scientific reports

Check for updates

OPEN

Microdroplet-based one-step RT-PCR for ultrahigh throughput single-cell multiplex gene expression analysis and rare cell detection

Jennifer Ma[1], Gary Tran[2], Alwin M. D. Wan[3], Edmond W. K. Young[1,3], Eugenia Kumacheva[1,3,4], Norman N. Iscove[2,5,8] & Peter W. Zandstra[6,7,8]✉

Gene expression analysis of individual cells enables characterization of heterogeneous and rare cell populations, yet widespread implementation of existing single-cell gene analysis techniques has been hindered due to limitations in scale, ease, and cost. Here, we present a novel microdroplet-based, one-step reverse-transcriptase polymerase chain reaction (RT-PCR) platform and demonstrate the detection of three targets simultaneously in over 100,000 single cells in a single experiment with a rapid read-out. Our customized reagent cocktail incorporates the bacteriophage T7 gene 2.5 protein to overcome cell lysate-mediated inhibition and allows for one-step RT-PCR of single cells encapsulated in nanoliter droplets. Fluorescent signals indicative of gene expressions are analyzed using a probabilistic deconvolution method to account for ambient RNA and cell doublets and produce single-cell gene signature profiles, as well as predict cell frequencies within heterogeneous samples. We also developed a simulation model to guide experimental design and optimize the accuracy and precision of the assay. Using mixtures of in vitro transcripts and murine cell lines, we demonstrated the detection of single RNA molecules and rare cell populations at a frequency of 0.1%. This low cost, sensitive, and adaptable technique will provide an accessible platform for high throughput single-cell analysis and enable a wide range of research and clinical applications.

Single-cell analysis techniques are critical to distinguish differences between individual cells within seemingly homogeneous populations, such as divergent cell cycle status, cell lineage bias, or other cellular processes. Single-cell analysis techniques can also be applied to the detection of rare cells within heterogeneous cell populations, which would be useful in both basic research and clinical applications[1,2]. Over the last decade, the development of advanced analysis techniques has enabled the study of complex biological systems and phenomena at single-cell resolution. Popular examples include single cell analysis of RNA expression using fluorescence in situ hybridization (FISH)[3–12], RNA sequencing (scRNA-seq)[13–23], and reverse-transcription polymerase chain reaction (scRT-PCR).

Despite the advantages of these techniques, they are not exempt from limitations, and can be technically and financially demanding. The sensitivity of FISH and scRNAseq depends on the specific protocols used. While single-molecule FISH (smFISH) is generally regarded as the gold standard for RNA quantification in single cells, the molecular-detection limit of scRNA-seq methods measured using spiked-in transcripts varied over four orders of magnitude[14]. The sensitivity of scRNA-seq is also critically dependent on sequencing depth[14,24], which scales with the number of cells tested and can significantly drive up experimental cost. This cost (> $400

[1]Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada. [2]Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada. [3]Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON M5S 3G8, Canada. [4]Department of Chemistry, University of Toronto, Toronto, ON M5S 3H6, Canada. [5]Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 1L7, Canada. [6]School of Biomedical Engineering, University of British Columbia, 2222 Health Sciences Mall, Vancouver, BC V6T 1Z3, Canada. [7]Michael Smith Laboratories, University of British Columbia, Vancouver, BC V6T 1Z4, Canada. [8]These authors jointly supervised this work: Norman N. Iscove and Peter W. Zandstra. ✉email: peter.zandstra@ubc.ca

per 1000 cells[23,25]) can be prohibitive when screening large samples (> 1000 cells) or if frequent sampling is necessary. Moreover, the experimental implementation of FISH and scRNA-seq often requires empirical optimization of multiple sample-specific steps in their procedures, which is time-consuming, labour-intensive, and expensive[26,27]. To improve the sensitivity and throughput of the assays (which would otherwise be confined to thousands of cells), they also require high capital investments (> $150 K) for specialized instruments. Coupling FISH with flow cytometry[11,28] or fluorescent activated cell sorting[29,30] has also been explored to increase cellular throughput, concurrently measure gene and protein expression, as well as isolate specific cell populations for downstream analysis. However, these methods are time consuming and have a detection limit of ten copies of transcripts or above. Lastly, high throughput FISH and scRNA-seq assays generate large amounts of complex data, which in turn demand high levels of expertise, complex computational tools, and up to weeks of processing time for bioinformatic analysis[9,31–35].

RT-PCR is a widely adapted technique due to its low cost, high sensitivity, short turnaround time, and easily customizable assays. However, current single-cell RT-PCR (scRT-PCR) platforms are either limited in cellular throughput or multiplex capability. Microcircuit-based technologies such as the Fluidigm Biomark system allows analysis of up to 192 genes, but the number of single cells captured per run is confined to the hundreds[36–38]. In contrast, microdroplet-based platforms are capable of high-throughput analysis of tens of thousands of cells in a single experiment, but amplification of only one target gene has been shown using this method[39–42]. Moreover, to alleviate cell lysate-mediated inhibition of PCR, all of these techniques employ multiple dilution or reagent addition steps performed in microfluidic devices or manually, which sacrifice sensitivity, limit throughput, and complicate automation of these assays.

As a result, despite the benefit that single-cell analysis can offer in a variety of studies, these state-of-the-art technologies are often restricted to specific applications. Clearly, there remains a need for a low cost, easily adaptable platform that can sample large numbers of single cells and analyze multiple targets simultaneously. Here, we present a microdroplet-based, multiplex RT-PCR platform that is designed for ultrahigh-throughput single-cell analysis of differential gene expression as well as rare cell populations. This technique provides single-cell resolution information with high signal-to-noise ratio and acquisition of hundreds of thousands of data points, which allows the analysis of rare cell populations down to a frequency of at least 0.1%. An in-house reagent mix was developed to overcome amplification inhibition caused by high cell lysate concentration and enabled one-step RT-PCR in nanoliter droplets. The benefit of adding bacteriophage T7 gene 2.5 protein (gp2.5) on relieving cell-lysate mediated inhibition was demonstrated for the first time in this study. To date we have successfully performed simultaneous amplification of three targets in a single encapsulated cell. This multiplex capability makes it a powerful tool for both gene expression analysis and rare cell detection. To enable widespread use in both laboratory and clinical settings, this assay is relatively low cost, easy to perform, readily adaptable for a wide range of applications, and produces reliable results rapidly. The simplicity of our device and protocol allows for future automation of the droplet-based RT-PCR system. This platform will empower research groups by granting them access to single-cell gene expression analysis that would have otherwise been technically and financially difficult.

Results

A microfluidic platform for single cell encapsulation and RT-PCR analysis. We designed a simple and fully automatable workflow to perform high-throughput single cell RT-PCR. A schematic of the microfluidic (MF) platform is shown in Fig. 1a. Single cells were encapsulated with RT-PCR reagents with lysis buffer in nanoliter (nL) droplets using an MF device. The cell suspension and reagents were delivered through two separate channels, and mixed immediately before the emulsion in oil was formed. The cells rapidly lysed within their respective droplets. The droplets were collected in a PCR tube and subjected to thermal cycling. The droplets, now containing fluorescent amplification products, were loaded into microchambers designed to trap droplets in monolayers for analysis using an automated imaging platform.

Figure 1b shows the design of the MF device. A flow-focusing geometry was chosen due to its ability to form droplets at high capillary numbers and therefore encapsulate at high rates[43]. This feature is beneficial for applications that require large quantities of data points such as rare cell detection. A million droplets of 1 nL volume (approximately 124 µm diameter, see Supplementary Fig. S1 for characterization of droplet diameter in a representative sample) can be generated per hour using a single device.

To quickly image hundreds of thousands of droplets, we used automated fluorescence microscopy. The resulting images were analyzed with the publicly available CellProfiler software. The analysis pipeline identifies droplets by detecting object boundaries in brightfield images and determines their size and shape, which allows for exclusion of droplets that have been broken up or undergone coalescence. The algorithm then measures the average fluorescence signals within each droplet to classify the droplet as positive or negative, either based on a negative control or a clustering algorithm. When microdroplets containing fluorophores were mixed with droplets without fluorophores, the system was capable of enumerating the fluorescent droplets down to ratios of 1 in 10,000 (Supplementary Fig. S1). This demonstrated the capacity of the microscopy system to recognize droplets of various sizes, determine their dimensions, quantify fluorescence intensity within droplets, classify positive droplets, and provide a frequency read-out of rare events.

Development of multiplex one-step RT-PCR with high cell-lysate tolerance. Performing RT-qPCR in nano-volume increases the sensitivity of the reaction[44–47] and reduces the amount of reagents used on a per cell basis. However, previous reports have demonstrated that cell concentrations greater than 200 cells/µL had an inhibitory effect on product yield when performing conventional RT-qPCR and during attempts to downscale the reaction volume in microfluidic devices[39,44,48]. Despite the encapsulation of only a single cell in

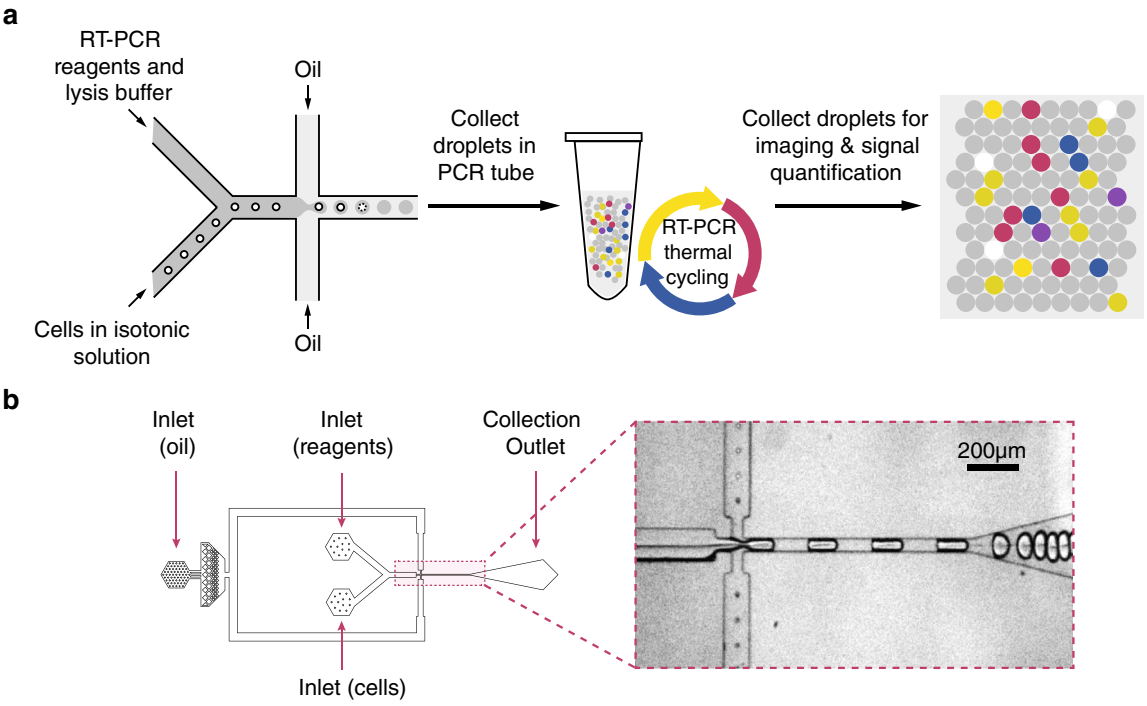


Figure 1. Workflow and devices for droplet-based, single-cell, one-step RT-PCR. (**a**) Process flow of the platform. Cells suspended in an isotonic solution enter a microfluidic device through a channel separate from the RT-PCR reagents and lysis buffer. These channels fuse to a single orifice that injects into a channel met by opposing streams of oil to form monodispersed, 1 nL volume droplets. These droplets are collected in a PCR tube and thermal cycled to generate fluorescent amplification products. The droplets are then deposited in monolayers and analyzed using an automated imaging platform, after which the fluorescent signal in each droplet is quantified. (**b**) Droplet generator design (to scale) and brightfield image of the orifice of the device during droplet formation.

