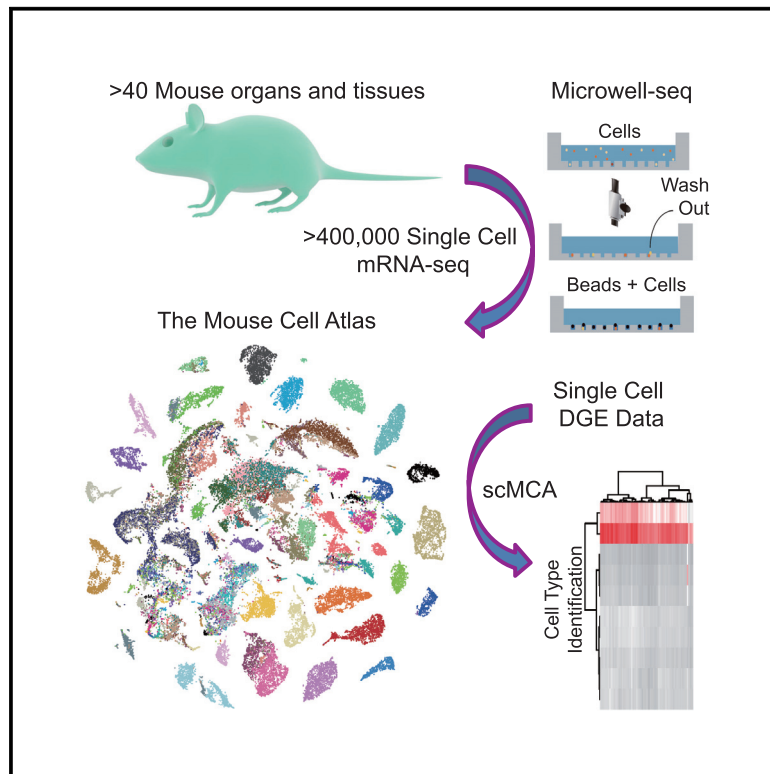


Mapping the Mouse Cell Atlas by Microwell-Seq

Graphical Abstract



Authors

Xiaoping Han, Renying Wang,
Yincong Zhou, ..., Guo-Cheng Yuan,
Ming Chen, Guoji Guo

Correspondence

xhan@zju.edu.cn (X.H.),
ggj@zju.edu.cn (G.G.)

In Brief

Development of Microwell-seq allows construction of a mouse cell atlas at the single-cell level with a high-throughput and low-cost platform.

Highlights

- Development of Microwell-seq, a high-throughput and low-cost scRNA-seq platform
- Construction of a single-cell mouse cell atlas (scMCA) covering major cell types
- Characterization of cellular heterogeneity with minimal batch effect
- Characterization of cross-tissue cellular network at the single-cell level

Data Resources

GSE108097



Mapping the Mouse Cell Atlas by Microwell-Seq

Xiaoping Han,^{1,12,13,*} Renying Wang,^{1,12,13} Yincong Zhou,^{2,12,13} Lijiang Fei,^{1,12,13} Huiyu Sun,^{1,12,13} Shujing Lai,^{1,12,13} Assieh Saadatpour,¹¹ Zimin Zhou,^{1,12} Haide Chen,^{1,12} Fang Ye,^{1,12} Daosheng Huang,¹ Yang Xu,¹ Wentao Huang,¹ Mengmeng Jiang,^{1,12} Xinyi Jiang,^{1,12} Jie Mao,³ Yao Chen,⁴ Chenyu Lu,⁵ Jin Xie,⁶ Qun Fang,⁷ Yibin Wang,⁸ Rui Yue,⁸ Tiefeng Li,³ He Huang,^{9,12} Stuart H. Orkin,¹⁰ Guo-Cheng Yuan,¹¹ Ming Chen,^{2,12} and Guoji Guo^{1,9,12,14,*}

¹Center for Stem Cell and Regenerative Medicine, Zhejiang University School of Medicine, Hangzhou 310058, China

²College of Life Sciences, Zhejiang University, Hangzhou 310003, China

³Institute of Applied Mechanics, Zhejiang University, Hangzhou 310027, China

⁴Department of Reproductive Endocrinology, Women's Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China

⁵Research Center of Infection and Immunity, Zhejiang University School of Medicine, Hangzhou 310058, China

⁶State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310058, China

⁷Institute of Microanalytical Systems, Department of Chemistry, Zhejiang University, Hangzhou 310058, China

⁸Institute for Regenerative Medicine, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China

⁹Institute of Hematology, the First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou 310003, China

¹⁰Division of Pediatric Hematology/Oncology, Dana Farber Cancer Institute and Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

¹¹Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Harvard School of Public Health, Boston, MA 02115, USA

¹²Stem Cell Institute, Zhejiang University, Hangzhou 310058, China

¹³These authors contributed equally

¹⁴Lead Contact

*Correspondence: xhan@zju.edu.cn (X.H.), ggj@zju.edu.cn (G.G.)

<https://doi.org/10.1016/j.cell.2018.02.001>

SUMMARY

Single-cell RNA sequencing (scRNA-seq) technologies are poised to reshape the current cell-type classification system. However, a transcriptome-based single-cell atlas has not been achieved for complex mammalian systems. Here, we developed Microwell-seq, a high-throughput and low-cost scRNA-seq platform using simple, inexpensive devices. Using Microwell-seq, we analyzed more than 400,000 single cells covering all of the major mouse organs and constructed a basic scheme for a mouse cell atlas (MCA). We reveal a single-cell hierarchy for many tissues that have not been well characterized previously. We built a web-based “single-cell MCA analysis” pipeline that accurately defines cell types based on single-cell digital expression. Our study demonstrates the wide applicability of the Microwell-seq technology and MCA resource.

INTRODUCTION

Cellular identity is defined by a unique combination of expressed genes. Since the discovery of the cell as the fundamental unit of life, investigators have sought to characterize and classify cell types based on their properties (Regev et al., 2017). At first, cells were mainly characterized by their position, shape, and cellular components; cell-type definition heavily depended on improvements in microscopy. Later, immunohistochemistry,

fluorescence-activated cell sorting (FACS), and fluorescence *in situ* hybridization (FISH) facilitated the use of gene expression markers for cell classification (Regev et al., 2017). These technologies revealed heterogeneity within morphologically similar cell populations. Subsequent studies associated molecular phenotypes with cellular functions and achieved remarkable advances in distinguishing cell types. Nonetheless, current cell-type classification systems depend on markers that are largely chosen based on serendipitous discovery. The choice of markers often varies among different laboratories, leading to difficulties in comparing results. Cross-tissue comparison is challenging as a given marker system, and preferred assay may differ for different tissues.

Recent advances in single-cell gene expression analysis offer an opportunity to greatly enhance cell identification and classification. Emerging methods, such as high-throughput single-cell qPCR (Dalerba et al., 2011; Guo et al., 2010; Han et al., 2017), single-cell mass cytometry (Bendall et al., 2011), and single-cell mRNA-seq (Hashimshony et al., 2012; Ramsköld et al., 2012; Shalek et al., 2013; Tang et al., 2009; Treutlein et al., 2014) allow for dissection of cell heterogeneity at high resolution. Massively parallel assays can process thousands of single cells simultaneously to measure transcriptional profiles with rapidly decreasing costs (Fan et al., 2015; Gierahn et al., 2017; Klein et al., 2015; Macosko et al., 2015). Other methodologies, such as single-cell genome analysis (Hou et al., 2012; Navin et al., 2011; Wang et al., 2012; Xu et al., 2012), epigenome analysis (Buenrostro et al., 2015; Cusanovich et al., 2015; Guo et al., 2017; Jin et al., 2015; Lorthongpanich et al., 2013; Mooijman et al., 2016; Nagano et al., 2013; Smallwood et al., 2014), and *in situ* analysis (Chen et al., 2015; Ke et al., 2013; Lee et al.,



2014), may provide further information regarding genetic and epigenetic regulation of gene expression as well as cellular and subcellular localization of biomolecules. With ongoing technical advances, a consensus has emerged that it is now time to build a comprehensive single-cell genomic database that includes all mammalian cell types (Regev et al., 2017).

Here, we describe Microwell-seq, a simple method to profile thousands of single cells by transcriptome analysis utilizing an agarose-constructed microwell array and barcoded beads. Microwell-seq has advantages in convenience and simplicity, which should make the method widely accessible. Using Microwell-seq, we constructed a first stage “mouse cell atlas” with more than 400 k single-cell transcriptomic profiles from 51 mouse tissues, organs, and cell cultures. The resulting map covers more than 800 major cell types and potentially more than 1,000 cell subtypes in the mouse system. We integrated published high-throughput single-cell data and built the MCA website database for the scientific community. Finally, we constructed a “single-cell MCA (scMCA)” tool that accurately defines cell types based on single-cell digital expression. Future efforts in data accumulation and integration will eventually lead to creation of a comprehensive mammalian cell map that will facilitate related basic research and clinical applications.

RESULTS

Microwell-Seq: A Convenient, Low-Cost, and Robust Platform for High-Throughput scRNA-Seq

To establish a cost-effective single-cell technology that is widely accessible, we combined the advantages of existing methodologies (Fan et al., 2015; Klein et al., 2015; Macosko et al., 2015) in designing the Microwell-seq method. In Microwell-seq, individual cells are trapped in an agarose microarray and mRNAs are captured on magnetic beads. The processes for bead synthesis and microarray fabrication are shown in Figure S1. Barcoded beads are synthesized by 3 split-pool rounds (Figure S1A; Table S1). Each oligonucleotide consists of a primer sequence, a cell barcode, a unique molecular identifier (UMI), and a poly T tail (Fan et al., 2015; Islam et al., 2014). Fabrication of the agarose microarray is simple and inexpensive (Figure S1B). The silicon and PDMS chips are reusable, meaning that a single silicon chip can be employed to generate many agarose microarrays. The size of the agarose chip can be readily adjusted by making different-sized PDMS chips for a wide range of input sample sizes and concentrations. Only minutes are required to make an agarose chip for each experiment.

The workflow of Microwell-seq is shown in Figure 1A. An agarose plate with 10^5 microwells is used to trap 5–10 K individual cells, similar to the Cytoseq platform (Fan et al., 2015). After cells (50–100 K) are loaded into the wells, the microwell array is inspected under a microscope and rare cell doublets are washed out with a capillary tube. The capture efficiency and cell quality are estimated. Barcoded magnetic beads are then loaded and trapped into each well by size. Each single bead is conjugated with 10^7 – 10^8 oligonucleotides, which share the same cellular barcode (Figures S1C and S1D). After incubation of beads and cells in a soft flow of lysis buffer, beads with captured mRNA

are retrieved with a magnet. The procedure from cell loading to cell lysis takes ~15 min. Beads are collected in a 1.5 mL tube in which reverse transcription and template switch steps are performed using the Smart-seq2 protocol (Picelli et al., 2013). Amplified cDNA is fragmented by a customized transposase that carries two identical insertion sequences (see the STAR Methods). The 3' ends of the transcripts are then enriched during library generation using PCR and sequenced using the Illumina HiSeq platform (Table S2).

To assess the fidelity of the method, we performed mixed species experiments with cultured human (293T) and mouse (3T3) cells. We identified the ratio of reads mapped to both the human and mouse genome in each single cell (Figure 1B). We found that Microwell-seq produced high-fidelity single-cell libraries with no more than 1.2% cell doublets. Approximately 6,500 genes and 55,000 transcripts can be detected, on average, by saturated sequencing (Figure 1C). Low reads versus the gene number ratio was observed in large-scale experiments (Figure 1D). Cell-cycle scores were calculated for each human 293T cell based on previously reported phase-specific genes and methods (Macosko et al., 2015) (Figure 1E). Cells at different cell-cycle stages were clearly separated based on their cell-cycle scores. By integrating our Cj7 mES Microwell-seq data with the published comparative analysis for different single-cell mRNA-seq methods (Svensson et al., 2017; Ziegenhain et al., 2017), we observed a sensitivity and accuracy that were comparable with those of other available methodologies (Figures S2A–S2C). Notably, Microwell-seq showed advantages in doublet rate and cost (Figures S2D and S2E). It detected more genes than other high-throughput single-cell mRNA-seq methods in the range of middle to low sequencing depth (Figures S2A and S2B). We then analyzed 4,323 thawed single cells of CD34⁺ and CD34⁻ compartments from mobilized human peripheral blood (mPB). We observed a clear distinction between the two populations (Figure S2F). Thawed mPB CD34⁺ cells from batch 1 and batch 2 showed little batch effects on a t-distributed stochastic neighbor embedding (t-SNE) map (Figure S2G). Notably, Microwell-seq worked reliably with thawed cells.

Construction of Mouse Cell Atlas Using Microwell-Seq

By harnessing the power of Microwell-seq, we embarked on creating an atlas of all major mouse cell types with minimal input from traditional classification schemes (e.g., FACS). We collected mammary gland (virgin, pregnant, lactation, and involution), uterus, bladder, ovary, intestine, kidney, lung, testis, pancreas, liver, spleen, muscle, stomach, bone marrow, thymus, prostate, cKit⁺ bone marrow, bone marrow mesenchymal cells, and peripheral blood samples from 6- to 10-week-old C57BL/6 mice. We collected E14.5 fetal liver, fetal lung, fetal stomach, fetal gonad, fetal brain, fetal intestine, fetal placenta, and mesenchymal tissues in addition to neonatal brain, neonatal skin, neonatal calvaria, neonatal rib, and neonatal muscle samples. Tissues were carefully washed and prepared into single-cell suspensions with optimized protocols (Table S3). We also included several cultured cells derived from mouse tissues: 3T3 cells, embryonic stem (ES) cells, trophoblast stem (TS) cells, and mesenchymal stem cells (MSCs). Single cells were then processed with Microwell-seq.

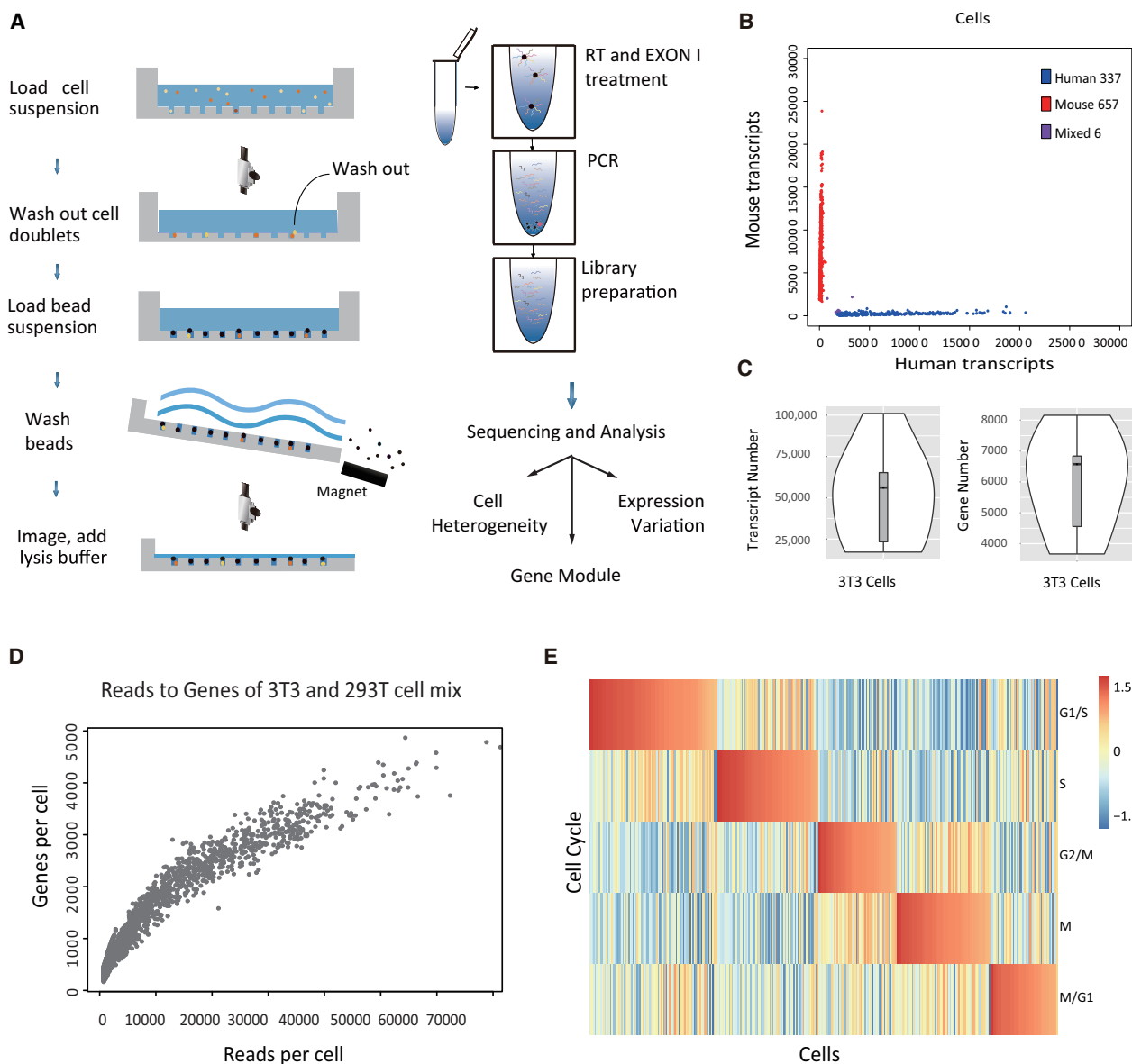


Figure 1. Workflow and Evaluation of Microwell-Seq

(A) A schematic of the basic workflow for Microwell-seq.

(B) Human-mouse mix test using Microwell-seq. Human (293T) and mouse (3T3) cells were mixed at equal concentrations. The scatterplot shows the number of human and mouse transcripts associated with each cellular barcode. Blue dots indicate single cells that were human-specific; red dots indicate single cells that were mouse-specific. Only 0.6% (purple dots) are human-mouse mixed cells.

(C) The distribution of gene number and transcript number in 9 deeply sequenced beads (3T3 cell captured) are shown by violin plots.

(D) Detected gene number versus read number of each individual cell in the species-mixing experiment.

(E) Cell-cycle state of 337 human (293T) cells measured by Microwell-seq. The score for each phase was calculated using a previously reported method (see data analysis in the [STAR Methods](#)). The cells were ordered by their phase scores.

