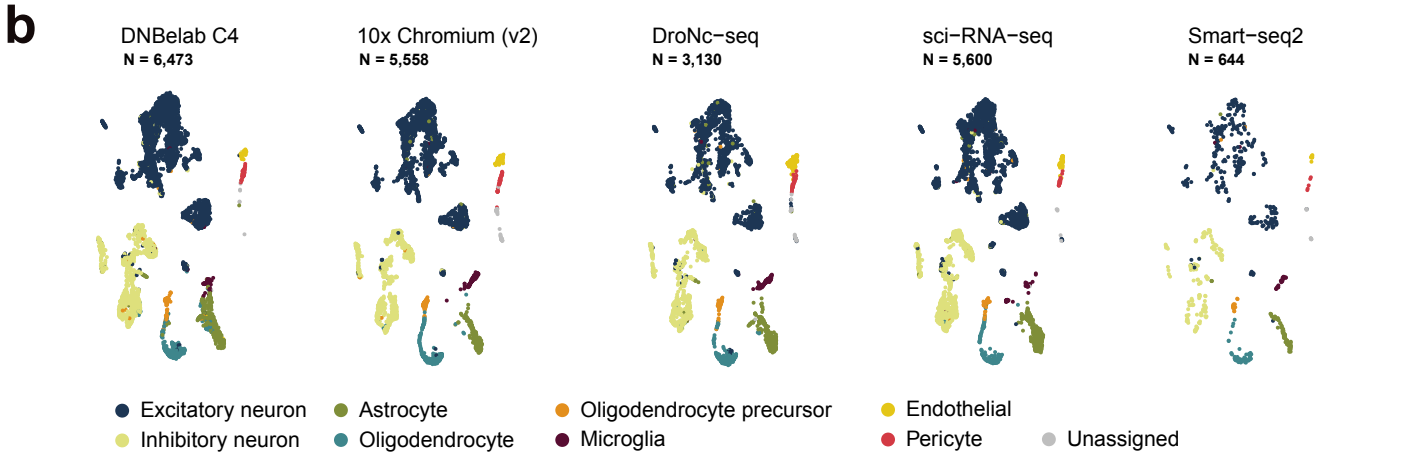
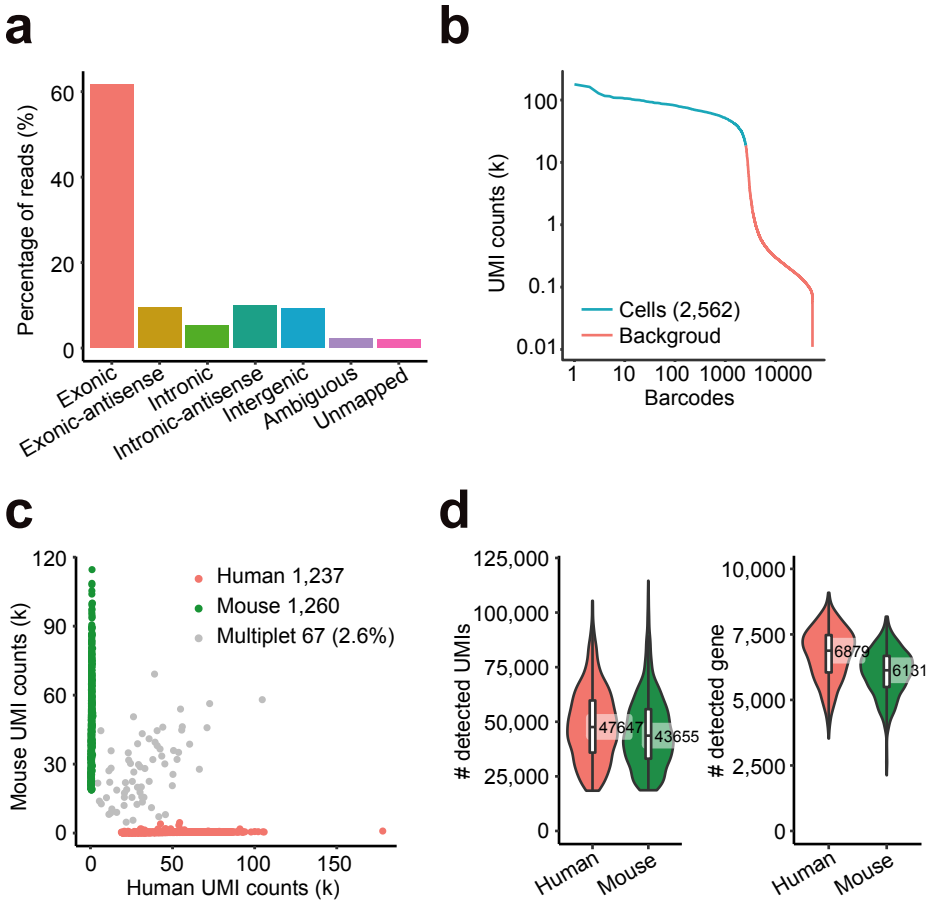

a
Rep 1
Rep 2
Multiplet rates(%)
cells
DNBelab C4
10x Chromium (v2)
Drop-seq
inDrops
Seq-Well
sci-RNA-seq
CEL-Seq2
Smart-seq2
b
Rep 1
Rep 2
detected UMIs
Human
Mouse
DNBelab C4
10x Chromium (v2)
Drop-seq
inDrops
Seq-Well
sci-RNA-seq
