Nucleic Acids Research, 2025, **53**, gkae1312
https://doi.org/10.1093/nar/gkae1312
Advance access publication date: 11 January 2025
Methods

inDrops-2: a flexible, versatile and cost-efficient droplet microfluidic approach for high-throughput scRNA-seq of fresh and preserved clinical samples

Simonas Juzenas[1,†], Karolis Goda[1,†], Vaidotas Kiseliovas[2,†], Justina Zvirblyte[1], Alvaro Quintinal-Villalonga[3], Juozas Siurkus[4], Juozas Nainys[5] and Linas Mazutis[1,6,*]

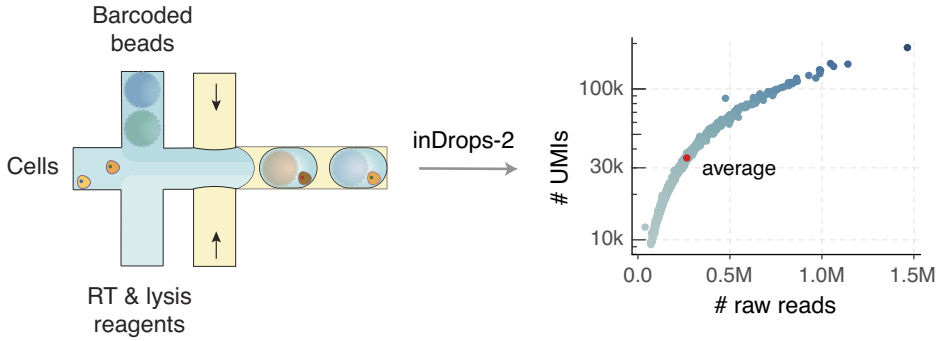
[1]Institute of Biotechnology, Life Sciences Center, Vilnius University, Vilnius, 10257, Lithuania
[2]Computational and Systems Biology, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, NY, 10065, USA
[3]Department of Medicine, Thoracic Oncology Service, Memorial Sloan Kettering Cancer Center, NY, 10065 USA
[4]Thermo Fisher Scientific Baltics, Research and Development, Vilnius, 02241, Lithuania
[5]Droplet Genomics, Vilnius, 08114, Lithuania
[6]Department of Molecular Biology, Umea University, Umea, 901 87, Sweden

[*]To whom correspondence should be addressed. Tel: +370 5 223 4356; Email: linas.mazutis@bti.vu.lt
[†]The first three authors should be regarded as Joint First Authors.

Abstract

The expansion of single-cell analytical techniques has empowered the exploration of diverse biological questions at the individual cells. Droplet-based single-cell RNA sequencing (scRNA-seq) methods have been particularly widely used due to their high-throughput capabilities and small reaction volumes. While commercial systems have contributed to the widespread adoption of droplet-based scRNA-seq, their relatively high cost limits the ability to profile large numbers of cells and samples. Moreover, as the scale of single-cell sequencing continues to expand, accommodating diverse workflows and cost-effective multi-biospecimen profiling becomes more critical. Herein, we present inDrops-2, an open-source scRNA-seq technology designed to profile live or preserved cells with a sensitivity matching that of state-of-the-art commercial systems but at a 6-fold lower cost. We demonstrate the flexibility of inDrops-2, by implementing two prominent scRNA-seq protocols, based on exponential and linear amplification of barcoded-complementary DNA, and provide useful insights into the advantages and disadvantages inherent to each approach. We applied inDrops-2 to simultaneously profile multiple human lung carcinoma samples that had been subjected to cell preservation, long-term storage and multiplexing to obtain a multiregional cellular profile of the tumor microenvironment. The scalability, sensitivity and cost efficiency make inDrops-2 stand out among other droplet-based scRNA-seq methods, ideal for large-scale studies on rare cell molecular signatures.

Graphical abstract



Introduction

Comprehensive molecular characterization of biological samples increasingly relies on single-cell technologies (1). Over the last few years, a large array of platforms and methods for single-cell analysis have been introduced, thereby ushering in a new era of single-cell omics (2). Of these -omics approaches, single-cell RNA sequencing (scRNA-seq) techniques have been particularly impactful and have gained increasing popularity. The rich biological information in individual cells that is captured by scRNA-seq methods has played a critical

Received: April 25, 2024. **Revised:** November 28, 2024. **Editorial Decision:** December 3, 2024. **Accepted:** December 26, 2024

role in identifying new cell types in the human body (3–5), delineating cancer heterogeneity (6–8) and patient response to therapy (9–11), and advancing our understanding of various biological systems (12–16). To date, scRNA-seq represents a leading technology for building cell atlases of the human body and diseases (17–20) and is likely to remain indispensable in the foreseeable future.

Among the scRNA-seq platforms developed to date (21–28), the most widely used are plate-based and droplet-based systems, each of which has unique strengths and weaknesses. Plate-based scRNA-seq techniques are advantageous for use in targeted applications, for instance, when cells of interest are isolated by fluorescence activated cell sorting (FACS) into microtiter plates for subsequent full-length scRNA-seq or copy number variation (CNV) analysis (29–31). These platforms often provide superior sensitivity, although at higher cost and limited throughput relative to droplet-based methods, which offer a few orders of magnitude greater throughput at a significantly lower cost. Historically, two droplet-based techniques, which were originally reported side-by-side in 2015 (21,22), have paved the way for high-throughput single-cell transcriptomics. Commercial systems based on these innovations, such as 10× Chromium™ (25) (analogue to inDrops) and Nadia™ (32) (analogue to drop-seq) have provided broad access to scRNA-seq technology.

Commercial systems ensure operational reproducibility and quality, making them a primary choice for single-cell transcriptomics studies. However, profiling single cells at the $>10^5$ scale using commercially available droplet-based systems, while feasible, can lead to unsustainable financial burdens, especially among research groups with limited resources. Open-source systems such as inDrops (22) and drop-seq (21), or their modifications (33–35), offer lower operation costs and can accommodate the diverse needs of researchers, such as processing unconventional samples (12,36). However, open-source systems often exhibit reduced sensitivity (e.g. transcript capture) when compared to commercial alternatives (37). Furthermore, barcoding 10^5–10^6 individual cells requires extended encapsulation, during which the cells of interest may undergo undesirable transcriptome changes. Therefore, for a scRNA-seq method to be broadly applicable, in addition to high sensitivity, the native state of the cell transcriptome needs to be preserved during the workflow.

In this work, we present inDrops-2, an open-source droplet microfluidics platform for performing high-throughput scRNA-seq studies of live or fixed cells with transcript and gene detection sensitivity similar to that of a state-of-the-art commercial platform (10× Chromium v3), yet at a 6-fold lower cost and a throughput of 5000 cells min^{-1}. To expand the applicability of inDrops-2, we implemented two highly sensitive scRNA-seq protocols: one based on linear amplification of complementary DNA (cDNA) by *in vitro* transcription (IVT) and the other based on exponential cDNA amplification by polymerase chain reaction (PCR) following template switching (TS). Furthermore, we developed a cell preservation protocol for processing clinical samples comprising as few as 20 000 cells that is compatible with inDrops-2 and other droplet-based scRNA-Seq platforms. We showed that dissociated cells acquired from clinical specimens can be stored in a dehydrated state for extended periods and later multiplexed by covalent conjugation to DNA oligonucleotides (38) for subsequent transcriptomic analysis. In summary, we present inDrops-2, a sensitive and cost-efficient scRNA-seq method for capturing clinically relevant cell phenotypes from human specimens that underwent preservation, long-term storage and multiplexing.

Materials and methods

Cell lines

Cryopreserved K-562 cells (ATCC, CCl-243) were stored in vapor phase nitrogen until use. Murine KP lung adenocarcinoma (LUAD) cell line (39) was a kind gift by Dr Stella Paffenholz (MSKCC). Cell culture was maintained in 25 cm^2 culture flask in 5 ml volume of Iscove's modified Dulbecco's medium (IMDM) (Gibco, 31 980 030) supplemented with 10% fetal bovine serum (FBS) (Gibco, 10270–106) and 1× penicillin–streptomycin (Gibco, 15 140 122) under 5% CO_2 and at 37°C. Cells were harvested at ∼10^6 cells/ml, collected into 15 ml conical tubes, pelleted at 300 × *g* for 5 min and washed twice in ice-cold 1× Dulbecco's phosphate bufferd saline (DPBS) with 0.05% (*w*/*v*) bovine serum albumin (BSA). The cell count and viability were quantified using the Countess II cell counter and 0.2% trypan blue staining.

Human peripheral blood mononuclear cell and bone marrow derived CD34 ± stem/progenitor cells

Cryopreserved primary peripheral blood mononuclear cells (PBMC) from healthy donors were purchased from ATCC (PCS-800–011), while bone marrow stem/progenitor CD34 + cells from healthy donors were purchased from All-Cells, LLC. (ABM022F) and stored in vapor phase nitrogen until use. Prior to scRNA-seq, a vial with frozen cells was removed from the liquid nitrogen tank and thawed at 37°C in a water bath for 2–3 min. Next, the vial content (∼1 ml) was transferred to a 50-ml conical tube and slowly diluted with 1 ml of warm (∼37°C) cell culture medium (IMDM with 10% FBS). To prevent osmolysis, warm medium was added dropwise while gently rocking the 50-ml tube with a hand. The thawed cells were serially diluted in five steps with 1:1 volume addition of warm medium and 2-min incubation between each step, until a final 32-ml volume was reached. The cell suspension was then pelleted at 300 × *g* for 5 min in a swinging bucket centrifuge. Supernatant was discarded and the cell pellet was washed twice in ice-cold 1× DPBS with 0.05% (*w*/*v*) BSA. The cell count and viability were determined with Countess II cell counter.

Human lung cancer biospecimen acquisition and processing

The patients with LUAD or small cell lung carcinoma (LC) undergoing a surgical resection at Memorial Sloan Kettering Cancer Center (MSKCC) provided informed consent through an Institutional Review Board-approved biospecimen collection and analysis protocol. Experiments using human subjects were performed in accordance with the ethical standards of the Helsinki Declaration. Each specimen was cut into three pieces, ∼5–10 mm^3 in size, and processed according to the dissociation protocol reported in the past (40). Each sample was dissociated for 15 min at 37°C on the GentleMACS Octo Dissociator with Heaters (Miltenyi) using Human Tumor Dissociation Kit (Miltenyi Biotec). Following tissue dissociation, the cell suspension was passed through 35 μm Cell Strainer Snap Cap and treated with red blood lysis (ACK buffer, Lonza)

for 2 min at room temp. One LUAD sample was resuspended in phosphate-buffered saline (PBS) with 0.04% BSA and processed fresh for single-cell encapsulation, while for the rest of the specimens, the cells were stained with live dye (Calcein AM) and PE anti-human CD45 antibody (BioLegend, cat no 368 510) mixture, and using BD FACS Aria II instrument sorted into CD45 + and CD45- compartments. The sorted cells were spun down for 5 min at 300 × *g* in a swinging bucket centrifuge at 4°C, and resuspended in 90% methanol. The methanol-preserved cells were transferred to −80°C until all specimens were acquired.

Methanol-based cell preservation

Cell preservation in methanol was adopted following the reports by Alles *et al*., 2017 and Chen *et al*., 2018 (38,41), with some modifications. Specifically, the live cells were first transferred to DNA LoBind tube (Eppendorf, 0 030 108 051), pelleted at 300 × *g* for 5 min at 4°C and gently resuspended in 100 µl ice-cold 1× DPBS. Next, 900 µl of ice-cold methanol was slowly added in a dropwise manner while gently rocking the tube; this prevents cells from clumping and osmolysis. Once suspended in 90% methanol, the cells were incubated on ice for another 15 min and then transferred to −20°C or −80°C for a long-term storage.

Rehydration of methanol-preserved cells

Briefly, the tube with cells preserved in methanol was placed on ice for 15 min and then centrifuged at 1000 × *g* for 10 min in a swinging bucket centrifuge set at 4°C. Most of the supernatant was removed leaving ∼50 µl on top of the cell pellet. Next, the cell pellet was resuspended in 400 µl of ice-cold Rehydration Buffer 1 (3× saline-sodium citrate buffer (SSC), 80 mM dithiothreitol (DTT), 0.2% BSA, 1 U/µl RNase Inhibitor) and entire suspension transferred onto a centrifugal tube filter (Millipore, UFC30DV25), that was earlier pretreated with 1% BSA. The column was centrifuged at 50 × *g* for 45 s using 4°C centrifuge. The flow through fraction was discarded. The cell suspension that was retained on top of the filter (∼50 µl volume) was washed two more times with an ice-cold Rehydration Buffer 1 and once with an ice-cold Rehydration Buffer 2 (1× SSC, 40 mM DTT, 0.1% BSA, 1 U/µl RNase Inhibitor). After final wash, the rehydrated cells were retrieved from the filter membrane, counted under hemocytometer and processed on 10× Chromium or inDrops-2 (TS) platform.

scRNA-seq using 10× Chromium platform

Single-cell encapsulation and messenger RNA (mRNA) barcoding on 10× Genomics Chromium instrument was performed with Single Cell 3′ Library and Gel Bead Kit V3 reagent kit, following the vendor's manual (CG00183 rev B). Briefly, for each sample a suspension of cells (viability 70–95%) were loaded onto Chromium microfluidics chip targeting for recovery of ∼5000 single-cells with 3.9% multiplet rate. For cells resuspended in rehydration buffer 2 (1× SSC, 40 mM DTT, 0.1% BSA, 1 U/µl RNase Inhibitor), the cells were first concentrated by centrifugation to ∼2000 cells/µl and 4 µl were mixed with the corresponding V3 reagents as an input for encapsulation. Following the reverse transcription (RT), the emulsion droplets were broken and barcoded cDNA purified with Dynabeads, followed by 12 cycles of PCR: 98°C for 180 s, 12 cycles (98°C for 15 s, 67°C for 20 s, 72°C for 60 s), and 72°C for 60 s. The PCR-amplified barcoded cDNA was diluted to 50 ng, fragmented with the reagents provided in the kit, purified with SPRIselect beads (Beckman Coulter, B23318), and ligated to the sequencing adapters. The ligation product was amplified by PCR: 98°C for 45 s, 14 cycles (98°C for 20 s, 54°C for 30 s, 72°C for 20 s) and 72°C for 60 s. The final DNA library was double-size purified (0.6–0.8×) with solid-phase reversible immobilization (SPRI) beads and sequenced on the Illumina NovaSeq 6000 platform (R1: 26 cycles; i7: 8 cycles; R2: 70 or more cycles) at a depth of 10 000–50 000 reads per cell.

Single-cell encapsulation and mRNA barcoding using inDrops platform

Single-cell suspensions (∼400 cells/µl) were prepared in 1× DPBS supplemented with 0.04% (*w*/*v*) BSA and 16% OptiPrep (M1248-100, BioVision). The OptiPrep is used to increase the density (ρ) of 1× PBS buffer to $\rho_{sol} = 1.044$ g/ml, thereby suppressing cell sedimentation. Single-cell suspensions were loaded onto microfluidics chip (Supplementary Figure S1) along with barcoded hydrogel beads (either V1 or V2 design, see Supplementary Table S1) and 2× RT-lysis mixture (see below). The barcoded hydrogel beads were suspended in 1× First Strand buffer (TFS, 18 080 044) supplemented with 0.3% (*v*/*v*) IGEPAL CA-630 (Sigma–Aldrich, 18896–50ML). The 2× RT-lysis mixture for inDrops-1 protocol comprised: 24 U/µl SuperScript III Reverse Transcriptase (TFS, 18 080 044), 1.3 U/µl SUPERase-In (TFS, AM2696), 0.83× First Strand buffer (TFS, 18 080 044), 0.6% (*v*/*v*) IGEPAL CA-630 (Sigma–Aldrich, 18896–50ML), 5 mM DTT (TFS, 00 561 515), 11 mM $MgCl_2$ (Ambion, AM9530G), 65 mM Tris-HCl (Invitrogen, 15568–025) and 1 mM dNTPs (TFS, R0192). The 2× RT-lysis mixture for inDrops-2 (IVT) protocol comprised: 24U/µl of Maxima H minus (TFS, EP0751), 2U/µl of RiboLock (TFS, EO0381) and 1U/µl Superase In (TFS, AM2696) RNase inhibitor, 2× RT buffer (TFS, EP0752), 1 mM dNTP, 0.6% (*v*/*v*) IGEPAL CA-630. The 2× RT-lysis mix for inDrops-2 (TS) protocol comprised: 24U/µl of Maxima H minus, 2U/µl of RiboLock and 1U/µl Superase In (TFS, AM2696) RNase inhibitor, 2× RT buffer, 1 mM dNTP, 0.6% (*v*/*v*) IGEPAL CA-630 and 50 µM template switching oligonucleotide (TSO) (Supplementary Table S1). When testing the performance of Super Script IV enzyme, the Maxima H minus enzyme and RT buffer was replaced with 24 U/µl SS-IV enzyme and corresponding SS-IV buffer (Invitrogen, 18 090 010), while keeping other ingredients in the reaction mixture the same. The microfluidics platform was operated under two flow regimes; standard and high-throughput. For a standard run, the flow rates were set at 250, 250 and 50–70 µl/h for cells (diluted at ∼500 cells/µl), 2× RT-lysis mixture and barcoded hydrogel beads, respectively. The droplet stabilization oil was set at 550 µl/h. For the high-throughput run, the flow rates were set at 100, 900, 150 and 1200 µl/h for cells (diluted at ∼2000 cells/µl), RT-lysis mixture, barcoded beads and carrier oil, respectively. Emulsion droplets were collected on-ice for 20–60 min. Next, the barcoded RT primers were photo-released by exposing the tube with an emulsion to a 350-nm light either using LED device (Droplet Genomics, MHT-LAS1) for 20 s, or UV lamp (UVP, cat. no. 95–0127-01) for 5 min. The emulsion was then transferred onto a heat block to initiate cDNA synthesis at either 42°C or 50°C for 60 min followed by the heat inactivation at 75°C for 15 min (for inDrops-1) or at 85°C for