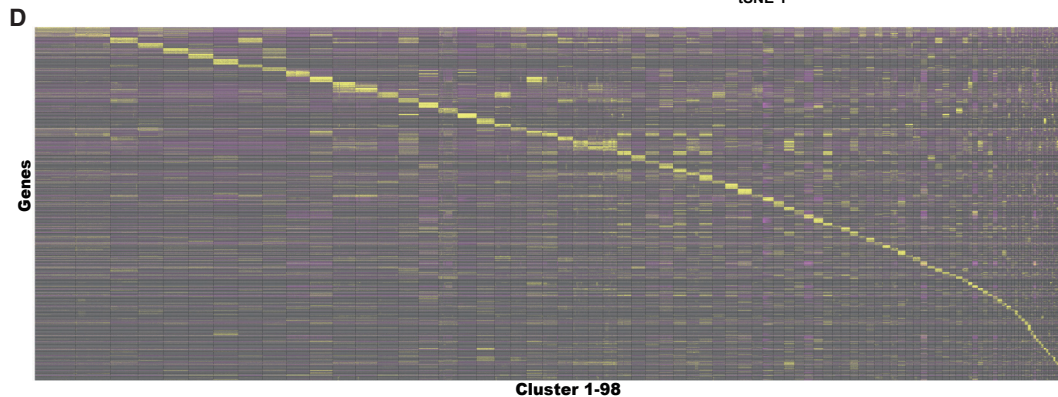
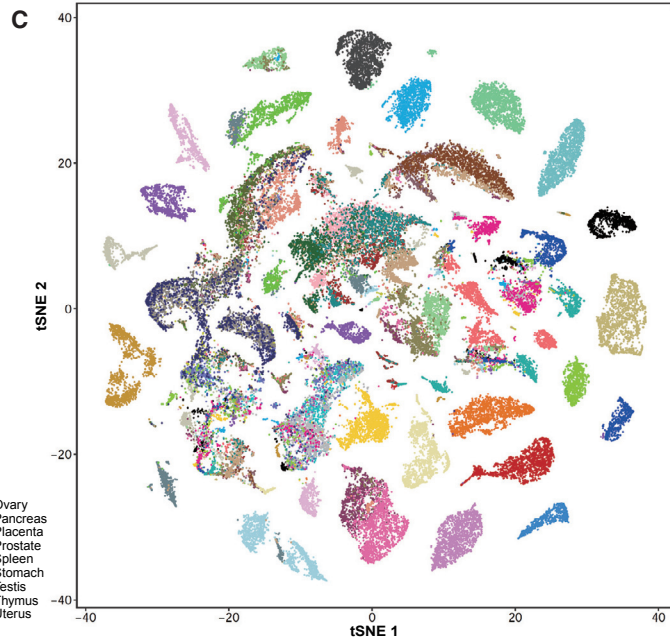
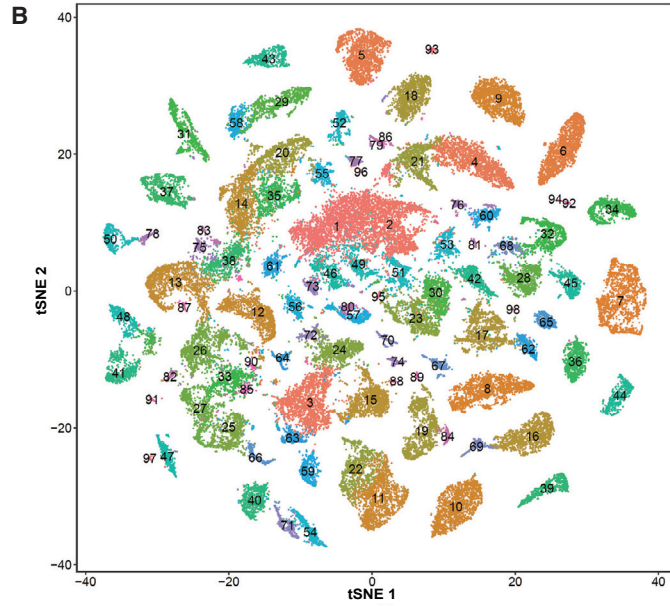
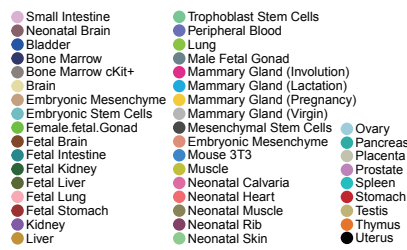
See also [Figures S1](#) and [S2](#) and [Tables S1](#) and [S2](#).

The sequencing data were processed using published pipelines ([Macosko et al., 2015](#); [Satija et al., 2015](#)). In total, we analyzed >400,000 single cells from >50 mouse tissues and cultures ([Figure 2A](#)). In a global view, we identified >800 cell types grouped into 98 major clusters using 60,000 cells sampled

from the complete dataset ([Figure 2B](#); [Table S4](#)). We found that multiple tissues, including liver, muscle, and placenta, contributed to the defined hematopoietic cell clusters, such as Clusters 3, 14, and 27 (C3, C14, and C27), suggesting that the cell identity defined by single-cell transcriptome data is unaffected by batch

A Currently Processed Cells at MCA

Tissues/Cell lines	Cell number
3T3	645
Bladder	3143
Bone Marrow	11632
Bone Marrow (c-Kit+)	26815
Bone Marrow Mesenchyme	7390
Brain	4778
Embryonic Mesenchyme (E14.5)	4475
Embryonic Stem Cells(Cj7)	1891
Embryonic Stem Cells (E14)	9994
Female Fetal Gonad (E14.5)	9312
Fetal Brain (E14.5)	4727
Fetal Intestine (E14.5)	8658
Fetal Kidney (E14.5)	18051
Fetal Liver (E14.5)	3730
Fetal Lung (E14.5)	8395
Fetal Stomach (E14.5)	7911
Kidney	6665
Liver	6426
Lung	9116
Mammary Gland (Lactation)	15699
Mammary Gland (Pregnancy)	9737
Mammary Gland (Virgin)	5681
Mammary Gland (Virgin,CD45-)	18714
Mammary Gland (Involution)	6301
Mammary Gland (Involution,CD45-)	5064
Male Fetal Gonad (E14.5)	3872
Mesenchymal Stem Cells (Cultured)	14688
Muscle (Leg)	1647
Neonatal Brain	2460
Neonatal Calvaria	8248
Neonatal Heart	5075
Neonatal Muscle (Leg)	6683
Neonatal Rib	8371
Neonatal Skin	4079
Ovary	4410
Pancreas	4405
Peripheral Blood	7972
Placenta (E14.5)	5423
Prostate	2948
Small Intestine	8984
Small Intestine (CD45-)	2938
Spleen	2417
Stomach	3509
Testis	19659
Thymus	5758
Trophoblast Stem Cells	19693
Uterus	3761
Cj7 WT Differentiation	9677
Cj7 Ezh2 ^{-/-} Differentiation	9940
Cj7 WT EB	14695
Cj7 Ezh2 ^{-/-} EB	9534
Total	405796



Cluster 1-98

(legend on next page)

effect in Microwell-seq experiments from different tissues (Figures 2B and 2C). Other clusters with significant multi-tissue contributions correspond to stromal cells (C28), endothelial cells (C17), neurons (C21), and myocytes (C2). These data encompass the most comprehensive genetic module repertoire yet described for the mouse (Figure 2D; Table S5). We performed t-SNE analysis and differential gene expression analysis for each specific organ type and uncovered previously unrecognized cell heterogeneity in a wide range of mouse tissues (Table S5).

To enable public access to these data, we constructed the mouse cell atlas website at <http://bis.zju.edu.cn/MCA>. The website enables browsing of processed single-cell data for all tissues and searching for genes of interest in the dataset (Figures S3A and S3B).

Microwell-Seq Dissects Cellular Heterogeneity in Diverse Mouse Tissues

Initial MCA data were collected from mouse embryonic stem (mES) cells and mouse trophoblast stem (mTS) cells. In mES cell culture, we detected *Actb_high*, *Rps28_high*, *Nedd4_high*, and 2-cell-like mES subpopulations. The rare 2-cell-like mES cluster specifically expresses *Zscan4d*, *Zscan4a*, and *Zsan4c*, as well as groups of predicted genes, such as *Gm5662*, *Gm8994*, and *Gm8300*; these signatures strongly resemble middle and late 2-cell-stage mouse blastomeres (Figure S3C; Tables S4 and S5). In mTS cell culture, we detected *Mrpl12_high*, *Mrpl55_high*, and *Rps28_high* subpopulations, as well as a *Krt8⁺* epithelium that appears to resemble an *in vivo* trophoblast progenitor cell type expressing *Gjb3*, *Hand1*, *Rhox6*, and *Rhox9* (Figure S3D; Tables S4 and S5). Interestingly, for both mES and mTS cells, there was an *Rps28_high* subpopulation expressing *Rps28*, *Rps29*, *Ppia*, and *Cox7c*; these signatures strongly resemble 8-cell-stage mouse blastomeres in the preimplantation embryo dataset (Table S5). With success in mouse cell line experiments, we then moved our analysis to more complex organs, including adult mouse mammary gland, kidney, lung, and E14.5 placenta.

The mammary gland provides a unique model for studying organ tissue specification, as it is the only glandular organ that reaches full development after birth. In our study, we digested the gland and were able to create a single-cell expression map for the entire tissue (Figures 3A–3C; Table S5). Two independent Microwell-seq experiments analyzing mammary glands collected from virgin mice exhibited a minimal batch effect (Figure 3A). Several distinct cell types emerged in analysis of pregnant mammary glands compared with virgin samples (Figure 3B). C18 is defined as luminal cells, due to its specific expression of *Krt8*, *Krt18*, and *Fgg*, whereas myoepithelial cells (C17) express

Krt5 and *Krt14*, as well as the smooth muscle markers *Mylk* and *Myl9* (Shackleton et al., 2006) (Figure 3B; Table S5). During pregnancy, glands of virgin adult mice develop an extensive network of secretory alveoli (C1) derived from the ductal luminal that highly express *Csn1s1*, *Csn3*, *Csn2*, and *Wfdc18*. We found that the *Elf5⁺*, *Krt8⁺*, and *Krt18⁺* population, which is considered to contain a luminal progenitor pool (C7), expands remarkably during pregnancy (Rios et al., 2014). The ratio of secretory alveoli cells increases dramatically in the lactating mammary gland and then drops to a normal level during mammary gland involution (Table S5). Moreover, we observed a large category of mammary gland resident immune cells, including T cells (C2, C3, C6, C11, C13), B cells (C4), natural killer (NK) cells (C10, C16), dendritic cells (C8, C12), monocytes (C9), and macrophages (C14). We also found two major types of stromal cells (C5, C15) that may play different roles in supporting mammary gland development.

The kidney is a complex blood filtration system that eliminates toxic products from the body and maintains fluid homeostasis. Previous studies used microdissection to identify markers for different regions (Cheval et al., 2011); however, systematic cell-type classification at the single-cell level has not been achieved in the adult mouse kidney. Our study covers kidney cells related to the whole process of urine production, from glomeruli to ureters (Figures 3D and S3E; Table S5). C12 and C15 are fenestrated endothelial cells expressing *Plvap* and *Tm4sf1*. As presented in Figure 3D, the tSNE map revealed an expression trajectory from C1 to C11 and then to C6. C1 is defined as proximal tubule brush border cells that express *Miox*, *Gsta2*, *Ass1*, and *Rida*. C6 represents a type of S1 proximal tubule cell characterized by high expression of *Alpl*, *Slc5a2*, *G6pc*, and *Nox4*. C11 might be a novel cell type physically between brush border cell and S1 cells. The *Osgin1⁺* C4 is suspected to represent S2 cells. S3 proximal tubule cells (C8) express *Kap*, *Keg1*, *Napsa*, and *Slc22a13*. The loop of Henle (C2) connects the proximal convoluted tubule with the distal convoluted tubule. Gene expression of C3, C9, and C14 forms another trajectory: distal tubule cells (C3 and C9) are characterized by expression of *Slc12a3*, *Pgam2*, and *Wnk1* (Cheval et al., 2011), while connecting tubule cells (C14) share features of both the distal tubule and collecting duct. The collecting duct contains two cell types, principal cells (C13) and intercalated cells (C5 and C7). C18 and C23 are two epithelial cell types in the kidney. We also identify an uncharacterized endothelial cell type C17 with high expression of *Epcam*, *Slc4a11*, and *Slc31a2*. The lineage specificity and spatial expression pattern of these identified markers can be further verified through other publicly available datasets, such as the Mouse Anatomical Atlas (Diez-Roux et al., 2011). As examples, we obtained *in situ* data for C6, as

Figure 2. Mapping Mouse Cell Atlas Using Microwell-Seq

(A) Number of cells currently processed at MCA.

(B) t-Distributed stochastic neighbor embedding (t-SNE) analysis of 60,000 single cells sampled from mouse cell atlas data. Ninety-eight main cell type clusters are labeled in the t-SNE map.

(C) t-SNE analysis of 60,000 single cells sampled from mouse cell atlas data. Tissue types are labeled in the t-SNE map.

(D) A hierarchical clustering heatmap showing differentially expressed genes (row) across 60,000 mouse cells clustered into 98 cell types (column). Yellow corresponds to a high expression level; purple and black correspond to low expression levels.

See also Figure S3 and Tables S3, S4, and S5.

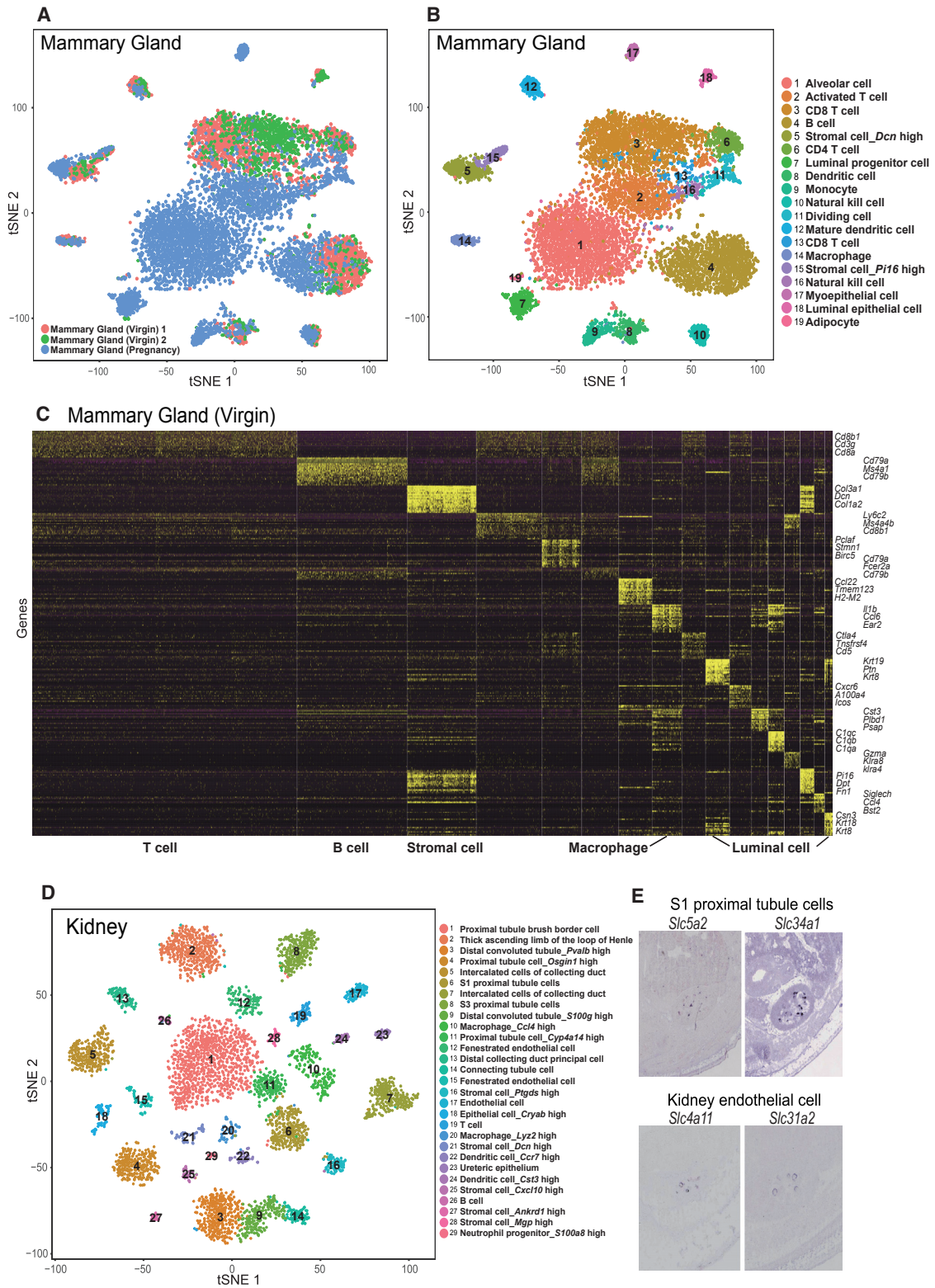


Figure 3. Resolving Cellular Heterogeneity in the Mammary Gland and Kidney

(A) t-SNE maps of mouse mammary gland single-cell data. Cells are colored by experimental batch. (B) A t-SNE map of mouse mammary gland single-cell data. Cells are colored by cell-type cluster.

(legend continued on next page)

well as C17, and show their distinct positions in the E14.5 kidney (Figure 3E). In addition, we detected five stromal cell populations (C16, C21, C25, C27, and C28), T cells (C19), B cell (C26), *Ccr7*-high dendritic cells (C22), *Cst3*-high dendritic cells (C24), *S100a8*-high neutrophil progenitors (C29), and two types of macrophages (C9, C20). Cell-type correlation analysis of our data with pre-published mouse kidney single-cell data suggests a strong overlap with markers and cell types from the two studies (Figure S4A). Functional experiments with reporter mice will be needed to verify the biological significance of these cell types.

The lung is the primary organ for gas exchange in mammals. Previous studies have focused on profiling embryonic alveolar epithelium using single-cell RNA sequencing (RNA-seq) (Treutlein et al., 2014). Here, by analyzing the whole adult mouse lung tissue, we defined 32 distinct clusters with specific molecular markers (Figures 4A and 4B; Table S5). Batch effects from three independent experiments were minimal (Figure 4C). The cell subpopulations inferred from this analysis were readily matched to known lung cell types, including four previously reported alveolar epithelial cell types: alveolar type 2 (AT2) cells (enrichment of surfactant-associated protein family genes *Sftpc*, *Sftpa1*, and *Sftpb* in C1, Figure 4D), alveolar type 1 (AT1) cells (enrichment of *Pdpr*, *Ager*, and *Clic5* in C14), Clara cells (enrichment of secretoglobulin family genes *Scgb1a1*, *Scgb3a2* in C12), and ciliated cells (enrichment of *1110017D15Rik* and *Foxj1* in C18). Notably, we identified alveolar bipotent-like cells (C27) coexpressing the AT1 markers *Ager*, *Emp2*, and *Aqp5*, as well as the AT2 markers *Sftpd* and *Sftpa1*. In addition, C27 is very different from AT1 and AT2 because of its high expression of the epithelial markers *Krt8* and *Krt18*. This is the first evidence to show that bipotent progenitors might exist in the adult mammalian lung. As an example, we show that two Clara cell markers, *Aldh1a1* and *Cyp2f2*, are already expressed in the trachea as early as embryonic day 14.5 (Figure 4E). Lung-resident immune cells play important roles during lung infection and tissue repair. We characterized both *F4/80*-low, *Siglecf*⁺, *Marco*⁺ alveolar macrophages (C3 and C26) and *F4/80*-high, *MHC II*⁺ interstitial macrophages (likely to be C8), which confirms the heterogeneity of lung-resident macrophages. The previously uncharacterized *Pclaf*-high alveolar macrophages (C26) correspond to cycling cells with high expression of *Ccna2*. Dendritic cells (DCs) are derived from various origins and can be categorized as conventional DCs (cDCs) or plasmacytoid DCs (pDCs). Both Cluster 15 and 9 express integrin *Cd11c* and *MHC II* and thus can be inferred as cDCs. These can be further divided into *Cd11b*-expressing cDCs (C15) and *Cd103*-positive cDCs (C9), as previously reported (Kopf et al., 2015). C22 and C32 correspond to dividing cDCs due to their high expression of *Ccnb2*. C5 is predicted to be pDCs because of its expression of *Irf5*, *Irf7*, and *Bst2*. We identified two other DC subtypes, namely, *Gngt2*-high DCs (C13) and *H2-M2*-high DCs (C29).

Among 32 clusters, 3 populations shared common endothelial cells markers (*Pecam1*, *Flt1*, *Chd5*, and *Kdr*), while each cluster possessed distinct markers used to categorize them as artery (cluster 17), vein (cluster 19), and capillary (cluster 20) endothelial cells (Figures 4A, 4B, and 4D). Clusters 10, 11, and 23 broadly expressed genes of the lung mesenchyme program. However, C10 is marked by high *Dcn* expression; C11 is marked by high *Inmt* and *Cxcl14* expression; C23 highly expresses *Acta2* (Figure 4D). Comparison of the adult and embryonic lung single-cell data suggests that at least two types of stromal cells (C10 and C23) are already present in the E14.5 developing lung (Figure S4B). The *Inmt*-high stromal cells (C11) are suspected to be a progeny of the *Dcn*-high population. Based on known markers, additional clusters correspond to *Cd8*-positive T cells (C4), dividing T cells (C24), B cells (C2), Ig-producing B cells (C28), natural killer (NK) cells (C6), eosinophils (C7), nuocytes (C16), neutrophils (C21), basophils (C30), and monocytes (C31) (Figure 4B; Table S4).