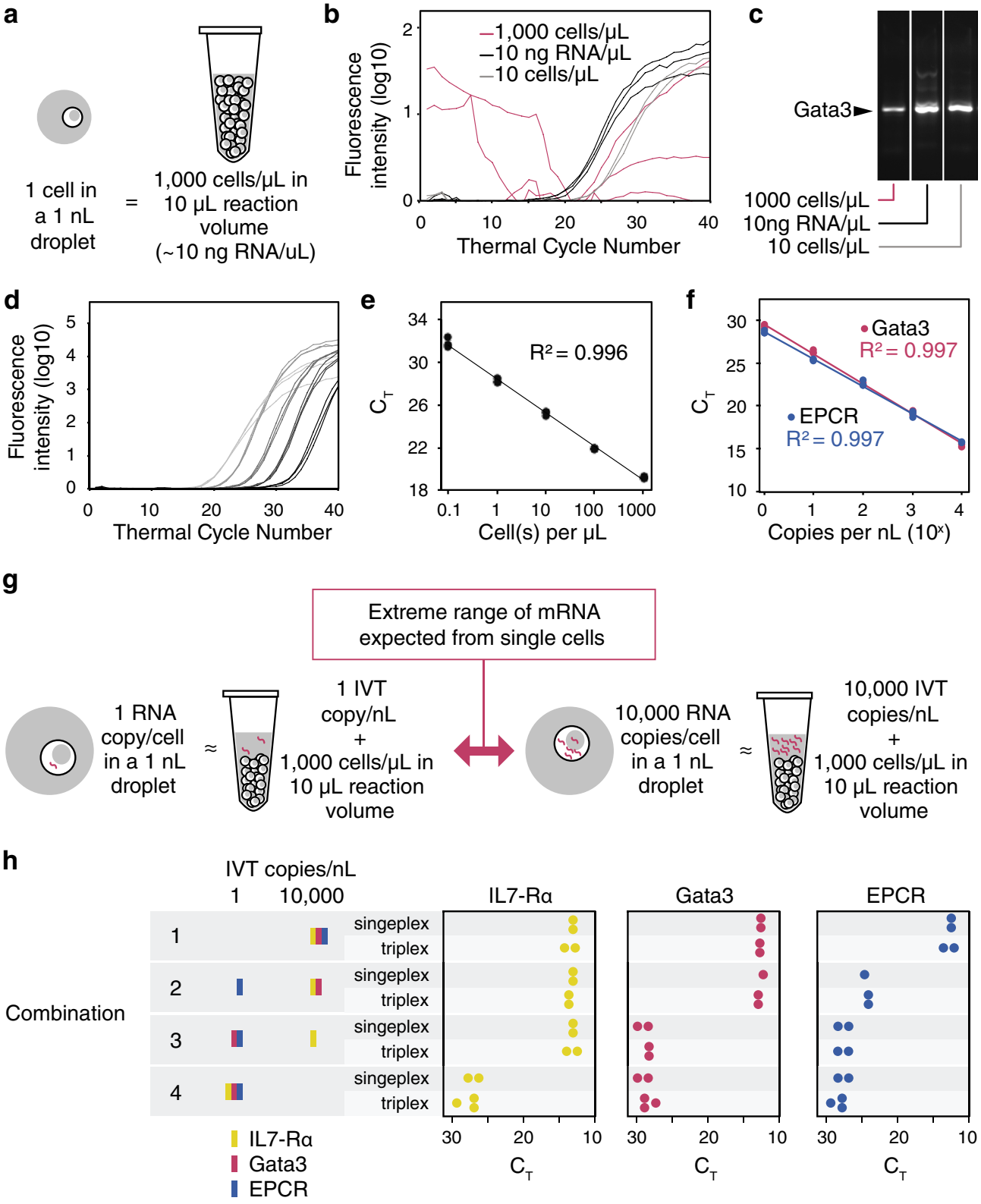
each 1 nL microdroplet, the cell lysate concentration is equivalent to approximately 1000 cells/µL (Fig. 2a). We found this high cell lysate concentration to inhibit the reaction in conventional RT-qPCR when performed using a sample concentration of 1000 cells/µL. Conventional RT-qPCR reactions were performed with 10 or 1000 cells/µL (Fig. 2b). With the higher cell concentration, generation of fluorescence was chaotic. In contrast, fluorescence developed in the expected cycle-dependent and quantitative pattern in reactions performed with 10 cells/µL. Fluorescence development was similarly quantitative in reactions performed with 10 ng purified total RNA/µL, which is the amount expected in 1000 cells. Agarose gel electrophoresis confirmed that the amount of product generated in the 1000 cells/µL sample was indeed less than that obtained with either 10 cells or purified RNA (Fig. 2c). The results thus confirmed the inhibitory effect of high cell lysate concentration on conventionally performed qRT-PCR reactions.

In an attempt to avoid additional sample dilution or purification steps that would negatively impact the complexity, sensitivity, or efficiency of our platform, we developed an in-house one-step RT-PCR reagent mix that could overcome the lysate-mediated inhibition. To optimize the reaction, RT-qPCR components were tested systematically by altering one component in each reaction. 10,000 cells were added to 10 µL RT-qPCR reactions to simulate the cell lysate concentration in the microdroplets and the effects on C_T value, fluorescence plateau, and amplification efficiency were observed. To address inhibition of amplification at high lysate concentration, we explored the effects of adding single-stranded DNA (ssDNA) binding proteins (SSBs) to our RT-qPCR reactions. SSBs have been shown to alleviate inhibitory effects of enzymes and other proteins on PCR reactions[49,50]. The addition of gp2.5 alone was able to partially rescue RT-qPCR performed using either our in-house mix or the CellsDirect One-Step RT-qPCR kit (Invitrogen) from lysate-mediate inhibition (Supplementary Fig. S2). Our resulting RT-qPCR mix was capable of quantifying Gata3 mRNA transcripts directly across a range of 1 to 10,000 EL4 cell(s) in 10 µL RT-qPCRs (Fig. 2d). Based on the slope of the standard curve (α) in Fig. 2e, the reaction efficiency E was calculated to be 1.08 (acceptable range = 0.90–1.10)[51] using the formula

$$E = 10^{-1/\alpha} - 1 \quad (1)$$

Accordingly, gp2.5 was included in the PCR mix for all subsequent experiments.

A broad range of murine Gata3 and EPCR transcript abundances at high cell lysate concentrations were then tested by RT-qPCR (Fig. 2f). Serial tenfold dilutions of Gata3 or EPCR in vitro transcripts (IVT) were added to 10 µL reactions based on the range of transcript concentrations expected in a microdroplet containing a single cell[52,53], while the cell concentration in the reaction was kept at 1000 cells/µL (using a Gata3 and EPCR negative cell line) to simulate the conditions in the droplets. RT-qPCR analysis performed on the dilution series of the


a
1 cell in a 1 nL droplet
=
1,000 cells/µL in 10 µL reaction volume (~10 ng RNA/uL)
b
Fluorescence intensity (log10)
1,000 cells/µL
10 ng RNA/µL
10 cells/µL
Thermal Cycle Number
c
Gata3
1000 cells/µL
10ng RNA/µL
10 cells/µL
d
Fluorescence intensity (log10)
Thermal Cycle Number
e
R² = 0.996
C_T
Cell(s) per µL
f
Gata3
R² = 0.997
EPCR
R² = 0.997
C_T
Copies per nL (10^x)
g
Extreme range of mRNA expected from single cells
1 RNA copy/cell in a 1 nL droplet
≈
1 IVT copy/nL + 1,000 cells/µL in 10 µL reaction volume
10,000 RNA copies/cell in a 1 nL droplet
≈
10,000 IVT copies/nL + 1,000 cells/µL in 10 µL reaction volume
h
IVT copies/nL
1
10,000
IL7-Rα
Gata3
EPCR
Combination
singeplex
triplex
IL7-Rα
Gata3
EPCR
C_T

◂**Figure 2.** High cell lysate concentrations were shown to be inhibitory to RT-qPCR, thus modifications to current RT-qPCR conditions were required. (**a**) The lysate concentration of a single cell encapsulated in a 1 nL droplet is equivalent to approximately 1000 cells/μL on the scale of conventional RT-qPCR volumes. (**b**,**c**) To demonstrate the inhibitory effect, one-step RT-qPCR was performed to detect Gata3 expression in EL4 cells at 10 and 1000 cells/μL concentration in a 10 μL reaction volume, as well as the amount of total RNA equivalent to 1000 cells/μL (10 ng RNA/μL). Inhibition was observed as either a C_T delay or sporadic fluorescence generation with no exponential phase at 1000 cells/μL (**b**). Gel electrophoresis also showed reduced amplification products at higher cell concentration (**c**). See Supplementary Figure S2 for full-length gel image. (**d**) Gata3 mRNA was quantified efficiently on a range of 1 to 10^4 EL4 cell(s) (shown by darkest to lightest colour plots) in a 10 μL RT-qPCR using our in-house one-step RT-PCR mix with gp2.5. (**e**) Dilutions of cells or (**f**) Gata3 and EPCR IVT doped in constant cell lysate concentrations (1000 cells/μL) were analyzed by RT-qPCR and plotted against their C_T values to demonstrate high amplification efficiencies using our in-house one-step RT-PCR mix. (**g**) To emulate the range of mRNA expected from single cells, IL-7Rα, Gata3, and EPCR IVT were doped in constant cell lysate concentrations (1,000 cells/μL) at either 1 or 10,000 copies/nL. (**h**) Using our in-house, multiplex, one-step RT-PCR mix, the three target genes were simultaneously quantified in four different combinations of transcript levels. The difference in C_T values obtained using our single and triplex assays were found to be insignificant after 40 thermal cycles ($P > 0.1$). Each dot represents one technical replicate. Data was analyzed and plotted using JMP (version 15.2.1 www.jmp.com)

Gata3 and EPCR transcripts yielded R^2 values greater than 0.99 and efficiencies of 0.94 and 1.06 respectively, indicating that the reactions took place with high efficiency despite the high concentration of cell lysate.

The inability to multiplex has been a critical limitation in current droplet-based RT-PCR platforms. Capability to perform multiplex RT-qPCR would allow for simultaneous detection of multiple surrogate markers in single cells. In single eukaryotic cells, the detectable mRNA transcript abundance in a single cell can range from 1 to 10,000 copies[53]. To mimic the range of mRNA concentrations expected in a 1 nL microdroplet containing a single cell, synthetic mixtures were assembled using four combinations of EPCR, Gata3, and IL7-Rα IVTs at either 1 or 10,000 copies/nL and added into 10 μL RT-qPCRs as template (as illustrated in Fig. 2g and the table in Fig. 2h). The markers Gata3, EPCR, and IL-7Rα were chosen as relevant model genes as they in combination significantly enrich for primitive mouse hematopoietic stem cells (57). Three sets of Taqman probes with different fluorophores were used to quantify the three markers. Into each 10 μL reaction, 10,000 B62c cells were spiked to simulate the amount of lysate in the microdroplets except for combination 4, in which the template composition (low level of IL7-Rα transcripts) conflicted with the high IL7-Rα expression of the B62c cell line. The results of each triplex reaction were compared directly to three individual singleplex reactions performed in parallel to evaluate their performance. The efficiency of the multiplex reaction was determined by comparing the exponential phase of its amplification curve and its C_T value to its singleplex counterparts, and the specificity by agarose gel electrophoresis.