5 min (for inDrops-2), or otherwise as indicated. The post-RT droplets were aliquoted into separate tubes (each containing ∼20.000 droplets) and then broken by adding 1H,1H,2H,2H-perfluorooctanol up to 10% (*v*/*v*). After a quick spin for 30 s at 300 × *g*, the supernatant was transferred onto filter column (Zymo, C1004-250). The flow-through fraction containing barcoded-cDNA was collected into a new 1.5 ml DNA LoBind tube (Eppendorf, 0 030 108 051) by centrifugation for 1 min at 1000 × *g*. After this step, the barcoded-cDNA can be stored at 4°C overnight, or processed further to construct the sequencing library as indicated below.

inDrops-1 sequencing library preparation

The barcoded-cDNA was diluted to 80 μl with nuclease-free water and treated with an enzyme cocktail comprising exonuclease I, restriction endonuclease HinfI and alkaline phosphatase. Specifically, the inDrops-v1 libraries were digested with 40U of ExoI (TFS, EN0581), 4 μl of FastDigest HinfI (TFS, FD0804) and 1U of FastAP (TFS, EF0654) for 15 min at 37°C and purified using 1.2× SPRIselect beads (Beckman Coulter, B23318). The inDrops-v1.1 libraries were digested with 40U of ExoI and 1U of FastAP for 15 min at 37°C, and purified using 1.2× SPRIselect beads. The inDrops-v1.2 iteration excluded enzymatic treatment and instead the libraries were purified with 0.8× SPRIselect beads. The purified cDNA was converted to double-stranded DNA (dsDNA) by performing second strand synthesis with NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module (NEB, E6112) in 20 μl reaction volume for 2.5 h at 16°C and 20 min at 65°C. The dsDNA was then linearly amplified using HiScribe T7 High Yield RNA Synthesis Kit (NEB, E2040S) for 15 h at 37°C. Reaction products, in the form of RNA, were purified using 1.2 × SPRIselect beads and their quality as well as yield was evaluated with RNA Pico Assay on Agilent Bioanalyzer 2100 instrument. The amplified material was fragmented with RNA fragmentation reagent (Ambion, AM8740) at 70°C for 2.5 min, terminated with 20 mM ethylenediaminetetraacetic acid (EDTA) and purified with 1.2 × SPRIselect beads. After purification, fragmented RNA library was mixed with 5 μM of PE2-N6 primer (Supplementary Table S1) and reverse transcribed using PrimeScript RTase (Takara, SD0418) for 60 min at 42°C. The resulting cDNA library was purified with 1.2 × SPRIselect beads and amplified by 12 cycles of PCR: 98°C for 2 min, 2 cycles (98°C for 20 s, 55°C for 30 s, 72°C for 40 s), 10 cycles (98°C for 20 s, 65°C for 30 s, 72°C for 40 s) and 72°C for 5 min. DNA amplification was conducted with Kapa HiFi HotStart PCR mix (Kapa Biosystems) using PE1 and PE2 indexing primers (Supplementary Table S1). The amplified and indexed libraries were further purified using SPRIselect dual-size selection (0.6–0.8×). The library size was evaluated using HS DNA assay (Agilent Technologies, 5067–4626). The libraries were adjusted to 10 pM spiked in with 15% PhiX Control v3 Library and sequenced on the NextSeq550 and HiSeq2500 (Illumina) platform (R1: 54 cycles; i7: 8 cycles; R2: 35 cycles or more) at a depth of ∼20 000 reads per cell.

inDrops-2 (IVT) sequencing library preparation

The detailed protocol for constructing inDrops-2 (IVT) libraries is provided as Supplementary Protocol 1. The barcoded-cDNA was diluted to 80 μl and purified using 0.8 × SPRIselect beads (B23318, Beckman Coulter), and subjected to second-strand synthesis using NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module in 20 μl reaction volume for 2.5 h at 16°C and 20 min at 65°C. The second-strand synthesis reaction product was then linearly amplified using HiScribe T7 High Yield RNA Synthesis Kit for 15 h at 37°C. The reaction products, in the form of RNA, were purified using 0.8 × SPRIselect beads and their quality as well as yield was evaluated using RNA Pico Assay (Agilent, 5067–1513). The RNA concentration was estimated based on ultraviolet absorption at 260 nm (NanoDrop) and diluted to 1000 ng/μl. Next, 9 μl of amplified RNA was fragmented with 1 μl RNA fragmentation reagent (Ambion, AM8740) at 70°C for 1.5 min, and purified with ice-cold STOP mixture [9 μl of 10 mM Tris (pH 7.0), 25 μl of SPRIselect beads, 2 μl of 200 mM EDTA]. The purified RNA was mixed with 5 μM of PE2-N6 primer (Supplementary Table S1), heated to 70°C for 2 min and allowed to hybridize for 3 min on ice. The RNA was reverse transcribed using Maxima H minus enzyme at 30°C for 10 min followed by 60 min at 42°C and heat inactivation for 15 min at 70°C. The cDNA was purified with 1.0 × SPRIselect beads and amplified by Kapa HiFi HotStart PCR mix using PE1 and PE2 Illumina index primers (Supplementary Table S1). The PCR was set for 10 cycles: 98°C for 2 min, 2 cycles (98°C for 20 s, 55°C for 30 s, 72°C for 40 s), 8 cycles (98°C for 20 s, 65°C for 30 s, 72°C for 40 s), and 72°C for 5 min. The amplified libraries were purified using SPRIselect dual-size selection (0.6–0.8×), inspected on a High Sensitivity DNA Chip (Agilent Technologies, 5067–4626), and library yield quantified with Qubit dsDNA HS Assay kit. The final inDrops-2 (IVT) libraries were diluted to 10 pM, mixed with 7.5% PhiX Control v3 Library, and sequenced on the NextSeq550 and HiSeq2500 (Illumina) instrument (R1: 54 cycles; i7: 8 cycles, R2: 35 cycles or more), at a depth of 10 000–50 000 reads per cell.

inDrops-2 (TS) sequencing library preparation

The detailed protocol for constructing inDrops-2 (TS) libraries is provided as Supplementary Protocol 2. The barcoded-cDNA was purified with 0.8 × SPRIselect beads and subjected to PCR (2 × KAPA HiFi HotStart Ready mix, KK2500) using cDNA amplification primers (Supplementary Table S1) at 0.5 μM, and following PCR program: 98°C for 3 min, 14 cycles (98°C for 15 s, 67°C for 20 s, 72°C for 60 s), 72°C for 1 min. The amplified DNA was purified with 0.8 × SPRIselect beads and assessed with DNA HS assay on Agilent Bioanalyzer 2100. Next, the library for sequencing was constructed with FS DNA Library Prep Kit (NEB, E7805) by following the manual instructions. Specifically, 50 ng of amplified cDNA was fragmented for 8 min at 37°C, followed by end-repair and dA tailing for 30 min at 65°C. Then, the modified double-stranded ligation adapter (Supplementary Table S1) supplied at 55 nM was ligated to fragmented DNA for 15 min at 20°C. The ligation product was purified using 0.8 × SPRIselect beads, eluted in 40 μl nuclease-free water, and half of the material was subjected to amplification by PCR (NEBNext Ultra II Q5 Master Mix) using indexing primers (Supplementary Table S1) at 0.5 μM. The PCR was set for 14 cycles: 98°C for 45 s, 14 cycles (98°C for 20 s, 54°C for 30 s, 72°C for 20 s) and 72°C for 1 min. The amplified DNA was purified using SPRIselect dual-size selection (0.6–0.8×) and the library size distribution was evaluated with DNA HS assay on Agilent Bioana-

lyzer 2100. The inDrop-2 (TS) libraries were sequenced on the MiSeq, HiSeq2500, NextSeq550 and NovaSeq6000 (Illumina) platforms, without PhiX spike-in. The sequencing parameters were R1: 28 cycles; i7: 8 cycles, R2: between 35 and 92 cycles (depending on instrument and sequencing reagent kit), at a depth of >10 000–50 000 reads per cell.

Barcoded hydrogel bead synthesis

Synthesis of V1 beads. The hydrogel beads carrying DNA barcode sets were synthesized following the previously described protocol (42) and employing the Agilent Bravo Automated Liquid Handling Platform. At first, the 58 ± 2 μm size acrylamide-based hydrogel beads carrying 50 μM of photo-cleavable DNA stub: 5′-/5Acryd/PC/CGATTGATCAACG TAATACGACTCACTATAGGGATACCATCTACACTCT TTCCCTACACGACGCTCTTCCG-3′, where 5Acryd is an acrydite moiety, PC is a photo-cleavable spacer) were generated using a microfluidics chip (Droplet Genomics, MCN-G5) and stored in the dark at 4°C in a Washing Buffer [10 mM Tris-HCl (pH 8.0), 0.1 mM EDTA and 0.1% Tween-20] until further use. Next, the DNA barcodes were attached to the DNA stub on hydrogel beads by conducting two rounds of primer extension reaction in a combinatorial split-and-pool manner. In the first round the hydrogel beads were washed two times in isothermal buffer [20 mM Tris-HCl (pH 8.8), 10 mM $(NH_4)_2SO_4$, 50 mM KCl, 2 mM $MgSO_4$, 0.1% Tween 20] and one time in the isothermal buffer supplemented with 0.3 mM dNTP (each). After diluting hydrogel bead suspension to ∼10.000 beads/μl, the Agilent Bravo Liquid Handling Platform was used to distribute the beads in the four 96-well plates (10 μl of beads per single well), preloaded with 5-μl of barcoded oligonucleotides (first set of sub-barcodes) at 50 μM concentration. (Supplementary Table S2). After heating the plates at 85°C for 2 min and 60°C for 1 h, the 5-μl of Bst 2.0 (NEB, M0537S) enzyme mixture (1 × isothermal buffer, 0.3 mM dNTP and 1.8U Bst 2.0 enzyme) was added to each well, gently mixed, and the primer extension reaction initiated at 60°C. After 30 min of incubation the reaction was terminated by adding 25 μl of STOP solution [10 mM Tris-HCl (pH 8.0), 50 mM EDTA; 0.1% Tween 20 and 100 mM KCl]. The beads were collected into a single 50 ml tube and the second strand was removed by alkaline denaturation. For that purpose, the beads were washed five times in Denaturation Solution (0.1 M NaOH, 0.5% Brij 35P) with 5 min incubations between the washes. The alkaline solution was neutralized with one volume of Neutralization Buffer [100 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM EDTA, 0.1% Tween 20], and then washed twice in a Washing Buffer until further use. At this step the hydrogel beads had first set of sub-barcodes attached to them in a single-stranded form, thus making them suitable for the second round of barcoding. Following the same procedure as described above the hydrogel beads were washed twice in the isothermal buffer, aliquoted in four 96-well plates preloaded with second set of sub-barcodes (Supplementary Table S2), and then subjected to the second round of barcoding followed by dsDNA denaturation, neutralization and washing. To remove the single-stranded DNA primers that lack poly(dT) tails the beads were resuspended in 1 × ExoI buffer (TFS, EN0581), hybridized to 50 μM poly(dA)24 primer for 10 min and treated with Exonuclease I enzyme (TFS, EN0581) for 30 min at room temperature. Next, Klenow Exo-minus reaction mixture comprising 1 × FastDigest (FD) buffer (TFS, B64), 0.25U Klenow Exo minus fragment (TFS, EP0421) and 8.7 mM dNTPs was added to the hydrogel bead suspension at a dilution 1:10 (Klenow reaction mixture:hydrogel bead suspension), and incubated at 37°C for 60 min. The hydrogel beads were then washed six times in Denaturation Solution with 3 min incubations between the washes and two times in Neutralization Buffer. The beads were filtered through a 70 μm strainer, resuspended in a Washing Buffer and stored at 4°C until further use. The quality and yield of fully barcoded DNA primers was evaluated with total RNA Pico Assay (Agilent Technologies, 5067–1513) and by fluorescent *in situ* hybridization (FISH) targeting 3′ poly(dT) tail, as described by Zilionis *et al.*, 2017. The full-length primer sequence on V1 hydrogel beads was: 5′-/5Acryd/PC/CG ATGACGTAATACGACTCACTATAGGGATACCACCAT GG CTCTTTCCCTACACGACGCTCTTCCGATCT[12 345 678 901] GAGTGATTGCTTGTGACGCCTT[12 345 678] NNNNNNNNTTTTTTTTTTTTTTTTT-3′, where 5Acryd is an acrydite moiety, PC is a photo-cleavable spacer, the letters in bold indicate T7 RNA promoter sequence, and underlined letters indicate the site for Illumina PE Read 1 Sequencing primer. The numbers indicate cell barcodes, which were designed to have 50% GC content and Hamming distance of ≥ 3 between each pair of barcodes. The barcode whitelist is provided in Supplementary Table S3.

Synthesis of V2 beads. The hydrogel beads 63 μm in size and carrying the photo-cleavable DNA stub (/5Acryd/CGATGACG(PC)CTACACGACGCTCTTC-3′, where Acryd is an acrydite moiety, PC is a photo-cleavable spacer) at 50 μM concentration (Cellanthe Labs) were subjected to two rounds of combinatorial ligation reaction. The liquid handling operations were conducted using Agilent Bravo Liquid Handling Platform. At first, the first set of sub-barcodes (Supplementary Table S4) were made double-stranded by mixing P_Bcd1 and RC_Bcd1 primers in equimolar amount, denaturating for 5 min at 95°C and cooling to 4°C at a rate 0.3°C/min. The oligonucleotide duplexes (200 μM) were aliquoted into 4 × 96-well plates, 6.25-μl per well. Then, a master mix containing hydrogel beads carrying DNA stub and ligation mixture was distributed in the same 4 × 96-well plates such that each well would contain 12.5 μl of close-packed hydrogel beads, 1.25 μl of 10 × ligation buffer (NEB, B0202S), 0.3 μl of T4 DNA ligase I (NEB, M0202L), 6.25-μl of double-stranded barcoded oligonucleotides (first set of sub-barcodes) at 200 μM concentration and 4.7 μl of nuclease-free water. The ligation reaction was conducted overnight at room temperature while slowly rotating the plates on a vertical tube rotator (Biosan). The reaction was stopped by adding 50 μl of STOP buffer [50 mM Tris-HCl (pH 8.0), 50 mM EDTA, 0.1% Tween 20] to each well, gently mixed by pipetting and all hydrogel beads pooled in a single 50-ml tube. After washing five times with 25 ml of 1 × ligation buffer, the ligation reaction was conducted with a second set of barcodes. The hydrogel beads carrying first set of barcodes were evenly distributed in four 96-well plates preloaded with 6.25 μl of single-stranded barcoded oligonucleotides (second set of sub-barcodes) at 200 μM concentration (Supplementary Table S4). The 25-μl volume reaction composition per single well comprised: 1 × ligation buffer, ∼500.000 hydrogel beads, 50 μM bcd2 primer and 120 U of T4 DNA ligase I. The ligation reaction proceeded overnight at room temperature on a rotating platform to pre-

vent hydrogel beads from settling. After completing two-step ligation, the beads were pooled and washed five times with 25-ml of Washing Buffer. To remove, the oligonucleotides lacking poly(dT) tails the hydrogel beads were resuspended in 1 × Exo I buffer (TFS, EN0581), hybridized to 50 μM poly(dA)24 primer for 10 min and treated with Exonuclease I enzyme (TFS, EN0581) for 30 min at room temperature. The hydrogel beads were then washed six times in Denaturation Solution with 3 min incubations between the washes and two times in Neutralization Buffer. The beads were filtered through a 70 μm strainer, resuspended a Washing Buffer, and the quality as well as yield of fully barcoded DNA primers was evaluated with total RNA Pico Assay (Agilent Technologies, 5067–1513) and by FISH targeting 3′ poly(dT) tail. The full-length V2 primer sequence on the beads was as follows:

5′-/5Acryd/CGATGACG/PC/CTACACGACGCTCT TCCGATCT[12 345 678]CATG[12 345 678] NNNNNN NNTTTTTTTTTTTTTTTTTT-3′, where 5Acryd is an acrydite moiety, PC is a photo-cleavable spacer, underlined letters indicate the site for Illumina P7 Read 1 Sequencing primer. The numbers indicate cell barcodes, which were designed to have 50% GC content and Hamming distance of ≥ 3 between each pair of barcodes. The barcode whitelist is provided in Supplementary Table S5. If required, the oligonucleotides can be modified to incorporate dual-indexing, compatible with exAMP chemistry sequencing platforms as previously described by Southard-Smith *et al.* (43). Supplementary Protocol 3 outlines a detailed, step-by-step procedure for hydrogel bead barcoding. Alternatively, fully barcoded hydrogel beads can be obtained from Cellanthe Labs or other vendors.