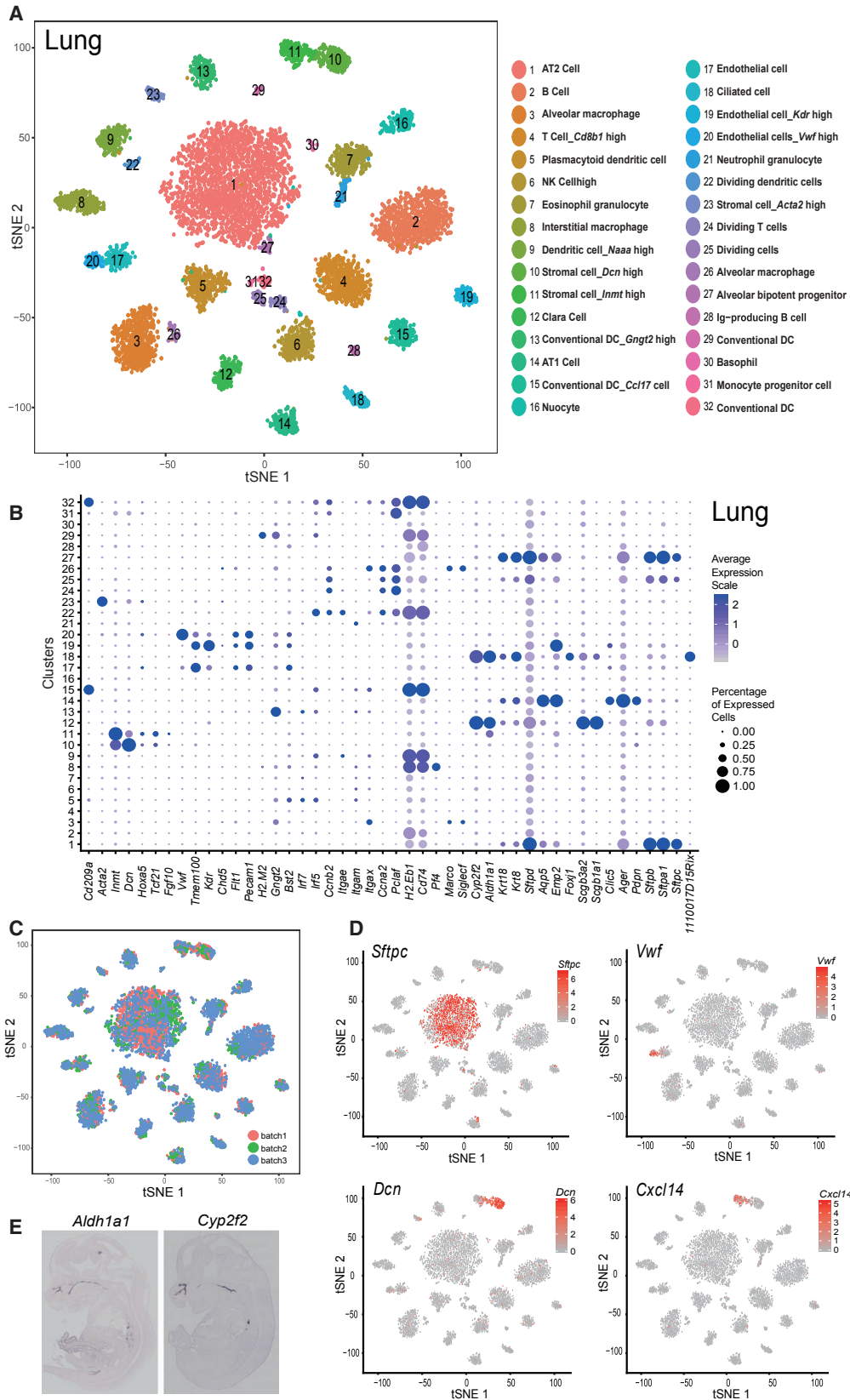
The placenta is an organ that connects the developing fetus to the uterine wall. Analysis of E14.5 mouse placenta tissue revealed 28 cell clusters (Figures 5A and 5B; Table S5). C1, C2, C5, C10, C15, C17, and C25 have trophoblast characteristics. The trophoblast stem cell marker *Tfap2c* was highly expressed in C2, C5, C17, and C25. C2 expresses *Gjb3*, *Hand1*, *Ldoc1*, *Phlda2*, and *Lad1*, resembling recently reported *Gjb3*⁺ progenitor trophoblasts (Nelson et al., 2016) (Figure S4C). C5 is a type of spiral artery trophoblast giant cell (SpA-TGCs). C17 appears to be an uncharacterized trophoblast progenitor type with high expression of *Taf71*, *Isg20*, and *Foxo4*. C25 is a type of labyrinthine trophoblast progenitor that expresses *Tfap2c*, *Epcam*, and *Ly6e*. C1 expresses *Pri7d1*, *Pri7a2*, *Pri3b1*, *Pri8a1*, and *Tpbbp* (Figure 5C). C10 and C15 are both spongiotrophoblasts with high expression of *Dio3* and *Phlda2* (Plasschaert and Bartolomei, 2014). In addition, we identify B cells (C24), NKT cells (C26), uterine natural killer (uNK) cells (C19), erythroid cells (C16), macrophages (C3, C9), neutrophils (C18), basophils (C21), monocytes, DCs (C8), megakaryocytes (C28), hematopoietic stem and progenitor cells (C23), granulocyte monocyte progenitors (C27), and an unknown immune progenitor type (C20). C13 has high expression of decidual stromal cell markers, including *Pri8a2*, *Cryab*, *Adm*, and *Angpt4*. C11 and C22 are two other types of stromal cells. C6 and C7 represent two endodermal cell types that share both epithelial and mesenchymal characteristics. They express *Sox17*, *Lama1*, *Lamb1*, and *Gata4*, suggesting that they may originate from a primitive endoderm lineage. Surprisingly, we found an endodermal cell cluster (C14) that highly expressed *Afp*, *Ttr*, *Apoa1*, and *Apoa2*, a signature that strongly resembles hepatocytes (Figure 5C). Trajectory analysis revealed distinct developmental branches of the endodermal and hepatocyte-like cells in the E14.5 placenta (Figure S4D). Immunocytochemistry located the position of SOX17- and HNF4A-positive

(C) A gene expression heatmap showing the top differentially expressed genes for each cell cluster in mouse virgin mammary gland single-cell data. Yellow corresponds to a high expression level; purple and black correspond to low expression levels.

(D) A t-SNE map of mouse kidney single-cell data. Cells are colored by cell-type cluster.

(E) *In situ* hybridization assay validated expression of the S1 proximal tubule cell markers *Slc5a2* and *Slc34a1*, as well as the kidney endothelial cell markers *Slc4a11* and *Slc31a2* in E14.5 mouse embryonic kidney.

See also Figure S4.



(legend on next page)

cells to the surface of the fetal-placental interface (Figure 5D). The function of the placental hepatocyte-like cells and their influence on other placental cell lineages require further investigation.

Similarly, in the mouse fetal, neonatal, and adult brain, we identified different types of oligodendrocytes, astrocytes, neurons, Schwann cells, ependymal cells, radial glia, microglia, endothelial cells, macrophages, and granulocytes (Figure S4E; Table S5) and then correlated the cell types with the published single-cell signatures from the mouse cortex and hippocampus (Zeisel et al., 2015). In the mouse adult bone marrow, we identified all hematopoietic lineage cells, including neutrophils, monocytes, macrophages, dendritic cells, mast cells, erythroblasts, eosinophils, B cells, T cells, NK cells, and hematopoietic stem and progenitor cells (Table S5) and then constructed a lineage trajectory (Figure S4F). In the same way, we defined cellular components in all the other tissues (Figures S5 and S6). The defined cell-type clusters and cluster-specific markers are summarized in Table S5 and on the MCA website.

Mouse Cell Atlas Reveals the Cross-Tissue Cellular Network

An asset of our study is the use of a single platform for generation of all MCA datasets. Consistency in technology is important for comparison of datasets from different tissues. Using MCA data, we were able to reveal the previously uncharacterized cross-tissue cellular hierarchy.

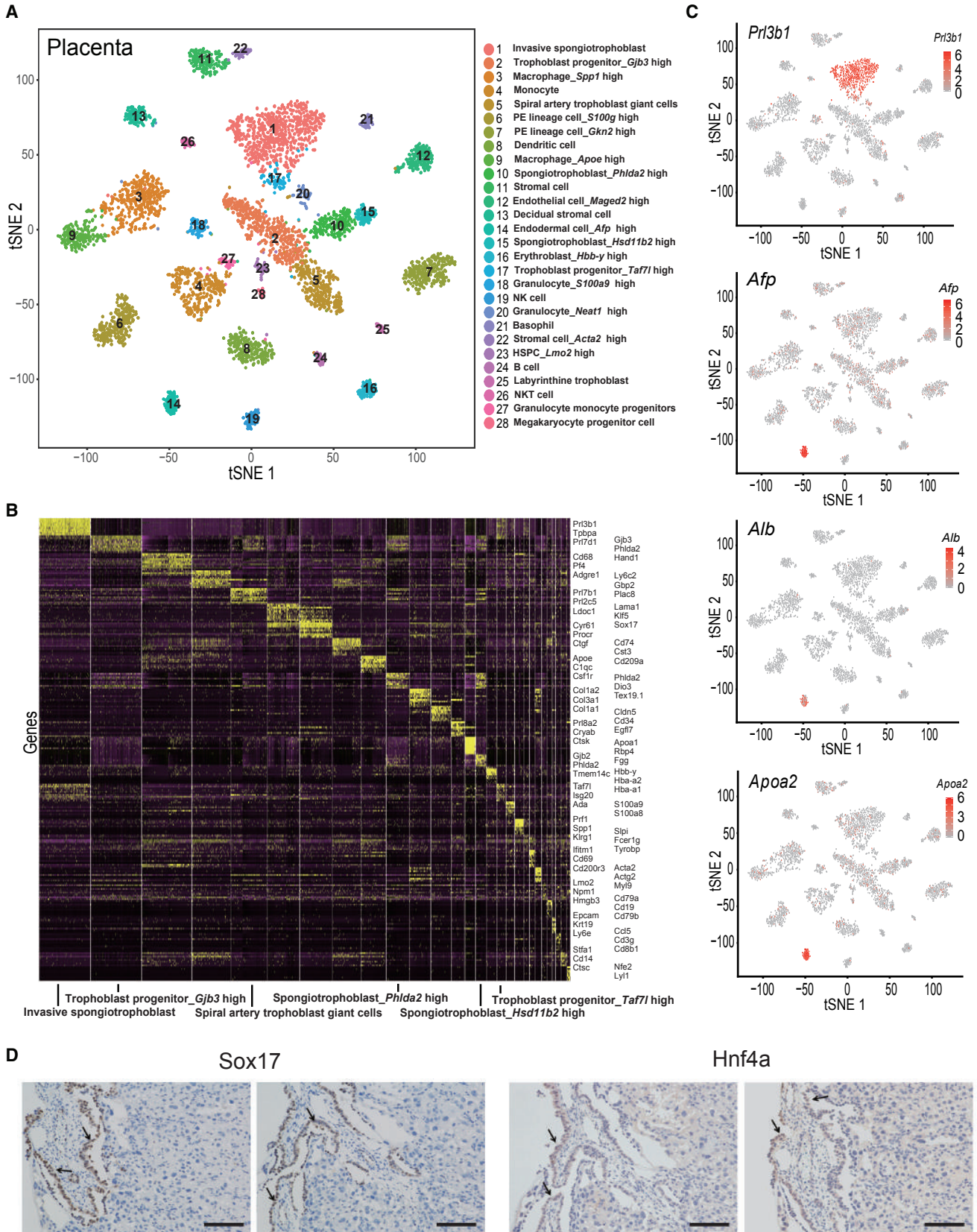
To systematically understand the relationships between different cell types, we built correlation-based networks at the cell-type level and tissue level. To this end, we used the reduced dataset with 60,000 cells grouped into 98 cell clusters defined in Figure 2A (Table S4). To reduce noise, we averaged the expression of every 100 cells within each cluster. We then formed a correlation network using pairwise Spearman correlation between these averaged cells. The resulting correlation map mimics a cellular landscape for different mouse tissue types (Figure 6A). In brief, embryonic stem cells and testicular cells form isolated networks. The mesenchymal-epithelial axis and endothelial-hematopoietic axis are the two most interconnected network groups in the mouse system. The extensive edges along these two axes may help to explain transitions between mesenchymal and epithelial cell types, as well as hematopoiesis emerging from endothelial tissues. Other notable cellular correlations include stomach and small intestine epithelial cells, trophoblasts and primitive endoderm epithelial cells, and astrocytes and oligodendrocytes, as well as different kidney tubule cells (Figure 6A).

In the MCA cellular network, there are two types of cells that contribute to the vast majority of organs and tissues: stromal cells and tissue-resident immune cells. We aimed to investigate

the previously uncharacterized hierarchy of stromal cells in the mammalian system. Stromal cells are connective tissue cells that support the function of parenchymal cells. We first defined stromal cell populations in individual tissues based on expression of collagens, laminins, elastin, and fibronectin. We then integrated the stromal cell data and observed heterogeneity across diverse mouse organs and tissues (Figures 6B and 6C; Table S6). The merged stromal cell data can be divided into 21 groups. C1 cells come from neonatal calvaria and neonatal rib, and they are marked by *Col9a1*, *Col9a3*, and *Col2a1*; they should be related to cartilage development. C21, which is close to C1 on the tSNE map, comes from embryonic mesenchymal tissues. C21 shares similar signatures with C1, but expresses higher levels of cartilage-specific genes, such as *Matn1*, *Matn4*, and *Sox9*. C5 stromal cells have higher levels of imprinted genes, such as *Dlk1*, *H19*, and *Igf2*; they come from neonatal muscle and neonatal skin tissues. C15, which is connected with C5, exhibits myogenic signatures with high expression of *Acta1* and *Tnnc2*. C8 and C20 stromal cells are mainly contributed by mammary gland tissues and are marked by metalloproteinase and fibronectin expression, respectively. C7 contains contributions from a wide range of tissues, such as the lung, bladder, testis, stomach, liver, and small intestine. This type of stromal cell is marked by high *Inmt* expression; it also expresses the antioxidant enzyme *Sod3*, which is thought to protect tissues from oxidative stress. C12 and C17 are two types of bladder-specific stromal cells with high expression of *Bmp4* and *Wnt2*. C12 expresses high levels of *Cxcl12* and the proliferation marker *Ifitm1*, while C17 expresses a high level of *Bmp5*. Both the testis and neonatal heart contribute to C9 stromal cells. C6 represents pancreas-specific stromal cells that express *Col15a1*, *Gdf10*, and *Col4a1*. Uterus-specific C2 is a novel stromal cell population that is marked by *Col6a4* and *Col6a3*, as well as many predicted genes, such as *Gm11361*, *Gm12248*, and *Gm8797*. C10 also comes from the uterus; it highly expresses *Mmp11* and *Cxcl12*. C19 are placental decidual stromal cells that express *Prf8a2* and *Cryab*. C4 and C13 are both from bone marrow mesenchymal samples. C4 corresponds to osteoblasts with high levels of *Bglap*, *Bglap2*, *Col1a1*, and *Col11a1*. C13 is marked by chondrocyte genes, such as *Col10a1*, *Col2a1*, and *Matn3*. With contributions from the most diverse tissues, C3 is identified as pericytes due to *Acta2*, *Pdgfrb*, *Mcam* (*Cd146*), and *Vim* expression. C14 also represents pericytes; it differs from C3 by high *Cspg4* (*NG-2*) expression. Notably, C3 and C14 (including bone marrow-specific pericytes) express high levels of important signaling molecules, such as *Kitl*, *Pdgfra*, and *Tgfb2*, suggesting that they have important roles in regulating tissue microenvironments. C16 is a kidney-specific stromal type. C11 cells correspond to myoblasts. C18 corresponds

Figure 4. Resolving Cellular Heterogeneity in Lung

- (A) A t-SNE map of mouse lung single-cell data. Cells are colored by cell-type cluster.
 (B) Dot plot visualization of each cell type in lung single-cell data. The size of the dot encodes the percentage of cells within a cell type, and the color encodes the average expression level.
 (C) A t-SNE map of mouse lung single-cell data. Cells are colored by experimental batch.
 (D) t-SNE maps of mouse lung single-cell data with cells colored based on the expression of marker genes for particular cell types. Gene expression levels are indicated by shades of red.
 (E) *In situ* hybridization assay for the two identified Clara cell markers *Aldh1a1* and *Cyp2f2* in E14.5 mouse embryos.



(legend on next page)

to high-passage mouse embryonic fibroblasts in culture. Many of these stromal cells are already present in the tissues at embryonic stages (Table S5). They play important roles in proper organ development, regeneration, and function.

Tissue-resident macrophages are unique myeloid cells that self-renew independent of hematopoietic stem cells (Gomez Perdiguero et al., 2015). However, cross-tissue heterogeneity of tissue-resident macrophages has not been well characterized. We merged single-cell transcriptome data of macrophages from diverse tissues to search for subclasses. We found that macrophages can be classified into 13 subtypes (Figures 6D and 6E; Table S6). They show preferential expression of different C-type lectin domain family proteins. C1 macrophages, which come from the uterus, ovary, and mammary gland, highly express *Cd74*, *Ms4a7*, *C1qc*, *Clec7a*, and *F4/80* (*Adgre1*). C4 macrophages can be found in a wide range of tissues, such as the mammary gland, testis, bladder, and pancreas; they express *Cd209f*, *Clec10a*, and *Cd209g*. Macrophages from the placenta (C5) specifically express *Spp1*, *Arg1*, *Clec4d*, and *Clec4n*. In the lung, two types of macrophages were identified: interstitial macrophages were merged into C1; alveolar macrophages, enriched with the markers *Ear1*, *Ear2*, *Marco*, and *Siglecf.*, were grouped into C7. C8 was identified as Kupffer cells from the liver, with specific expression of *Clec4f*, *Clec1b*, and *Macro*. C12, which is another liver-specific macrophage subtype, expresses *Clec4e* and *Clec4d*. C9 macrophages are of peripheral blood origin and express *Adgre4*, *Adgre5*, *Clec4a3*, and *Clec4a1*. C13 cells are mammary gland-resident macrophages that highly express *Retnla*, *Clec4b1*, and *Ear2*. Through organ maturation, tissue-resident macrophages appear to become increasingly specialized to adapt to the organ environment and organ function.

In the same way, we performed cross-tissue analysis for both endothelial cells and dendritic cells (Table S6). Our analyses revealed a wide diversity of tissue-specific mesenchymal and immune cells. The accurate identification of these cellular subtypes should contribute to improved understanding and monitoring of tissue regeneration. Stromal cells, endothelial cells, macrophages, and dendritic cells are integral parts of the tissue microenvironment. Successful organ engineering depends not only on efficient generation of parenchymal cells but also on proper integration of the supporting cells.