CEL-Seq2
c
detected genes
Human
Mouse
DNBelab C4
10x Chromium (v2)
Drop-seq
inDrops
Seq-Well
sci-RNA-seq
CEL-Seq2
Smart-seq2

Figure 4

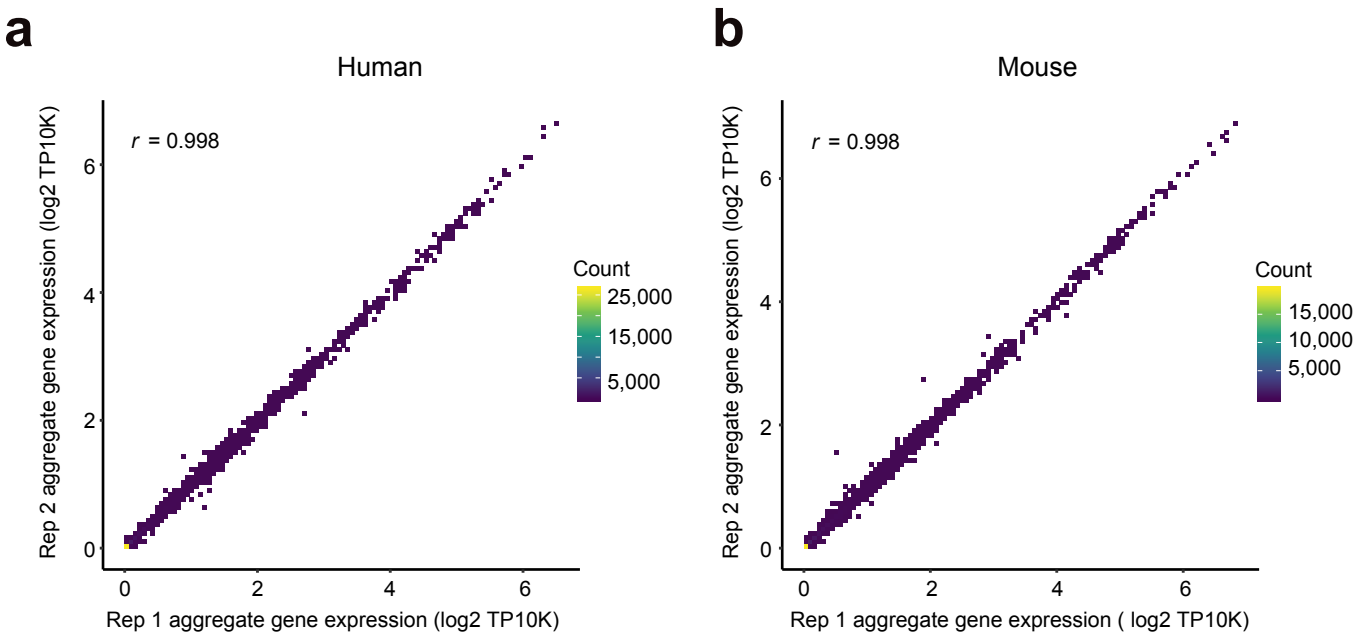

a
Mouse cortex
detected UMIs
20,000
15,000
10,000
5,000
0
4097
4020
2382
2470
DNBelab C4
10x Chromium (v2)
DroNc-seq
sci-RNA-seq