Cell hashing using click-chemistry oligonucleotide tags

Cell hashing with DNA oligonucleotides was adapted from the work by Gehring *et al.*, 2020 (44) with following modifications:

- ClickTag preparation. The 5′-amino modified oligos (Supplementary Table S6) were activated by 1 mM NHS-methyltetrazine (Click Chemistry Tools) in 50% DMSO for 60 min at 21°C. The activated ClickTags were precipitated by ethanol and resuspended in 10 mM HEPES (pH 7.2) to yield the final concentration of 40 μM. Activated ClickTags were stored in the dark at −20°C for up to 3 days, before use.
- Cell hashing with ClickTags. To hashtag the cells with ClickTags, 25 μl of methanol-preserved cells (∼30 000 cells in total) were mixed with 1 mM NHS-TCO (Click Chemistry Tools) and incubated in the dark at 21°C for 5 min. Next, the cell suspension was mixed with 3 μl of 40 μM activated hashtag oligonucleotide and incubated for 30 min on a rotating platform at room temperature. The reaction was terminated by quenching with a 10 mM Tris-HCl (pH 8.0) buffer supplemented with 50 μM methyltetrazine-DBCO (Click Chemistry Tools). The methanol-preserved cells conjugated with ClickTags were rehydrated following Supplementary Protocol 4.
- scRNA-seq of hash-tagged cells. The rehydrated hash-tagged cells were resuspended in a rehydration buffer (1× SSC, 40 mM DTT, 0.1% BSA, 1U/μl Superase In RNase Inhibitor) with 20% Optiprep at a dilution of ∼2000 cells/μl. Then, cell suspension was loaded onto inDrops-2 microfluidics chip along with 1.1× RT-lysis mixture and barcoded hydrogel beads. The flow rates of the microfluidics platform were adjusted to compensate for the inhibitory effect of citric acid that is present in the rehydration buffer. As such, the flow rates were set at 100 μl/h for cells, 900 μl/h for 1.1× RT-lysis mixture and 100–150 μl/h for barcoded hydrogel beads, respectively. The droplet stabilization oil was set at 1200 μl/h. The emulsion was collected on-ice for 20 min and after release of barcoded RT primers the cDNA synthesis was initiated at 42°C for 90 min followed by 85°C for 5 min. The post-RT droplets were broken with 10% perfluorooctanol and supernatant containing barcoded-cDNA was collected by passing through a filter column (Zymo, C1004-250) at 1000 × *g* for 1 min.
- ClickTag sequencing library preparation. Following inDrops-2 (TS) approach the cDNA derived from hashtags and mRNA was co-purified using 2 × AMPure beads. Then, purified material was amplified by 14 cycles of PCR [98°C for 3 min, 14 cycles (98°C for 15 s, 67°C for 20 s, 72°C for 60 s), 72°C for 1 min] using 0.5 μM of cDNA amplification oligos (Supplementary Table S1) that targeted mRNA-derived cDNA, and 2 μM of Hash_fwd_oligo (Supplementary Table S6) that targeted hashtag-derived cDNA. The amplified material was mixed with 0.6 × volume of SPRIselect beads and eluted in two fractions. The magnetic bead bound fraction comprising mRNA-derived cDNA was processed following the inDrops-2 (TS) protocol (Supplementary Protocol 2). The supernatant fraction comprising hashtag-derived cDNA was mixed with SPRI beads to reach a final SPRI ratio of 1.6×, incubated at room temperature for 5 min, washed twice with 80% ethanol and eluted in 20 μl of nuclease-free water. The hashtag libraries were spiked with mRNA-derived cDNA library at a ratio of 1:10, and sequenced on the NovaSeq6000 (Illumina) platform (R1 – 28 cycles; i7: 8 cycles, R2: 70 cycles or more), at a depth of >10 000 reads per cell.

RNA extraction and evaluation

Total RNA extraction was performed with TRIzol reagent (TFS, 15 596 026) following manufacturer's instructions. The extracted RNA was further purified with RNA clean and Concentrator kit (Zymo Research, R1060). The RNA integrity number (RIN) was estimated using total RNA Pico Assay (Agilent Technologies, 5067–1513) on the Agilent Bioanalyzer 2100 instrument.

Optimization of cDNA synthesis and purification

To arrive at optimal reaction conditions that can consistently generate high cDNA yields, the effect of various cell lysis agents, the RT enzymes and temperature, TSO concentration, and cDNA purification strategies were first tested in a bulk format (Supplementary Figure S2). The RT reaction was performed in 20 μl and comprised a total of ∼20.000 K562 cells including 1× RT buffer, 0.5 μM RT primer (5′-CTACACGACGCTCTTCCGATCTNNNNNNTTTTTTTT TTTTTTTTTTTTTTTTTTTT-3′), 0.5 mM dNTP, 1 U/μl RiboLock RNase inhibitor, 0.1% (*v*/*v*) of corresponding lysis agent, varying amount of TSO, varying amount and type of RT enzyme. The reaction was carried out at 42°C for 60 min followed by 5 min at 85°C. The lysis agents tested in

this work included IGEPAL CA-630 (Sigma–Aldrich, 18896–50ML), Tween 20 (Sigma–Aldrich, P9416-100mL), Triton X-100 (Sigma–Aldrich, 93426–100mL), Brij58 (Thermo Scientific, 28 336), Digitonin (Invitrogen, BN2006), n-octyl-β-glucopyranoside (Roth, CN23.1), all at 0.1% (*w*/*v*). The impact of TSO amount was evaluated in the range of 5–50 μM (Supplementary Figure S2C). The RT enzymes tested included Maxima H minus (TFS, EP0751) and SuperScript IV (Invitrogen, 18 090 050). In all cases, the cDNA was purified with 1.2 × SPRIselect beads except when evaluating different purification strategies, which included; 1.2 × SPRIselect beads, Zymo DNA Clean & Concentrator kit (Zymo Research, R1060), GndHCl buffer [7.3 M guanidinium chloride, 100 mM Tris-HCl (pH 6.2)] and GndSCN buffer [3 M guanidinium isothiocyanate, 33% isopropanol, 4% Triton X-100, 20 mM Tris-HCl (pH 6.2)] (Supplementary Figure S2E) (45). cDNA purification with SPRIselect beads and Zymo DNA Clean & Concentrator kit was performed according to manufacturer's recommendations. To purify cDNA with in Dynabeads first, Dynabeads were mixed with either GndHCl or GndSCN buffers in 1:24 ratio. Then post-RT reaction mix was purified by adding 2.5 × volumes of Dynabeads in binding buffer and incubating for 5 min at room temperature. After incubation tubes were placed on a magnetic stand for magnetic beads to settle. Supernatant was removed and the bead pellet was washed two times with 180 μl of 80% ethanol. After washing, the beads were left to dry for 2 min at room temperature. To elute cDNA, the bead pellet was resuspended in 20 μl nuclease-free water, incubated for 2 min at room temperature and the supernatant collected after settling the magnetic beads on a magnetic stand. Once purified, the cDNA was amplified by 14 cycles PCR (KAPA): 98°C for 3 min, 14 cycles (98°C for 15 s, 67°C for 20 s, 72°C for 60 s), 72°C for 1 min, using 0.5 μM cDNA amplification primers (Supplementary Table S1). The amplified cDNA was purified with 1.2 × SPRIselect beads, diluted six times in pure MQ-water and analyzed with DNA HS assay on Agilent Bioanalyzer 2100.

Optimization of inDrops-2 (TS) library construction

To arrive at the final inDrops-2 (TS) library construction protocol, we tested multiple reaction conditions in bulk format. Specifically, for each condition tested we performed cDNA synthesis in 20 μl volume comprising ∼20.000 cells (K-562), 1× RT buffer, 0.1% (*v*/*v*) IGEPAL CA-630, 0.5 mM dNTP, 25 μM TSO, 0.5 μM RT primer (see RT-trim3 primer on Supplementary Table S1), 1 U/μl RiboLock RNase inhibitor and 10 U/μl Maxima H-minus RTase. The RT reaction was carried out at 42°C for 60 min, terminated at 85°C for 5 min and cDNA purified with 1.2 × SPRIselect beads. Following 14 cycles PCR (KAPA) the amplified cDNA was purified with 1.2 × SPRIselect beads, and 50 ng of material was fragmented for 8 min at 37°C and A-tailed for 30 min at 65°C using NEBNext Ultra II FS DNA Library Prep Kit (NEB, E7805S). The fragmented and dA-tailed DNA library was ligated to dsDNA adapter having both forward (Ligation FWD primer) and reverse primers (Ligation REV primer) blocked at 3′ and 5′ ends, respectively (Supplementary Table S1). The adapter ligation reaction was performed for 15 min at 20°C using a Ligation Master Mix provided with NEBNext Ultra II FS DNA Library Prep Kit. The libraries were purified with 0.8 × SPRIselect beads, PCR amplified and analyzed on a High Sensitivity DNA Chip. The results of these efforts are provided in Supplementary Figure S2.

Microfluidics platform setup

We used a custom-built microfluidics platform reported previously in details (42) as well as open-source microfluidics platform Onyx (Droplet Genomics) both of which provide experimental flexibility for encapsulating cells at a desirable throughput and reaction conditions. The microfluidic chips were made of the polydimethylsiloxane bound to a microscope glass slide, and having rectangular microchannels 80 μm height. When performing scRNA-seq, the samples were injected into the microfluidics chip via PTFE tubing [int. 0.56 mm; ext. 1.07 mm, Droplet Genomics (MAN-TUB2)] connected to 1 ml syringes (Omnifix-F, BRAUN) and 0.6 × 25 mm Neolus needles (Terumo). The syringe and tubing for barcoded hydrogel beads was wrapped in aluminium foil to protect photo-cleavable primers from illumination by ambient light. The flow rates of liquids and carrier oil were controlled by syringe pumps (PHD 2000, Harvard Apparatus). Emulsions were collected off-chip into 1.5 ml DNA LoBind tubes (Eppendorf, 0 030 108 051) placed on a cooling rack.

scRNA-seq data pre-processing

Each pair of the index 1 (i7) demultiplexed read 1 and 2 fastq files (containing cell barcode and transcript sequences, respectively) were processed using STAR v2.7.10a (46). Specifically, the reads were mapped to the human GRCh38 genome (GENCODE v41) or the mouse GRCm38 genome (GENCODE M30) with default parameters using –soloFeatures *GeneFull* to count genes (incl. exons and introns) and to generate cell × gene count matrices. To reduce the gene overlap-derived read loss (mapping to multiple features), read through transcripts and transcripts within pseudoautosomal regions were excluded from the GENCODE annotations and were limited to protein coding, lncRNA, IG/TR gene and pseudogene biotypes. Since only uniquely mapped genes were counted, for some libraries, transcript reads (read 2) were trimmed before the alignment to achieve an equal number of bases among compared libraries using seqkit v2.0.0 (47). To define barcode geometry, barcode correction and unique molecular identifiers (UMI) deduplication for different protocols, the parameters were the following: (i) for inDrops-1 and inDrops-2 libraries generated with barcodes v1: –soloType *CB_UMI_Complex*, –soloCBmatchWLtype *EditDist_2*, –soloUMIdedup *Exact*, –soloAdapterSequence *GAGTGATTGCTTGTGACGCCAA*, –soloCBposition *0_0_2_-1 3_1_3_8*, –soloUMIposition *3_9_3_16* (barcode list is provided in the Supplementary Table S3); (ii) for inDrops-2 (TS) with barcodes v2: –soloType *CB_UMI_Complex*, –soloCBmatchWLtype *EditDist_2*, –soloUMIdedup *Exact*, –soloCBposition *0_0_0_7 0_12_0_19*, –soloUMIposition *0_20_0_27* (barcode list is provided in the Supplementary Table S5); (iii) for 10× Genomics v3: –soloType *CB_UMI_Simple*, –soloCBmatchWLtype *1MM_multi*, –soloUMIdedup *1MM_CR*. Complete examples of the STAR parameters for each protocol are provided in the supplementary code (see the 'Data availability' section). The processing of the methanol-fixed hashtag scRNA-seq data was done using a combination of SEQC (48) and CITE-seq-Count (49) pipelines. The SEQC with default parameters for inDrops-2 (TS) (barcode version 2 (Supplementary Table S5)

was used to obtain cell × gene count matrix, while the CITE-seq-Count (v1.4.4) with the default parameters and *–no_umi_correction* was used to count hashtag sequences (Supplementary Table S6) and to generate cell × hashtag count matrix. Cells were classified as singlets or doublets and assigned to corresponding samples with HashSolo (50).

scRNA-seq data downsampling

The resulting bam files, tagged with corrected cell barcode (CB), UMI (UB) and gene (GN) information were parsed using the pysam v0.20.0 and down-sampled employing random sampling with the numpy v1.19.2 random.rand() approach. More explicitly, each cell barcode was either downsampled to a given number of raw reads (20 000 or 15 000) or to a given proportion of raw reads (5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90% and 100%), while counting the detected UMIs and unique genes. To generate saturation curves for each protocol version, the top quarter of cell barcodes based on UMI counts ($n > 1000$) were included. To evaluate whether the observed differences in UMI and gene count were statistically significant, two-tailed *t*-test from the rstatix v0.7.0 was applied in a pairwise manner on downsampled UMI and gene counts of cells that achieved required sequencing depth. The resulting *P*-values were adjusted for multiple corrections by Benjamini–Hochberg false discovery rate (FDR). Custom scripts used for the down-sampling are provided in supplementary code (see the 'Data availability' section).

Quality control and doublet detection

To remove cell barcodes with low complexity, the cell x gene count matrices were filtered by UMI counts (>2300 UMIs for PBMC samples and >1500 UMIs for CD34$^+$ bone marrow cells) and mitochondrial gene count fraction (>20% for all samples). For LC samples cells were filtered based on hashtag data, removing cells that were assigned multiple hashtags. Doublets were removed from PBMC and CD34$^+$ samples using Scrublet (51) algorithm by calculating doublet scores for each cell in each library, clustering cells in high resolution using Spectral Clustering in scikit-learn package, evaluating mean doublet score and fraction of predicted doublets per cluster, and removing clusters with high doublet score and doublet fraction.

Uniform manifold approximation and projection construction and cell type annotation

The filtered matrices from PBMC and LC samples were normalized to 10 000 total counts per cell (CP10k), log-transformed and scaled. After normalization, genes with at least 10 CP10k in at least five cells (10 cells for LC) were considered abundant and retained. Followed by exclusion of mitochondrial and ribosomal genes, top 3000 genes for PBMC and top 2000 genes for LC samples, based on Fano factor (22) were used for PCA. Dataset integration was performed using scanpy.external.pp.harmony_integrate() function in scanpy package (52) k-nearest neighbor graph was constructed using the adjusted principal components (number of nearest neighbors for PBMC samples k = 20, for LC samples k = 30) and was used to build uniform manifold approximation and projection (UMAP) representation. The resulting representation was used for exploration in interactive SPRING application (53). For PBMCs, initial clustering using Spectral Clustering revealed a cluster ($n = 1075$) with lower complexity which was removed from further analysis. Final clustering for PBMCs was performed using the number of clusters k = 10. Differential gene expression analysis (cluster versus rest of cells, Mann–Whitney U test with Benjamini–Hochberg correction) was performed and top 50 marker genes of each cluster (adjusted P-value < 0.05) were used for manual cell type annotation (Supplementary Table S7). Markers to identify cell types included *MS4A1* and *CD79A* (B cells), *LEF1* and *CD8B* (CD8 T cells), *IL7R* and *LTB* (CD4 T cells), *GNLY*, *NKG7*, (NK cells), *LYZ*, *S100A8* (CD14$^+$ Monocytes), *FCGR3A*, *LST1* (FCGR3A$^+$ Monocytes), *CLEC10A*, *FCER1A* (cDC), *LILRA4*, *PLD4* (pDC), *PPBP*, *PLD4* (Megakaryocytes). To annotate cell types in the LC data, first clustering using PhenoGraph Leiden algorithm (54) with parameter resolution = 2 revealed a cluster of lower complexity cells ($n = 1670$) with mitochondrial gene enrichment which were removed from further analysis. Following the removal of the cluster, a UMAP embedding was constructed again as described above. As initially, the graph was clustered using PhenoGraph Leiden implementation with resolution = 2 and resulting clusters were assigned into myeloid, lymphoid or non-immune groups based on FACS CD45 status (positive/negative) and expression of canonical markers (7). Finally, UMAPs were constructed as described above for each group separately (myeloid, lymphoid, epithelial/stromal) and clustered using PhenoGraph Leiden implementation with resolution = 0.5 for myeloid, resolution = 0.8 for lymphoid and resolution = 0.6 for epithelial/stromal compartment. Within each group, differential gene expression analysis was performed as described above. In total, 38 cellular phenotypes were annotated based on marker genes (list is provided in Supplementary Tables S8–S10) and the labels were then transferred to the original UMAP embedding containing all cells ($n = 32\,937$). For plotting purposes, cells were grouped by broad phenotype labels (i.e. epithelial, endothelial, etc.).