Single-Cell MCA Analysis Pipeline Defines Cell Types with Accuracy and Efficiency

Having constructed a basic scheme for a mouse cell atlas database, we then sought to utilize the defined cell types as references for an unbiased classification of single-cell transcriptome data. Previous studies have used population gene expression profiling for determination of engineered cell types (Cahan

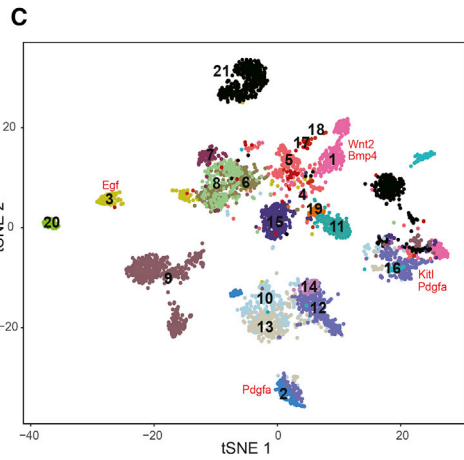
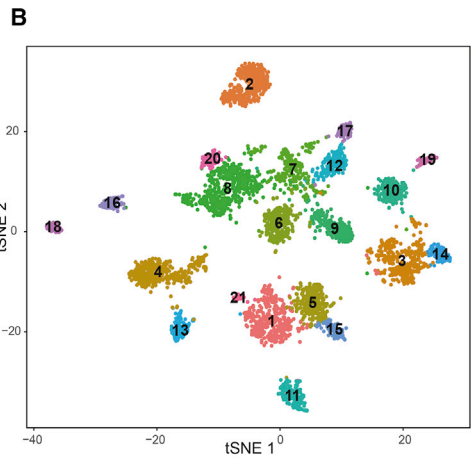
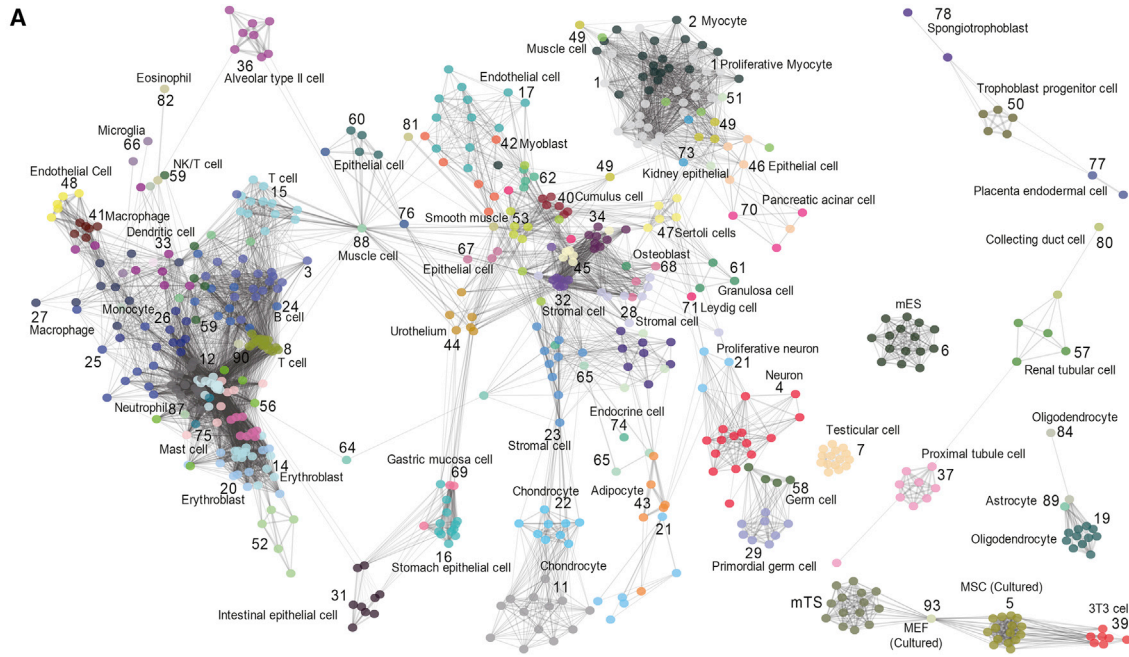
et al., 2014; Morris et al., 2014). However, these methods overlook heterogeneity within cell populations, and the reference cell types are not comprehensive. In the mouse cell atlas, we seek to establish a pipeline for accurate cell type determination at the single-cell level. We built the scMCA analysis tool for this purpose.

In brief, the scMCA analysis was set up through following steps. First, we integrated Microwell-seq data with published Drop-seq and 10 × Genomics data (see the STAR Methods). We then clustered the data into 894 cell-type clusters and determined the average expression in each cluster for transcriptome references. Finally, input single-cell DGE was compared with each transcriptome reference to provide a match score based on gene expression correlation (Figure 7A). In high-throughput scRNA-seq experiments, such as Microwell-seq, Drop-seq, and 10 × Genomics, sequencing depth is usually sacrificed; the average gene number detected for each cell is ~1,000. However, by adding a cluster and averaging step, we obtained cell-type transcriptome references with digital expression of more than 10 K genes. Our transcriptome references perform well in scMCA experiments; a combination of the top 10 markers from each cell type was sufficient as a gene set for efficient scMCA calculation (Figures S7A–S7C). As an example, we show that the placental *Afp*⁺ endodermal cells we identified are clearly mapped to fetal liver hepatocytes, with only one spongiotrophoblast contamination (Figure S7D).

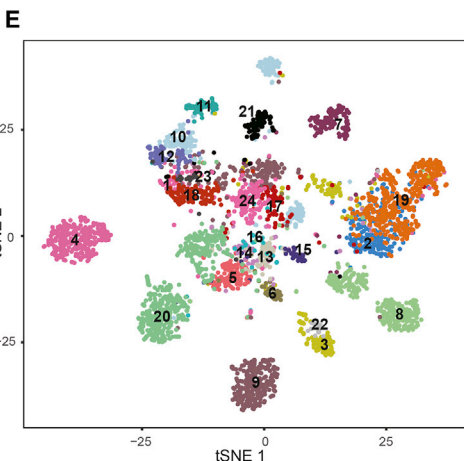
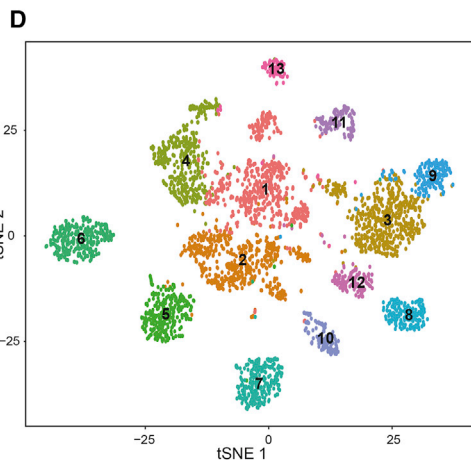
Importantly, the scMCA analysis is compatible with different technologies. We first examined C1 single-cell data of mouse embryonic distal lung epithelium (Treutlein et al., 2014). Figure 7B shows all correlated cell-type clusters for input single cells. We found that AT1, AT2, Clara, and Ciliated cells defined by Treutlein et al. (2016) exhibit the best correlation with our adult AT1, AT2, Clara, and Ciliated cell-type references, respectively. The bipotential progenitors express both AT1 and AT2 signature genes and correlate with our bipotential progenitor reference. The only discrepancy between our analysis and the published study was a single AT1 cell that appeared to show higher correlation to stromal cells. Next, we examined single-cell data of mouse hematopoietic stem cell formation (Zhou et al., 2016). The data were generated using the classical polyA tailing method for single-cell mRNA amplification (Tang et al., 2009). The scMCA results provide a description for the process of HSC emergence (Figure 7C). Endothelial cells from the study of Zhou et al. (2016) strongly correlate with endothelial cell clusters in our database. E12 HSCs and E14 HSCs correlate best with our fetal liver hematopoietic stem and progenitor cell cluster. Interestingly, T1 and T2 PreHSCs exhibit both endothelial and HSC signatures. The analysis revealed a single ganglion cell as well as a single macrophage contamination in the dataset. Similarly, we

Figure 5. Resolving Cellular Heterogeneity in Placenta

- (A) A t-SNE map of E14.5 mouse placenta single-cell data. Cells are colored by cell-type cluster.
 (B) A gene expression heatmap showing top differentially expressed genes for each cell cluster in E14.5 mouse placenta single-cell data. Yellow corresponds to high expression level; purple and black correspond to low expression levels.
 (C) t-SNE maps of E14.5 mouse placenta single-cell data with cells colored based on the expression of marker genes for particular cell types. Gene expression levels are indicated by shades of red.
 (D) Immunohistochemistry assay of endodermal cell marker SOX17 and HNF4A in E14.5 mouse placenta samples. Arrow, endodermal cells. Scale bars, 100 μm. See also Figures S5 and S6 and Table S6.



- 1 Bladder
- 2 Embryonic Mesenchyme
- 3 Kidney
- 4 Liver
- 5 Lung
- 6 Mammary Gland (Involution)
- 7 Mammary Gland (Pregnancy)
- 8 Mammary Gland (Virgin)
- 9 Bone Marrow Mesenchyme
- 10 Neonatal Calvaria
- 11 Neonatal Heart
- 12 Neonatal Muscle
- 13 Neonatal Rib
- 14 Neonatal Skin
- 15 Pancreas
- 16 Placenta
- 17 Small Intestine
- 18 Stomach
- 19 Testis
- 20 MEF (Cultured)
- 21 Uterus



- 1 Bladder
- 2 Bone Marrow
- 3 Bone Marrow cKit
- 4 Brain
- 5 Embryonic Mesenchyme
- 6 Fetal Liver
- 7 Kidney
- 8 Liver
- 9 Lung
- 10 Mammary Gland (Involution)
- 11 Mammary Gland (Pregnancy)
- 12 Mammary Gland (Virgin)
- 13 Neonatal Heart
- 14 Neonatal Muscle
- 15 Neonatal Rib
- 16 Neonatal Skin
- 17 Ovary
- 18 Pancreas
- 19 Peripheral Blood
- 20 Placenta
- 21 Small Intestine
- 22 Spleen
- 23 Testis
- 24 Uterus

(legend on next page)

reanalyzed CEL-seq single-cell data by Rizvi et al. (2017) and observed the transition from pluripotency to neuron lineages. We also analyzed C1 single-cell data from fibroblast to neuron transdifferentiation (Treutlein et al., 2016) and confirmed activation of a myogenetic program in the process and simultaneous generation of muscle progenitor cells.

We propose using the scMCA pipeline for dissection of more complex systems. Ezh2 is an important component of Polycomb repressive complex 2, which mediates H3K27me3 and regulates pluripotency in mESCs (Shen et al., 2008). Differentiation of Ezh2^{-/-} ES cells is impaired. We aimed to utilize the single-cell approach for comprehensive analysis of the roles of Ezh2 in different developmental lineages. We compared day 14 embryonic body (EB) cells from wild-type and Ezh2^{-/-} mESCs (Shen et al., 2008) using Microwell-seq coupled with scMCA methodology (Figures 7D and 7E). We found that in wild-type EBs, differentiated cells are mainly mapped to mesoderm MEF cells, smooth muscle cells, ectoderm neuronal cells, and endoderm hepatocytes. However, in Ezh2^{-/-} EBs, the majority of the differentiated cells are mapped to two types of placental cells that appear to be of primitive endoderm origin, suggesting differentiation to extraembryonic tissues (Figure 7E; Table S5). The two differentiated cell types express high levels of *Sox17*, *Lamb1*, *Lama1* and *Aqp8*, and *Vim* and *Gata4*, respectively. Immunostaining results suggested a significantly higher percentage of GATA4- and SOX17-positive cells in the Ezh2^{-/-} EBs (Figure 7F). Pooled gene expression analysis further confirmed aberrant activation of the extraembryonic lineage program during differentiation of Ezh2^{-/-} ES cells (Figures S7G and S7H). We infer that Ezh2 plays essential roles in regulating the epigenetic barrier between embryonic and extraembryonic tissues. Deletion of Ezh2 enables mESCs to cross the barrier and transdifferentiate into extraembryonic cell types.

In summary, we have developed a scMCA application that faithfully predicts cell types using single-cell data generated from a wide range of technologies.

DISCUSSION

The Microwell-seq method holds advantages over other related technologies, largely related to superiority in cost and convenience. A silicon wafer containing ~100,000 microwells can be used to make hundreds of polydimethylsiloxane (PDMS) micropillar arrays, which can be used multiple times to create hundreds of agarose-constructed microwell arrays. The magnetic property of barcoded beads allows for their efficient collection. The beads that remain outside of the microwells can be reused. The cost of sequencing library generation for each cell is estimated to be under 0.02 USD. In addition, microwells with too

many cell doublets can be washed to reduce doublet rate, ensuring that Microwell-seq yields high-fidelity single-cell libraries. Experiments can be readily scaled up by simultaneous handling of multiple microwells. Finally, the washing and quality checking steps before bead capture also ensure removal of cell debris and clumps that might cause contamination in droplet-based methods. As exemplified by our study, Microwell-seq can be used with all tissue types. It is a portable, efficient, faithful, and inexpensive high-throughput scRNA-seq platform.

Using Microwell-seq, we profiled more than 50 mouse organ, tissue, and cell lines. We dissected cellular components in many tissues that are not well characterized. It is important to mention that for every organ and tissue, we profiled not only the tissue-specific cell lineages but also the tissue-resident stromal and immune cell types to provide information on tissue microenvironments. Due to space limitations, here, we have only presented results for the mammary gland, lung, kidney, testis, and placenta in the main text. Additional data are deposited on the MCA website. The complete annotation of the dataset is a continuous project. Although we have defined most of the major clusters on MCA, there are many subclusters that remain to be defined by more focused analysis of a particular population (Table S6). The MCA website welcomes public discussion, correction, and validation of annotated cell types. Moreover, the MCA database will be updated with newly published high-throughput single-cell data from relevant technologies, such as 10xGenomics, Drop-seq, In-Drop, Seq-well, and ddSEQ. Future directions also include integration of proteomic data (such as CyTOF data) and spatial information (such as *in situ* data) with the MCA database to provide a more complete atlas with multi-omic information.

In conclusion, Microwell-seq provides a highly accessible and inexpensive platform for single-cell RNA-seq technology. Its simplicity should facilitate broad use in the community. In addition, we present a mouse cell atlas, which is the most comprehensive mammalian single-cell data resource to date. We introduced the scMCA analysis for accurate cell-type identification based on single-cell transcriptome data. These new methodologies should accelerate progress in dissecting cell types and functions in the mammalian system.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Mouse strains and husbandry

Figure 6. Cross-Tissue Cellular Network

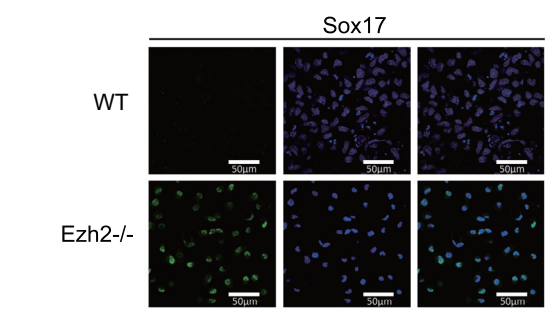
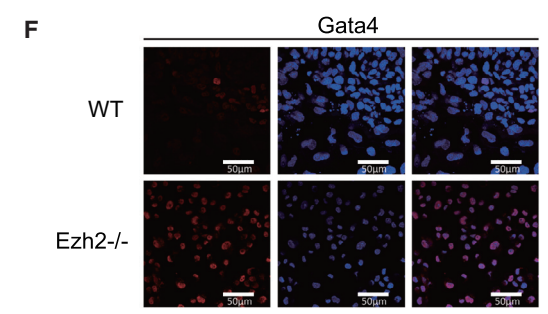
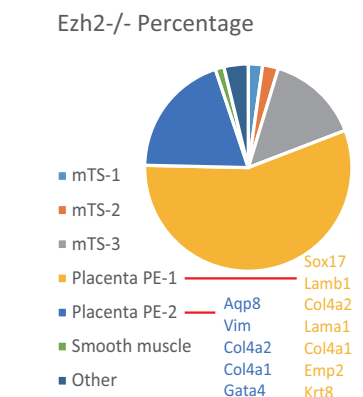
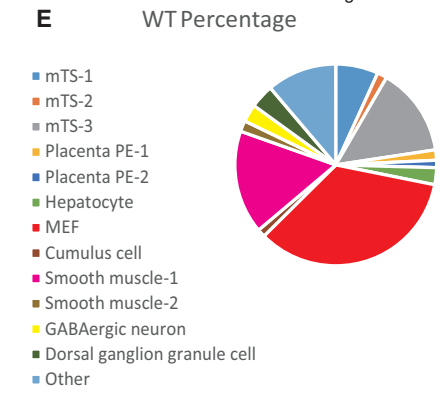
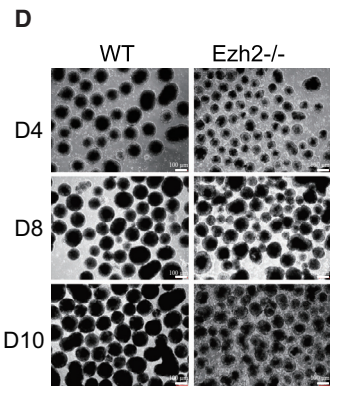
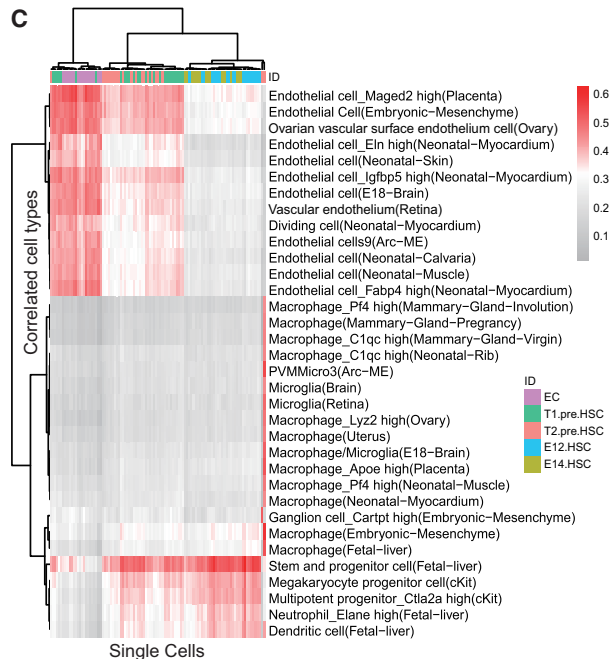
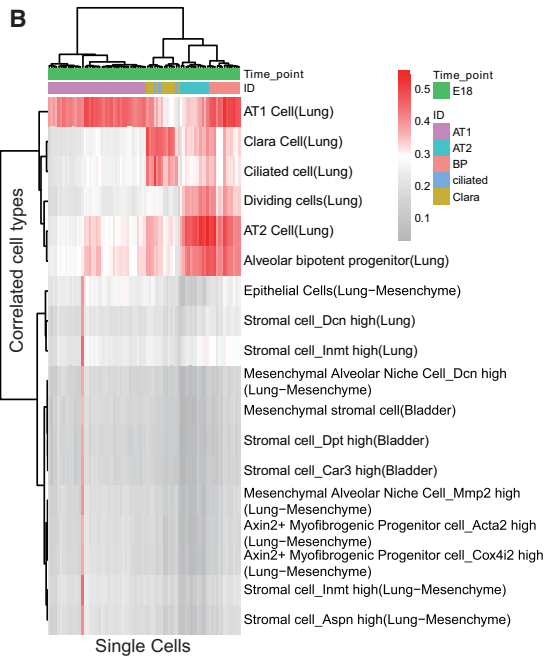
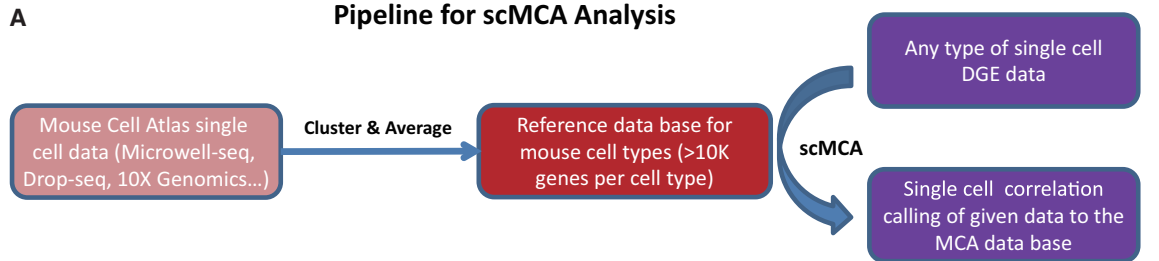
(A) A correlation network showing relationships among 98 cell groups defined using MCA data. Each node represents the averaged expression of every 100 cells within each of the 98 cell groups defined in Figure 2B. Each edge corresponds to a network correlation between two nodes. Each node is colored by their cluster ID.