High ratio differences in abundance levels between the different genes in the synthetic mixtures resulted in unreliable multi-gene quantification (Supplementary Fig. S3). The amplification of sparse targets was hindered by the presence of abundant targets as they competed for many of the same reagents[54]. This effect was most apparent in IVT combination 3, where IL-7Rα was faithfully amplified but the less abundant Gata3 and EPCR transcripts were not accurately quantified. This IVT combination was, therefore, chosen as the baseline template for optimizing the multiplex reaction by systematically testing the reaction components.

The concentrations of the singleplex RT-qPCR components including binding proteins, salts, dNTPs, and Taq polymerase were reassessed to accommodate simultaneous amplification of three genes, as detailed in Table 1. Primers and Taqman probes were also redesigned to minimize primer-primer interactions across all primer pairs[54,55]. Figure 2h and Supplementary Fig. S4 show the C_T values and exponential phases of the triplex reactions using the modified RT-qPCR mix superimposed onto their respective singleplex controls. The difference in C_T values obtained using the triplex and singleplex assays were found to be insignificant (two-sample t-test, $P > 0.1$), demonstrating that the identified multiplex RT-qPCR parameters supported efficient quantification of EPCR, Gata3, and IL7-Rα levels in all 4 IVT combinations despite 10,000-fold differences.

Simultaneous detection of 3 target genes at single molecule level in nanoliter droplets. Our next step was to implement the identified multiplex RT-PCR conditions in nanoliter droplets and evaluate the sensitivity of the assay. Limiting dilutions of IL-7Rα, Gata3, and EPCR IVTs were encapsulated with the multiplex RT-PCR mix (Fig. 3a). Compared to the no template control (NTC), substantial increase in fluorescence intensities in the three channels corresponding to the Taqman probes was evident after 50 PCR cycles (Fig. 3b,c). This indicated that the reaction successfully amplified all three transcripts in the droplets. 50 PCR cycles were used to ensure that the reaction in the droplets that contained the target transcripts had reached the plateau phase to maximize the signal-to-noise ratio.

The fluorescence signal was quantified and the droplets were clustered into populations expressing different marker combinations using the density-based spatial clustering algorithm (DBSCAN)[57] (Fig. 3c), which can identify clusters of arbitrary shapes and does not require prior knowledge of the number of clusters present in each sample. The clusters were labeled as either positive or negative for the three IVTs and the concentrations of the three transcripts in each sample was then estimated utilizing Poisson statistics[58]

$$P(0) = e^{-\lambda} \tag{2}$$

RT-PCR component	Modification	Singleplex	Multiplex
Binding protein gp2.5	The concentration of binding protein was raised to achieve 0.2 µg per pmol of primers and probes to sequester the additional amount of primer pairs and TaqMan probes in multiplex reaction	0.51 µg/µL	0.95 µg/µL
Taq polymerase	The quantity of Taq polymerase was increased for efficient amplification of multiple targets[54,56]	0.3 U/µL	0.6 U/µL
KCl	KCl facilitates annealing of DNA molecules and increase in KCl increases product yield of shorter amplicons[55]	55 mM	90 mM
$MgCl_2$	Facilitates annealing of DNA molecules and acts as a cofactor of Taq polymerase[51]	2.0 mM	2.6 mM
dNTP	More dNTPs are required to produce more amplification products. As increasing amounts of dNTPs can bind to Mg^{2+} electrostatically, reduce the amount of Mg^{2+} available and inhibit the reaction[55], concentrations of $MgCl_2$ and dNTPs were tested in a factorial experimental design	400 µM	500 µM
Primers and Taqman probes	Primers and Taqman probes were redesigned and verified by gel electrophoresis to minimize dimer formation and off-target amplification	See Supplementary Table S1 for sequences	See Supplementary Table S1 for sequences

Table 1. Summary of modifications to RT-qPCR components for efficient multiplex amplification.

where $P(0)$ is the fraction of droplets that did not express a target gene, and λ is the average number of that transcript per droplet. We attained excellent correlation between the empirically determined dilution factors and the sample dilution factors (Fig. 3d,e). This illustrated the ability of the droplet RT-PCR assay to amplify and detect transcripts of 3 target genes simultaneously at single molecule level with high specificity.

Deconvolution and simulation model for predicting cellular composition of heterogeneous samples. We developed an analytical pipeline that enables deconvolution of single-cell gene signature profile and cellular composition from our droplet RT-PCR assay. The presence or absence of marker genes make up the gene signature of a droplet or a single cell. With three marker genes, there are eight unique gene signatures, as listed in the table in Fig. 2a. A single-cell gene signature profile describes the proportion of single cells in the sample that express each gene signature. As demonstrated in the IVT experiments, a single molecule of RNA can be amplified and produce a positive signal. While this speaks to the sensitivity of our assay, it also means that ambient RNA or cell-free transcripts in the solution released by intact or damaged cells will also be detected. This ambient RNA can be co-encapsulated in droplets with cells, resulting in a false positive signal. Additionally, the encapsulation of cells in droplets is a Poisson process, and the distribution of cells is dependent on the starting cell concentration and droplet volume. Keeping the volume constant, the fraction of droplets that contain more than one cell (multiplets) rises as cell concentration increases. Both ambient RNA and multiplets must be accounted for in order to determine the true gene signature profiles of single cells. The corrected profiles can then be used to estimate the proportions of constituent populations when analyzing heterogeneous samples consisting of multiple cell types.

Figure 4 illustrates our analytical pipeline that consists of three steps. Briefly, Step 1 classifies droplets as positive or negative for the presence of cells as well as each target gene based on their fluorescence intensities using the Variational Bayesian Gaussian Mixture with a Dirichlet process prior model[59]. The model assumes that all the data points are generated from a mixture of Gaussian distributions with unknown parameters and does not require a predefined total number of clusters. It then determines the proportions of droplets with cells displaying each gene signature ($\boldsymbol{d}$), the proportions of empty droplets displaying each gene signature caused by ambient RNA ($\boldsymbol{n}$), and the average number of cells each droplet contains (λ) estimated based on the proportions of empty droplets.

In Step 2, the single-cell gene signature profile of the sample ($\boldsymbol{s}$) is estimated. $\boldsymbol{s}$, along with the empirically determined proportions $\boldsymbol{d}$ and $\boldsymbol{n}$, are regarded as probability distributions. $\boldsymbol{d}$ is then modeled as the sum of all possible combinations of ambient RNA and cells (up to having two cells in each droplet to simplify the model) that can generate each signature, allowing the computation of the single-cell profile $\boldsymbol{s}$.

In Step 3, the proportions of constituent populations in the heterogeneous sample ($\boldsymbol{w}$) are predicted. Given a heterogeneous sample, which is a physical mixture of its constituent cell populations, with gene signature profile $\boldsymbol{s}$, its cellular composition $\boldsymbol{w}$ can be estimated using reference gene signature profiles $\boldsymbol{r}$ obtained by assaying the pure constituent populations. Based on the non-negative least squares (NNLS) model[60], the mixture profile $\boldsymbol{s}$ is modeled as a positively weighted sum of the reference profiles, where weight w_k represents the proportion of reference population k within the mixture (assuming that all constituent populations are represented in the reference profiles).

To evaluate our deconvolution model as well as illustrate the effects of various input parameters on the accuracy and precision of our predictions, a simulation of the droplet assay was built to generate thousands of datasets for testing (see Supplementary Fig. S5 for details). We simulated mixtures of three cell lines, B62c (murine pre-B lymphocytes), D4T (murine endothelial cells), and EL4 (murine T lymphocytes), which were assayed using our droplet platform to provide real-world single-cell gene signature profiles and ambient RNA levels as input for our simulations. We also compared the results obtained with and without Step 2 of our analysis pipeline to demonstrate the influence of ambient RNA and multiplets. Representative results from the simulation that reveal

the effects of altering average cell number per droplet λ and cellular composition of the input cell mixture $\boldsymbol{c}$ are shown in Fig. 5. The predicted proportions of each cell line fell within the same order of magnitude as the input in all conditions where Step 2 was implemented. It was observed that increasing λ boosted the precision of our predictions, as demonstrated by the decrease in confidence intervals. This was due to the higher number of cells "encapsulated" and sampled even though the total number of droplets remained constant (200,000 droplets per experiment). The accuracy of the predictions of the rarest population was improved when Step 2 was incorporated in the analysis, but only when λ was low (≤ 0.2). This result was expected due to the increasing proportion of droplets that contain more than 2 cells in samples where λ was high (0.11% for $\lambda = 0.2$, 1.44% for $\lambda = 0.5$, 8.02% for $\lambda = 1$), which our model did not account for. As for various cellular compositions, we also observed an improvement in accuracy when Step 2 was applied, whereas the rare populations were often undetected without this correction step. As one would expect, the less overlap between the gene signature profiles of the rare (0.1%) and majority (90%) populations, the higher the precision and accuracy attained. Taken together, the simulation results informed the utility and limitations of our assay and analysis pipeline and guided the selection of parameters of our experiments in the next section.

Single-cell gene expression profiling and rare population detection. To test our platform's capability to perform multiplex single-cell RT-PCR on living cells encapsulated in microdroplets, three murine cell lines with various gene expression profiles, namely B62c, D4T, and EL4, were assayed. No issues were observed when encapsulating the cell lines despite the differences in their cell size and adhesion properties, indicating that our droplet generator can be used on a wide range of cell types. We analyzed the cell lines using the droplet RT-PCR platform based on the expression of murine Gata3, EPCR, and either IL-7Rα or Gusb, a housekeeping gene that is expressed in all three cell lines. Substantial increases in fluorescent amplification products were again observed in all three channels after thermal cycling (Fig. 6a). As expected, Gusb was detected in over 90% of the cells in all cell lines (Fig. 6b), while Gata3, EPCR, and IL-7Rα were present only in subsets of the cell lines, consistent with results obtained by conventional RT-qPCR (Fig. 6c,d). The less than 100% detection rate of Gusb, which concurred with a previous single cell transcriptomic study where Gusb was undetected in 18.6% of cells in the "Mouse Atlas"[61,62], could be explained by transcriptional bursts of individual genes[63]. Notably, switching out the marker IL-7Rα for Gusb or other genes (data not shown) in the triplex reaction did not require additional optimization of the reaction conditions, suggesting that the assay can be used to target any mRNA targets. This signified the ability of our droplet RT-PCR platform to detect multiple transcript expressions with high specificity and sensitivity in not only purified RNA samples, but also in living cells.