Cell trajectory reconstruction

To determine cell fate probabilities, we employed Palantir algorithm as described in Setty *et al.*, 2019 (55) To prepare the CD34$^+$ data for Palantir, briefly, the filtered data was normalized, log-transformed and scaled as described above. Highly variable genes were selected using the scanpy highly_variable_genes() function with flavor argument set to 'cell_ranger'. Cell cycle regression was performed using cell cycle genes defined in Tirosh *et al.*, 2015 (56) as an input for scanpy score_genes_cell_cycle() function, first, to calculate cell cycle scores and then to regress out G2M and S scores using regress_out() function. Next, PCA was performed, data was integrated using scanpy.external.pp.harmony_integrate() function in scanpy package and k-neighbor graph (k = 50) was constructed, as described above. The graph was used to build force-directed layout representation. Clustering and differential gene expression was performed as previously described above. Markers used for cell type annotation were *SPINK2*, *AVP* (HSC), *SPINK2*, *SMIM24* (MPP), *CD79B*, *VPREB1* (CLP), *ELANE*, *AZU1* (NMP), *KLF1*, *HBB* (Ery), *SCT*, *IRF8* (pDC), *PLEK*, *PPBP* (Mega), *LYZ*, *S100A9* (cDC), *CLC*, *PRG2* (Mast) (complete marker list is provided at Supplementary Table S11). Once data was prepared, Palantir package was used to construct diffusion maps using 10 components, and impute data using MAGIC. Early cell was determined in the HSC cluster employing the

early_cell() function and was used as a starting point to run the Palantir algorithm with 1200 waypoints. Main differentiation trajectories of hematopoiesis were identified using the following markers: *CD34* for hematopoietic stem cells, *IRF8* for dendritic cells, *KLF1* for erythroid, *MYADM* for myeloid, *CD79A* for lymphoid and *ITGA2B* for megakaryocytes.

Comparison of IVT- and TS-based inDrops-2

To calculate sequence alignment and quality control metrics, including gene-body coverages, mapping distributions across genomic features and read GC contents from the bam files, the ReSQC v5.0.1 (57) was employed with default parameters and GRCh38 genome (GENCODE v42) annotations. To compare gene length distributions between IVT- and TS-based protocols, gene lengths were retrieved from the .gtf annotation file (GENCODE v41, as described above) and were either binned into 10 or 3 categories, so that the number of genes is approximately equal in each of the category. The binning was performed using the ntile() function from dplyr v1.0.9. Filtered cell × gene expression data was size factor normalized and $\log_2(x + 1)$-transformed using normalizeCounts() with default parameters from the scuttle v1.6.3. Fractions of each gene length category per cell were calculated by dividing the sum of counts per category by the total counts. Differential gene expression analysis between IVT- and TS-based scRNA-Seq data was performed using the MAST package (58) More explicitly, hurdle models were fitted on filtered, size factor normalized and $\log_2(x + 1)$-transformed expression data using protocol type and number of detected genes (centered) as covariates to adjust for the cellular detection rate. Gene set enrichment analysis was performed with a sorted gene list (by descending $\log_2$ fold change values) and gene ontology (GO) (40) terms using the gseGO() function from ClusterProfiler package (59). The resulting *P*-values were adjusted using Benjamini–Hochberg FDR correction.

Results

Optimized inDrops for improved transcript and gene detection

To identify critical parameters that may improve transcript capture and detection, we re-examined the workflow of the original inDrops technique (22). Using a microfluidics setup reported in the past (42) and further detailed in Supplementary Figure S1, we encapsulated lymphoblast cells (K-562) in 1 nanolitre droplets together with barcoding hydrogel beads carrying photoreleasable RT primers comprising the T7 RNA polymerase promoter (T7p), cell barcode, unique molecular identifier (UMI) and poly(dT19) sequence (Supplementary Table S1). We set the flow rate of the microfluidics platform to achieve a high throughput of 1 million droplets/hour while maintaining hydrogel bead loading >85% and stable droplet formation (Supplementary Figure S1D). With this setup, ∼300 000 single cells can be encapsulated in less than 1 h, with a low (∼3%) doublet rate.

We subjected the encapsulated cells to RT reaction and prepared sequencing libraries following the workflow outlined in Figure 1A while systematically examining each step of the scRNA-seq protocol. Specifically, we profiled 1000–3000 cells per experimental condition by collecting ∼10 000 droplets, photoreleasing RT primers from the hydrogel beads and performing RT to produce barcoded cDNA, followed by second strand synthesis and IVT-mediated linear amplification. The IVT libraries were fragmented with Lewis acid (Zn^{2+} ions), transcribed into single-stranded cDNA and PCR-amplified with sequencing adapters to obtain final scRNA-seq libraries compatible with Illumina sequencers.

After a series of optimizations, we arrived at the inDrops-2 (IVT), described in Supplementary Protocol 1, which produced markedly improved transcript and gene detection (Figure 1B). The most important findings can be summarized as follows. Compared to the original inDrops (22), purification of barcoded cDNA and primer dimer removal by SPRI, rather than digestion with a nuclease cocktail, had the most noticeable impact (Figure 1B). At a sequencing depth of 15 000 reads/cell, inDrops-2 produced a 2.72-fold increase in UMI (mean ± s.d., 8166 ± 350 versus 2999 ± 461 UMIs; *t*-test, $P_{FDR} < 1 \times 10^{-300}$) and a 1.86-fold increase in gene capture (mean ± standard deviation (s.d.), 3399 ± 143 versus 1829 ± 207 genes; *t*-test, $P_{FDR} < 1 \times 10^{-300}$). Replacing SuperScript III with Maxima H-minus reverse transcriptase further increased UMI and gene capture 1.19-fold and 1.18-fold, respectively (*t*-test, UMI $P_{FDR} = 1.62 \times 10^{-281}$, gene $P_{FDR} = 5.82 \times 10^{-283}$). The improved UMI and gene detection using inDrops-2 was also observed in primary cells (Supplementary Figure S1E and F) and was not due to a preference for a specific RNA biotype (Supplementary Figure S1G), and overall displayed lower technical variability compared to original inDrops (Figure 1C). The inDrops-2 libraries prepared with SuperScript IV (SS-IV) enzyme had a slightly higher UMI (SS-IV 9108 ± 198 versus Maxima H- 9108 ± 709, *t*-test, $P_{FDR} = 6.22 \times 10^{-317}$) (Figure 1D). However, given the significantly higher cost of the SS-IV enzyme and the only relatively mild UMI improvements observed, this enzyme was excluded from our subsequent analyses. The inDrops-2 tended to show slightly higher UMI and gene counts (Figure 1D) as well as to contain lower proportion of ribosomal protein (RP) transcripts (Supplementary Figure S1H) when RT was performed at 42°C than at 50°C, irrespective of the RT enzyme used. No significant effect on UMI and gene recovery was observed when using different commercially available second-strand synthesis or IVT kits, indicating that the critical steps for obtaining high UMI counts are indeed related mainly to the RT and subsequent purification of barcoded cDNA molecules. In summary, compared to original inDrops, single-cell transcriptional profiling with inDrops-2 shows markedly improved UMI and gene detection, 20-fold greater throughput (275 droplets s^{-1} versus 15 droplets s^{-1}), and higher-quality data (Figure 1E–G). The step-by-step protocol incorporating the aforementioned improvements accompanies this manuscript as Supplementary Protocol 1.

inDrops-2 based on a TS reaction for rapid construction of sequencing libraries

While improved inDrops-2 based on linear amplification enables a substantially higher UMI and gene recovery per single cell, an alternative scRNA-seq strategy commonly used, often referred to as SMART, relies on an RT enzyme-driven TS reaction (60–62). Owing to its intrinsic terminal transferase activity, RTase tends to add a few nontemplated nucleotides (predominantly cytidines) to the 3′ end of cDNA, which occurs preferentially when the RNA template is G-capped (63–65). A large variety of scRNA-seq methods have exploited this unique RT enzyme feature to incorporate PCR adapters at

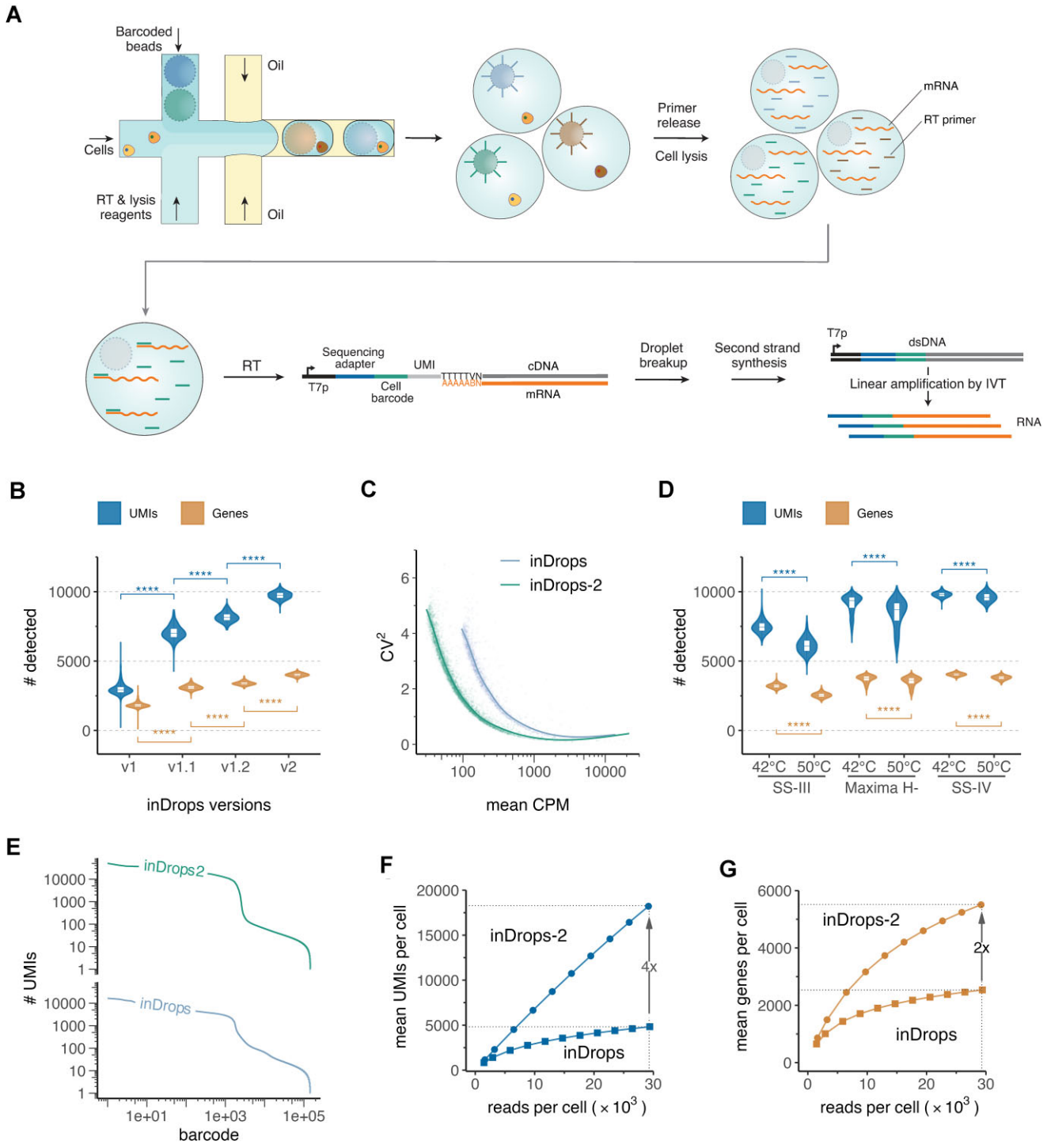


Figure 1. The overview of inDrops-2 performance. (**A**) Schematics of inDrops-2 technique using linear amplification by IVT of barcoded cDNA. (**B**) Detection of transcripts (UMIs) and genes in K-562 cell line, using different versions of inDrops (see the 'Materials and methods' section). To normalize for the sequencing depth, each cell barcode was downsampled to 15 000 raw reads. (**C**) Comparison of technical variability between inDrops and inDrops-2, where CPM and CV^2 refers to counts per million and squared coefficient of variation, respectively. (**D**) Evaluation of RT enzymes (SuperScript III Reverse Transcriptase [SS-III], Maxima H minus Reverse Transcriptase [Maxima H-] and SuperScript IV Reverse Transcriptase [SS-IV] and temperature (42 and 50°C) on the UMI and gene detection. Downsampled to 15 000 raw reads. (**E**) Barcode rank plots derived from scRNA-seq data acquired with inDrops-1 and inDrops-2 techniques. Barcodes having at least one UMI are arranged from the highest to the lowest UMI counts. (**F**) Mean UMI count per cell as a function of sequencing depth. (**G**) Mean gene count per cell as a function of sequencing depth. (**B**,**D**) Boxplots within density violins show median (center line), first and third quartiles (lower/upper hinges), 1.5 × interquartile range (lower/upper whiskers); ****P-value < 0.0001 (two-sided t-test, Benjamini–Hochberg correction).

the 5′ mRNA end and facilitate library preparation for next-generation sequencing (25,29,61,66). Motivated by these and other reports, we next sought to adopt a TS-based reaction with inDrops and to compare the UMI/gene capture results obtained using this approach to those obtained by linear amplification. To differentiate the two scRNA-seq approaches, we hereafter refer to them as inDrops-2 (IVT) and inDrops-2 (TS).

First, we aimed to maximize the yield of barcoded cDNA using the inDrops-2 (TS) approach (Figure 2A). For that purpose, we encapsulated K-562 cells together with barcoding hydrogel beads and RT/lysis reaction mix supplemented with TSO, which is required for barcoded cDNA amplification by PCR (see the 'Materials and methods' section). We thoroughly optimized each step of the workflow, namely, cell lysis and RT reaction, TSO concentration, temperature, barcoded cDNA purification, library fragmentation, A-tailing and adapter ligation (Supplementary Figure S2) to obtain a robust and reproducible scRNA-seq protocol (Supplementary Protocols 2 and 3).

Next, we evaluated the sensitivity of inDrops-2 (TS) by profiling human PBMCs and compared the results to those obtained with a commercial analogue (10× Genomics, Chromium v3 chemistry). At the same sequencing coverage, inDrops-2 (TS) detected nearly the same UMI and gene count in single cells as did the current gold-standard in the field (Figure 2B). The cell types comprising the biospecimens that were identified with the two techniques showed almost identical cell compositions (Figure 2C and D) and high correlations in gene expression were observed among the respective cell types (Figure 2E). Furthermore, deep sequencing (mean reads per cell: ∼221 000) revealed that inDrops-2 (TS) was sensitive enough to detect up to ∼140 000 UMIs (mean: 39 388) and up to ∼7500 genes (mean: 4909) in a murine LUAD cell line (39) at a sequencing saturation of 0.81 (Figure 2F and G). Taken together, these results confirmed that inDrops-2 (TS) can serve as a highly efficient and cost-effective method for profiling cells of different origins.

Comparison of scRNA-seq protocols based on linear amplification versus exponential amplification of cDNA

Having established inDrops-2 (TS), we then asked which scRNA-seq approach, IVT-based or TS-based, can deliver a greater number of unique transcripts and genes. While previous benchmarking studies have indicated that scRNA-seq libraries constructed by linear amplification (e.g. CEL-seq2) recover a higher diversity of genes than do those constructed by PCR-based approaches such as SMART-seq2 (67), to our knowledge, head-to-head comparisons of IVT- and TS-based approaches are lacking. To perform such a comparison, we first formed an emulsion comprising ∼5000 cells, which were compartmentalized into 1 nl droplets along with hydrogel beads carrying barcoding RT primers with the T7 RNA polymerase promoter sequence and 25 μM TSO. Upon completion of cDNA synthesis at 42°C for 90 min, the emulsion was divided into two equal fractions, which were processed separately following either the inDrops-2 (IVT) or inDrops-2 (TS) protocols (Figure 3A). Following this strategy, we prepared and sequenced lymphoblast cells (K-562) and primary human LUAD samples. After downsampling the sequencing depth of each library to 20 000 raw reads per cell, we found that both inDrops-2 (IVT) and inDrops-2 (TS) approaches recovered a similar number of UMIs per cell (Figure 3B). However, a significant fraction of the transcripts in the TS-based libraries corresponded to RP genes (Figure 3C), which are often excluded from downstream analyses. Removing the RP genes enhanced the differences between the two protocols, with the libraries constructed by linear amplification revealing ∼25% higher UMI and gene detection (Supplementary Figure S3A).

Interestingly, inDrops-2 (IVT) was also better at capturing unspliced RNAs, as confirmed by the greater fraction of reads aligning to introns and 3′ UTRs, which play a role in posttranscriptional gene expression regulation (Figure 3D). As a drawback, the IVT-based approach exhibited slightly higher technical noise and run-to-run variability (Supplementary Figure S3B). The TS-based libraries exhibited a slightly higher GC content than did the IVT-based libraries (Supplementary Figure S3C), in accordance with a higher fraction of coding sequences (68).