(B) A t-SNE map of single-cell data for mouse tissue-specific stromal cells. Cells are colored by stromal cell subtype.

(C) A t-SNE map of single-cell data for mouse tissue-specific stromal cells. Cells are colored by tissue type.

(D) A t-SNE map of single-cell data for mouse tissue-resident macrophages. Cells are colored by macrophage subtype.

(E) A t-SNE map of single-cell data for mouse tissue-resident macrophages. Cells are colored by tissue type.



(legend on next page)

- **METHOD DETAILS**
 - Fabrication of microwell device
 - Synthesis of barcoded beads
 - Cell collection and lysis
 - Reverse transcription
 - Exonuclease I treatment
 - cDNA amplification
 - Transposase fragmentation and selective PCR
 - Cell preparation
 - qPCR analysis
 - Immunofluorescence
 - Immunohistochemistry
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Processing of the Microwell-seq Data
 - Cell cycle analysis of 293T Cells
 - Batch removal for cross tissue comparison
 - Cell-cell interaction network
 - Comparative analysis of different scRNA-seq methods
 - Cell type analysis
 - Single cell trajectory analysis
 - MCA website construction
 - Single cell MCA analysis
 - Comparison of scMCA, scmap and randomForest
- **DATA AND SOFTWARE AVAILABILITY**
 - Data Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables and can be found with this article online at <https://doi.org/10.1016/j.cell.2018.02.001>.

ACKNOWLEDGMENTS

We thank S.R. Quake, L. Wang, L. Lu, D. Zhang, L. Shen, S. Ying, B. Liu, F. Tang, J. Yu, T. Cheng, G. Fan, X. Li, X. Yu, and L. Sun for help with the study. We thank Annoroad, VeritasGenetics, G-BIO, Novogene, and Geneseeq for deep sequencing experiments. We thank Vazyme for supplying customized enzymes for the study. G.G. is a participant of the Cell Atlas Project (Zhejiang University Stem Cell Institute), Alliance for Atlas of Blood Cells (China), and Human Cell Atlas Project (International). This work was supported by grants from the National Natural Science Foundation of China (31722027, 81770188, and 31701290), the National Key Program on Stem Cell and Translational Research (2017YFA0103401), the 973 Program (2015CB964900), Zhejiang Provincial Natural Science Foundation of China (R17H080001), and Fundamental Research Funds for the Central Universities (2016XZZX002-04).

AUTHOR CONTRIBUTIONS

The project was designed by G.G. Experiments were carried out by X.H., R.W., S.L., Z.Z., H.C., F.Y., M.J., X.J., J.M., Y.C., C.L., Y.W., R.Y., and G.G. Compu-

tational analyses were performed by Y.Z., L.F., H.S., A.S., D.H., Y.X., and W.H. The paper was written by G.G., X.H., R.W., J.X., Q.F., T.L., H.H., S.H.O., G.-C.Y., and M.C.

DECLARATION OF INTERESTS

G.G. has submitted a patent application related to the Microwell-seq method reported in this paper.

Received: September 25, 2017

Revised: December 15, 2017

Accepted: January 30, 2018

Published: February 22, 2018

REFERENCES

- Bendall, S.C., Simonds, E.F., Qiu, P., Amir, A.D., Krutzik, P.O., Finck, R., Bruggner, R.V., Melamed, R., Trejo, A., Ornatsky, O.I., et al. (2011). Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* 332, 687–696.
- Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* 523, 486–490.
- Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. *Cell* 158, 903–915.
- Chen, K.H., Boettiger, A.N., Moffitt, J.R., Wang, S., and Zhuang, X. (2015). RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348, aaa6090.
- Cheval, L., Pierrat, F., Dossat, C., Genete, M., Imbert-Teboul, M., Duong Van Huyen, J.P., Poulain, J., Wincker, P., Weissenbach, J., Piquemal, D., and Doucet, A. (2011). Atlas of gene expression in the mouse kidney: new features of glomerular parietal cells. *Physiol. Genomics* 43, 161–173.
- Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348, 910–914.
- Dalerba, P., Kalisky, T., Sahoo, D., Rajendran, P.S., Rothenberg, M.E., Leyrat, A.A., Sim, S., Okamoto, J., Johnston, D.M., Qian, D., et al. (2011). Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat. Biotechnol.* 29, 1120–1127.
- de Jonge, H.J., Fehrmann, R.S., de Bont, E.S., Hofstra, R.M., Gerbens, F., Kamps, W.A., de Vries, E.G., van der Zee, A.G., te Meerman, G.J., and ter Elst, A. (2007). Evidence based selection of housekeeping genes. *PLoS ONE* 2, e898.
- Deng, Q., Ramsköld, D., Reinis, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* 343, 193–196.
- Diez-Roux, G., Banfi, S., Sultan, M., Geffers, L., Anand, S., Rozado, D., Magen, A., Canidio, E., Pagani, M., Peluso, I., et al. (2011). A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol.* 9, e1000582.

Figure 7. Broad Applicability of scMCA Pipeline

(A) A diagram showing the pipeline for scMCA analysis.

(B) scMCA results of mouse embryonic distal lung epithelium data (Treutlein et al., 2014). Each row represents one cell type in our reference. Each column represents one single cell in the customer dataset. Red means high correlation; gray means low correlation.

(C) scMCA results of mouse HSC emergence data (Zhou et al., 2016). Each row represents one cell type in our reference. Each column represents one single cell in the customer dataset. Red means high correlation; gray means low correlation.

(D) Pictures showing the morphology of EBs generated from wild-type and *Ezh2*^{-/-} mES cells.

(E) A summary of scMCA results for differentiated cells from wild-type and *Ezh2*^{-/-} EBs. Each cell was labeled according to the maximum correlation with one of our reference cell types.

(F) Immunofluorescence assay for the PE cell markers SOX17 and GATA4 in the EB cells generated from wild-type and *Ezh2*^{-/-} ES cells. Scale bars, 50 μ m.

See also Figure S7.

- Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.
- Fan, H.C., Fu, G.K., and Fodor, S.P. (2015). Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science* 347, 1258367.
- Gierahn, T.M., Wadsworth, M.H., 2nd, Hughes, T.K., Bryson, B.D., Butler, A., Satija, R., Fortune, S., Love, J.C., and Shalek, A.K. (2017). Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* 14, 395–398.
- Gomez Perdiguero, E., Klapproth, K., Schulz, C., Busch, K., Azzoni, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M.F., Geissmann, F., and Rodewald, H.R. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythroid progenitors. *Nature* 518, 547–551.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.
- Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. *Dev. Cell* 18, 675–685.
- Guo, F., Li, L., Li, J., Wu, X., Hu, B., Zhu, P., Wen, L., and Tang, F. (2017). Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Res.* 27, 967–988.
- Han, X., Yu, H., Huang, D., Xu, Y., Saadatpour, A., Li, X., Wang, L., Yu, J., Pineilo, L., Lai, S., et al. (2017). A molecular roadmap for induced multi-lineage trans-differentiation of fibroblasts by chemical combinations. *Cell Res.* 27, 386–401.
- Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep.* 2, 666–673.
- Herring, C.A., Banerjee, A., McKinley, E.T., Simmons, A.J., Ping, J., Roland, J.T., Franklin, J.L., Liu, Q., Gerdes, M.J., Coffey, R.J., and Lau, K.S. (2017). Unsupervised trajectory analysis of single-cell RNA-seq and imaging data reveals alternative Tuft cell origins in the gut. Published online November 7, 2017. [10.1016/j.cels.2017.10.012](https://doi.org/10.1016/j.cels.2017.10.012).
- Hou, Y., Song, L., Zhu, P., Zhang, B., Tao, Y., Xu, X., Li, F., Wu, K., Liang, J., Shao, D., et al. (2012). Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* 148, 873–885.
- Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lönnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods* 11, 163–166.
- Jin, W., Tang, Q., Wan, M., Cui, K., Zhang, Y., Ren, G., Ni, B., Sklar, J., Przytycka, T.M., Childs, R., et al. (2015). Genome-wide detection of DNase I hypersensitive sites in single cells and FFPE tissue samples. *Nature* 528, 142–146.
- Ke, R., Mignardi, M., Pacureanu, A., Svedlund, J., Botling, J., Wählby, C., and Nilsson, M. (2013). In situ sequencing for RNA analysis in preserved tissue and cells. *Nat. Methods* 10, 857–860.
- Kiselev, V.Y., Yiu, A., and Hemberg, M. (2017). scmap - A tool for unsupervised projection of single cell RNA-seq data. [bioRxiv. https://doi.org/10.1101/150292](https://doi.org/10.1101/150292).
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201.
- Kopf, M., Schneider, C., and Nobs, S.P. (2015). The development and function of lung-resident macrophages and dendritic cells. *Nat. Immunol.* 16, 36–44.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26.
- Lee, J.H., Daugharthy, E.R., Scheiman, J., Kalhor, R., Yang, J.L., Ferrante, T.C., Terry, R., Jeanty, S.S., Li, C., Amamoto, R., et al. (2014). Highly multiplexed subcellular RNA sequencing in situ. *Science* 343, 1360–1363.
- Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News* 2, 18–22.
- Lorthongpanich, C., Cheow, L.F., Balu, S., Quake, S.R., Knowles, B.B., Burkholder, W.F., Solter, D., and Messerschmidt, D.M. (2013). Single-cell DNA methylation analysis reveals epigenetic chimerism in preimplantation embryos. *Science* 341, 1110–1112.
- Macosko, E.Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214.
- Mooijman, D., Dey, S.S., Boisset, J.C., Crosetto, N., and van Oudenaarden, A. (2016). Single-cell 5hmC sequencing reveals chromosome-wide cell-to-cell variability and enables lineage reconstruction. *Nat. Biotechnol.* 34, 852–856.
- Morris, S.A., Cahan, P., Li, H., Zhao, A.M., San Roman, A.K., Shivdasani, R.A., Collins, J.J., and Daley, G.Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. *Cell* 158, 889–902.
- Nagano, T., Lubling, Y., Stevens, T.J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E.D., Tanay, A., and Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64.
- Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., Cook, K., Stepansky, A., Levy, D., Esposito, D., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature* 472, 90–94.
- Nelson, A.C., Mould, A.W., Bikoff, E.K., and Robertson, E.J. (2016). Single-cell RNA-seq reveals cell type-specific transcriptional signatures at the maternal-foetal interface during pregnancy. *Nat. Commun.* 7, 11414.
- Park, J., Shrestha, R., Qiu, C., Kondo, A., Huang, S., Werth, M., Li, M., Barasch, J., and Susztak, K. (2017). Comprehensive single cell RNAseq analysis of the kidney reveals novel cell types and unexpected cell plasticity. [bioRxiv. https://doi.org/10.1101/203125](https://doi.org/10.1101/203125).
- Picelli, S., Björklund, A.K., Faridani, O.R., Sagasser, S., Winberg, G., and Sandberg, R. (2013). Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* 10, 1096–1098.
- Plasschaert, R.N., and Bartolomei, M.S. (2014). Genomic imprinting in development, growth, behavior and stem cells. *Development* 141, 1805–1813.
- Posfai, E., Petropoulos, S., de Barros, F.R.O., Schell, J.P., Jurisica, I., Sandberg, R., Lanner, F., and Rossant, J. (2017). Position- and Hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. *eLife* 6, e22906.
- Ramsköld, D., Luo, S., Wang, Y.C., Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtkova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* 30, 777–782.
- Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al.; Human Cell Atlas Meeting Participants (2017). The Human Cell Atlas. *eLife* 6, e27041.
- Rios, A.C., Fu, N.Y., Lindeman, G.J., and Visvader, J.E. (2014). In situ identification of bipotent stem cells in the mammary gland. *Nature* 506, 322–327.
- Rizvi, A.H., Camara, P.G., Kandror, E.K., Roberts, T.J., Schieren, I., Maniatis, T., and Rabadan, R. (2017). Single-cell topological RNA-seq analysis reveals insights into cellular differentiation and development. *Nat. Biotechnol.* 35, 551–560.
- Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502.
- Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. *Nature* 439, 84–88.
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublot, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
- Shen, X., Liu, Y., Hsu, Y.J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G.C., and Orkin, S.H. (2008). EZH1 mediates methylation on histone H3 lysine 27 and

- complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* 32, 491–502.
- Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nat. Methods* 11, 817–820.
- Svensson, V., Natarajan, K.N., Ly, L.H., Miragaia, R.J., Labalette, C., Macaulay, I.C., Cvejic, A., and Teichmann, S.A. (2017). Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* 14, 381–387.
- Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. *Science* 282, 2072–2075.
- Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382.
- Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* 509, 371–375.
- Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., and Quake, S.R. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* 534, 391–395.
- Wang, J., Fan, H.C., Behr, B., and Quake, S.R. (2012). Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150, 402–412.
- Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., Brown, P.O., and Botstein, D. (2002). Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* 13, 1977–2000.
- Xu, X., Hou, Y., Yin, X., Bao, L., Tang, A., Song, L., Li, F., Tsang, S., Wu, K., Wu, H., et al. (2012). Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 148, 886–895.
- Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138–1142.
- Zhou, F., Li, X., Wang, W., Zhu, P., Zhou, J., He, W., Ding, M., Xiong, F., Zheng, X., Li, Z., et al. (2016). Tracing haematopoietic stem cell formation at single-cell resolution. *Nature* 533, 487–492.
- Ziegenhain, C., Vieth, B., Parekh, S., Reinius, B., Guillaumet-Adkins, A., Smets, M., Leonhardt, H., Heyn, H., Hellmann, I., and Enard, W. (2017). Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* 65, 631–643.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-Human SOX17 (Polyclonal)	R&D Systems	Cat# AF1924
Anti-Human HNF-4 α (clone C19)	Santa Cruz Biotechnology	Cat# sc-6556
Anti-Human GATA4 (clone G-4)	Santa Cruz Biotechnology	Cat# sc-25310
Anti-mice CD117 (c-Kit) (2B8)	eBioscience	Cat# 12-5981-82
Chemicals, Peptides, and Recombinant Proteins		
0.25% Trypsin	ThermoFisher Scientific	Cat# 15050057
FGF4	Peptidech	Cat# AF-100-31
ESGRO [®] Leukaemia inhibitory factor (LIF)	Millipore	Cat# ESG1106
Collagenase type I	Worthington	Cat# LS004196
Critical Commercial Assays		
PrimeScript II Reverse Transcriptase	Takara	Cat# 2690A
Exonuclease I	NEB	Cat# M0293L
TruePrep DNA Library Prep Kit V2 for Illumina	Vazyme Biotech	Cat# TD513
EasyPure RNA Kit	Transgen Biotech	Cat# ER101
TransScript All-in-One First-Strand cDNA Synthesis SuperMix for qPCR (One-Step gDNA Removal)	Transgen Biotech	Cat# AT341-01
KAPA HiFi HotStart ReadyMix (2X)	Kapa Biosystems	Cat# KK2602
2 × Phanta Master Mix	Vazyme Biotech	Cat# P511-AA
Hieff qPCR SYBR Green Master Mix (No Rox)	Yeasen	Cat# 11201ES03
MojoSort Mouse CD45 Nanobeads	BioLegend	Cat# 480027
Magnetic beads	Knowledge & Benefit Sphere Tech.	Cat# MC20000
AMPure XP beads	Beckman Coulter	Cat# A63881
Deposited Data		
Single cell RNA-seq raw, processed and DGE data	NCBI Gene Expression Omnibus	GSE108097
Single cell RNA-seq DGE data	Figshare	https://figshare.com/s/865e694ad06d5857db4b
Experimental Models: Cell Lines		
Mouse: 3T3	Stuart Orkin Lab	RRID:CVCL_0594
Mouse: Embryonic stem cell (Cj7)	Stuart H. Orkin Lab	RRID:CVCL_C316
Mouse: Embryonic stem cell (E14)	George Q. Daley Lab	RRID:CVCL_C320
Mouse: Ezh2 ^{-/-} Embryonic stem cell (Cj7)	Stuart H. Orkin Lab	Shen et al., 2008
Mouse: Trophoblast stem cell	Janet Rossant Lab	Tanaka et al., 1998
Human: 293T	Stuart Orkin Lab	RRID:CVCL_0063
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	Shanghai SLAC Laboratory Animal	N/A
Oligonucleotides		
Primer for Synthesis of barcoded beads, see Table S1	Sangon Biotech	A: Lot# 9306358586– Lot# 9306358681 B: Lot# 8306023817–Lot# 8306023912 C: Lot# 8307373546–Lot# 8307373641
indexed bead seqA:5'-Bead-Linker-TTTAGGGATAAC AGGGTAATAAGCAGTGGTATCAACGCAGAGTACGT jjjjjjCGACTCACTACAGGGjjjjjjTCGGTGACACGATCG jjjjjjnnnnnnTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
indexed bead seqB:5'-Bead-Linker-TTTAGGGATAAC AGGGTAATAAGCAGTGGTATCAACGCAGAGTACjjjjj CGACTCACTACAGGGjjjjjTCGGTGACACGATCGjjjjj nnnnnTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	This paper	N/A
Primer for Reverse transcription:AAGCAGTGGTATCA ACGCAGAGTGAAT/rG/rG/iXNA_G	Sangon Biotech	Lot# 7308193208
Primer for cDNA amplification:AAGCAGTGGTATCAA CGCAGAGT	Sangon Biotech	Lot# 9308193206
Primer for library construction and sequencing, see Table S2	This paper	N/A
Software and Algorithms		
BBMap	https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/	https://github.com/BioInfoTools/BBMap
Dropseq_tools-1.12	Macosko et al., 2015	http://mccarrolllab.com/dropseq/
Seurat	Satija et al., 2015	http://satijalab.org/seurat/
p-Creode	Herring et al., 2017	https://github.com/KenLauLab/pCreode
Circlize	Gu et al., 2014	http://zuguang.de/circlize_book/book/
randomForest	Liaw and Wiener, 2002	https://cran.r-project.org/web/packages/randomForest/index.html
Scmap	Kiselev et al., 2017	http://www.hemberg-lab.cloud/scmap/
Cytoscape	Shannon et al., 2003	http://www.cytoscape.org/
Cell-cycle gene list	Macosko et al., 2015	N/A
caret	Kuhn, 2008	https://cran.r-project.org/web/packages/caret/index.html
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
R: The R Project for Statistical Computing	The R Foundation	https://www.r-project.org
Other		
MCA website	This paper	http://bis.zju.edu.cn/MCA/gallery.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to, and will be fulfilled by the Lead Contact, Guoji Guo (ggj@zju.edu.cn). Commercialized Microwell-seq Kit is available at G-BIO (http://www.igeneseq.com/research_detail/dxbcp/microwell.html).

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Mouse strains and husbandry**

Wild-type C57BL/6J mice were ordered from Shanghai SLAC Laboratory Animal. All mice were housed at Zhejiang University Laboratory Animal Center in a Specific Pathogen Free (SPF) facility with individually ventilated cages. The room has controlled temperature (20–22°C), humidity (30%–70%) and light (12 hour light-dark cycle). Mice were provided *ad libitum* access to a regular rodent chow diet.

For adult samples, testis and prostate tissues were collected from 8–10 week-old male mice, all the other tissues were collected from 6–10 week-old female mice. For embryonic (E14.5 embryos) and neonatal (1 day-old pups) samples (except for gonads), sex is not determined before tissue pooling due to small sample size. More details are listed in Table S3.

All experiments performed in this study were approved by the Animal Ethics Committee of Zhejiang University. All experiments conformed to the relevant regulatory standards at Zhejiang University Laboratory Animal Center.