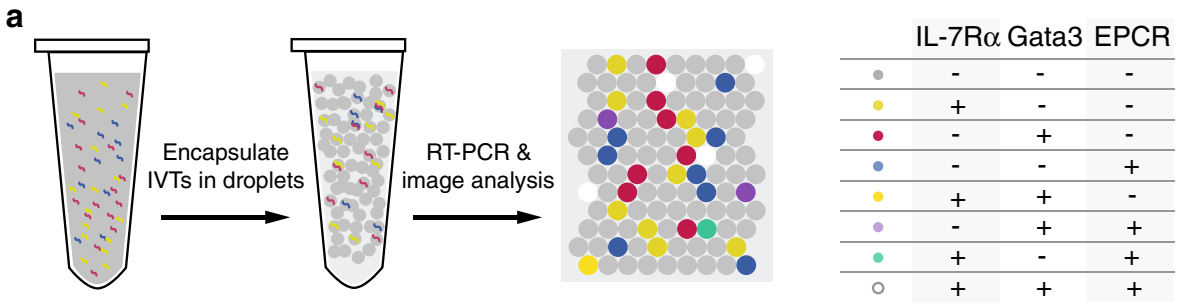
Next, we mixed the three cell lines at predetermined ratios to illustrate the platform's capability to deconvolve cell populations based on their expression levels. Three different sets of input parameters were selected based on the simulation results (Fig. 6e). Mixture 1 and 2 differed in cellular composition, while Mixture 3 shared the same composition with Mixture 2, but had a higher starting cell concentration thus higher average cell number per droplet. 100,000–200,000 droplets were collected per sample and over 130,000 cells were analyzed per experiment. Conforming to our simulation results, the more abundant the cell type, the more accurately the platform was able to estimate their proportions. The rarest population (0.1%) was undetected in Mixture 1, which was also in agreement with our simulation. In contrast, in Mixture 3, the estimations fell far above the 95% confidence interval of the simulated results. During our analysis, we observed that the increase in cell concentration contributed to an increased background level of the cell tracking dye. This rendered the separation of empty droplets from droplets containing cells more difficult and unreliable, and likely affected the output of the analysis. Mixture 2, on the other hand, consistently produced estimates of the rare population with the highest accuracy, congruent with our simulated results. This demonstrated not only our platform's ability to enumerate rare cells, but also the power of the simulation for optimizing experimental parameters.

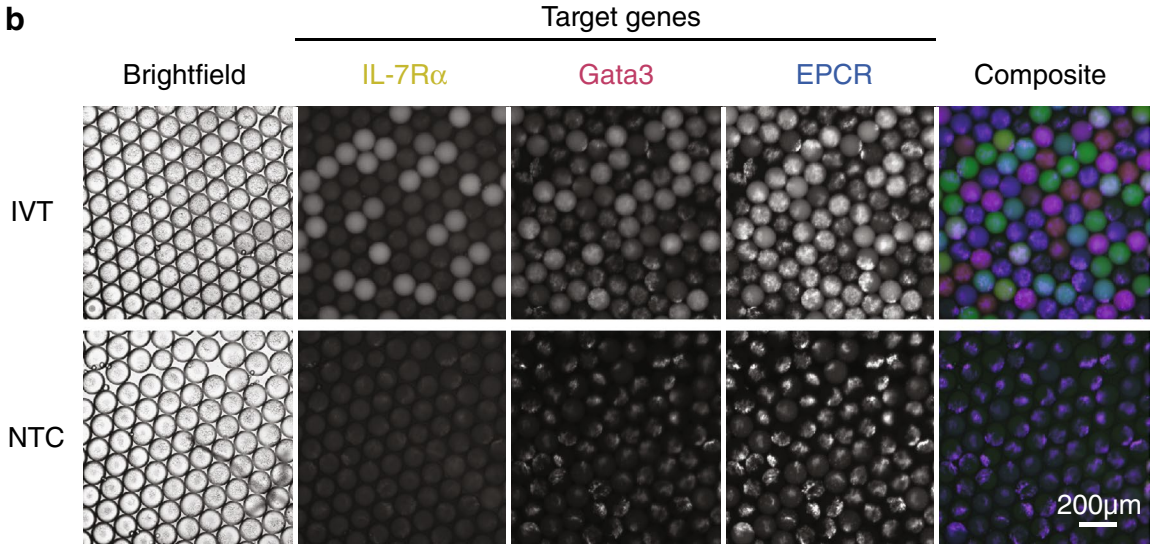
Discussion

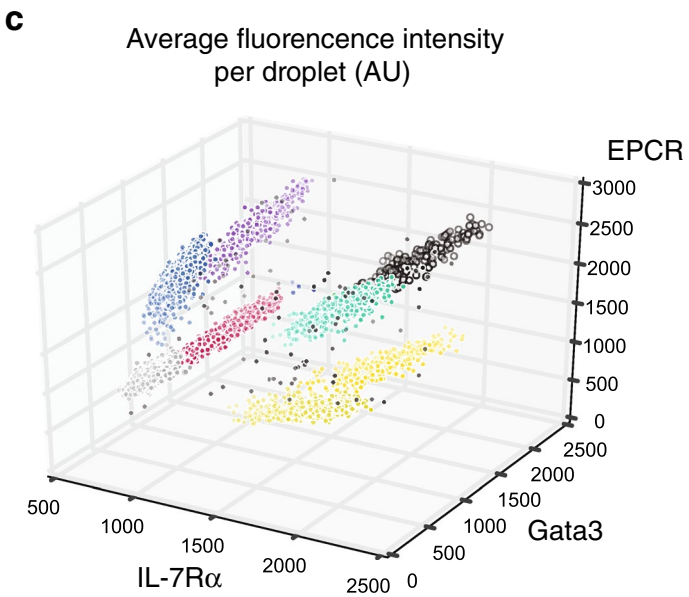
Over the past decade, there has been a surge in technological advancements associated with single-cell transcriptomics. The selection of an appropriate assay for any given application depends on technical considerations, such as throughput, sensitivity, and multiplex capability, which directly affect the information collected, as well as *accessibility* from the perspective of cost, processing time, and any inherent expertise needed to operate and analyze the assay. A survey of existing techniques, considering both technical and accessibility parameters, revealed a wide gap for researchers and clinicians to cross in order to feasibly embark on single-cell studies.

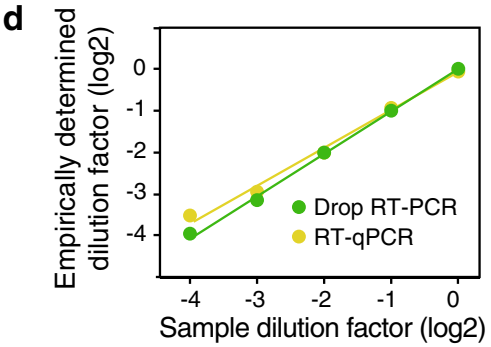
For instance, FISH has the advantage of producing high resolution spatial information, allowing transcript quantification and localization studies within single cells and in the context of tissue structures[3,9,34,40], and is used as a prognostic and diagnostic tool for diseases such as cancer[2]. scRNA-seq provides extensive information of the transcriptome and is often used for exploratory studies to uncover novel cell populations and their molecular properties[64]. Unfortunately, both of these techniques require expensive equipment and reagents, as well as experienced technicians to perform the experiments and analyze the resulting data[9,25,65]. Neither of these methods have yet to be integrated as standard lab procedures because of the restrictions with respect to the types of applications to which they are most suited.

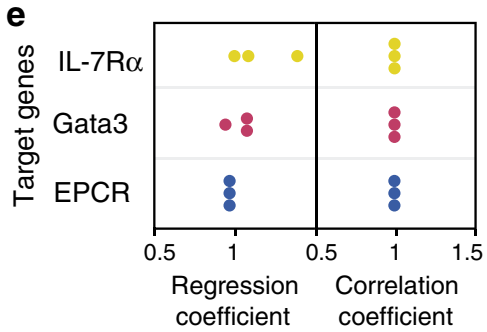
In contrast, RT-PCR is a well-established, convenient, and economical technique that is broadly used. Advances in high-throughput single-cell capture methods have allowed RT-PCR to be applied to individual cells, but they either suffer from low cellular throughput or are limited to targeting one gene per assay[38,39,42,66–70]. In this paper, we have described an ultrahigh-throughput, multiplex one-step RT-PCR platform that is capable of evaluating gene expression of up to three targets at single-cell resolution. To enable miniaturization of the single-cell assay and massive parallelization, an important breakthrough was identifying an optimal mix of RT-PCR reagents to address the problem of cell lysate-mediated inhibition, a long-standing obstacle to achieving


a
Encapsulate
IVTs in droplets
RT-PCR &
image analysis
IL-7Rα Gata3 EPCR
- - -
+ - -
- + -
- - +
+ + -
- + +
+ - +
+ + +


b
Target genes
Brightfield
IL-7Rα
Gata3
EPCR
Composite
IVT
NTC
200μm


c
Average fluorencence intensity
per droplet (AU)
EPCR
3000
2500
2000
1500
1000
500
0
2500
2000
1500
1000
500
0
Gata3
500
1000
1500
2000
2500
IL-7Rα


d
Empirically determined
dilution factor (log2)
0
-1
-2
-3
-4
Drop RT-PCR
RT-qPCR
-4
-3
-2
-1
0
Sample dilution factor (log2)


e
Target genes
IL-7Rα
Gata3
EPCR
0.5
1
0.5
1
1.5
Regression
coefficient
Correlation
coefficient

◂**Figure 3.** Simultaneous detection of three targets by our droplet RT-PCR platform at single molecule level is demonstrated using limiting dilutions of IVTs. (**a**) Samples containing various dilutions of IL-7Rα, Gata3, and EPCR IVTs were encapsulated for droplet RT-PCR analysis. A substantial increase in fluorescent signal was observed and quantified. The concentration of IVT in each sample was then estimated based on the fraction of droplets with significant increases in fluorescent signal and Poisson statistics. (**b**) Images of droplets containing a mixture of IL-7Rα, Gata3, and EPCR IVTs after amplification targeting the three genes. *NTC* No template control. (**c**) The distribution of fluorescence intensity of each droplet. The colours represent different populations clustered using DBSCAN, and were assigned based on the table in (A) to indicate the presence or absence of the three transcripts in each subpopulation. (**d**) A representative plot showing the correlation between input and estimated concentration of IL-7Rα IVTs using RT-qPCR and droplet RT-PCR. (**e**) The empirically determined dilution factors highly correlated with the sample dilution factors with both regression coefficient and correlation coefficient close to 1 ($P < 0.01$). Each dot represents one technical replicate. Data was analyzed and plotted using JMP (version 15.2.1 www.jmp.com).

one-step single-cell droplet RT-PCR. This allows robust detection of transcripts in nanoliter volume without downstream manipulation of the droplets, such as dilution or washing. As a result, we do not compromise the sensitivity of the assay due to analyte loss, or complicate the workflow in a way that may negatively impact the potential for system integration and automation.

A critical component of our in-house RT-PCR mix is the bacteriophage T7 gene 2.5 protein. Gp2.5 is a ssDNA binding protein that stimulates DNA synthesis and plays an important role in T7 DNA replication, recombination, and repair[71–74]. Gp2.5 has also been employed as a hot-start strategy by sequestering primers and probes to prevent non-specific binding and amplification at lower temperatures[75]. The addition of SSBs to our RT-PCR was first considered due to their ability to protect RT-PCR reactions from inhibitory cellular contaminants[49,50]. Diverse cellular proteins can interfere with reverse transcription or polymerase reactions, or interact, degrade, and sequester nucleic acids[49,76–78]. Ratnamohan et al.[48] demonstrated that heat treatment alone was not sufficient to remove cell lysate-mediated inhibition. While proteinase K digestion was able to reduce the effect, subsequent ethanol precipitation that eliminated the denatured proteins and other substances further removed the inhibition. To enable one-step, uncoupled RT-PCR, Chandler et al. experimented with SSBs to relieve inhibition of PCR amplification by residual reverse transcriptase, which was shown to interact directly with the specific primer-template complex[49]. Including SSBs during, but not after, the RT phase, increased the RT-PCR product yield by almost fivefold. This suggested that first-strand cDNA synthesis was more efficient in the presence of SSBs. SSBs such as gp2.5 are known to disrupt the secondary structure of ssDNA[79] and may hold the nucleotides in a favourable conformation for pairing with complementary nucleotides[80]. A recent study also showed that gp2.5 can facilitate fast template-primer hybridization, increasing the hybridization rate of complementary ssDNA strands by 36-fold[81]. Furthermore, given their high affinity to ssDNA strands[79,82], binding proteins may displace inhibitory substances from the templates and primers, allowing them to participate in RT. These suggest a potential use of gp2.5 as a protectant and facilitator of template-primer hybridization in high cell lysate concentrations, where PCR inhibitors are abundant.