In contrast to the sharp enrichment at the 3′ end observed in the IVT-based approach, the gene body coverage in the TS-based approach was shifted (Figure 3E), implying a trend towards capturing the transcripts of shorter genes. Indeed, a striking difference between the two scRNA-seq protocols became evident when all of the detected genes were binned according to their length (Figure 3F and G). This analysis clearly revealed the bias of the TS-based approach towards shorter genes, a trend that was also observed in independent experiments on PBMCs and when using the 10× Genomics (v3) platform (Supplementary Figure S3D–G). Some of the observed differences could be attributed to an increased probability of stalling and drop-off events by reverse transcriptase while synthesizing cDNA of long RNA templates (Supplementary Figure S3H), to which the IVT-based approach is likely less sensitive since the truncated cDNAs can still be linearly amplified and subsequently sequenced. Additionally, cDNA products corresponding to the internal priming and transcripts that may lack a G-cap and A-tail, such as non-coding RNAs, were more frequently found in inDrops-2 (IVT) libraries (Supplementary Figure S3I). Therefore, single-cell transcriptome libraries prepared with TS-based or IVT-based approaches are prone to specific technical biases that, when unaccounted for, could negatively impact the interpretations of gene expression dynamics in individual cells, as shown by gene set enrichment analysis performed on differentially expressed genes (Figure 3H and Supplementary Tables S12 and S13). Nevertheless, both approaches provide high confidence in identifying different cell types (Supplementary Figure S3J) and demonstrate robustness in reproducibility (Figure 4 and Supplementary Figure S4), making them suitable for use in cellular composition profiling and tissue atlas construction.

Cell preservation for long-term storage and transcriptomic studies of primary cells

Besides improved transcript and gene capture, an important task in single-cell transcriptomics studies is to ensure that native transcriptional state is minimally affected by sample processing. To mitigate changes in the transcript profile that do not reflect their physiological state at the time of collection, it is desirable to safeguard the cellular transcriptome by fixing the cells so that no transcriptional changes occur during the encapsulation process. In this regard, cell preservation in methanol represents an appealing option because it retains

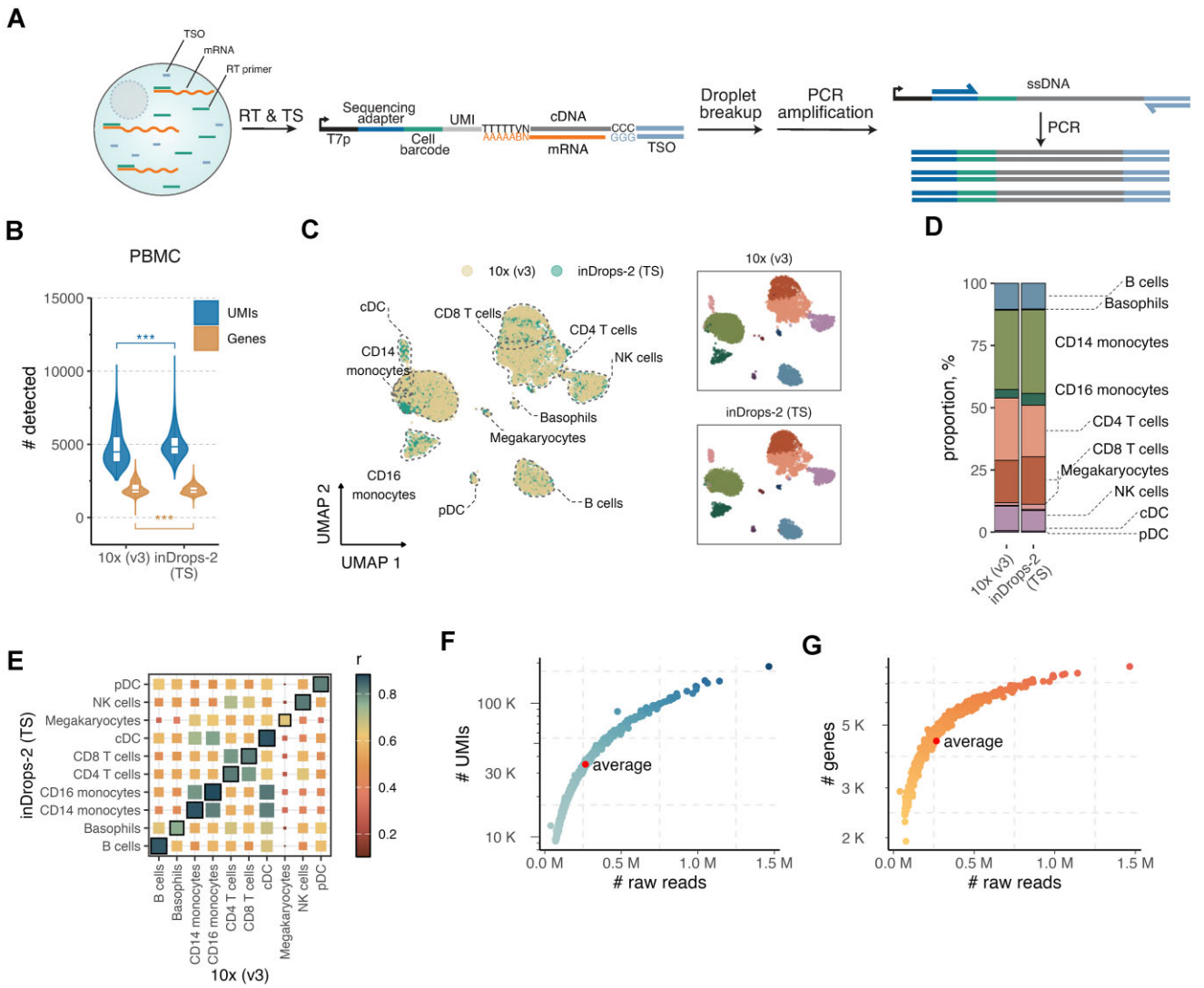


Figure 2. The overview of inDrops-2 using TS approach. (**A**) Schematics of inDrops-2 (TS) technique based on exponential cDNA amplification following TS reaction. (**B**) Comparison of transcript (UMIs) and gene detection in human PBMCs between 10× Genomics (v3) and inDrops-2 (TS) platforms at sequencing depth of 20 000 reads per cell. Boxplots within density violins show median (center line), first and third quartiles (lower/upper hinges), 1.5 × interquartile range (lower/upper whiskers); ***P-value < 0.001 (two-sided t-test, Benjamini–Hochberg correction). (**C**) Dimensionality reduction (UMAP) of human PBMCs profiled with 10× Genomics (v3) ($n = 4803$) and inDrops-2 (TS) ($n = 6025$) and colored by platform (left panel) and annotated cell type (right top and bottom panels). The UMAP representation is based on 3000 highly variable genes after data integration by Harmony (see the 'Materials and methods' section). (**D**) Comparison of cell type fractions recovered with 10× Genomics (v3) and inDrops-2 platforms. (**E**) Pearson's correlation analysis between highly variant genes in cell type clusters (cluster ellipsoids within percentile of 0.5) inferred by inDrops-2 (TS) [rows] and inferred by 10× Genomics (v3) (columns) platforms. (**F**) and (**G**) Deep sequencing results of scRNA-seq libraries prepared using inDrops-2 (TS) approach. Murine KP cell line was used as a model system. The plots show raw reads of a given cell barcode on the x-axis and UMI (**F**) as well as gene (**G**) counts on the y-axis in a $\log_{10}$ scale.

the transcriptional signatures of cells and is compatible with droplet microfluidics methods (38,41,69). Unfortunately, our attempts to adopt previously reported methanol-based cell preservation protocols were unsatisfactory for primary cells (e.g. PBMCs), as we witnessed significant RNA degradation and cell loss due to clumping (Supplementary Figures S5A and S5D). Primary cell recovery was particularly problematic when handling clinical samples comprising a low number (n ≤ 100 000) of cells. We reasoned that the excessive centrifugal force required to pellet the cells during rehydration might cause cell clumping and damage. Accordingly, after a series of independent tests, we found that methanol-fixed cells placed on 0.65 μm pore size filters (see Supplementary Protocol 4) can be effectively rehydrated without excessive centrifugal force. We confirmed that the RNA integrity of cells in rehydration buffer containing citrate remained high (RIN > 8.0) after 30 min on ice (Supplementary Figure S5B) and was only minimally affected by preservation time (up to 30 days) in methanol (Supplementary Figure S5C). Importantly, when rehydration columns were used, cell clumping was negligible (Supplementary Figure S5D), and the cellular morphology was consistently retained (Supplementary Figure S5E), thereby increasing the reproducibility of cell recovery.

We then applied inDrops-2 to compare UMI and gene capture between methanol-preserved and fresh human PBMCs from the same individual (Figure 5A–E). To avoid RT reaction inhibition by citrate present in the rehydration buffer, we adjusted the flow rates such that the final dilution of

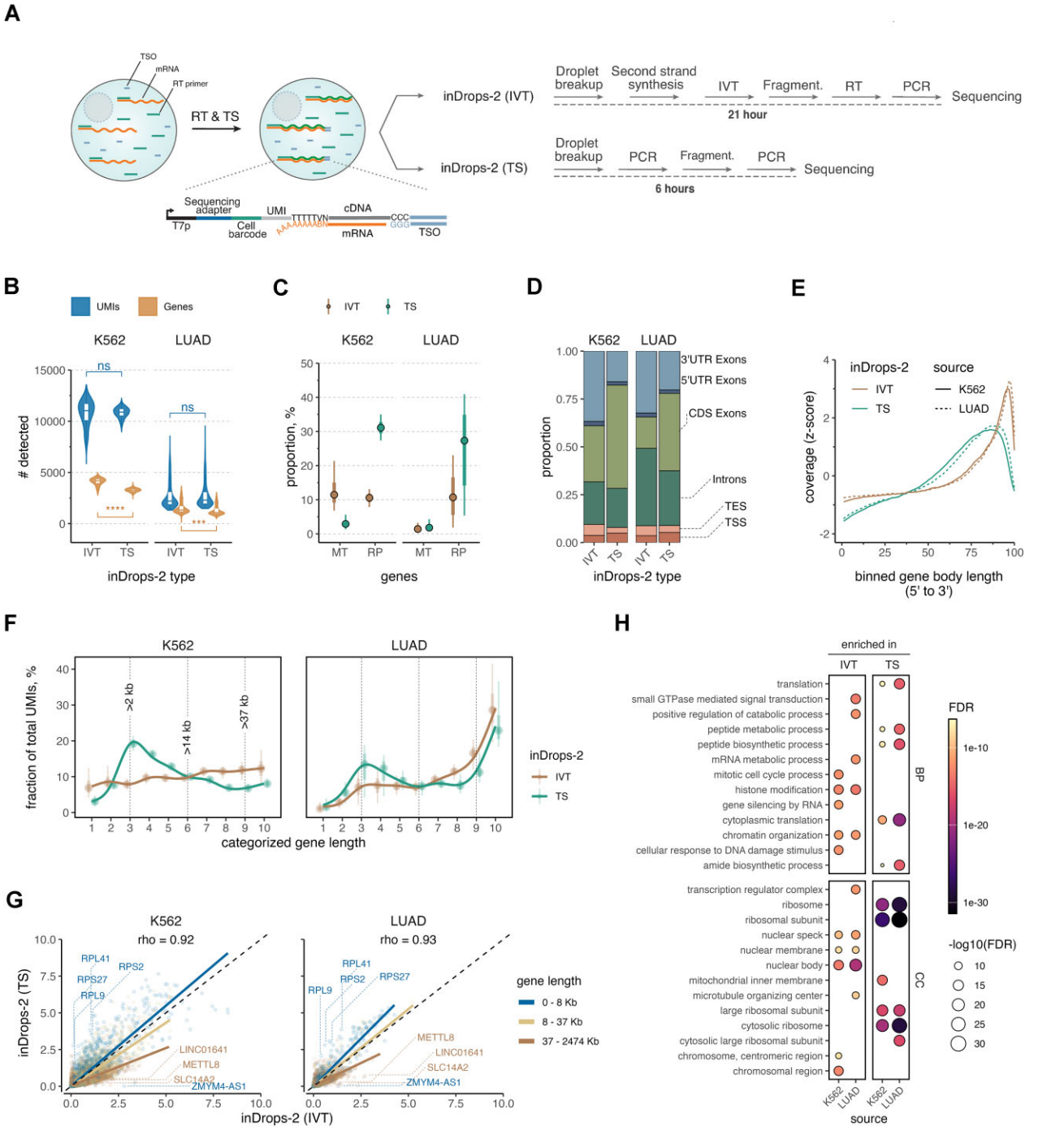


Figure 3. Comparative analysis of scRNA-seq libraries prepared with linear and exponential amplification of cDNA in primary and cultured cells. (**A**) Schematics of the experiment. A droplet-based RT reaction is performed in the presence of barcoded RT primers (comprising T7 promoter) and TSO. The post-RT emulsion droplets were split in two equal fractions and sequencing libraries prepared according to inDrops-2 (IVT) and inDrops-2 (TS) protocol. (**B**) Number of UMIs and genes detected in K-562 and primary LUAD cells in scRNA-seq libraries prepared by inDrops-2 (IVT) and inDrops-2 (TS). Sequencing depth was normalized to 20 000 raw reads per cell. Boxplots display median (center point), first and third quartiles (lower/upper hinges), 1.5 × interquartile range (lower/upper whiskers); ****P-value < 0.0001, ***P-value < 0.001, ^{ns}P-value > 0.05 (**C**) Fraction of genes corresponding to mitochondrial and RP in IVT-based and TS-based scRNA-seq libraries. (**D**) Fraction of reads mapping to different regions of a gene. (**E**) Sequencing coverage across the gene body. To adjust for the sequencing depth, raw coverage values were scaled and centered to a z-score. (**F**) Fraction of UMIs as a function of binned gene length. Fractions per cell, alongside with the boxplots, are displayed as curves fitted using loess smoothing and colored by the inDrops-2 protocol. (**G**) Correlation between inDrops-2 (IVT) and inDrops-2 (TS) protocols. Spearman's coefficient (rho) is depicted at the top of the scatter plot. Each dot represents normalized expression levels of detected genes. Dashed line (diagonal) divides panels in two equal parts, whereas blue, yellow and brown lines display linear regression curves corresponding to short (0–8 kb), medium (8–37 kb) and long gene length (37–2474 kb) categories, respectively. Top differentially expressed genes (lowest P-value, $n = 20$ in each cell type, based on MAST) that overlap in both primary and cultured cells are annotated using gene symbols. (**H**) GSEA performed on ordered gene list by the level of differential expression ($\log_2$ fold change) between IVT-based and TS-based scRNA-seq libraries. GO terms BP and CC refers to biological process and cellular compartment, respectively. FDR-adjusted P-values are presented as color gradient as well as are shown proportionally to dot size in $-\log_{10}$ scale. (**F**,**G**), Gene categories were determined by binning all genes based on their length into groups with approximately equal number of genes per category.

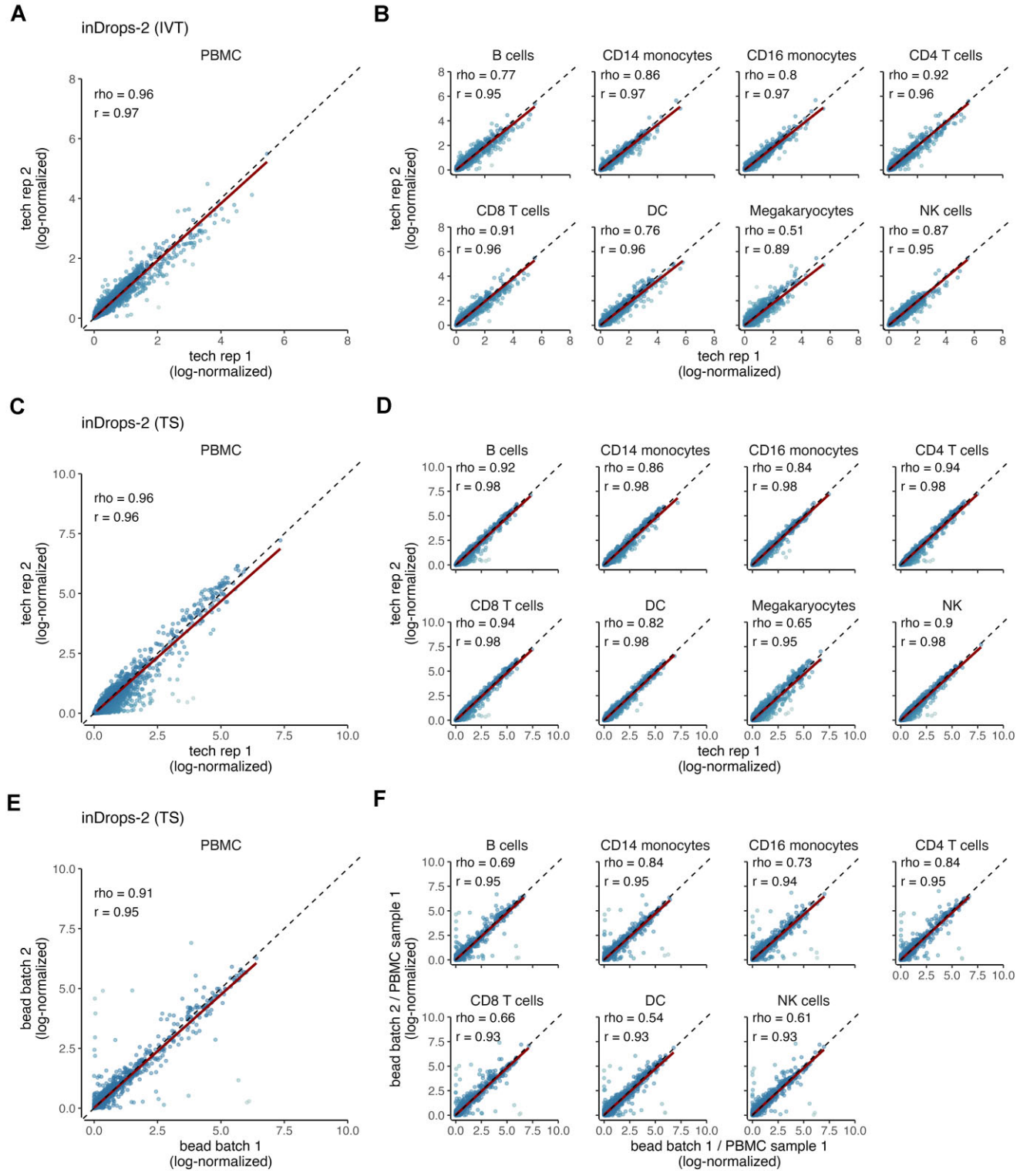


Figure 4. Reproducibility of inDrops-2 in PBMC samples. Scatterplots display the overall correlation in PBMC samples as well as correlations among different cell populations, comparing (**A, B**) technical replicates of inDrops-2 (IVT), (**C, D**) technical replicates of inDrops-2 (TS) prepared with the same batch of beads, and (**E, F**) technical replicates of inDrops-2 (TS) prepared with two separate batches of V2 beads. Log-normalized gene expression values were mean aggregated per sample [in panels (A), (C) and (E)] or mean aggregated per cell population within the sample [panels (B), (D) and (F)]. Correlation was evaluated using Spearman's correlation coefficient (rho) and Pearson's correlation coefficient (r).