METHOD DETAILS**Fabrication of microwell device**

The diameter and depth of the microwells were 28 μm and 35 μm , respectively. First, a silicon plate with 100,000 microwells was manufactured by Suzhou Research Materials Microtech. The silicon microwell plate was then used as a mold to make a PDMS plate

that had the same number of micropillars. Prior to experiments, a disposable agarose microwell plate was made by pouring 5% agarose solution onto the surface of the PDMS plate. Both the silicon and PDMS plates are reusable. One silicon microwell plate allows for almost permanent use.

Synthesis of barcoded beads

Magnetic beads coated with carboxyl groups were provided by Zhiyi (diameter 20–25 μm). The barcoded oligonucleotides on the surface of the beads were synthesized by three rounds of split-pool. All the sequences used are listed in supplementary information [Table S1](#).

For each batch of bead synthesis, 300–350 μL of carboxyl magnetic beads (50 mg/ml) were washed twice with 0.1 M MES (2-[N-morpholino]ethanesulfonic acid). The beads were then suspended in 0.1 M MES at a final volume of 635 μL . 3.08 mg of EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) were added to the beads. 6.2 μL of beads were then distributed into each well of a 96-well plate. 2.5 μL of amino modified oligonucleotide (50 μM in 0.1 M MES) were then added into every well. After vortexing and incubation for 20 minutes at ambient temperature, 0.5 μL mix (Add 6 mg of EDC in 100 μL of 0.1 M MES) was distributed into every well. After another round of vortexing and incubation for 20 minutes at ambient temperature, 0.5 μL more mix (Add 6 mg of EDC in 100 μL of 0.1 M MES) was distributed into every well. After vortexing and incubate for 80 minutes at ambient temperature, the beads were collected in 1 mL of 0.1 M PBS containing 0.02% Tween-20. After centrifugation, supernatant was removed carefully. The beads were then washed two times in 1 mL of TE (pH 8.0).

In the second split-pool, the beads were washed with water and split into each well of another 96 well plate containing the PCR mix: 1 \times Phanta Master Mix (Vazyme) and 5 μM oligonucleotides. The oligonucleotides in every tube encoded a sequence that was reverse complementary to linker 1, a unique barcode and a linker 2 sequence. PCR program was as follows: 94°C 5 min; 5 cycles of 94°C 15 s, 48.8°C 4 min, and 72°C 4 min; 4°C hold. The third split-pool procedure was the same as the second one. PCR program was as follows: 94°C for 5 min, 48.8°C for 20 min, 72°C for 4 min and 4°C hold. The oligonucleotides used in every tube encoded a linker 2 reverse complementary sequence, a unique barcode, a UMI sequence and a poly T tail. All the oligos were synthesized by Sangon Biotech with HPLC purification. Beads were resuspended in 1 mL of ddH₂O. To remove complementary chains, put beads into 95°C water bath for 6 min, separate beads with magnetic separator and remove the supernatant quickly for 2 times. Beads could be stored in TE-TW (10 mM Tris pH 8.0, 1 mM EDTA, 0.01% Tween20) for 4 weeks at 4°C.

Cell collection and lysis

Cell concentration should be carefully controlled in Microwell-seq. Both cell and bead concentrations were estimated with a hemocytometer. The proper cell concentration is \sim 100,000/ml (with 10% of wells occupied by single cells). The proper bead concentration is \sim 1,000,000/ml (with every well will occupied by single beads). Evenly distributed cell suspension was pipetted onto the microwell array, and extra cells were washed away. To eliminate cell doublets, the plate was inspected under a microscope. Cell doublets were removed with a capillary tube blowing around the doublet well. Bead suspension was then loaded on the microwell plate, which was placed on a magnet. Excess beads were washed away slowly. Cold lysis buffer (1 M Tris-HCl pH 7.5, 0.5 M LiCl, 1% SDS, 10 mM EDTA, and 5 mM dithiothreitol) was pipetted over the surface of the plate and removed after 12 min of incubation. Then, beads were collected and transferred to an RNase-free tube, washed once with 1 mL of 6 \times SSC, once with 500 μL of 6 \times SSC and then once with 200 μL of 50 mM Tris-HCl pH 8.0. Finally, \sim 50,000 beads were collected into a 1.5 mL tube.

Reverse transcription

In this procedure, the instructions from the Smart-seq2 protocol were followed. Briefly, 20 μL of RT mix was added to the collected beads. The RT mix contained 200 U SuperScript II reverse transcriptase, 1 \times Superscript II first-strand buffer (Takara), 20 U RNase Inhibitor (Sangon), 1 M betaine (Sigma), 6 mM MgCl₂ (Ambion), 2.5 mM dithiothreitol, 1 mM dNTP and 1 μM TSO primer. The beads were incubated at 42°C for 90 minutes (the tube was shaken every 6 min), then washed with 200 μL of TE-SDS (1 \times TE + 0.5% sodium dodecyl sulfate) to inactivate reverse transcriptase.

Exonuclease I treatment

After being washed with 200 μL of TE-TW and 200 μL of 10 mM Tris-HCl pH 8.0, beads were suspended in 200 μL of exonuclease I mix (containing 1 \times exonuclease I buffer and 1 U/ μL exonuclease I (NEB)), and incubated at 37°C for 60 minutes (shake beads every 10 min) to remove oligonucleotides that did not capture mRNA. Beads were pooled and washed once with TE-SDS, once with 1 mL of TE-TW and once with 200 μL of 10 mM Tris-HCl pH 8.0.

cDNA amplification

Beads were then distributed to 4 tubes. To every tube, 12.5 μL of PCR mix was added, which included 1 \times HiFi HotStart Readymix (Kapa Biosystems) and 0.1 μM TSO_PCR primer (Supplementary information, [Table S2](#)). The PCR program was as follows: 98°C for 3 min; 4 cycles of 98°C 20 s, 65°C 45 s, and 72°C 6 min; 10–14 cycles of 98°C 20 s, 67°C 20 s, and 72°C 6 min; 72°C 10 min and 4°C hold. After pooling all PCR products, AMPure XP beads (Beckman Coulter) were used to purify the cDNA library.

Transposase fragmentation and selective PCR

The purified cDNA library was fragmented by a customized transposase that carries two identical insertion sequences. The customized transposase was from TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme). The fragmentation reaction was performed following the instructions of the manufacturer. We replaced the index 2 primers (N5 ×) in the kit with our P5 primer (Supplementary information, [Table S2](#)) to specifically amplify fragments that contain the 3' end of transcripts. Other fragments will form self-loops, which impede their binding to PCR primers. The PCR program was as follows: 72°C 3 min; 98°C 30 s; 14 cycles of 98°C 15 s, 60°C 30 s, and 72°C 3 min; 72°C 5 min and 4°C hold. To eliminate primer dimers and large fragments, AMPure XP beads were then used to purify the cDNA library. Then, size distribution was analyzed on an Agilent 2100 bioanalyzer, and a peak at approximately 400~700 bp range should be observed. Finally, the samples were ready for sequencing on an Illumina HiSeq system.

Cell preparation

293T, 3T3 and mesenchymal stem cells (MSC) were cultured in Dulbecco's Modified Eagle Medium (DMEM, ThermoFisher) with 10% Fetal Bovine Serum (FBS, ThermoFisher) and 1% Penicillin-streptomycin (ThermoFisher). E14 and Cj7 embryonic stem (ES) cells were grown in DMEM supplemented with 15% FBS, 2 mM GlutaMax supplement (ThermoFisher), 1% nonessential amino acids (ThermoFisher), 0.1 mM β-mercaptoethanol (Sigma), 1% Penicillin-streptomycin and 1,000 U/ml leukemia inhibitory factor (LIF, Millipore) under feeder-free conditions. Trophoblast stem (TS) cells were cultured in Roswell Park Memorial Institute Medium (RPMI, ThermoFisher) supplemented with 20% FBS, 2 mM GlutaMax supplement, 0.05 mM β-mercaptoethanol, 1 mM sodium pyruvate (ThermoFisher), 1% penicillin-streptomycin, 25 ng/ml FGF4 (Peprotech) and 1 μg/ml heparin (Sigma) on Mouse Embryonic Fibroblast (MEF) feeders ([Tanaka et al., 1998](#)). Cells were harvested by trypsinization and resuspended in cold Dulbecco's Phosphate-Buffered Saline (DPBS, Corning) with 2 mM Ethylenediaminetetraacetic acid (EDTA, ThermoFisher) at a density of 1×10^5 cells/ml.

The organs and tissues were isolated from E14.5 embryos, 1-day-old pups or 6- to 10- week-old adult mice (C57BL/6). The samples were quickly transferred into cold DPBS, and then minced into ~1 mm pieces on ice with scissors. Tissue pieces were transferred to a 15 mL centrifuge tube, rinsed twice with cold DPBS and then re-suspended with 5 mL of dissociation enzymes. Samples were treated with different enzymes for different durations ([Table S3](#)). During the dissociation, tissue pieces were pipetted up and down gently for several times until there was no visible tissue fragment. The methods for single cell isolation from different tissues are listed in [Table S3](#). Dissociated Cells were centrifuged at $300 \times g$ for 5 min at 4°C, and then re-suspended in 3 mL of cold DPBS. After passing through a 40 μm strainer (Biologix), cells were washed twice, centrifuged at $300 \times g$ for 5 min at 4°C, and re-suspended in cold DPBS with 2 mM EDTA at a density of 1×10^5 cells/ml.

Bone marrow was isolated from femur and tibia bones (C57BL/6). Muscles were removed from bones with Delicate Task Wipers (KIMTECH). Both ends of bones were carefully trimmed to expose the interior marrow shaft. We then used DPBS with 2mM EDTA to flush the marrow using a 1 mL syringe (with 26G needle). Marrow cavity was flushed 2-3 times to collect the most of the cells. After gentle pipetting, cells were filtered using 40 μm strainers (Biologix) and collected into a 50 mL centrifuge tube. Cells were centrifuged at $300 \times g$ for 5 min at 4°C and then re-suspended with 5 mL of red blood cell lysis buffer (Biolegend). Cell suspension was incubated on ice for 5 min with occasional shaking. Lysis reaction was quenched by adding 30 mL DPBS with 2 mM EDTA. Cells were collected at $300 \times g$ for 5 min at 4°C and washed 3 times to remove the lysis buffer. Finally, cells were counted with hemocytometer and diluted with cooled DPBS with 2 mM EDTA to a density of 1×10^5 cells/ml.

Bone marrow mesenchyme was isolated from femur and tibia bones (C57BL/6). Muscles were removed from bones with Delicate Task Wipers. Both ends of bones were carefully trimmed to expose the interior marrow shaft. We then used DPBS with 2 mM EDTA to flush the marrow using 1 mL syringe (with 26G needle). Marrow was collected into a 15 mL centrifuge tube. The remaining bones were cut into pieces (except for the distal end of the tibia) and grounded in mortar. Bone samples were transferred to a 15ml centrifuge tube with 1ml of DPBS, centrifuged at $300 \times g$ for 5 min at 4°C, and re-suspended with 1ml of Hank's Balanced Salt Solution (HBSS, GIBCO) containing 3 mg/ml collagenase type I (Worthington), 4 mg/ml neutral protease and 200 U/ml DNase (Sigma). After incubation at 37°C for 15 min, the tube was vortexed 5 s for three times, and settled on ice for 1 min. The supernatant was transferred to another 15 mL centrifuge tube with 10 mL of cold DPBS. The remaining tissues were re-suspended with 1 mL of HBSS containing 3 mg/ml collagenase type I, 4 mg/ml neutral protease and 200 U/ml DNase. After incubation at 37°C for 15 min, the tube was vortexed 5 s for three times, and settled on ice for 1 min. The supernatant was then collected to the same 15 mL centrifuge tube. Finally the cells were collected and re-suspended with 3 mL of DPBS. After passing through a 40 μm strainer, cells were washed twice, counted with a hemocytometer and diluted to 1×10^5 cells/ml in DPBS with 2 mM EDTA.

For embryonic mesenchymal tissues, E14.5 embryo head, limbs and visceral tissues were removed; the remaining tissues were minced before single cell dissociation (similar to MEF generation).

For peripheral blood samples, erythroid cells were removed by red blood cell lysis buffer (Batch 2), or Ficoll separation (Batch 1, 3, 4, 5). Note that neutrophil is enriched in Batch 2.

Thymus and spleen were squeezed to pass through a 40 μm strainer using a plunger. Cells were collected into a 15 mL centrifuge tube, and then centrifuged at $300 \times g$ for 5 min at 4°C. Supernatant was discarded. Cells were re-suspended with 5 mL of red blood cell lysis buffer. Cell suspension was incubated on ice for 5 min with occasional shaking. Lysis reaction was quenched by adding 30 mL DPBS. Cells were centrifuged at $300 \times g$ for 5 min at 4°C and washed 3 times to remove the lysis buffer. Then cells were counted with a hemocytometer and diluted in cold DPBS with 2 mM EDTA at a density of 1×10^5 cells/ml.

Mesenchymal Stem Cells (MSCs) were derived from C57BL/6 mouse bone marrow. In brief, bone marrow cells from femur and tibia bones (C57BL/6) were collected and prepared into single cell suspensions. Cells were cultured in L-DMEM media at a density of 1×10^6 /ml on 10mm tissue culture plate (Corning). Culture media were changed after 12 h, and non-attached cells were discarded. Subsequently, media were changed every 24 h. After 16 days, cells were trypsinized and passaged at the ratio of 1:3. The MSCs were passaged every 3-4 days. Single cell analysis was performed with MSC culture at passage 15.

Embryoid Body (EB) differentiation for mESCs: Cells were harvested by trypsinization, re-suspended in EB Medium (ESCs medium without LIF) and seeded into 6-Well Ultra-Low Adherent Plates (Corning) at a density of 2×10^5 cells per well. Medium was changed every three days. On day 14, EB single cells were harvested by trypsinization and resuspended in cold DPBS with 2 mM EDTA at a density of 1×10^5 cells/ml.

Attached differentiation for mESCs: Day 7 EBs were plated on gelatin coated dishes and cultured in differentiation medium (ESC medium without LIF) for 7 days. Medium was changed every other day. On day 14, cells were harvested by trypsinization and re-suspended with cold DPBS containing 2 mM EDTA at a density of 1×10^5 cells/ml.

CD45- cells were enriched with MojoSort Mouse CD45 Nanobeads (Biolegend) following the manufacturer's protocol. For cKit+ cells, bone marrow cells were suspended in 100 μ L of DPBS with 5% FBS for exposure with cKit antibodies (eBioscience). cKit+ cells were then sorted with a BD FACSAria II cell sorter. Cells were diluted in DPBS with 2 mM EDTA and filtered through a 40 μ m size strainer before Microwell-seq experiments.

qPCR analysis

EasyPure RNA Kit (Transgen) was used to extract total RNA from cell samples. Total RNA was reverse transcribed into complementary DNA (cDNA) by TransScript All-in-One First-Strand cDNA Synthesis SuperMix for qPCR kit (Transgen). The diluted cDNA were used as templates in qPCR performed using Hieff qPCR SYBR[®] Green Master Mix (Yeast). The qPCR was performed with the LightCycler 480 (Roche) system.

Immunofluorescence

EBs were seeded into 0.1% Gelatin-coated dishes (NEST, 35/15mm) for attachment. Attached EBs were fixed in 4% paraformaldehyde at room temperature for 30 min. Then permeabilized treatment was performed at room temperature for 30 min with PBS + 0.2% Triton X-100. After blocking with PBS + 1% BSA + 4% FBS + 0.4% Triton X-100 at room temperature for 1 h, EBs were incubated with primary antibodies, diluted in PBS+0.2% BSA+0.1% Triton X-100, at 4°C overnight. After washing, EBs were incubated with AlexaFluor secondary antibodies (Invitrogen) for 1 h at room temperature. After being washed, EBs were incubated with DAPI for 5 min at room temperature. After the second round of fixation for 30 min, cells were imaged by Olympus FV3000. The primary antibodies used were anti-SOX17 (AF1924, 1:100, R&D) and anti-GATA4 (sc-25310, 1:100, Santa).

Immunohistochemistry

The harvested placentas were fixed in 4% paraformaldehyde at 4°C overnight. The entire placentas were sectioned with a thickness of 6 μ m. Following dewax in xylene and rehydration in graded alcohol solution, the placentas sections were sent to Servicebio (Wuhan) for immunohistochemistry services. The primary antibodies used were anti-SOX17 (AF1924, 1:100, R&D) and anti-HNF4A (sc-6556, 1:100, Santa).

QUANTIFICATION AND STATISTICAL ANALYSIS

Processing of the Microwell-seq Data

The drop-seq core computational tool was used for preprocessing of the Microwell-seq data. The implementation is described in the Drop-seq computational cookbook (<http://mccarrolllab.com/wp-content/uploads/2016/03/Drop-seqAlignmentCookbookv1.2Jan2016.pdf>). We first filtered out reads without two linkers or poly T sequences using the bduk2 function in bbmap. A cellular barcode and unique molecular identifier (UMI) were then extracted from Read One. We discarded the paired reads if the quality of any base in the barcode was below 10. We used STAR (version 2.5.2a) with default parameters for mapping (Dobin et al., 2013). Reads from 3T3 cells and 293T cells were aligned to a merged hg19-mm10 genome reference (provided by Drop-seq group, GSE63269). Reads from MCA data were aligned to the Mus_musculus.GRCm38.88 genome. All multi-aligned reads were removed and GTF annotation files from Gencode were used to tag aligned reads. For UMI count, molecular barcodes with one edit distance were merged to one within a gene. For cell quality control, we excluded cells in which less than 500 transcripts were expressed. A high proportion (> 10%) of transcript counts derived from mitochondria-encoded genes may indicate low cell quality, and we removed these unqualified cells from the downstream analysis. After obtaining the digital gene expression (DGE) data matrix, we used Seurat for dimension reduction, clustering and differential gene expression analysis (Satija et al., 2015).

Cell cycle analysis of 293T Cells

We carried out cell cycle analysis of single cell using the method according to dropseq (Macosko et al., 2015). Gene sets reflecting five phases of the HeLa cell cycle (G1/S, S, G2/M, M and M/G1) were referred from Whitfield et al. (2002). To find the genes that fit our 293T data, we calculated the correlation between the expression level of each gene and the average expression level of all genes in

that phase-specific gene set. We excluded genes with a low correlation ($R < 0.3$). We first calculated the mean of all gene expression values in each gene set and used the mean value as the score for that phase. Then, the phase-specific scores were normalized twice. First, each phase scores were centered and divided by their standard deviation. Second, we normalized the phase score across all the phases within each cell by centering and normalizing. The cells were assigned to a cell phase by their maximal phase scores. The cells were first reordered by cell phase (G1/S, S, G2/M, M and M/G1) and then by the descending phase score.

Batch removal for cross tissue comparison

For cross tissue comparison in [Figures 2B, 6B, and 6D](#), we removed the batch gene background to improve presentation. We assume that for each batch of experiment, the cell barcodes with less than 500UMI correspond to the empty beads exposed free RNA during the cell lysis, RNA capture and washing steps. The batch gene background value is defined as the average gene detection for all cellular barcodes with less than 500 UMI, multiplied by a coefficient of 2, and then rounded to the nearest integer. Genes detected in less 25% of all cells are removed from the batch gene background list. We subtract the batch gene background for each cell from the digital expression matrix before making the cross tissue comparison figures.

Cell-cell interaction network

To systematically understand the relationships between different cell types and tissues, we built cell-cell correlation-based networks. For [Figure 6A](#), we used the reduced dataset with 60,000 cells and 6,298 marker genes of 98 cell clusters. To reduce noise, we averaged every 100 cells within each cell type. We then formed a correlation network using pairwise Spearman correlation between these averaged cells. Edges with $r > 0.75$ were considered significant. The network was visualized using Cytoscape ([Shannon et al., 2003](#)) with the “edge-weighted spring embedded” layout.