Additionally, single transcripts of IVT and ambient mRNA were detected at limiting dilutions using our droplet platform even in the presence of high cell lysate concentrations. This demonstrates high sensitivity of the PCR chemistry capable of detecting low abundance transcripts in the lysate. However, it is noted that the procedures used here for disrupting the cells do not ensure that every expressed transcript in a cell is accessible. For example, transcripts located within the nucleus may not be liberated, and cytoplasmic transcripts could be physically unavailable for other reasons. Future studies that explore transcript availability and compare our platform with others such as smFISH will allow better insight into the detection limit for endogenous mRNA.

We are also the first to demonstrate multiplex RT-PCR of single cells in nanoliter droplets, greatly enhancing the capability of our platform to perform gene expression profiling of single cells and identify highly specific populations of cells. Our assay can potentially be engineered to detect more genes via further optimization and approaches such as amplitude modulation, where the relative TaqMan probe concentrations amongst targets are varied to yield distinctive PCR curves[83,84]. Since PCR primers and probes can be designed for any DNA and RNA targets, this technique can be readily adapted to analyze any cell population of interest.

Our technique does not rely on solid-phase RNA capture, ensuring high sensitivity and cell capture rate. This makes it suitable for assaying low-input samples such as clinical specimens with minimal cell loss. Potential applications include dissecting the composition of a tumor sample for diagnosis and prognosis. We have also demonstrated the platform's ability to handle large samples that consist of hundreds of thousands of cells. The current flow rate can be raised by at least fivefold without causing cell damage[85], and multiple devices can operate in parallel to further improve the speed of droplet generation. Additionally, automated imaging analysis enables rapid signal quantification and visualization of the samples, which can be manually inspected to ensure correct interpretation of the results. Combined with the high throughput, this feature is especially beneficial for applications that have low tolerance for false positives, such as detecting and characterizing rare cell types based on their gene expression profiles. As demonstrated in this study, our platform can detect rare populations at a frequency of 0.1%, which is equivalent to the number of blood stem cells found in primitive populations derived from umbilical cord blood[86]. An exciting application of our platform would be for enumerating these highly regenerative cells for therapeutic and research purposes. Other potential applications include assessing minimal residual disease, detection of rare cells with viral infection or genetic abnormalities, and detection of fetal cells in maternal blood for non-invasive prenatal diagnosis[1]. Similar to the calcein dye used in this study that tagged live cells, other biomolecules such as proteins can also be fluorescently labeled prior to cell encapsulation and visualized in the droplets after cell lysis. This allows analysis of the co-expression of different transcripts and

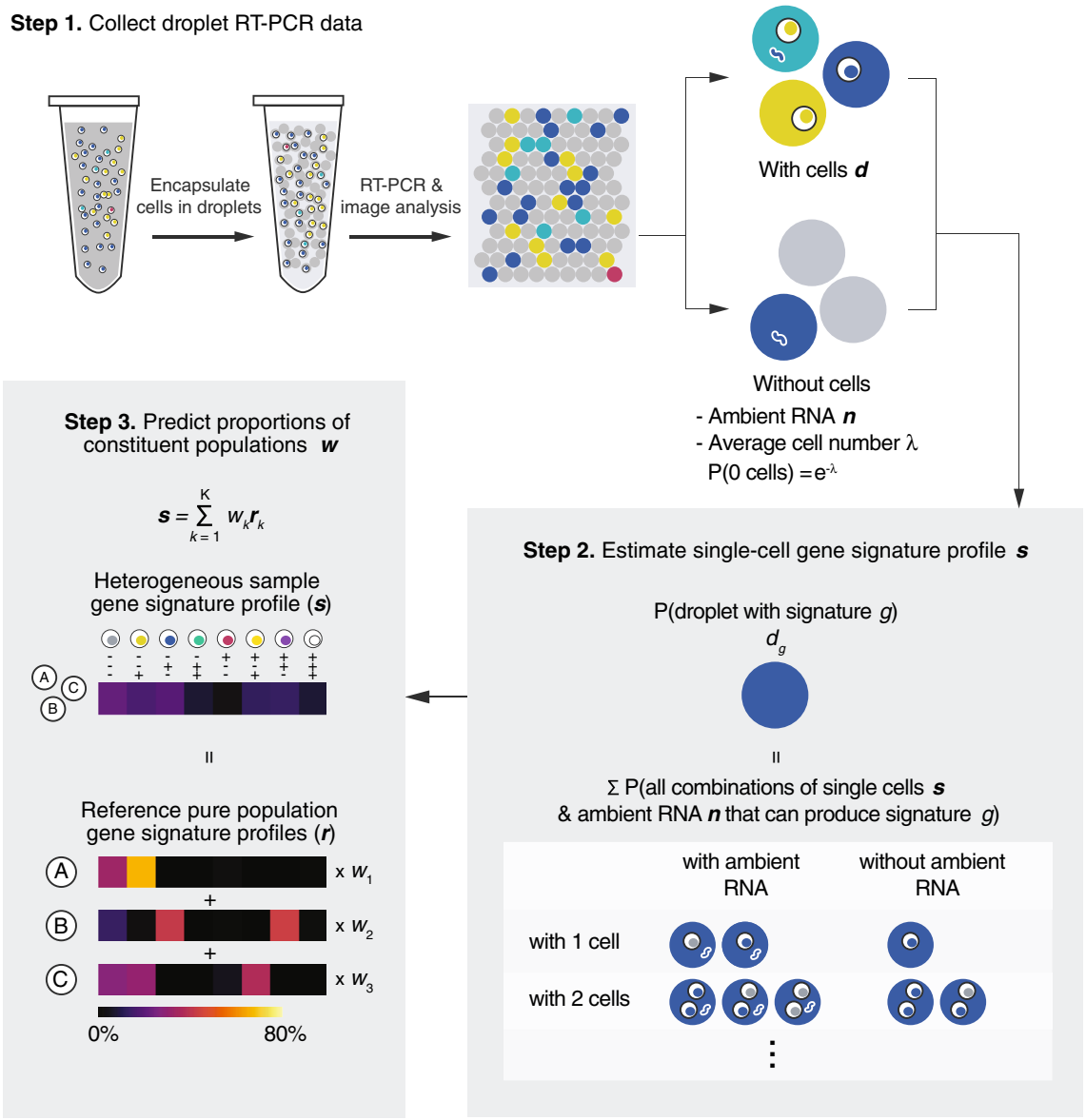


Figure 4. Schematic of deconvolution model to predict single-cell gene signature profiles and the cellular composition of heterogeneous samples. Step 1: Cells are encapsulated and assayed using the droplet RT-PCR platform. The fluorescent signal in every droplet is quantified and the droplet is classified as positive or negative for the presence of cells as well as each target gene. The proportions of droplets that contain cells displaying each gene signature (***d***) are determined. Droplets without cells (empty droplets) inform the probability of observing a certain gene signature in a droplet caused by ambient RNA (***n***). The average number of cells each droplet contains (λ) is estimated based on Poisson statistics and the proportions of empty droplets. Step 2: Single-cell gene signature profile of the sample (***s***) is estimated by correcting for ambient RNA and cell doublet effects. Due to the presence of ambient RNA and droplets that contain more than 1 cells, the droplet profile ***d*** does not represent the single-cell gene signature profile of the sample. Considering all the combinations of ambient RNA and cells that can generate a certain signature (up to having 2 cells in each droplet), the proportion of cells expressing each gene signature (***s***) is computed based on the data collected in Step 1. Step 3: The proportions of constituent populations in the heterogeneous sample (***w***) is predicted using the sample (***s***) and reference (***r***) gene signature profiles. Given that the heterogeneous sample is a physical mixture of its constituent cell populations with reference gene signature profiles (***r***) (obtained by assaying pure populations), the composition of the sample mixture (***w***) can be predicted based on the non-negative least squares model using the mixed profile (***s***) obtained from Step 2 and the reference profiles (***r***).