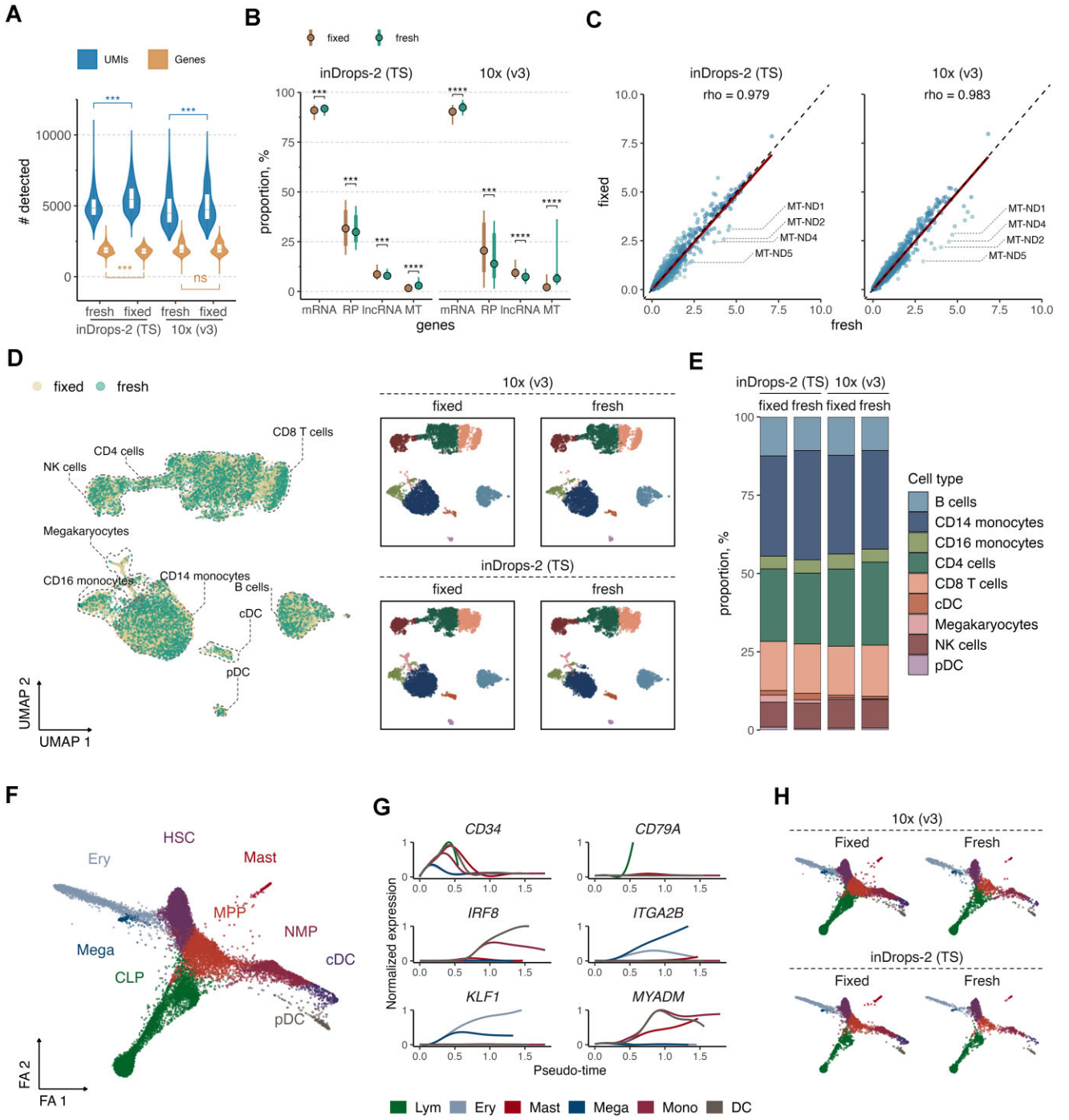


Figure 5. scRNA-seq of fresh and methanol-fixed PBMCs and CD34⁺ cells using inDrops-2 and 10× Genomics (v3) platforms. (**A**) UMI and gene detection in fresh and methanol-fixed PBMCs. (**B**) Fraction of UMIs along RNA biotypes. (**C**) Spearman's correlation analysis between fresh and methanol-fixed PBMCs from the same individual sequenced with inDrops-2 and 10× Genomics (v3) platforms. Correlation coefficient (rho) is shown at the top of the scatter plot. Size factor-normalized and $\log_2(x + 1)$-transformed gene expression levels are displayed as dots. The black dashed line represents diagonal, while the red solid line displays fitted linear regression. (**D**) UMAP of fresh and methanol-fixed PBMCs sequenced with inDrops-2 (TS) ($n = 12\,279$) and 10× Genomics (v3) ($n = 9582$), based on 3000 highly variable genes, and integrated using Harmony. The UMAP is colored by cell preservation type (left panel) and annotated cell type, faceted by platform and preservation type (right top and bottom panels). (**E**) Cell types and their fractions recovered in fresh and fixed PBMC samples in inDrops-2 (TS) and 10× Genomics (v3) methods. (**F**) Force Atlas (FA) embedding of human CD34⁺ profiled with 10× Genomics (v3) ($n = 10\,343$) and inDrops-2 (TS) ($n = 5029$) colored by annotated cell type (HSC, hematopoietic stem cell; MPP, multipotent progenitor; Ery, erythroid progenitor; Mega, megakaryocyte; CLP, common lymphoid progenitor; NMP, neutrophil-myeloid progenitor; cDC, conventional dendritic cell; pDC, plasmacytoid dendritic cell; Mast, Mast cell). The FA is based on 2000 highly variable genes with data integration by Harmony. (**G**) Gene expression trends for characteristic lineage genes along Palantir pseudo-time. Genes selected for each lineage were *CD34* for HSC, *CD79A* for CLP, *IRF8* for DC cells, *ITGA2B* for megakaryocytes, *KLF1* for erythroid cells and *MYADM* for myeloid cells. (**H**) FA from panel (F) colored by annotated cell type and faceted by platform as well as cell preservation method. (a), (b) median (center line or dot), first and third quartiles (lower/upper hinges), and 1.5 × interquartile range (lower/upper whiskers); $^{****}P$-value < 0.0001, $^{***}P$-value < 0.001, ^{ns}P-value > 0.05.

rehydration buffer in a droplet corresponded to 0.1 or 1.5 mM sodium citrate (see the 'Materials and methods' section). The UMI and gene detection results revealed that the inDrops-2 (TS) transcript recovery in fixed cells was efficient and, even though nominally statistically significant, closely matched that observed in live PBMCs and those obtained using a commercial platform (10× Chromium V3) (Figure 5A). As expected, a greater fraction of mitochondrial genes was found in live cells than in fixed cells (Figure 5B), indicative of transcriptional activity in fresh (unfixed) cells during cell handling procedures. The average gene expression levels in live and fixed cells exhibited a high correlation (Figure 5C) and similar gene mapping characteristics (Supplementary Figure S6A), showing remarkable reproducibility. Feature selection, dimensionality reduction and clustering of PBMCs revealed all expected cell populations, including CD4 T, CD8 T, NK, CD14 and CD16 monocytes, dendritic cells and megakaryocytes with similar cell proportions in fixed and fresh samples (Figure 5D and E, and Supplementary Figure S6B). The expression levels and detection rates of the PBMC marker genes resembled those of freshly profiled libraries; thus, indicating that there was no noticeable ambient RNA contamination in the methanol-fixed cells (Supplementary Figure S6C).

In addition to benchmarking fresh and methanol-fixed human PBMCs, we also tested CD34-positive ($CD34^+$) bone marrow cells. A transcriptional map of $CD34^+$ cells reconstructed haematopoietic stem cell differentiation into all known progenitor lineages (Figure 5F) and, analysis of marker gene expression, recapitulated the expected trends in haematopoiesis (Figure 5G). All major lineages, such as common lymphoid progenitors, erythroid precursor cells, megakaryocytes, neutrophil-myeloid progenitors, dendritic cell precursors and mast cell precursors, were present, and comparison of the fixed and fresh cells indicated that the cell type compositions were concordant in both (Figure 5H).

As a final quality assessment, we compared the mapping statistics in fresh and preserved cells and observed no significant differences (Supplementary Figure S6D). Overall, these results demonstrate that column-based rehydration of methanol-preserved cells ensures (i) minimal cell loss during handling, (ii) high singlet recovery, (iii) efficient UMI/gene detection, (iv) accurate recapitulation of the transcriptional signature matching that of live cells, and (v) alleviation of the adverse effects of cell viability decline (i.e. increased mitochondrial gene expression) caused by extended workflows.

Leveraging click chemistry hashtags to increase the scale of inDrops-2

Sample multiplexing with hashtags provides an appealing option for increasing the scale of scRNA-seq experiments (44,70–73). Exploiting methanol-based cell preservation, we applied multiplexing with methyltetrazine-modified DNA oligonucleotides, also known as 'ClickTags' (44). A distinct feature of ClickTags is that they do not rely on specific cell epitopes for labelling and can be chemically attached to cellular proteins by the Diels–Alder reaction, thus making them suitable for use in a broad range of cells and biospecimens. We sought to profile the human lung tumor microenvironment by conducting multiregional cell composition and gene expression analyses. The surgical samples acquired from LC patients were cut into three pieces, then the cells were dissociated and sorted into CD45-positive and CD45-negative compartments (74). Following FACS, the cell suspensions were preserved in methanol and transferred to −80°C for long-term storage. After 30 days, the cells were retrieved and, while in methanol, hashed with ClickTags (see the 'Materials and methods' section). In total, 18 samples were hashed, pooled, rehydrated (see Supplementary Protocol 4) and barcoded following the inDrops-2 (TS) workflow. After sequencing, filtering and quality control steps (see the 'Materials and methods' section), we obtained 32 937 high-quality cells with a consistent number of cells observed across hashtags (Figure 6A and Supplementary Figure S7A). The average UMI and gene count were high at 6959 and 1966, respectively (Figure 6B and C), and closely matched the sequencing statistics of fresh LUAD samples of our previous work (e.g. 5000 UMIs and 2100 genes, on average, at 40 000 reads per cell) (75). As expected, there was an inverted U-shaped dependency on ClickTag and UMI counts per cell ($R^2 = 0.24$; $P < 2\ 10^{-16}$), showing that excessive counts of hasthags leads to reduced UMI counts (Figure 6D).

To further characterize lung tumor samples we applied data normalization, feature selection, dimensionality reduction, clustering and visualization with UMAP. inDrops-2 recovered all major specialized lung epithelial, infiltrating stromal and immune cell phenotypes, including patient-specific cell populations, consistent with those reported previously using 10× Chromium platform (7,76,77) (Figure 6E–H). As expected, transcriptomes of FACS-sorted cells showed clear separation to CD45-positive and CD45-negative compartments (Figure 6G). As is common in tumors, there was high inter-patient variability in cellular composition (Figure 6F), while interregional differences within individual tumors were not as pronounced (Figure 6I). High-resolution analysis of the nonimmune cell compartment revealed lung-specialized epithelial cells, such as alveolar epithelial cells (AECs, markers *SFTPA1* and *HOPX*), club (*SCGB1A1, SCGB3A2*), ciliated (*CAPS, PIFO*), neuroendocrine (*CALCA, UCHL1*) and basal (*KRT17, KRT15*) cells. This analysis also recovered a patient-specific club cell population expressing several factors previously identified to be associated with cancer progression (78,79), namely, *SPINK1* and *CEACAM6* (Figure 6E and J).

Other nonimmune cells in our LC cohort included lymphatic endothelial (expressing *CCL21* and *NR2F2*) and tumor endothelial cells (*CLDN5, CLEC14A*), mesothelial cells (*MSLN, UPK3B*), smooth muscle cells (*ACTA2, TAGLN*) and diverse population of fibroblasts (*COL1A2, FN1*), which included two transcriptionally distinct groups involved in inflammation (Figure 6E). Complement-high fibroblasts had an unusually high expression of complement system constituents (i.e. *C7, C3* and *CFD*), indicative of inflammatory processes in the tumor microenvironment (Supplementary Figure S7B). Another fibroblast population was enriched for *HAS1* (Figure 6J), similar to an invasive fibroblast population recently discovered in fibrotic lungs (80). Moreover, this population of cells upregulated the expression of the potent chemokine for the attraction of monocytes, *CCL2*, and other inflammatory factors, such as *CXCL1, CXCL2* and *IL6* cytokines (Supplementary Figure S7B).

Consistent with the findings of previous reports (7,76,77), the myeloid compartment in our study included mast cells (*TPSB2, TPSAB1*), monocytes (*S100A9, FCN1*), conventional type 1 dendritic cells (*CLEC9A, CST3*), activated dendritic cells (*CCR7, CCL22*), monocyte-like dendritic cells (*CCL17, CLEC10A*), alveolar macrophages (*MARCO, FABP4*) as well

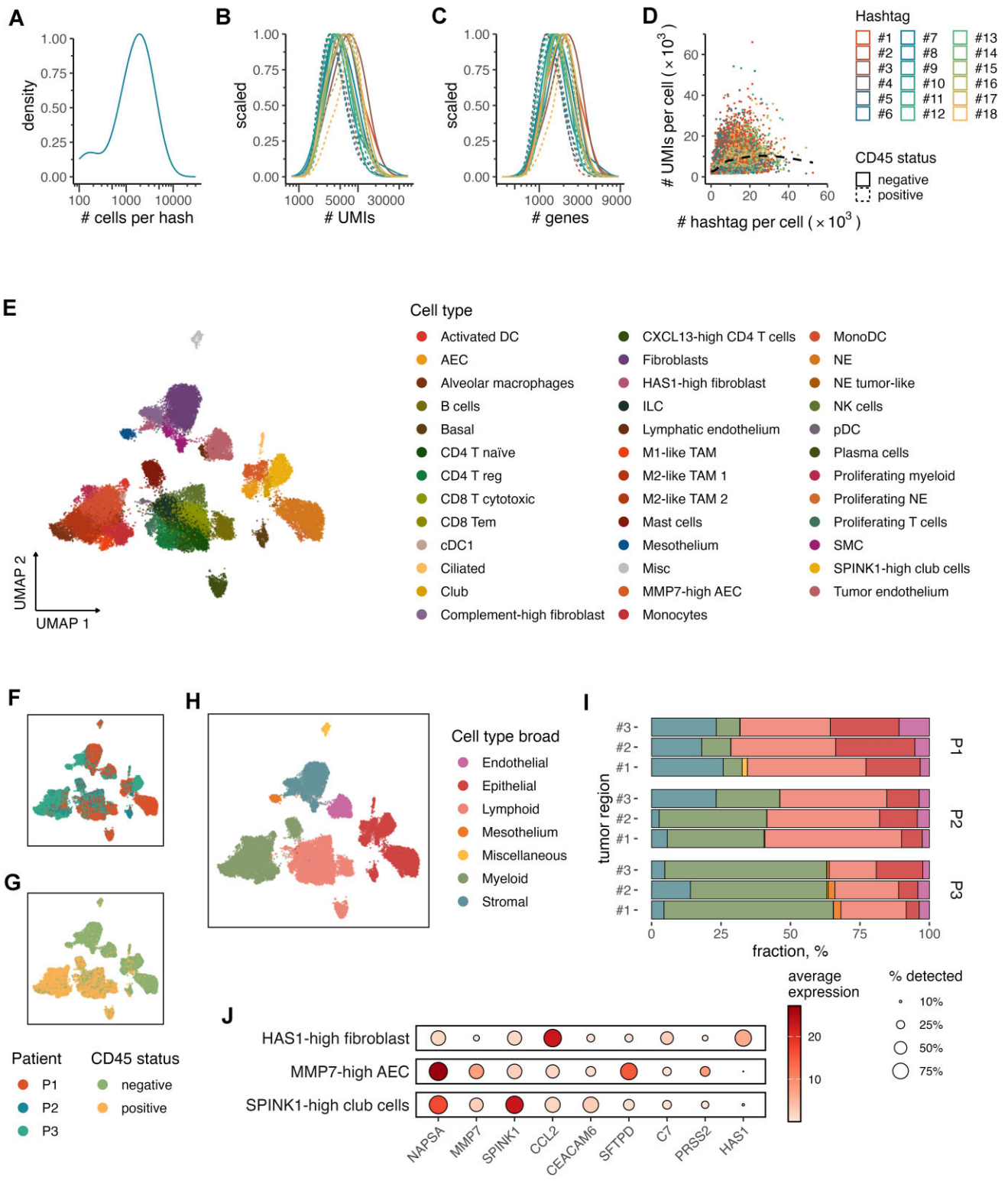


Figure 6. scRNA-seq of methanol-fixed and hashtag-indexed LC cells. The plots display the probability distributions of (**A**) cell number, (**B**) total UMI count, and (**C**) number of detected genes per individual hashtag, delineated by CD45 status. (**D**) Relationship between UMI and hashtag counts per cell, where x-axis displays hashtag counts and y-axis UMI counts of the same cell barcode (dot colored by an assigned hashtag). The dashed curve is fitted on the data using loess method and shows a decrease in UMI count, when more reads are mapped to hashtags. (**E**) An annotated UMAP of all LC cells (n = 32 937) displays high heterogeneity of cellular phenotypes in the tumor microenvironment. (**F**) and (**G**) UMAP colored by patient ID and CD45 status, respectively. (**H**) UMAP representation colored by broad cell type categories. (**I**) Sample composition analysis by broad cell type shows inter-patient compositional variability. (**J**) Marker gene expression in selected cell populations, displayed by CP10k-normalized mean expression (represented by color) and fraction of cells expressing those genes (indicated by dot size). AEC, alveolar epithelial cells; cDC1, conventional dendritic cells type I; ILC, innate lymphoid cell; NE, neuroendocrine cell; pDC, plasmacytoid dendritic cell; SMC, smooth muscle cell; TAM, tumor associated macrophage.