Comparative analysis of different scRNA-seq methods

We compared Microwell-seq with several other scRNA-seq methods about sensibility ([Figures S2A and S2B](#)), accuracy ([Figure S2C](#)) and mixed rate ([Figure S2D](#)). For [Figure S2A](#), our Cj7 mES data (EmbryonicStemCell.CJ7) were compared with mES data produced by different technologies ([Ziegenhain et al., 2017](#)) within the low (50000 raw reads/cell) to middle (100000 raw reads/cell) range sequencing depth. [Figure S2B](#) presents the number of genes detected (counts ≥ 1) per cell for different methodologies with middle range sequencing depth (30000-60000 mapped reads/cell). Double sequenced Cj7 mES data (EmbryonicStemCell.CJ7_Deep) is used in this comparison. 65 cells were randomly sampled within the selected data range ([Ziegenhain et al., 2017](#)) to make the boxplot. For [Figure S2C](#), to assess the quantification accuracy of different methods, we computed the Pearson Correlation Coefficient (R) between expression values of housekeeping genes ([de Jonge et al., 2007](#); [Zhou et al., 2016](#)) in each cell and the average expression of reference genes for all sampled cells (65 cells) from a method. Data at low (50000 raw reads/cell) sequencing depth from ([Ziegenhain et al., 2017](#)) and our mES data were selected for the comparison.

Cell type analysis

After selection of the significant differential genes ($\text{ave_diff} > 1$) from each cell type, we compared our subtypes with other published paper using top differentially expressed markers. The results of cell type analysis were displayed with circize in R ([Gu et al., 2014](#)). Each connective line represents the same marker genes across different cell subtypes identified by different methods. Data source: [Figure S4A](#): Kidney from MCA and [Park et al. \(2017\)](#), [Figure S4B](#): Fetal lung and Adult lung both from MCA, [Figure S4C](#): Placenta from MCA and [Nelson et al. \(2016\)](#), [Figure S4E](#): Bran from MCA and [Zeisel et al. \(2015\)](#).

Single cell trajectory analysis

We used p-Creode ([Herring et al., 2017](#)) to perform developmental trajectory analysis for cells in placenta and cKit+ bone marrow ([Figures S4D and S4F](#)). For placenta, hematopoietic cells, endothelial cells as well as low quality data (< 500 genes/cell) were removed. The remaining data were used to construct an unsupervised development trajectory (noise = 10 and target = 30) based on top 20 differentially expressed marker genes in each cluster. For cKit+ bone marrow, due to the amount of cells, we randomly sampled 200 cells from each cluster with gene number per cell above the median to construct an unsupervised development trajectory (noise = 12, target = 50). The top 20 differentially expressed marker genes in each cluster were used for the analysis.

MCA website construction

The main site of the MCA uses the bootstrap framework to improve overall adaptability and interactivity. Its back-end is completed by PHP, R language, and mysql. The main functions of MCA website is divided into three parts, Gallery, Search and scMCA. Gallery provides interactive t-SNE maps for 44 tissues to show the distribution of different clusters. Significant markers for each cluster are listed in a data table. The search part provides the expression of a given gene in different clusters from to any selected tissue. The scMCA provides the function of single-cell correlation analysis with the MCA database. After users upload their own DGE files, the data will be processed by the R script and compared with the MCA reference file. The scMCA result will be returned in JSON format and presented as an interactive heatmap.

Single cell MCA analysis

To build a scMCA reference, we first normalized each cell to 100K transcripts. And then for each cell type cluster, we randomly sampled 100 single cells (all cells for clusters with less than 100 cells), calculate the averaged expression, and round the DGE number to the nearest integer. We constructed the averaged cell type transcriptome data three times for each cell cluster. This resulted in 894 main cell type references in our scMCA pipeline. We then performed differential gene expression analysis for each cell type against all the other cells and select the top 10 marker genes for each cell type (log-fold change > 1). Markers for each cell types were merged to make the combined feature gene list. The Pearson correlations of the given single cell data against each MCA cell type reference were then calculated using the combined feature gene list. Single cell FPKM, RPKM and TPM DGE matrix were log-transformed before scMCA analysis.

Comparison of scMCA, scmap and randomForest

To compare the quality of different mapping methods, we calculated the accuracy of self-projections corresponding to randomForest, scMCA and scmap. We used randomForest R package (Liaw and Wiener, 2002), scMCA pipeline and scmap R Package (Kiselev et al., 2017) to build references and run cell type mapping analysis. Figure S7C is based on the 5 datasets including male fetal gonad, pancreas and spleen data from MCA, as well as hematopoietic stem cells (Zhou et al., 2016) and preimplantation embryo data (Deng et al., 2014; Posfai et al., 2017). For each dataset, we randomly chose 80% cells to build the training model, and then use the remaining 20% cells to fit the model, the pipeline was repeated 3 times for each dataset. The unassigned cells from scmap were removed. The mapping results were converted into confusion matrix and the accuracy was calculated by the R package caret (Kuhn, 2008).

DATA AND SOFTWARE AVAILABILITY

Data Resources

The accession number for the raw data files for the RNA sequencing analysis reported in this paper is GEO: GSE108097.

Digital Expression Matrixes are accessible through:

<https://figshare.com/s/865e694ad06d5857db4b>

An R package for scMCA analysis is available at GitHub (<https://github.com/ggjlab/scMCA>).

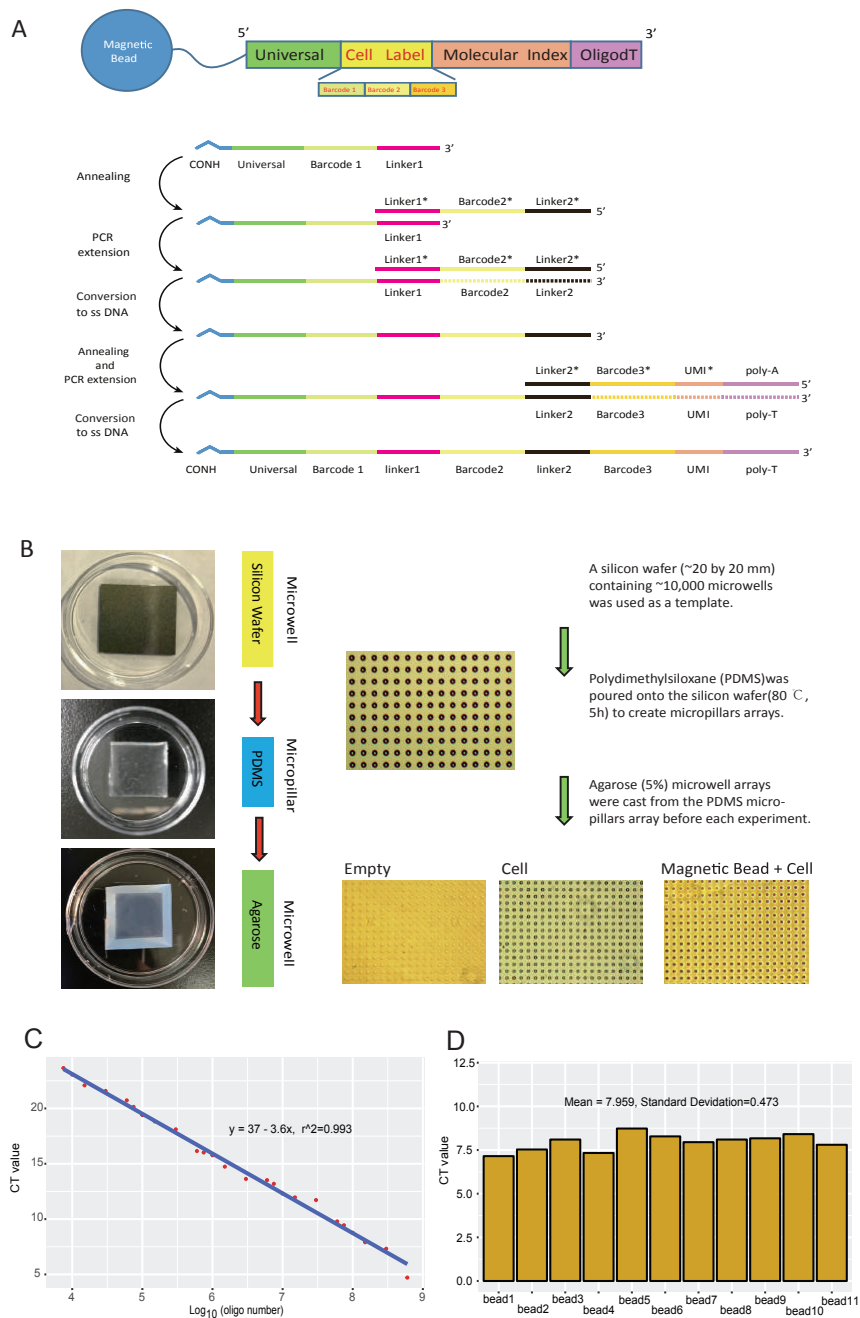


Figure S1. Microwell Fabrication and Bead Synthesis, Related to Figure 1

(A) The synthesis of barcoded beads. Three split-pool rounds are used to introduce the 3 parts of the oligonucleotides on the beads. In the first split-pool round, magnetic beads coated with carboxyl groups are distributed randomly into a 96-well plate in which 5' amino-modified oligonucleotides are conjugated to the beads. Oligonucleotides in each well have a unique barcode sequence. Beads are then pooled and split into another 96-well plate where a second barcode sequence is introduced by single-cycle PCR. In the final split-pool round, the third barcoded sequence, unique molecular identifier (UMI) and polyT tail are added. After the split-pool, all oligonucleotides on the same bead will have the same cell barcode but a different UMI, while oligonucleotides on different beads will have different cell barcodes.

(B) The microarray fabrication. The microarray fabrication strategy involves using a silicon microarray chip to construct a micropillar PDMS chip and then an agarose microarray chip. Both the silicon and PDMS chip are reusable, which greatly reduces cost and saves time.

(C) Full-length oligonucleotide solutions at different concentrations were used to draw a standard curve. There were three replicates for each concentration. The average CT value of three replicates is shown in the figure.

(D) The oligonucleotide numbers on synthesized beads were examined by single-bead qPCR. The results of a single-bead qPCR experiment. Eleven single beads were picked out for single-bead qPCR. The results showed that $\sim 10^8$ oligonucleotides were conjugated to every bead.

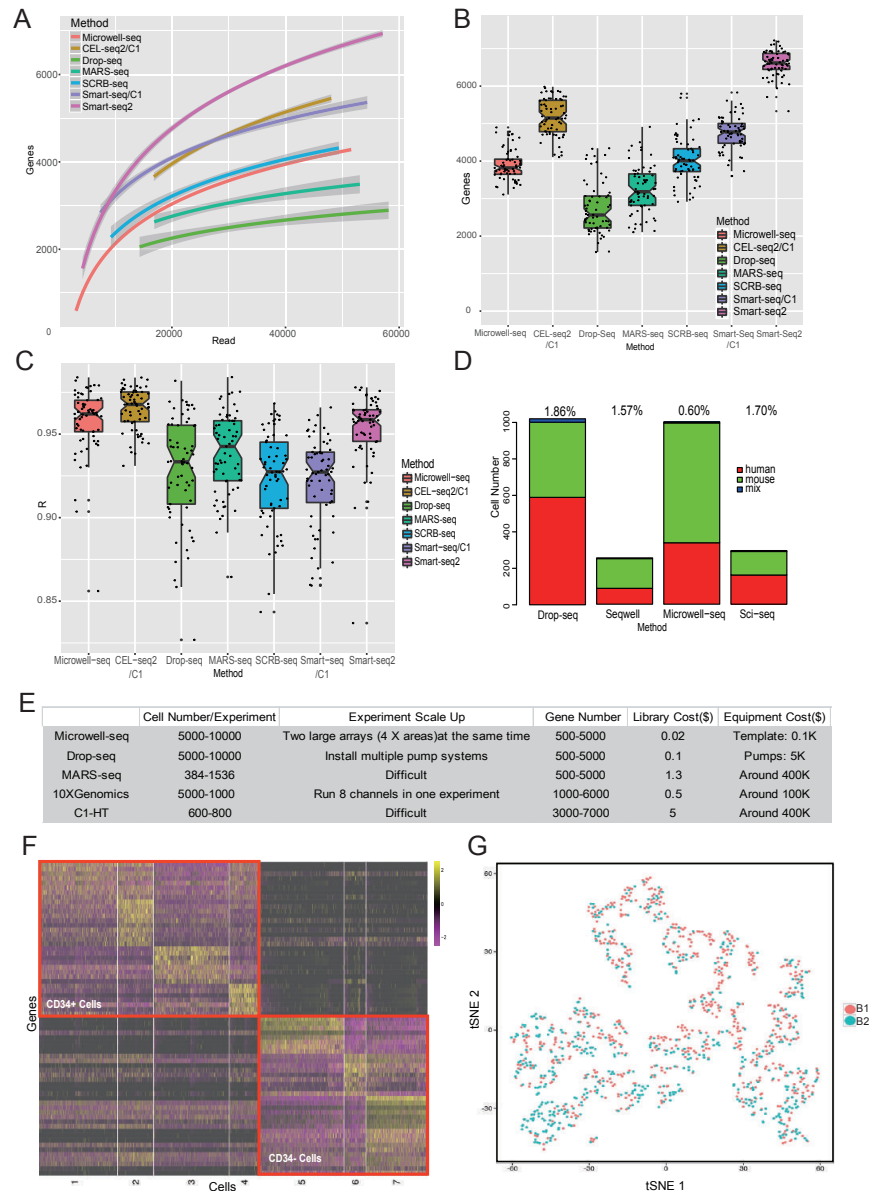


Figure S2. Validation of the Microwell-Seq Platform, Related to Figure 1

(A) Reads (mapped) to a gene plot for the mES experiments using different platforms. Our Cj7 mES microwell-seq data were compared with a series of mES single cell data obtained using different technologies (Ziegenhain et al., 2017). Microwell-seq shows an advantage in the low (50000 raw reads/cell) to middle (100000 raw reads/cell) range sequencing depth.

(B) Number of genes detected (counts ≥ 1) per cell for different methodologies with middle range sequencing depth (30000-60000 mapped reads/cell). Double sequenced Cj7 mES data were compared with a series of mES single cell data obtained using different technologies (Ziegenhain et al., 2017). Each dot represents a cell, and each box represents the median and first and third quartiles of replicates.

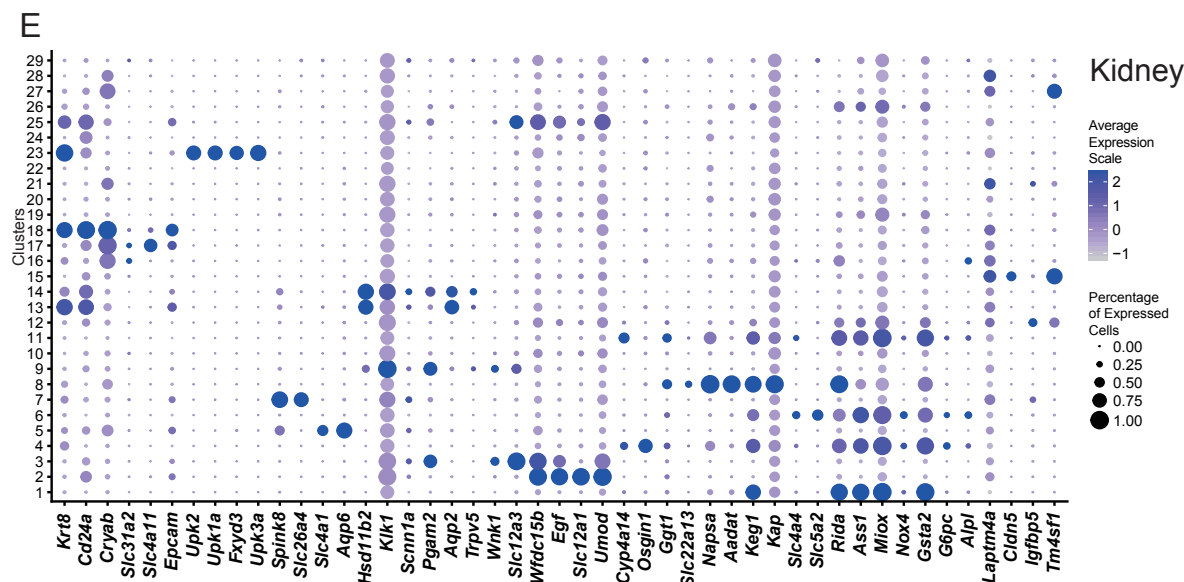
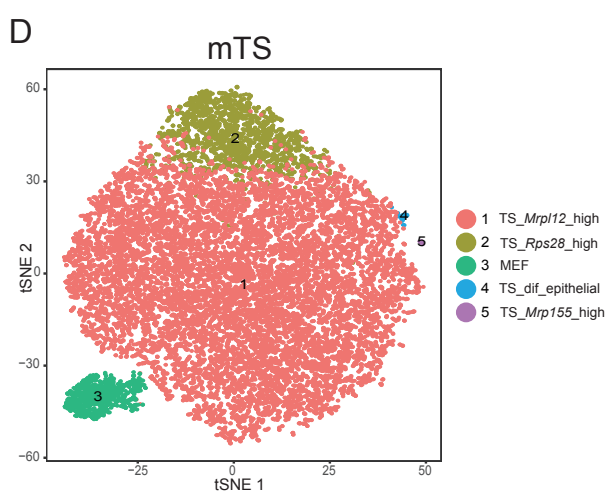
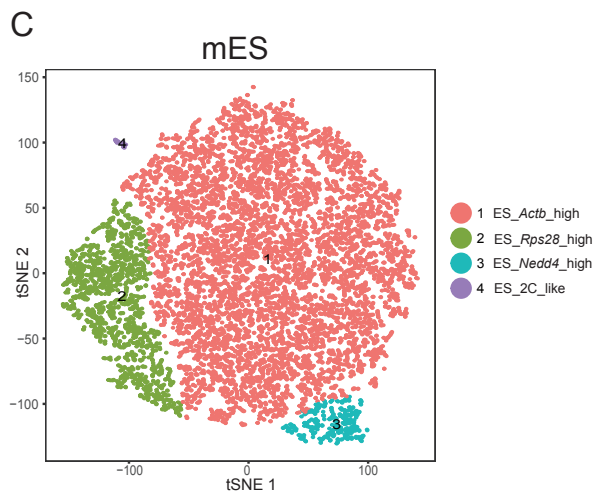
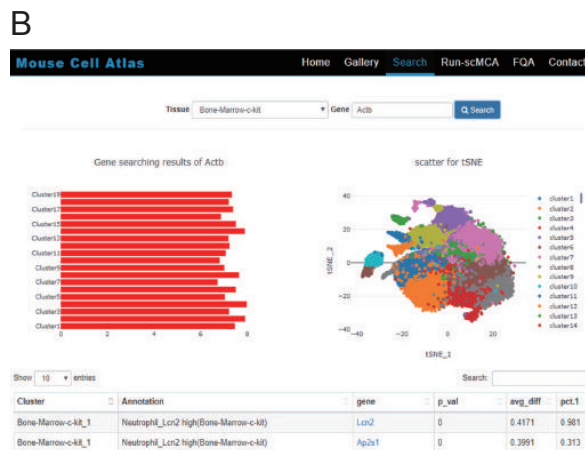
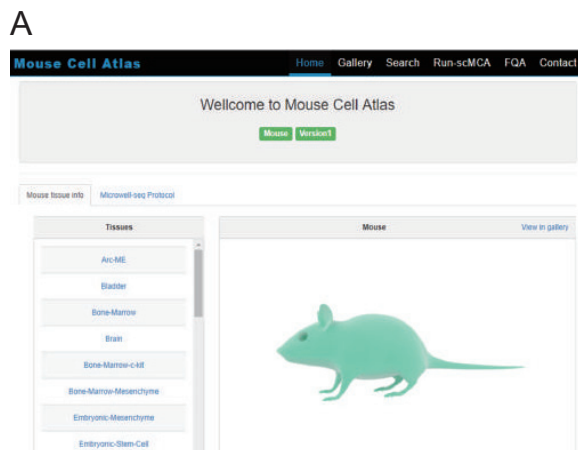
(C) Accuracy of scRNA-seq methods. Our Cj7 mES microwell-seq data were compared with a series of mES single cell data obtained using different technologies (Ziegenhain et al., 2017). Data at low (50000 reads/cell) sequencing depth were selected for the comparison. Correlations of reference gene expression values (counts per million reads for Smart-seq/C1 and Smart-seq2 and UMIs per million reads for all others) between single-cell expression and mean expression are calculated for each cell. Distributions of Pearson correlations (R) are shown for each method. Each dot represents a cell, and each box represents the median and first and third quartiles of replicates.