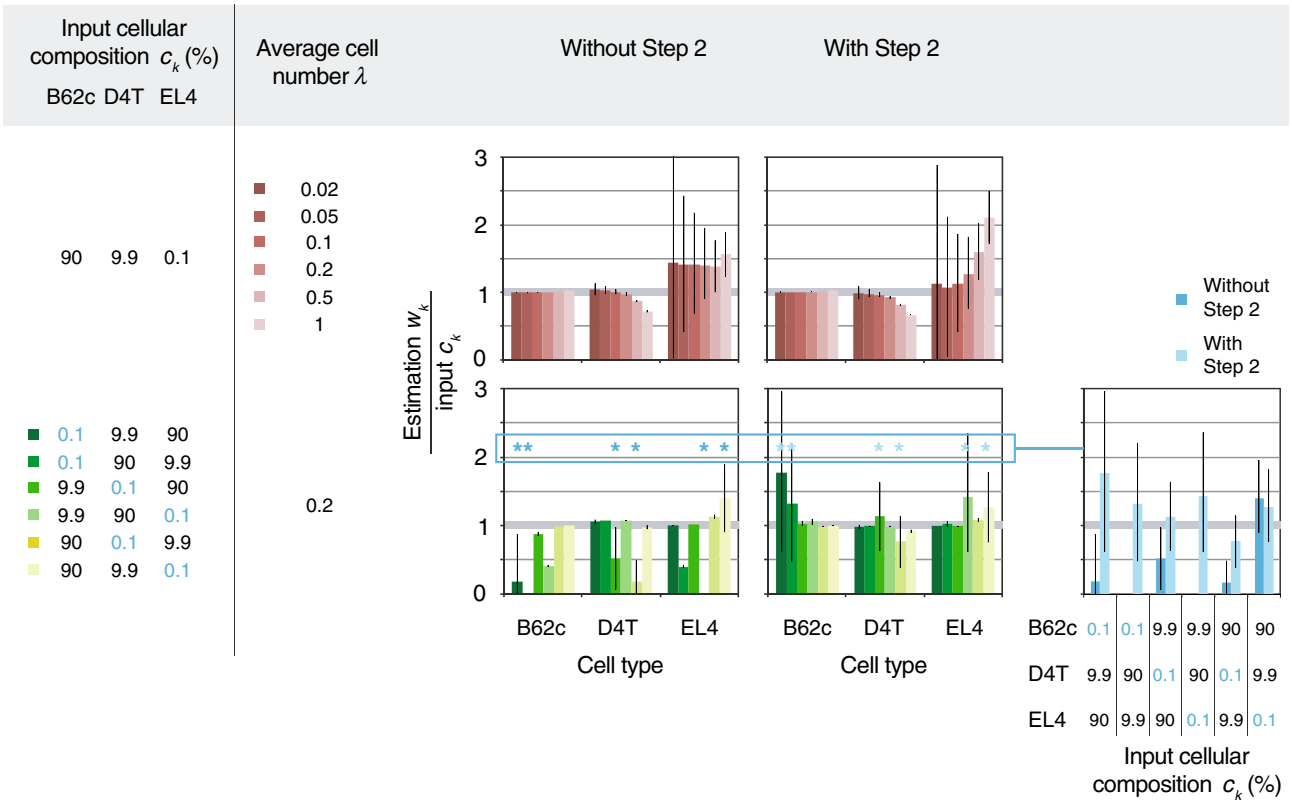


Figure 5. Computer simulation of the droplet assay provides insight into the effects of different input parameters on the accuracy and precision of the assay. Representative results from the simulation illustrating the effects of varying the average cell number per droplet λ and the cellular composition $\boldsymbol{c}$. Three cell lines (B62c, D4T, and EL4) were assayed using the droplet RT-PCR platform to acquire empirical, real-world single-cell gene signature profiles and ambient RNA levels, which were used as input reference profiles and ambient RNA levels ($\boldsymbol{r}$ and $\boldsymbol{n}$) for the simulation. The bar graphs show predictions made with or without implementing Step 2 of the deconvolution model (means of n = 1000 trials, error bars represent the 95% confidence intervals). As expected, an increase in λ (red bar graphs) resulted in higher precision of our predictions, as evidenced by the reduced error bars. The accuracy was improved by incorporating Step 2, which corrected for ambient RNA and duplets, but only when $\lambda \leq 0.2$. When varying cellular compositions (green bar graphs), an improvement in accuracy was also observed when Step 2 was applied, whereas the rarest populations (marked by asterisks and plotted separately in the blue bar graph for better comparison) were often undetected without this correction step.

biomolecules, whereas sorting for droplets with specific biomolecules can potentially be built into the microfluidic device. Another future development of our droplet system is to implement real-time imaging during the RT-PCR cycles to allow for transcript quantification. The ability to reliably quantify gene expression in hundreds of thousands of single cells would be highly valuable.

Our computational tools for simulating, predicting, and analyzing the results of our assay are also proven to be crucial for maximizing the accuracy of its readout, especially when applied to rare population detection. Our simulation model can be used to guide experimental design, including the selection of surrogate markers, sample size, and cell concentration, to achieve optimal results. Our deconvolution pipeline corrects for background noise from ambient RNA as well as doublets in the samples, which are issues commonly seen in droplet RT-PCR techniques but seldom addressed. The former contributes to a robust assay that is less sensitive to the quality of the cell sample. The latter allows us to achieve higher cellular throughput without increasing the number of droplets, therefore reduces reagent and time consumption without compromising the accuracy of our assay. This may be further improved by expanding our model to account for a higher number of multiplets. Additionally, our current model relies on pre-specified reference gene signature profiles that accurately portraits the constituent populations. More advanced deconvolution algorithms that can accommodate populations that are not included in the reference populations[87], or perturbations due to microenvironmental or developmental effects that alter the gene expression of the constituent populations[88] could improve the accuracy of the prediction and broaden the application of the assay. Our analysis pipeline requires minimal user supervision except for labeling identified clusters of populations, which can potentially be automated as well to further enhance usability.

At roughly $3 per 1000 cells (see Supplementary Table S3 for cost breakdown), our experimental cost falls within an affordable range that allows routine measurements. Our setup does not require expensive and highly specialized instrumentation and software, and can be assembled using pieces of apparatus readily available in a standard biology lab. The key to further drive down cost in the future such that millions or more cells can

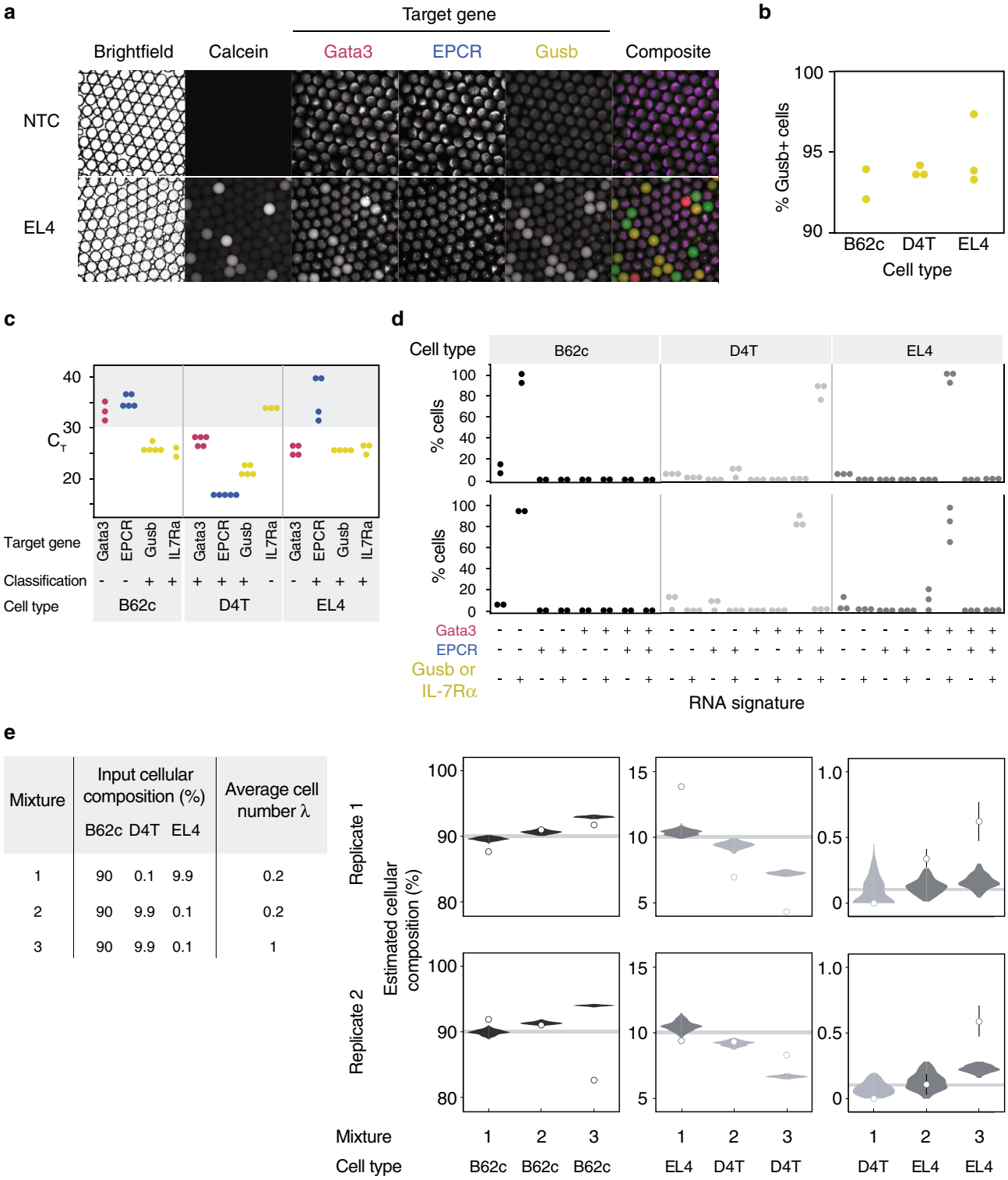


Figure 6. Implementation of droplet RT-PCR for single-cell gene expression profiling and rare cell detection. (**a**) Images of droplets containing EL4 crude cells after amplification targeting Gata3 (red), EPCR (blue), and Gusb (green). A substantial increase in fluorescent signal indicating the expressions of Gata3 and Gusb was observed in the EL4 sample but not in the no template control (NTC). (**b**) Three cell lines (A: B62c, B: D4T, and C: EL4) were assayed using the droplet RT-PCR platform. Detection rates were determined based on the percentage of cells classified as Gusb (housekeeping gene) positive. Each dot represents one technical replicate. (**c**) Population-level gene expression profiles of the three cell lines determined by conventional RT-qPCR. Targets with C_T value over 30 are classified as negative. Each dot represents one technical replicate. (**d**) Percentage of single cells expressing each gene signature determined by our droplet RT-PCR assay. Each dot represents one technical replicate. (**e**) Three cell lines were mixed at known ratios and assayed using the droplet RT-PCR platform. The parameters tested are shown in the left table. Results are grouped by the proportions of each cell line in the mixtures and the thick horizontal grey bars indicate the input proportions. The empirical predictions are plotted as dots with error bars representing standard errors, while the results obtained from our computer simulation (100 trials) are overlaid as violin plots.

Reagents	Unit	Conventional RT-qPCR			Droplet RT-PCR	
		Simplex (before optimization)	Simplex	Multiplex	IVT multiplex	Crude cell multiplex
TrisHCl, pH 8.0	mM	10	30	30	30	30
KCl	mM	55	55	90	70	70
$MgCl_2$	mM	2	2	2.5	2.5	2.5
SUPERase•In RNase Inhibitor (Invitrogen)	U/μL	1	1	1	1	1
Taq polymerase (produced in Iscove lab)	U/μL	0.1	0.3	0.6	0.6	0.6
HotStart-IT Binding Protein (Thermo Scientific)	ng/μL	510	510	950	950	950
dNTPs	mM	0.4	0.4	0.5	0.5	0.5
Forward & reverse primers (Sigma-Aldrich)	uM	0.3	0.9	0.6	0.6	0.6
TaqMan MGB Probe (Applied Biosystems)	nM	250	350	350	350	350
Nonidet P-40	% vol	0.01	0.01	0.01	0.01	0.14
Pluronic F-68 Non-ionic Surfactant (Gibco)	% vol	0	0	0	0.05	0.2
SuperScript III reverse transcriptase (Invitrogen)	U/μL	1	1	1	1	1
Bovine serum albumin (Roche)	ng/μL	150	450	450	1380	1380
DTT (Invitrogen)	mM	1	1	1	1	1
ROX passive reference dye	nM	50	50	50	0	0

Table 2. Formulation of the simplex and multiplex RT-PCR reagent mixes.

be assayed will be to decrease reagent consumption by reducing droplet size, increasing droplet stability, and improving encapsulation efficiency through new technologies that can enforce allocation of a single cell to each droplet in place of Poisson distribution.

In summary, our unique platform takes advantage of the miniaturization, compartmentalization, and precise liquid handling capabilities of the ultrahigh-throughput microdroplet system to develop an automatable, low-cost, and highly sensitive assay. As such, this novel technique provides a better single-cell analysis solution for many applications than other existing platforms, and has a high potential to be implemented in both clinical and research settings for rapid and direct evaluation of single-cell gene expression.

Materials and methods

Microfluidic device fabrication. The MF droplet generators were fabricated by soft lithography following standard protocols (SI Materials and Methods). The imaging chip with microchambers was fabricated from 1.5 mm thick sheets of poly(methyl methacrylate) (PMMA) (#8560K173, McMaster-Carr, Elmhurst, IL, USA). All parts were fabricated with computer numerical control (CNC) milling using a Tormach PCNC 770 vertical milling machine (Tormach Inc., Waunakee, WI). Microchannels and microfeatures (patterned into devices) were modelled with SolidWorks (Dassault Systemes, Velizy-Villacoublay, France). The CNC program was created with SprutCAM (SprutCAM, Naberezhnye Chelny, Russia). Devices were milled using a 1/32″ (794 μm) 4 flute carbide endmill (#89318919, MSC Industrial Supply Co., Melville, NY, USA), using procedures previously described[89]. Milled devices were sealed using solvent bonding procedures previously described[90].