as M1- and M2-like subpopulations of tumor-associated macrophages. The lymphoid compartment consisted of innate lymphoid cells (ILCs, expressing *CD3D*), B cells (*CD79A, MS4A1*), plasma cells (*IGHG4, JCHAIN*), plasmacytoid dendritic cells (*LILRA4, CLIC3*), NK cells (*NKG7, GZMB*) and a large group of diverse T-cell phenotypes. Specifically, within the T-cell population, we captured CD4 regulatory T cells (*FOXP3, CTLA4*), naïve CD4 T cells (*IL7R, CCR7*), effector memory CD8 T cells (*CD52, S100A4*), cytotoxic CD8 T cells (*GZMA, CCL4*) and CXCL13-high CD4 T cells (Figure 6E). Interestingly, the patient P1 samples comprising the CXCL13-high T-cell phenotype coincided with a high count of B cells, therefore, supporting recent findings that CXCL13 acts as a potent attractant of B cells and other immune cells (81) (Supplementary Figure S7C). Overall, these findings clearly illustrate the ability of inDrops-2 to obtain high resolution gene expression profiles of tens of thousands of single cells in an efficient and inexpensive manner and to recover clinically relevant cell phenotypes from biospecimens that have undergone preservation, long-term storage and chemical multiplexing.

Discussion

Advancements in high-throughput scRNA-seq technologies (1) and computational methods (82) have opened new possibilities for investigating the gene expression programs and cellular composition in both normal and pathological conditions at unprecedented resolution and scale. As the range of scRNA-seq applications continues to expand across different domains of biomedical and biological sciences (16,17,83), there is a constant need for systems that not only deliver high throughput and sensitivity but are also cost effective. This is particularly relevant for analysis of biospecimens characterized by high heterogeneity, such as human tissues or cancer, necessitating the profiling of a large number of cells. Moreover, diverse needs of researchers often require versatile scRNA-seq platforms that can accommodate a broad range of samples and workflows.

Here, we present inDrops-2, an open-source scRNA-seq platform, which enables high-throughput single-cell transcriptomic studies with transcript and gene detection matching that of state-of-the-art commercial platforms (i.e. 10× Genomics Chromium v3) but at a 6-fold lower cost (Supplementary Table S14). The system is highly flexible and customizable, providing a straightforward option to implement user-specific workflows. For instance, we implemented and compared two highly sensitive scRNA-seq protocols—one based on linear amplification of cDNA via IVT and the other based on a TS reaction followed by exponential cDNA amplification by PCR. While both protocols were well suited for identifying different cell types and detecting similar transcript counts, they also exhibited important differences. The TS-based workflow had lower technical variability and require less labour to obtain sequencing results. However, due to the limited efficiency of TS reaction, scRNA-seq libraries produced using this approach tend to be enriched in shorter (≤14 kb) genes, which include a significant fraction of RPs. Given that the median length of protein-coding genes in humans is 26 kb (84), the aforementioned technical bias might skew conclusions in certain biological contexts. For example, it has been reported that longer genes tend to be associated with cell development, complex diseases and cancer, while shorter genes are more commonly associated with biological processes that require a rapid response, such as the immune system (85). There are indications that transcript length also plays a role in ageing (86). The scRNA-seq libraries constructed following IVT-mediated linear amplification of cDNA is labour intensive and relies on advanced molecular biology skillset. However, these libraries have greater complexity and capture more genes per single cell, including non-coding RNAs. Therefore, scRNA-seq libraries produced following IVT-mediated linear amplification of cDNA may be a better choice for studying gene expression at the whole-genome level.

Another important factor relevant to droplet-based scRNA-seq methods, including inDrops-2, is the quality of barcoded hydrogel beads. Our experience, along with previous reports (37,87) suggest that batch-to-batch variability can compromise reproducibility, impacting UMI and gene capture. To ensure consistent results, we emphasize the importance of cautiously adhering to the guidelines provided in the Supplementary Protocol 3, where we outline step-by-step procedure for producing high-quality barcoded hydrogel beads.

In addition to improving the performance of scRNA-seq using live cells, we developed a cell preservation procedure that safeguards intracellular mRNA from degradation and alleviates the technical challenges of working with primary cells, which are prone to transcriptional changes during sample handling. Importantly, while our procedure is loosely based upon those described in previous reports (38,41,69), in contrast to others, we apply a gentle cell rehydration process to ensure minimal cell clumping and loss, making the process suitable for use with clinical samples of limited volume ($n \leq 20\,000$ cells). The quality of the data from rehydrated cells was consistent with current standards in the field, with UMI/gene capture similar to that of live cells using both inDrops-2 and 10× Chromium platforms.

Taking one step further, we demonstrated the possibility of chemical hashing (indexing) of dehydrated cells based on the Diels–Alder reaction (44). Using inDrops-2, we performed multiregional profiling of LCs by hashing 18 samples that were methanol preserved after their collection from patients. We not only captured all major specialized lung epithelial, infiltrating stromal and immune cell phenotypes (7,76,77) but also identified cells with potential significance in immunotherapy, such as CXCL13-producing CD4 + T cells. It has been shown that the abundance of PDCD1-high CXCL13-producing T cells can be used to predict the response to PD-1 blockade therapy and is correlated with increased overall survival in non-small cell lung cancer patients (88). Furthermore, we captured several patient-specific populations two of which exhibited clinically relevant phenotypes. The club cells (positive for *SCGB3A1* and *SCGB3A2* markers) featured upregulated *SPINK1* expression that is known to enhance tumor cell proliferation and invasion *in vitro* and leads to adverse outcomes in a multitude of cancers (78). Interestingly, the *SPINK1*^{high} club cells also expressed *CEACAM6*, which has been implicated in lung cancer progression and poor clinical outcomes (79). Another relevant phenotype was specific to the alveolar epithelium. We observed two distinct phenotypes of alveolar epithelial cells — both of which expressed canonical AEC markers (*SFTPA1, SFTPA2* and *SFTPC*); however, only one population was marked by elevated *MMP7* and *PRSS2* co-expression. *MMP7* is a widely used biomarker for pulmonary fibrosis (89), while *PRSS2* is associated with invasive and metastasis-promoting features (90); thus, our results suggest that the *MMP7*^{high} alveolar epithelial cell pheno-

type might be involved in disease progression and plasticity. These results underscore the broad potential of inDrops-2 in biomedical research, especially in scenarios where characterization of complex diseases at the single-cell level is highly relevant. Moreover, chemical cell hashing of methanol-fixed cells with DNA oligonucleotides (ClickTags) not only minimizes technical batch effects but also benefits data analysis by facilitating unbiased removal of cell doublets while preserving clinically relevant cellular phenotypes.

As an open platform, inDrops-2 can accommodate diverse samples and workflows matching user-specific needs and has the potential to further democratize single-cell technologies. The inDrops-2 protocols reported in this work (Supplementary Material) can benefit other platforms such as VASA-seq (35), which outperforms commercial and plate-based methods in terms of gene and transcript capture, and spinDrops (91) that offers target cell enrichment by on-chip sorting. In addition to transcriptomics, inDrops-2 can be adapted to probe other -omic modalities in individual cells, similar to open chromatin by HyDrop (33) or the genome by Microbe-seq (92). In conclusion, inDrops-2 represents an affordable, sensitive and adjustable open-source system that could expand the scope of single-cell omics applications and further enhance the scalability of scRNA-seq experiments.

Data availability

The single-cell RNA-seq data presented in this work have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB71611. The code utilized to process the data and generate the main results is accessible at the https://github.com/mazutislab/indrops-2 repository and Zenodo at https://doi.org/10.5281/zenodo.14245235.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

We thank the Integrated Genomics Operation at MSKCC and Genomics Core Facility at EMBL Heidelberg for their assistance. We are grateful to Ignas Masilionis for assistance and to the members of the Single Cell Research Initiative (SCRI) at the Memorial Sloan Kettering Cancer Center for valuable support.

Author contributions: S.J., K.G., V.K. and J.Z.: data analysis and interpretation; K.G., V.K. and J.N.: method development, single-cell RNA-seq experiments; A.Q.V.: biospecimen acquisition, logistics and processing; J.S. provision of analytical reagents and materials; L.M.: study design, supervision and funding acquisition. L.M. wrote the initial draft of the manuscript. S.J., K.G., J.Z. and L.M. revised the manuscript. All the authors have read and approved the final manuscript.

Funding

European Regional Development Fund [01.2.2-LMT-K-718-04-0002]; Research Council of Lithuania; Alan and Sandra Gerry Metastasis and Tumour Ecosystems Center; HORIZON EUROPE Marie Skłodowska-Curie Actions [Grant agreement ID: 101030265 to S.J.].

Conflict of interest statement

J.N. and V.K. are employees at Droplet Genomics. J.S. is an employee at Thermo Fisher Scientific. J.N. and L.M. are shareholders of Droplet Genomics.

References

1. Vandereyken,K., Sifrim,A., Thienpont,B. and Voet,T. (2023) Methods and applications for single-cell and spatial multi-omics. *Nat. Rev. Genet.*, **24**, 494–515.
2. Zhu,C., Preissl,S. and Ren,B. (2020) Single-cell multimodal omics: the power of many. *Nat. Methods*, **17**, 11–14.
3. Plasschaert,L.W., Zilionis,R., Choo-Wing,R., Savova,V., Knehr,J., Roma,G., Klein,A.M. and Jaffe,A.B. (2018) A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature*, **560**, 377–381.
4. Villani,A.-C., Satija,R., Reynolds,G., Sarkizova,S., Shekhar,K., Fletcher,J., Griesbeck,M., Butler,A., Zheng,S., Lazo,S., *et al.* (2017) Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science*, **356**, eaah4573.
5. Keren-Shaul,H., Spinrad,A., Weiner,A., Matcovitch-Natan,O., Dvir-Szternfeld,R., Ulland,T.K., David,E., Baruch,K., Lara-Astaiso,D., Toth,B., *et al.* (2017) A unique microglia type associated with restricting development of Alzheimer's disease. *Cell*, **169**, 1276–1290.
6. Azizi,E., Carr,A.J., Plitas,G., Cornish,A.E., Konopacki,C., Prabhakaran,S., Nainys,J., Wu,K.M., Kiseliovas,V., Setty,M., *et al.* (2018) Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell*, **174**, 1293.
7. Laughney,A.M., Hu,J., Campbell,N.R., Bakhoum,S.F., Setty,M., Lavallee,V.P., Xie,Y., Masilionis,I., Carr,A.J., Kottapalli,S., *et al.* (2020) Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.*, **26**, 259–269.
8. Nowicki-Osuch,K., Zhuang,L., Jammula,S., Bleaney,C.W., Mahbubani,K.T., Devonshire,G., Katz-Summercorn,A., Eling,N., Wilbrey-Clark,A., Madissoon,E., *et al.* (2021) Molecular phenotyping reveals the identity of Barrett's esophagus and its malignant transition. *Science*, **373**, 760–767.
9. Sade-Feldman,M., Yizhak,K., Bjorgaard,S.L., Ray,J.P., de Boer,C.G., Jenkins,R.W., Lieb,D.J., Chen,J.H., Frederick,D.T., Barzily-Rokni,M., *et al.* (2018) Defining T cell states associated with response to checkpoint immunotherapy in melanoma. *Cell*, **175**, 998–1013.
10. Jerby-Arnon,L., Shah,P., Cuoco,M.S., Rodman,C., Su,M.-J., Melms,J.C., Leeson,R., Kanodia,A., Mei,S., Lin,J.-R., *et al.* (2018) A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell*, **175**, 984–997.
11. Xue,J.Y., Zhao,Y., Aronowitz,J., Mai,T.T., Vides,A., Qeriqi,B., Kim,D., Li,C., de Stanchina,E., Mazutis,L., *et al.* (2020) Rapid non-uniform adaptation to conformation-specific KRAS(G12C) inhibition. *Nature*, **577**, 421–425.
12. Briggs,J.A., Weinreb,C., Wagner,D.E., Megason,S., Peshkin,L., Kirschner,M.W. and Klein,A.M. (2018) The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science*, **360**, 981–987.
13. Zilionis,R., Engblom,C., Pfirschke,C., Savova,V., Zemmour,D., Saatcioglu,H.D., Krishnan,I., Maroni,G., Meyerovitz,C.V., Kerwin,C.M., *et al.* (2019) Single-cell transcriptomics of Human and mouse lung cancers reveals conserved myeloid populations across individuals and species. *Immunity*, **50**, 1317–1334.
14. The Tabula Muris Consortium (2018) Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, **562**, 367–372.
15. Liao,M., Liu,Y., Yuan,J., Wen,Y., Xu,G., Zhao,J., Cheng,L., Li,J., Wang,X., Wang,F., *et al.* (2020) Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.*, **26**, 842–844.