(D) Mixed rate for the species mixing experiments using different platforms.

(E) Cost efficiency for different single-cell RNA-seq methodologies.

(F) Heatmap of 4323 thawed mPB CD34+ and mPB CD34- cells. CD34+ and CD34- cells were clearly clustered into two mega groups. Various subclusters can be observed in the CD34+ and CD34- mega groups.

(G) Thawed mPB CD34+ cells from batch 1 (B1) and batch 2 (B2) are visualized on a t-SNE map.



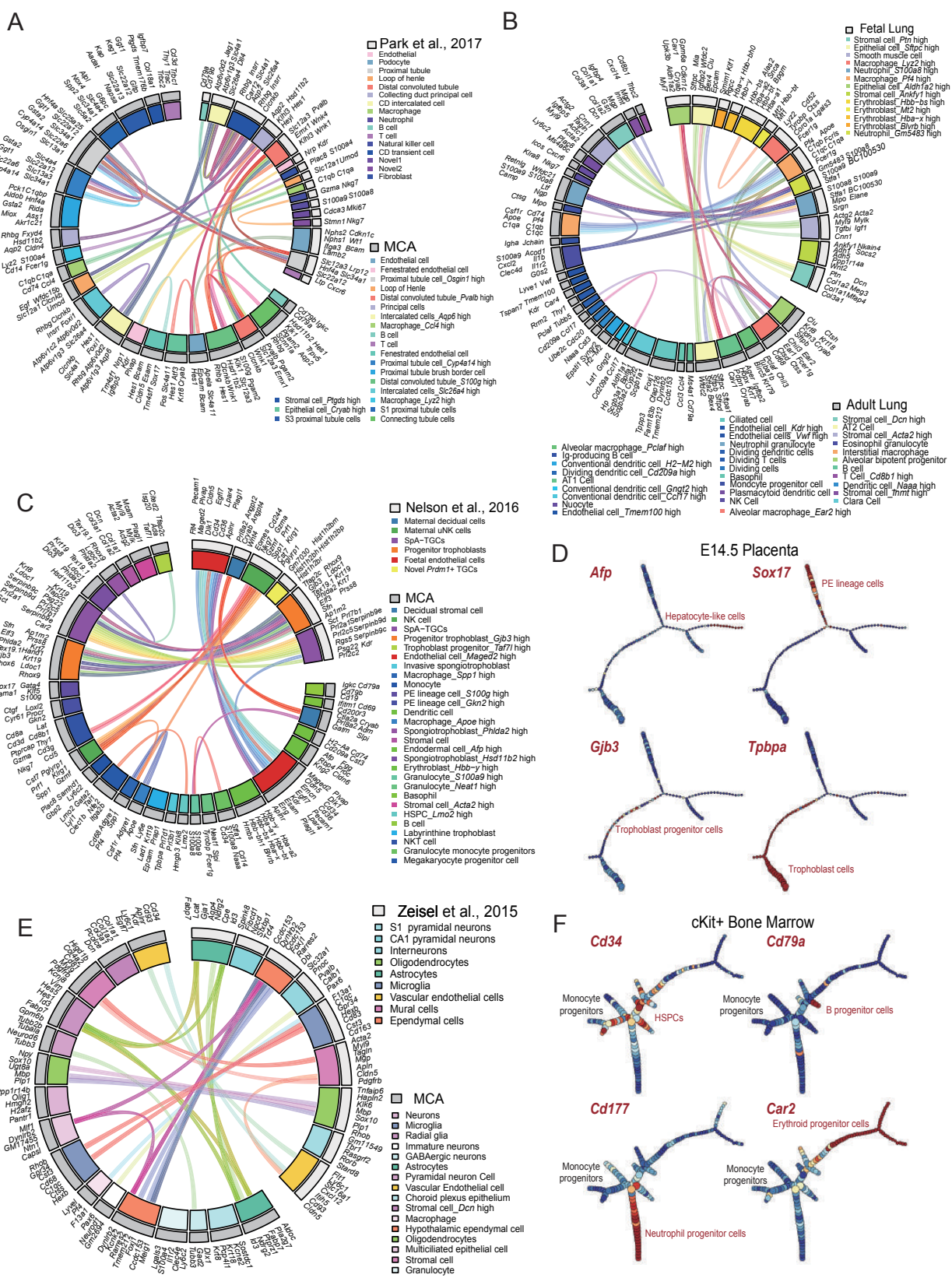
(legend on next page)

Figure S3. Snapshots of the Mouse Cell Atlas Website and Representative Data, Related to Figure 2

(A and B) Snapshots of the Mouse Cell Atlas website.

(C) and (D) t-SNE maps of mES (C) and mTS (D) single cell data. Cells are colored by cell type cluster.

(E) Dot plot visualization of each cell type in kidney single-cell data. The size of the dot encodes the percentage of cells within a cell type, and the color encodes the average expression level.



(legend on next page)

Figure S4. Comparison of MCA Data with Other Available Datasets, Related to Figure 3

(A) A Circos plot showing overlap of cell types and markers between MCA kidney data and the prepublished kidney single cell data (Park et al., 2017). Correlated cell types are connected by lines. Representative marker genes are listed outside of the circle.

(B). A Circos plot showing overlap of cell types and markers between MCA adult lung data and MCA fetal lung data. Similar cell types are connected by lines indicating same marker gene expression. Representative marker genes are listed outside of the circle.

(C) A Circos plot showing overlap of cell types and markers between MCA placenta data and the published placenta single cell data (Nelson et al., 2016). Similar cell types are connected by lines indicating same marker gene expression. Representative marker genes are listed outside of the circle.

(D) Development trajectory of placenta single cells constructed by p-Creode (Herring et al., 2017).

(E) A Circos plot showing overlap of cell types and markers between MCA neonatal brain data and the published cortex and hippocampus single cell data (Zeisel et al., 2015). Similar cell types are connected by lines indicating same marker gene expression. Representative marker genes are listed outside of the circle.

(F) Development trajectory of ckit+ bone marrow single cell data constructed by p-Creode.

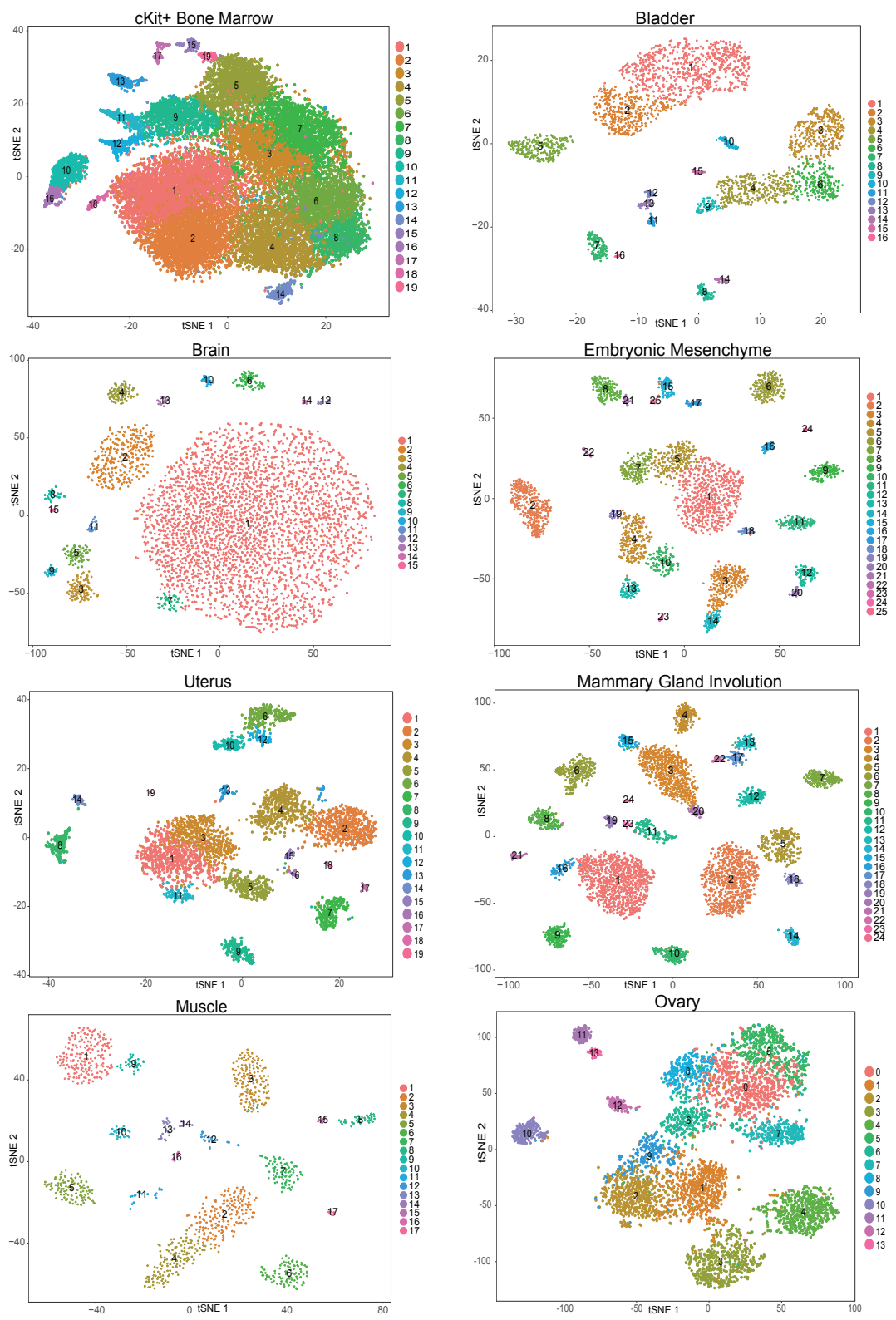


Figure S5. t-SNE Maps for Analyzed MCA Tissues, Related to Figure 5

t-SNE maps for single-cell data from cKit+ bone marrow, bladder, brain, embryonic mesenchyme, uterus, mammary gland involution, muscle, and ovary. Cells are colored by cell-type cluster.

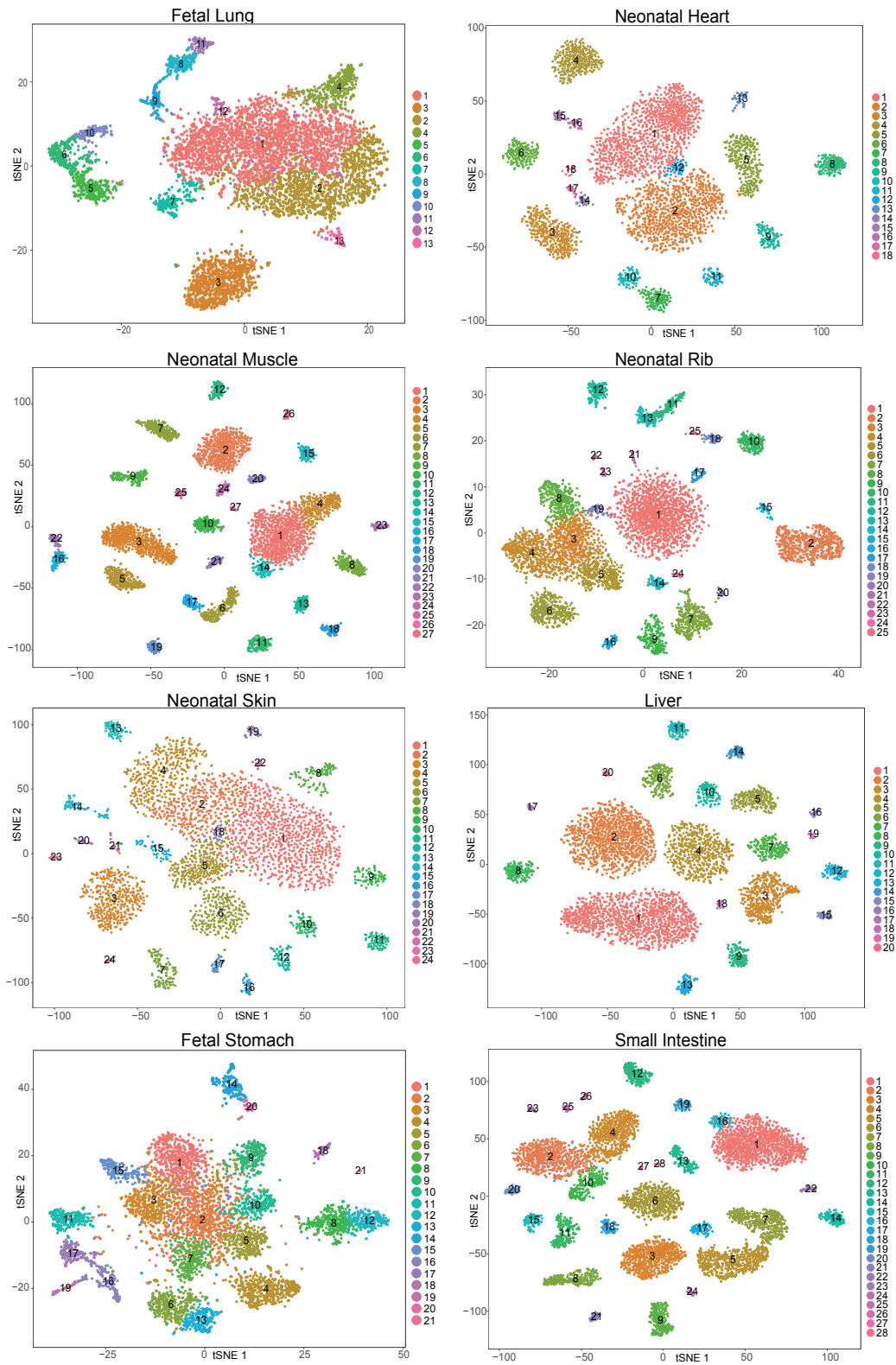
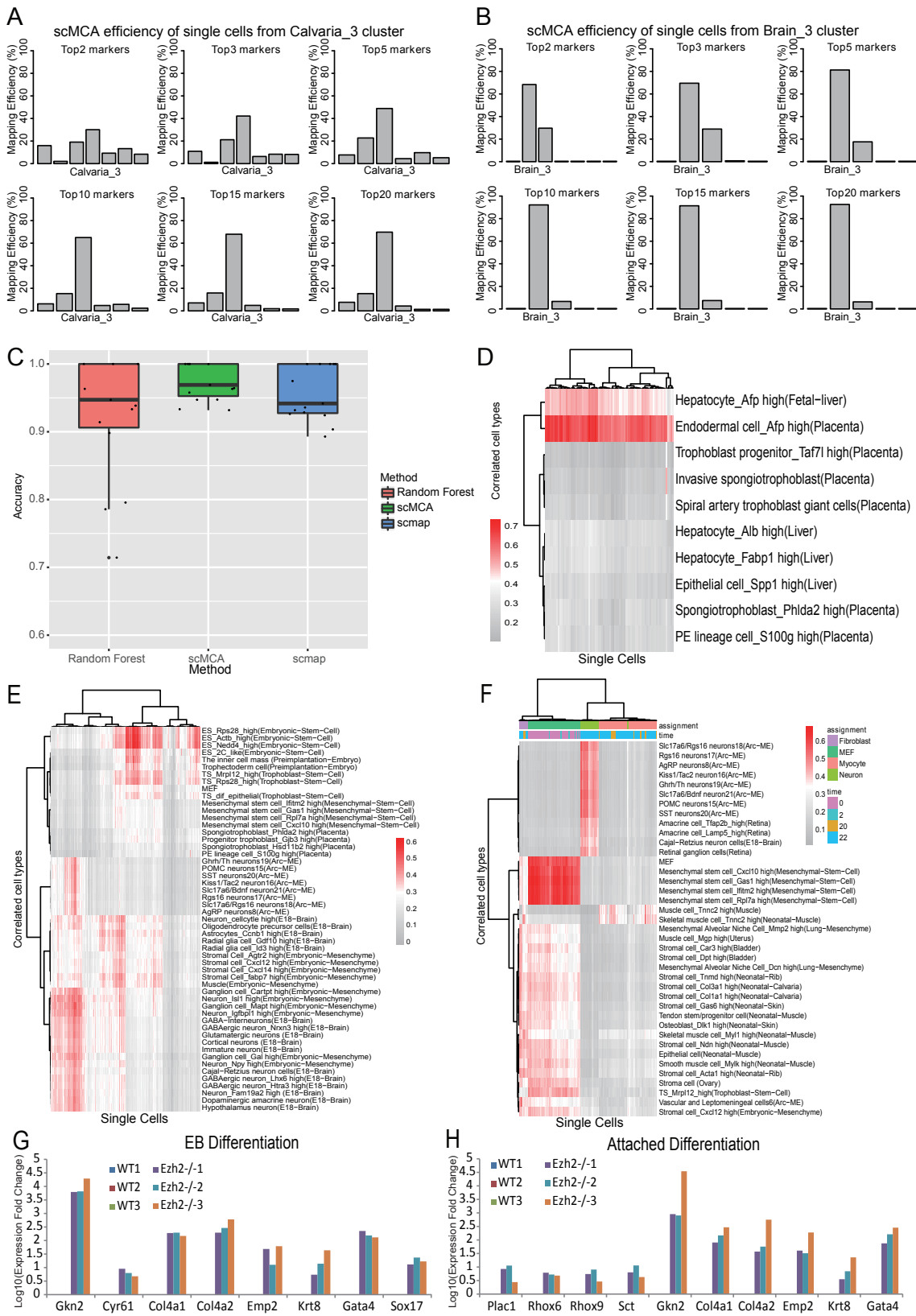


Figure S6. t-SNE Maps for Analyzed MCA Tissues, Related to Figure 5

t-SNE maps for single-cell data from fetal lung, neonatal heart, neonatal muscle, neonatal rib, neonatal skin, liver, fetal stomach, and small intestine. Cells are colored by cell-type cluster.



(legend on next page)

Figure S7. Testing of scMCA Pipeline, Related to Figure 7

(A and B) Mapping efficiency of single cells from defined cell-type clusters when using Top2, Top3, Top5, Top10, Top15 and Top20 markers for each of the cell type references.

(C) Accuracy of single-cell mapping algorithms for different available methods. RF: Random Forests (Liaw and Wiener, 2002). scmap: pipeline adopted from the prepublished study (Kiselev et al., 2017). Each box represents the median and first and third quartiles of replicates for different methods. RF: RandomForest.

(D) scMCA result of placenta *Afp*-high endodermal cell data. Each row represents one cell type in our reference. Each column represents one single cell in the customer dataset. Red means high correlation; gray means low correlation.

(E) scMCA result of mouse neuronal differentiation data (Rizvi et al., 2017). Each row represents one cell type in our reference. Each column represents one single cell in the customer dataset. Red means high correlation; gray means low correlation.

(F) scMCA result of mouse fibroblast to neuron transdifferentiation data (Treutlein et al., 2016). Each row represents one cell type in our reference. Each column represents one single cell in the customer dataset. Red means high correlation; gray means low correlation.

(G and H) QPCR analysis of extraembryonic lineage markers expressed during EB and attached differentiation of WT and *Ezh2*^{-/-} mES cells.