Droplet generation. Cell suspension and RT-PCR reagents were loaded into 0.5 mL syringes (B305620, BD) and injected at an equal flow rate of 0.5 mL/h into a droplet generator through polyethylene tubing (427405, BD) using a syringe pump (70–4505, Harvard Apparatus). QX200 Droplet Generation Oil for EvaGreen Assays (1864006, Bio-Rad) was used for the oil phase and loaded into a 3 mL syringe. The flow rate (~2 mL/h) of the fluorinated oil was adjusted and controlled using a syringe pump to achieve the desired droplet size of 124 μm diameter. The droplet generation process was monitored using an inverted microscope (AE2000, Motic). The emulsions were collected through the outlet and polyethylene tubing into 1.5 mL microcentrifuge tubes (MCT-150-A, Axygen) before transferring to 0.2 mL PCR tubes (PCR-02-C, Axygen) for thermal cycling.

One-step RT-PCR and primer design. RT-PCR reagents were assembled on ice from one-time use aliquots right before the experiments. Either the CellsDirect One Step RT-qPCR kit (Invitrogen) or an in-house RT-qPCR mix was used. See Table 2 for the composition of the simplex and multiplex in-house reaction mixtures. Either the forward or reverse primer of each target gene was designed to span an exon-exon junction to avoid amplifying genomic DNA. TaqMan hydrolysis probes were used for quantification of product generation.

PREMIER Biosoft Beacon Designer 8.0 was used as the first step to design the TaqMan assays with the following criteria:

- Primers: 18–25 bp; melting temperature $T_m = 56 \pm 4$ °C; amplicon length = 80–200 bp; exon junction spanning.
- Taqman probes: 18–30 bp; $T_m = (T_m$ of primers) + ~ 10 °C.

TaqMan probes were designed with a higher T_m to ensure that they hybridize to their targets before the primers. Beacon Designer and NetPrimer were used to identify regions of the primers prone to form primer-dimers, and particular attention was paid to avoid GC-rich complementary regions. Assays were examined for unintended targets using NCBI Primer-BLAST and known SNPs were avoided using NCBI dbSNP. The sequences of the primers and probes directed against murine mRNAs are listed in Supplementary Table S1. IVTs were prepared using the MEGAscript T7 Transcription Kit (Invitrogen) and purified using TRIzol Reagent (Invitrogen) (see Supplementary Materials and Methods for detailed protocol and Supplementary Table S2 for primer sequences).

The samples were combined with the reagents either in 384-well PCR plates (for conventional RT-qPCR) or in droplets collected in PCR tubes (for single-cell RT-PCR). They were then subjected to the following thermal cycles: 55 °C for 30 min, 95 °C for 2 min, and 40 cycles of 95 °C for 15 s, 55 °C or 58 °C for 30 s, and 72 °C for 30 s. RT-qPCR was performed using an Applied Biosystems 7900HT instrument, and droplet RT-PCR was performed with a Biometra T-Personal Thermal Cycler.

Cell culture and fluorescence staining. EL4 and D4T cells were cultured in IMDM (Gibco) supplemented with 10% FBS (Gibco). B62c cells were cultured in IMDM (Gibco) supplemented with 10% FBS (Gibco) and 0.63%(v/v) alpha-thioglycerol. Conditioned media with IL-7 was added directly to B62c culture every 3 days at a ratio of 1:200 total volume. The EL4, D4T, and B62c cell lines were kind gifts from the Mak lab, the Iscove lab, and the Paige lab in Toronto.

To stain the cells with eBioscience Calcein Violet 450 AM Viability Dye (Invitrogen), D4T cells were first dissociated using 0.25% Trypsin–EDTA (Gibco) at 37 °C for 2 min and then suspended in culture medium. Each cell line was resuspended at a concentration of 10^6 cells/mL in PBS with 100 μM of Calcein Violet after centrifugation (200*g* for 5 min for EL4 and D4T cells, 450 *g* for 10 min for B62c cells) and incubated at 37 °C for 20 min, protected from light. 5 times the PBS volume of culture medium was then added to the cells and incubated at 37 °C for 5 min. The cells were pelleted by centrifugation and resuspended in fresh culture medium. After incubating for 30 min at 37 °C, the cells were centrifuged and resuspended in a 2.2% (w/w) glycine solution with 0.1 μg/mL Ambion RNase A (Invitrogen). The concentration and mixture of different cell types were adjusted based on the experiments they were prepared for. The cells were then incubated on ice for 3 h to reduce the amount of ambient RNA prior to encapsulation.

High-throughput droplet imaging and analysis. Droplets were loaded into microchambers on an imaging chip using a micropipette. The chips were then placed on a slide holder for automated imaging using the Cellomics Arrayscan VTI platform (Thermo Scientific). The slides were imaged with a 5X widefield objective in brightfield and 4 fluorescence channels that corresponded to the Calcein dye and Taqman probes, rendering 64 field images per microchamber per channel. These field images were tiled into one image per chamber per channel using an in-house Python script and analyzed in CellProfiler to quantify the size and total fluorescence intensities of the droplets.

Analytical pipeline and deconvolution model. Our models are implemented in Julia and are available online at https://gitlab.com/ stemcellbioengineering/droplet-rtpcr along with the parameters tested. In the following descriptions, variables are in italics, vectors are in bold, constants are in uppercase. Our models were built with the following assumptions:

- Empty droplets can be distinguished from droplets with cells.
- The proportion of droplets with more than two cells is negligible. The efficiency of the reaction in each droplet is not affected by the number of cells it contains.
- Cells and ambient RNA in a droplet can combine to produce new signatures. The possible signatures of a droplet can range from 1 (−-) to $G = 2^3$ (+ + +), where each bit represents a marker, + means positive and— means negative.
- All constituent populations in heterogeneous cell mixtures are represented in the reference profiles. The reference profiles are obtained by assaying individual reference cell types with the droplet RT-PCR platform and provided as input to deconvolve the cellular composition of cell mixtures.

The model imports data exported by Cellprofiler, which consists of the area and fluorescence intensities of each droplet. The upper and lower limits for droplet size can be assigned to discard droplets that are either too big or too small. The Variational Bayesian Gaussian Mixture algorithm with a Dirichlet process prior model (DPGMM) from the scikit-learn Python package[91] is used to cluster the droplets based on their fluorescence intensity from the Calcein dye. The clusters are then manually assigned as positive or negative for Calcein to determine the presence of cells in the droplet. The average number of cells per droplet λ is estimated based on the proportion of empty droplets and Poisson statistics using Eq. (2).

DPGMM is applied to cluster droplets based on their fluorescence intensities from the 3 Taqman probes, and the clusters are assigned as either positive or negative for each target gene. The proportions of droplets with cells that display each gene signature are determined, where $\boldsymbol{d}$ is a vector of length G and d_g is the proportion of droplets that contain cells and display the gene signature g. The probabilities of observing different gene signatures caused by ambient RNA alone are also estimated, where $\boldsymbol{n}$ is a vector of length G, and n_g is the proportion of empty droplets in the sample that display the gene signature g.

$\boldsymbol{d}$ is then modeled as the sum of all possible combinations of ambient RNA and cells that can generate each signature. The single-cell gene expression profile, where $\boldsymbol{s}$ is a vector of length G, and s_g is the proportion of single cells in the sample that express the gene signature g, is solved iteratively from the following system of equations:

$$d_1 = p_1 n_1 s_1 + p_2 n_1 {s_1}^2 \tag{3}$$

$$d_2 = p_1[n_2 s_1 + (n_1 + n_2)s_2] + p_2\left[n_2 {s_1}^2 + (n_1 + n_2)\left({s_2}^2 + s_1 s_2\right)\right] \tag{4}$$

$$d_3 = p_1[n_3 s_1 + (n_1 + n_3)s_3] + p_2\left[n_3 {s_1}^2 + (n_1 + n_3)\left({s_3}^2 + s_1 s_3\right)\right] \tag{5}$$

$$\begin{aligned} d_4 = {} & p_1[n_4 s_1 + (n_3 + n_4)s_2 + (n_2 + n_4)s_3 + (n_1 + n_2 + n_3 + n_4)s_4] + p_2\left\{n_4 {s_1}^2 + (n_3 + n_4)\left({s_2}^2 + s_1 s_2\right)\right. \\ & \left. + (n_2 + n_4)\left({s_3}^2 + s_1 s_3\right) + (n_1 + n_2 + n_3 + n_4)\left[{s_4}^2 + (s_1 + s_2 + s_3)s_4 + s_2 s_3\right]\right\} \end{aligned} \tag{6}$$

$$d_5 = p_1[n_5 s_1 + (n_1 + n_5)s_5] + p_2\left[n_5 {s_1}^2 + (n_1 + n_5)\left({s_5}^2 + s_1 s_5\right)\right] \tag{7}$$

$$\begin{aligned} d_6 = {} & p_1[n_6 s_1 + (n_5 + n_6)s_2 + (n_2 + n_6)s_5 + (n_1 + n_2 + n_5 + n_6)s_6] + p_2\left\{n_6 {s_1}^2 + (n_5 + n_6)\left({s_2}^2 + s_1 s_2\right)\right. \\ & \left. + (n_2 + n_6)\left({s_5}^2 + s_1 s_5\right) + (n_1 + n_2 + n_5 + n_6)\left[{s_6}^2 + (s_1 + s_2 + s_5)s_6 + s_2 s_5\right]\right\} \end{aligned} \tag{8}$$

$$\begin{aligned} d_7 = {} & p_1[n_7 s_1 + (n_5 + n_7)s_3 + (n_3 + n_7)s_5 + (n_1 + n_3 + n_5 + n_7)s_7] + p_2\left\{n_7 {s_1}^2 + (n_5 + n_7)\left({s_3}^2 + s_1 s_3\right)\right. \\ & \left. + (n_3 + n_7)\left({s_5}^2 + s_1 s_5\right) + (n_1 + n_3 + n_5 + n_7)\left[{s_7}^2 + (s_1 + s_3 + s_5)s_7 + s_3 s_5\right]\right\} \end{aligned} \tag{9}$$

$$\begin{aligned} d_8 = {} & p_1[n_8 s_1 + (n_7 + n_8)s_2 + (n_6 + n_8)s_3 + (n_5 + n_6 + n_7 + n_8)s_4 + (n_4 + n_8)s_5 + (n_3 + n_4 + n_7 + n_8)s_6 \\ & + (n_2 + n_4 + n_6 + n_8)s_7 + s_8] + p_2\left\{n_8 {s_1}^2 + (n_7 + n_8)\left({s_2}^2 + s_1 s_2\right) + (n_6 + n_8)\left({s_3}^2 + s_1 s_3\right)\right. \\ & + (n_5 + n_6 + n_7 + n_8)\left[{s_4}^2 + s_4(s_1 + s_2 + s_3) + s_2 s_3\right] + (n_4 + n_8)\left({s_5}^2 + s_1 s_5\right) + (n_3 + n_4 + n_7 + n_8) \\ & \left[{s_6}^2 + s_6(s_1 + s_2 + s_5) + s_2 s_5\right] + (n_2 + n_4 + n_6 + n_8)\left[{s_7}^2 + s_7(s_1 + s_3 + s_5) + s_3 s_5\right] \\ & \left. + s_2 s_7 + s_3 s_6 + s_4(s_5 + s_6 + s_7) + s_6 s_7 + s_8\right\} \end{aligned} \tag{10}$$

where and p_1 and p_2 represent the normalized probability of capturing 1 and 2 cells in a droplet that contain cells respectively:

$$p_x = \frac{P(x\ cell)}{P(1\ cell) + P(2\ cells)} \tag{11}$$

and $P(x\ cells)$ represents the probability of capturing x cells in a droplet based on Poisson distribution given λ:

$$P(x\ cells) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{12}$$

The single-cell profile of the heterogeneous sample $\boldsymbol{s}$ is then modeled as a linear combination of signature profiles of the reference populations, $\boldsymbol{r}_k$, weighted by mixture proportions $\boldsymbol{w}$:

$$\boldsymbol{s} = \sum_{k=1}^{K} w_k \boldsymbol{r}_k \tag{13}$$

where weight w_k represents the proportion of reference population k in the sample. $\boldsymbol{r}_k$ is a vector of length G and $r_{k,g}$ is the proportion of cell type k that expresses the gene signature g. The nnls() function from the NNLS package in Julia, which implements the non-negative least squares solver from[60], is used to estimate the optimal non-negative values of w_k.