Mouse cortex
detected genes
15,000
10,000
5,000
0
2262
2138
1540
1354
5340
DNBelab C4
10x Chromium (v2)
DroNc-seq
sci-RNA-seq
Smart-seq2
c
Percent of cells (%)
100
75
50
25
0
DNBelab C4
10x Chromium (v2)
DroNc-seq
sci-RNA-seq
Smart-seq2


b
DNBelab C4
N = 6,473
10x Chromium (v2)
N = 5,558
DroNc−seq
N = 3,130
sci−RNA−seq
N = 5,600
Smart−seq2
N = 644
Excitatory neuron
Inhibitory neuron
Astrocyte
Oligodendrocyte
Oligodendrocyte precursor
Microglia
Endothelial
Pericyte
Unassigned

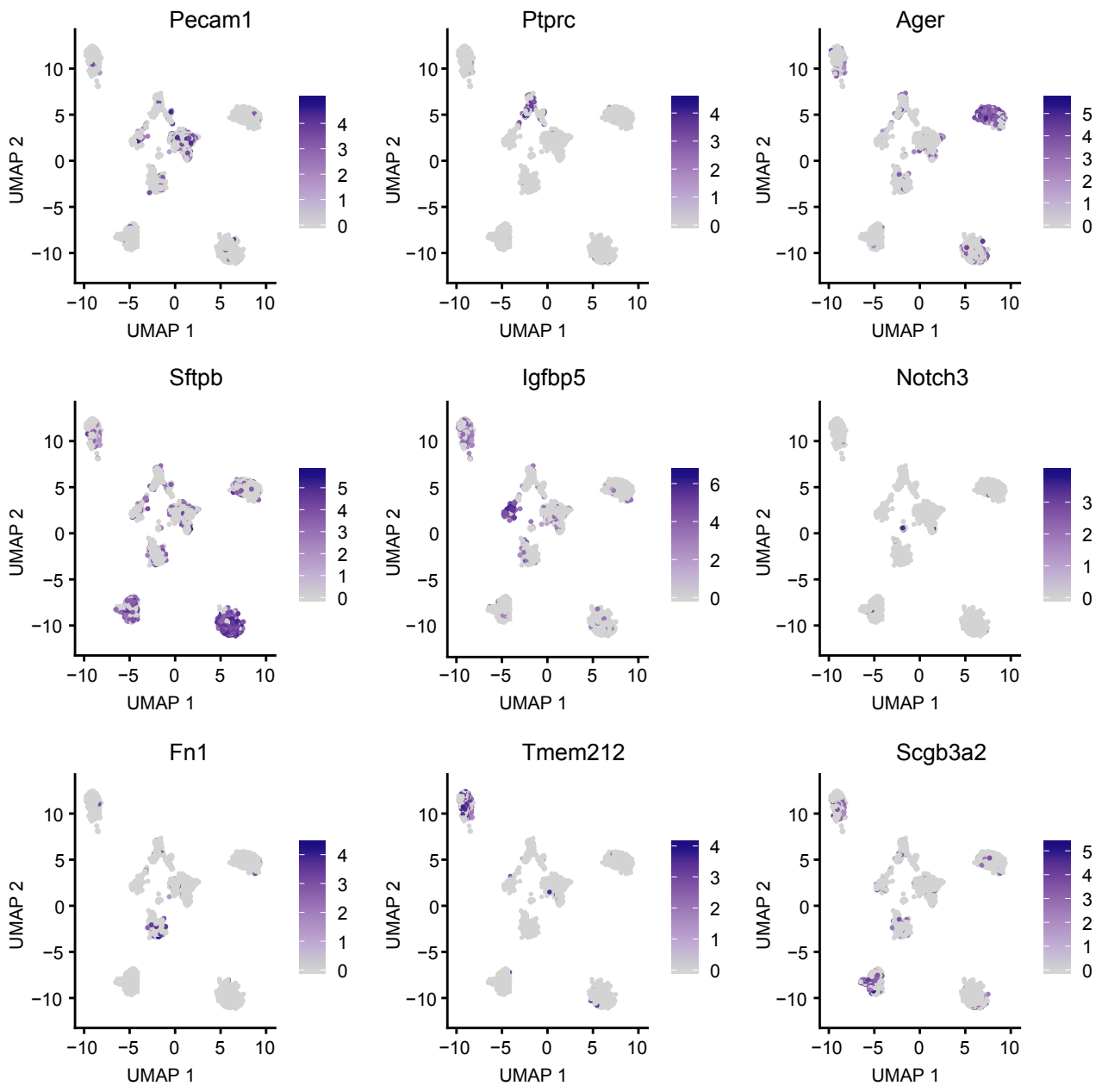
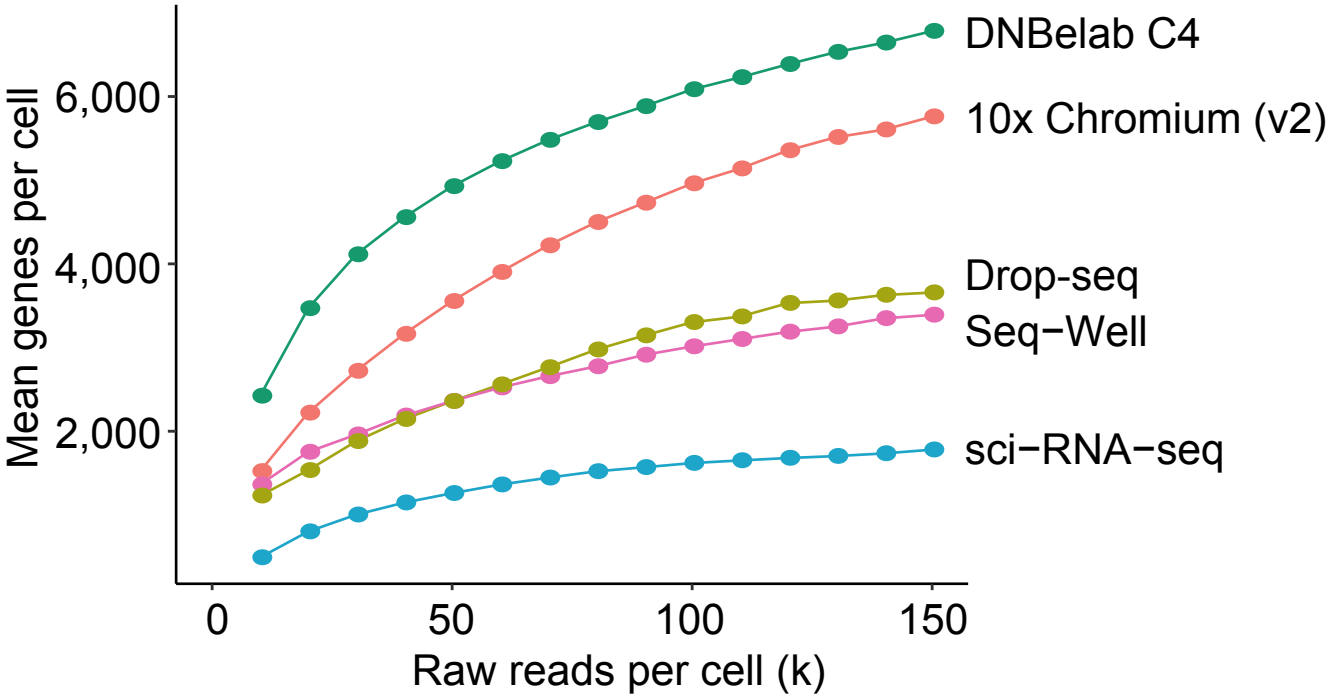
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4