16. Shaw,R., Tian,X. and Xu,J. (2021) Single-cell transcriptome analysis in plants: advances and challenges. *Mol. Plant*, **14**, 115–126.
17. Elmentaite,R., Domínguez Conde,C., Yang,L. and Teichmann,S.A. (2022) Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.*, **23**, 395–410.
18. Regev,A., Teichmann,S.A., Lander,E.S., Amit,I., Benoist,C., Birney,E., Bodenmiller,B., Campbell,P., Carninci,P., Clatworthy,M., *et al.* (2017) The Human cell atlas. *eLife*, **6**, e27041.
19. Rozenblatt-Rosen,O., Regev,A., Oberdoerffer,P., Nawy,T., Hupalowska,A., Rood,J.E., Ashenberg,O., Cerami,E., Coffey,R.J., Demir,E., *et al.* (2020) The Human Tumor Atlas network: charting tumor transitions across space and time at single-cell resolution. *Cell*, **181**, 236–249.
20. Jones,R.C., Karkanias,J., Krasnow,M.A., Pisco,A.O., Quake,S.R., Salzman,J., Yosef,N., Bulthaup,B., Brown,P., Harper,W., *et al.* (2022) The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science*, **376**, eabl4896.
21. Macosko,E.Z., Basu,A., Satija,R., Nemesh,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.
22. 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.
23. Picelli,S., Faridani,O.R., Björklund,Å.K., Winberg,G., Sagasser,S. and Sandberg,R. (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.*, **9**, 171–181.
24. 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.
25. Zheng,G.X.Y., Terry,J.M., Belgrader,P., Ryvkin,P., Bent,Z.W., Wilson,R., Ziraldo,S.B., Wheeler,T.D., McDermott,G.P., Zhu,J., *et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.*, **8**, 14049.
26. Gao,R., Kim,C., Sei,E., Foukakis,T., Crosetto,N., Chan,L.-K., Srinivasan,M., Zhang,H., Meric-Bernstam,F. and Navin,N. (2017) Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer. *Nat. Commun.*, **8**, 228.
27. Bose,S., Wan,Z., Carr,A., Rizvi,A.H., Vieira,G., Pe'er,D. and Sims,P.A. (2015) Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol.*, **16**, 120.
28. Drake,R.S., Villanueva,M.A., Vilme,M., Russo,D.D., Navia,A., Love,J.C. and Shalek,A.K. (2023) Profiling transcriptional heterogeneity with seq-Well S3: a low-cost, portable, high-fidelity platform for massively parallel single-cell RNA-Seq. In: *Single cell transcriptomics: Methods and protocols*. Springer US Humana, NY, pp. 57–104.
29. Hagemann-Jensen,M., Ziegenhain,C. and Sandberg,R. (2022) Scalable single-cell RNA sequencing from full transcripts with Smart-seq3xpress. *Nat. Biotechnol.*, **40**, 1452–1457.
30. Haber,A.L., Biton,M., Rogel,N., Herbst,R.H., Shekhar,K., Smillie,C., Burgin,G., Delorey,T.M., Howitt,M.R., Katz,Y., *et al.* (2017) A single-cell survey of the small intestinal epithelium. *Nature*, **551**, 333–339.
31. Gao,R., Bai,S., Henderson,Y.C., Lin,Y., Schalck,A., Yan,Y., Kumar,T., Hu,M., Sei,E., Davis,A., *et al.* (2021) Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat. Biotechnol.*, **39**, 599–608.
32. Davey,K., Wong,D., Konopacki,F., Kwa,E., Ly,T., Fiegler,H. and Sibley,C.R. (2021) A flexible microfluidic system for single-cell transcriptome profiling elucidates phased transcriptional regulators of cell cycle. *Sci. Rep.*, **11**, 7918.
33. De Rop,F.V., Ismail,J.N., Bravo González-Blas,C., Hulselmans,G.J., Flerin,C.C., Janssens,J., Theunis,K., Christiaens,V.M., Wouters,J., Marcassa,G., *et al.* (2022) Hydrop enables droplet-based single-cell ATAC-seq and single-cell RNA-seq using dissolvable hydrogel beads. *eLife*, **11**, e73971.
34. Habib,N., Avraham-Davidi,I., Basu,A., Burks,T., Shekhar,K., Hofree,M., Choudhury,S.R., Aguet,F., Gelfand,E., Ardlie,K., *et al.* (2017) Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods*, **14**, 955–958.
35. Salmen,F., De Jonghe,J., Kaminski,T.S., Alemany,A., Parada,G.E., Verity-Legg,J., Yanagida,A., Kohler,T.N., Battich,N., van den Brekel,F., *et al.* (2022) High-throughput total RNA sequencing in single cells using VASA-seq. *Nat. Biotechnol.*, **40**, 1780–1793.
36. Karaiskos,N., Wahle,P., Alles,J., Boltengagen,A., Ayoub,S., Kipar,C., Kocks,C., Rajewsky,N. and Zinzen,R.P. (2017) The *Drosophila* embryo at single-cell transcriptome resolution. *Science*, **358**, 194–199.
37. Zhang,X., Li,T., Liu,F., Chen,Y., Yao,J., Li,Z., Huang,Y. and Wang,J. (2019) Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-seq systems. *Mol. Cell*, **73**, 130–142.
38. Alles,J., Karaiskos,N., Praktiknjo,S.D., Grosswendt,S., Wahle,P., Ruffault,P.-L., Ayoub,S., Schreyer,L., Boltengagen,A., Birchmeier,C., *et al.* (2017) Cell fixation and preservation for droplet-based single-cell transcriptomics. *BMC Biol.*, **15**, 44.
39. Sánchez-Rivera,F.J., Diaz,B.J., Kastenhuber,E.R., Schmidt,H., Katti,A., Kennedy,M., Tem,V., Ho,Y.J., Leibold,J., Paffenholz,S.V., *et al.* (2022) Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. *Nat. Biotechnol.*, **40**, 862–873.
40. Aleksander,S.A., Balhoff,J., Carbon,S., Cherry,J.M., Drabkin,H.J., Ebert,D., Feuermann,M., Gaudet,P., Harris,N.L., Hill,D.P., *et al.* (2023) The gene ontology knowledgebase in 2023. *Genetics*, **224**, iyad031.
41. Chen,J., Cheung,F., Shi,R., Zhou,H. and Lu,W. (2018) PBMC fixation and processing for Chromium single-cell RNA sequencing. *J. Transl. Med.*, **16**, 198.
42. Zilionis,R., Nainys,J., Veres,A., Savova,V., Zemmour,D., Klein,A.M. and Mazutis,L. (2017) Single-cell barcoding and sequencing using droplet microfluidics. *Nat. Protoc.*, **12**, 44–73.
43. Southard-Smith,A.N., Simmons,A.J., Chen,B., Jones,A.L., Ramirez Solano,M.A., Vega,P.N., Scurrah,C.R., Zhao,Y., Brenan,M.J., Xuan,J., *et al.* (2020) Dual indexed library design enables compatibility of in-drop single-cell RNA-sequencing with exAMP chemistry sequencing platforms. *BMC Genomics*, **21**, 456.
44. Gehring,J., Hwee Park,J., Chen,S., Thomson,M. and Pachter,L. (2019) Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins. *Nat. Biotechnol.*, **38**, 35–38.
45. Zainabadi,K., Dhayabaran,V., Moideen,K. and Krishnaswamy,P. (2019) An efficient and cost-effective method for purification of small sized DNAs and RNAs from human urine. *PLoS One*, **14**, e0210813.
46. 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.
47. Shen,W., Le,S., Li,Y. and Hu,F. (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*, **11**, e0163962.
48. Azizi,E., Carr,A.J., Plitas,G., Cornish,A.E., Konopacki,C., Prabhakaran,S., Nainys,J., Wu,K., Kiseliovas,V., Setty,M., *et al.* (2018) Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell*, **174**, 1293–1308.
49. Stoeckius,M., Zheng,S., Houck-Loomis,B., Hao,S., Yeung,B.Z., Mauck,W.M. 3rd, Smibert,P. and Satija,R. (2018) Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.*, **19**, 224.
50. Bernstein,N.J., Fong,N.L., Lam,I., Roy,M.A., Hendrickson,D.G. and Kelley,D.R. (2020) Solo: doublet identification in single-cell RNA-seq via semi-supervised deep learning. *Cell Syst.*, **11**, 95–101.

51. Wolock,S.L., Lopez,R. and Klein,A.M. (2019) Scrublet: computational identification of cell doublets in single-cell transcriptomic data. Cell Syst., **8**, 281–291.
52. Wolf,F.A., Angerer,P. and Theis,F.J. (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.*, **19**, 15.
53. Weinreb,C., Wolock,S. and Klein,A.M. (2018) SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. *Bioinformatics*, **34**, 1246–1248.
54. Levine,J.H., Simonds,E.F., Bendall,S.C., Davis,K.L., Amir el,A.D., Tadmor,M.D., Litvin,O., Fienberg,H.G., Jager,A., Zunder,E.R., *et al.* (2015) Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell*, **162**, 184–197.
55. Setty,M., Kiseliovas,V., Levine,J., Gayoso,A., Mazutis,L. and Pe'er,D. (2019) Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.*, **37**, 451–460.
56. Kowalczyk,M.S., Tirosh,I., Heckl,D., Rao,T.N., Dixit,A., Haas,B.J., Schneider,R.K., Wagers,A.J., Ebert,B.L. and Regev,A. (2015) Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res.*, **25**, 1860–1872.
57. Wang,L., Wang,S. and Li,W. (2012) RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, **28**, 2184–2185.
58. Finak,G., McDavid,A., Yajima,M., Deng,J., Gersuk,V., Shalek,A.K., Slichter,C.K., Miller,H.W., McElrath,M.J., Prlic,M., *et al.* (2015) MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.*, **16**, 278.
59. Yu,G., Wang,L.G., Han,Y. and He,Q.Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*, **16**, 284–287.
60. Matz,M., Shagin,D., Bogdanova,E., Britanova,O., Lukyanov,S., Diatchenko,L. and Chenchik,A. (1999) Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Res.*, **27**, 1558–1560.
61. Ramsköld,D., Luo,S., Wang,Y.-C., Li,R., Deng,Q., Faridani,O.R., Daniels,G.A., Khrebtukova,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.
62. Zhu,Y.Y., Machleder,E.M., Chenchik,A., Li,R. and Siebert,P.D. (2001) Reverse transcriptase template switching: a SMART™ approach for full-length cDNA library construction. *BioTechniques*, **30**, 892–897.
63. Schmidt,W. (1999) CapSelect: a highly sensitive method for 5' CAP-dependent enrichment of full-length cDNA in PCR-mediated analysis of mRNAs. *Nucleic Acids Res.*, **27**, e31.
64. Tang,D.T., Plessy,C., Salimullah,M., Suzuki,A.M., Calligaris,R., Gustincich,S. and Carninci,P. (2013) Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Res.*, **41**, e44.
65. Wulf,M.G., Maguire,S., Humbert,P., Dai,N., Bei,Y., Nichols,N.M., Corrêa,I.R. and Guan,S. (2019) Non-templated addition and template switching by Moloney murine leukemia virus (MMLV)-based reverse transcriptases co-occur and compete with each other. *J. Biol. Chem.*, **294**, 18220–18231.
66. Hahaut,V., Pavlinic,D., Carbone,W., Schuierer,S., Balmer,P., Quinodoz,M., Renner,M., Roma,G., Cowan,C.S. and Picelli,S. (2022) Fast and highly sensitive full-length single-cell RNA sequencing using FLASH-seq. *Nat. Biotechnol.*, **40**, 1447–1451.
67. Mereu,E., Lafzi,A., Moutinho,C., Ziegenhain,C., McCarthy,D.J., Álvarez-Varela,A., Batlle,E.,,Sagar, Grün,D., Lau,J.K., *et al.* (2020) Benchmarking single-cell RNA-sequencing protocols for cell atlas projects. *Nat. Biotechnol.*, **38**, 747–755.
68. Zhang,L., Kasif,S., Cantor,C.R. and Broude,N.E. (2004) GC/AT-content spikes as genomic punctuation marks. *Proc. Natl Acad. Sci. U.S.A.*, **101**, 16855–16860.
69. Katzenelenbogen,Y., Sheban,F., Yalin,A., Yofe,I., Svetlichnyy,D., Jaitin,D.A., Bornstein,C., Moshe,A., Keren-Shaul,H., Cohen,M., *et al.* (2020) Coupled scRNA-seq and intracellular protein activity reveal an immunosuppressive role of TREM2 in cancer. *Cell*, **182**, 872–885.
70. McGinnis,C.S., Patterson,D.M., Winkler,J., Conrad,D.N., Hein,M.Y., Srivastava,V., Hu,J.L., Murrow,L.M., Weissman,J.S., Werb,Z., *et al.* (2019) MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat. Methods*, **16**, 619–626.
71. Peterson,V.M., Zhang,K.X., Kumar,N., Wong,J., Li,L., Wilson,D.C., Moore,R., McClanahan,T.K., Sadekova,S. and Klappenbach,J.A. (2017) Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.*, **35**, 936–939.
72. Shahi,P., Kim,S.C., Haliburton,J.R., Gartner,Z.J. and Abate,A.R. (2017) Abseq: ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci. Rep.*, **7**, 44447.
73. Stoeckius,M., Hafemeister,C., Stephenson,W., Houck-Loomis,B., Chattopadhyay,P.K., Swerdlow,H., Satija,R. and Smibert,P. (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods*, **14**, 865–868.
74. Quintanal-Villalonga,Á., Chan,J.M., Masilionis,I., Gao,V.R., Xie,Y., Allaj,V., Chow,A., Poirier,J.T., Pe'er,D., Rudin,C.M., *et al.* (2022) Protocol to dissociate, process, and analyze the human lung tissue using single-cell RNA-seq. *STAR Protoc.*, **3**, 101776.
75. Chan,J.M., Quintanal-Villalonga,Á., Gao,V.R., Xie,Y., Allaj,V., Chaudhary,O., Masilionis,I., Egger,J., Chow,A., Walle,T., *et al.* (2021) Signatures of plasticity, metastasis, and immunosuppression in an atlas of human small cell lung cancer. *Cancer Cell*, **39**, 1479–1496.
76. Bischoff,P., Trinks,A., Obermayer,B., Pett,J.P., Wiederspahn,J., Uhlitz,F., Liang,X., Lehmann,A., Jurmeister,P., Elsner,A., *et al.* (2021) Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene*, **40**, 6748–6758.
77. Sinjab,A., Han,G., Treekitkarnmongkol,W., Hara,K., Brennan,P.M., Dang,M., Hao,D., Wang,R., Dai,E., Dejima,H., *et al.* (2021) Resolving the spatial and cellular architecture of lung adenocarcinoma by multiregion single-cell sequencing. *Cancer Discov.*, **11**, 2506–2523.
78. Xu,L., Lu,C., Huang,Y., Zhou,J., Wang,X., Liu,C., Chen,J. and Le,H. (2018) SPINK1 promotes cell growth and metastasis of lung adenocarcinoma and acts as a novel prognostic biomarker. *BMB Rep.*, **51**, 648.
79. Son,S.M., Yun,J., Lee,S.H., Han,H.S., Lim,Y.H., Woo,C.G., Lee,H.C., Song,H.G., Gu,Y.M., Lee,H.J., *et al.* (2019) Therapeutic effect of pHLIP-mediated CEACAM6 gene silencing in lung adenocarcinoma. *Sci. Rep.*, **9**, 11607.
80. Habermann,A.C., Gutierrez,A.J., Bui,L.T., Yahn,S.L., Winters,N.I., Calvi,C.L., Peter,L., Chung,M.I., Taylor,C.J., Jetter,C., *et al.* (2020) Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.*, **6**, eaba1972.
81. Ukita,M., Hamanishi,J., Yoshitomi,H., Yamanoi,K., Takamatsu,S., Ueda,A., Suzuki,H., Hosoe,Y., Furutake,Y., Taki,M., *et al.* (2022) CXCL13-producing CD4+ T cells accumulate in the early phase of tertiary lymphoid structures in ovarian cancer. *JCI Insight*, **7**, e157215.
82. Stuart,T. and Satija,R. (2019) Integrative single-cell analysis. *Nat. Rev. Genet.*, **20**, 257–272.
83. Stubbington,M.J.T., Rozenblatt-Rosen,O., Regev,A. and Teichmann,S.A. (2017) Single-cell transcriptomics to explore the immune system in health and disease. *Science*, **358**, 58–63.
84. Piovesan,A., Caracausi,M., Antonaros,F., Pelleri,M.C. and Vitale,L. (2016) GeneBase 1.1: a tool to summarize data from NCBI gene datasets and its application to an update of human gene statistics. *Database (Oxford)*, **2016**, baw153.
85. Lopes,I., Altab,G., Raina,P. and de Magalhães,J.P. (2021) Gene size matters: an analysis of Gene length in the Human genome. *Front. Genet.*, **12**, 559998.
86. Stoeger,T., Grant,R.A., McQuattie-Pimentel,A.C., Anekalla,K.R., Liu,S.S., Tejedor-Navarro,H., Singer,B.D., Abdala-Valencia,H.,

Schwake,M., Tetreault,M.-P., *et al.* (2022) Aging is associated with a systemic length-associated transcriptome imbalance. *Nature Aging*, **2**, 1191–1206.
87. Gutiérrez-Franco,A., Ake,F., Hassan,M.N., Cayuela,N.C., Mularoni,L. and Plass,M. (2023) Methanol fixation is the method of choice for droplet-based single-cell transcriptomics of neural cells. *Commun. Biol.*, **6**, 522.
88. Thommen,D.S., Koelzer,V.H., Herzig,P., Roller,A., Trefny,M., Dimeloe,S., Kiialainen,A., Hanhart,J., Schill,C., Hess,C., *et al.* (2018) A transcriptionally and functionally distinct PD-1(+) CD8(+) T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat. Med.*, **24**, 994–1004.
89. Kobayashi,Y., Tata,A., Konkimalla,A., Katsura,H., Lee,R.F., Ou,J., Banovich,N.E., Kropski,J.A. and Tata,P.R. (2020) Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.*, **22**, 934–946.
90. Sui,L., Wang,S., Ganguly,D., El Rayes,T.P., Askeland,C., Borretzen,A., Sim,D., Halvorsen,O.J., Knutsvik,G., Arnes,J., *et al.* (2022) PRSS2 remodels the tumor microenvironment via repression of Tsp1 to stimulate tumor growth and progression. *Nat. Commun.*, **13**, 7959.
91. De Jonghe,J., Kaminski,T.S., Morse,D.B., Tabaka,M., Ellermann,A.L., Kohler,T.N., Amadei,G., Handford,C.E., Findlay,G.M., Zernicka-Goetz,M., *et al.* (2023) spinDrop: a droplet microfluidic platform to maximise single-cell sequencing information content. *Nat. Commun.*, **14**, 4788.
92. Zheng,W., Zhao,S., Yin,Y., Zhang,H., Needham,D.M., Evans,E.D., Dai,C.L., Lu,P.J., Alm,E.J. and Weitz,D.A. (2022) High-throughput, single-microbe genomics with strain resolution, applied to a human gut microbiome. *Science*, **376**, eabm1483.

Received: April 25, 2024. **Revised:** November 28, 2024. **Editorial Decision:** December 3, 2024. **Accepted:** December 26, 2024