Simulation model. The simulation model was designed to emulate the droplet RT-PCR experiment. Input parameters include the total number of droplets in each experiment, the average number of cells per droplet, as well as the signature profiles and concentration of each cell type. The simulation can be summarized as follows:

- The total number of droplets deposited per trial is D.

- The number of cells in any given droplet follows a Poisson distribution with a mean (average number of cells per droplet) of λ as predicted by Eq. (11).
- Cells of type k have a concentration of c_k in a well-mixed population. Cells are drawn randomly with probability equal to their concentration.
- Each cell type is characterized by two sets of signature profiles. $\boldsymbol{r}_k$ is a vector of length G and $r_{k,g}$ is the proportion of cell type k that expresses the gene signature g. Cell signatures are assigned randomly with probability equal to its proportion. $\boldsymbol{n}_k$ is a vector of length G, and $n_{k,g}$ is the proportion of empty droplets that display the gene signature g when assaying cell type k.
- The levels of ambient RNA of the sample $\boldsymbol{m}$ is modeled as a linear combination of the ambient RNA levels of the constituent cell populations, $\boldsymbol{n}_k$, weighted by mixture proportions $\boldsymbol{c}$:

$$\boldsymbol{m} = \sum_{k=1}^{K} c_k \boldsymbol{n}_k \tag{14}$$

  Ambient RNA signatures are assigned randomly with probability equal to its proportion.

- Cells and ambient RNA in a droplet are combined to produce the signature of the droplet.
- Once results are obtained from simulations, the cellular composition of the sample $\boldsymbol{w}$ is estimated using our deconvolution model described above and compared with the initial input concentrations $\boldsymbol{c}$.
- The experiment is run thousands of times with different random seeds to numerically obtain key statistics of the estimated cellular composition of the sample.

Data availability

The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Received: 3 December 2020; Accepted: 10 March 2021
Published online: 24 March 2021

References

1. Donnenberg, A. D. & Donnenberg, V. S. Rare-event analysis in flow cytometry. *Clin. Lab. Med.* **27**, 627–652 (2007).
2. Cui, C., Shu, W. & Li, P. Fluorescence in situ hybridization: Cell-based genetic diagnostic and research applications. *Front. Cell Dev. Biol.* **4**, 89 (2016).
3. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
4. Femino, A. M. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
5. Pichon, X., Lagha, M., Mueller, F. & Bertrand, E. A growing toolbox to image gene expression in single cells: Sensitive approaches for demanding challenges. *Mol. Cell* **71**, 468–480 (2018).
6. Player, A. N., Shen, L. P., Kenny, D., Antao, V. P. & Kolberg, J. A. Single-copy gene detection using branched DNA (bDNA) in situ hybridization. *J. Histochem. Cytochem.* **49**, 603–612 (2001).
7. Wang, F. *et al.* RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *J. Mol. Diagn.* **14**, 22–29 (2012).
8. Choi, H. M. T. *et al.* Programmable in situ amplification for multiplexed imaging of mRNA expression. *Nat. Biotechnol.* **28**, 1208–1212 (2010).
9. Moffitt, J. R. *et al.* High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11046–11051 (2016).
10. Eng, C.-H.L. *et al.* Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature* **568**, 235–239 (2019).
11. Arrigucci, R. *et al.* FISH-Flow, a protocol for the concurrent detection of mRNA and protein in single cells using fluorescence in situ hybridization and flow cytometry. *Nat. Protoc.* **12**, 1245–1260 (2017).
12. Battich, N., Stoeger, T. & Pelkmans, L. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nat. Methods* **10**, 1127–1133 (2013).
13. Gierahn, T. M. *et al.* Seq-well: Portable, low-cost RNA sequencing of single cells at high throughput. *Nat. Methods* **14**, 395–398 (2017).
14. Svensson, V. *et al.* Power analysis of single-cell RNA-sequencing experiments. *Nat. Methods* **14**, 381–387 (2017).
15. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
16. Sasagawa, Y. *et al.* Quartz-Seq: A highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. *Genome Biol.* **14**, R31 (2013).
17. Islam, S. *et al.* Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. *Nat. Protoc.* **7**, 813–828 (2012).
18. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
19. Hashimshony, T. *et al.* CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).
20. Jaitin, D. A. *et al.* Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* **343**, 776–779 (2014).
21. Macosko, E. Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
22. Klein, A. M. *et al.* Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
23. Zhang, X. *et al.* Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-Seq systems. *Mol. Cell* **73**, 130-142. e5 (2019).
24. Huang, M. *et al.* SAVER: Gene expression recovery for single-cell RNA sequencing. *Nat. Methods* **15**, 539–542 (2018).
25. Ziegenhain, C. *et al.* Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* https://doi.org/10.1101/035758 (2016).
26. Huber, D., von Voithenberg, L. V. & Kaigala, G. V. Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH?. *Micro Nano Eng.* **1**, 15–24 (2018).
27. Salomon, R. *et al.* Droplet-based single cell RNAseq tools: A practical guide. *Lab Chip* **19**, 1706–1727 (2019).

28. Lai, C., Stepniak, D., Sias, L. & Funatake, C. A sensitive flow cytometric method for multi-parametric analysis of microRNA, messenger RNA and protein in single cells. *Methods* **134–135**, 136–148 (2018).
29. Amamoto, R. *et al.* Probe-Seq enables transcriptional profiling of specific cell types from heterogeneous tissue by RNA-based isolation. *Elife* **8**, 2 (2019).
30. Klemm, S. *et al.* Transcriptional profiling of cells sorted by RNA abundance. *Nat. Methods* **11**, 549–551 (2014).
31. Stegle, O., Teichmann, S. A. & Marioni, J. C. Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* **16**, 133–145 (2015).
32. Bacher, R. & Kendziorski, C. Design and computational analysis of single-cell RNA-sequencing experiments. *Genome Biol.* **17**, 63 (2016).
33. Theodosiou, Z. *et al.* Automated analysis of FISH and immunohistochemistry images: A review. *Cytometry A* **71**, 439–450 (2007).
34. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
35. Zappia, L., Phipson, B. & Oshlack, A. Exploring the single-cell RNA-seq analysis landscape with the scRNA-tools database. *PLoS Comput. Biol.* **14**, e1006245 (2018).
36. Teles, J., Enver, T. & Pina, C. Single-cell PCR profiling of gene expression in hematopoiesis. *Methods Mol. Biol.* **1185**, 21–42 (2014).
37. White, A. K., Heyries, K. A., Doolin, C., Vaninsberghe, M. & Hansen, C. L. High-throughput microfluidic single-cell digital polymerase chain reaction. *Anal. Chem.* **85**, 7182–7190 (2013).
38. VanInsberghe, M., Zahn, H., White, A. K., Petriv, O. I. & Hansen, C. L. Highly multiplexed single-cell quantitative PCR. *PLoS ONE* **13**, e0191601 (2018).
39. Eastburn, D. J., Sciambi, A. & Abate, A. R. Ultrahigh-throughput Mammalian single-cell reverse-transcriptase polymerase chain reaction in microfluidic drops. *Anal. Chem.* **85**, 8016–8021 (2013).
40. Levsky, J. M., Shenoy, S. M., Pezo, R. C. & Singer, R. H. Single-cell gene expression profiling. *Science* **297**, 836–840 (2002).
41. Sun, H. *et al.* A bead-based microfluidic approach to integrated single-cell gene expression analysis by quantitative RT-PCR. *RSC Adv.* **5**, 4886–4893 (2015).
42. Kim, S. C., Clark, I. C., Shahi, P. & Abate, A. R. Single-cell RT-PCR in microfluidic droplets with integrated chemical lysis. *Anal. Chem.* **90**, 1273–1279 (2018).
43. Abate, A. R. *et al.* Impact of inlet channel geometry on microfluidic drop formation. *Phys. Rev. E. Stat. Nonlin. Soft Matter. Phys.* **80**, 26310 (2009).
44. White, A. K. *et al.* High-throughput microfluidic single-cell RT-qPCR. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13999–14004 (2011).
45. Bontoux, N. *et al.* Integrating whole transcriptome assays on a lab-on-a-chip for single cell gene profiling. *Lab Chip* **8**, 443–450 (2008).
46. Curry, J., McHale, C. & Smith, M. T. Low efficiency of the Moloney murine leukemia virus reverse transcriptase during reverse transcription of rare t(8;21) fusion gene transcripts. *Biotechniques* **32**, 755–768 (2002).
47. Lareu, R. R., Harve, K. S. & Raghunath, M. Emulating a crowded intracellular environment in vitro dramatically improves RT-PCR performance. *Biochem. Biophys. Res. Commun.* **363**, 171–177 (2007).
48. Ratnamohan, V. M., Cunningham, A. L. & Rawlinson, W. D. Removal of inhibitors of CSF-PCR to improve diagnosis of herpesviral encephalitis. *J. Virol. Methods* **72**, 59–65 (1998).
49. Chandler, D. P., Wagnon, C. A. & Bolton, H. Jr. Reverse transcriptase (RT) inhibition of PCR at low concentrations of template and its implications for quantitative RT-PCR. *Appl. Environ. Microbiol.* **64**, 669–677 (1998).
50. Kreader, C. A. Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Appl. Environ. Microbiol.* **62**, 1102–1106 (1996).
51. Bustin, S. A. *et al. AZ of Quantitative PCR* (International University Line La Jolla, 2004).
52. Schwanhäusser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
53. Marguerat, S. *et al.* Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151**, 671–683 (2012).
54. Persson, K., Hamby, K. & Ugozzoli, L. A. Four-color multiplex reverse transcription polymerase chain reaction–overcoming its limitations. *Anal. Biochem.* **344**, 33–42 (2005).
55. Henegariu, O., Heerema, N. A., Dlouhy, S. R., Vance, G. H. & Vogt, P. H. Multiplex PCR: Critical parameters and step-by-step protocol. *Biotechniques* **23**, 504–511 (1997).
56. Shuber, A. P., Grondin, V. J. & Klinger, K. W. A simplified procedure for developing multiplex PCRs. *Genome Res.* **5**, 488–493 (1995).
57. Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. in *KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining* (ed. Evangelos Simoudis Jiawei Han) 226–231 (Institute for Computer Science, University of Munich, München, Germany, 1996).
58. Haight, F. A. Handbook of the poisson distribution. (1967).
59. Blei, D. M. & Jordan, M. I. Variational inference for Dirichlet process mixtures. *Bayesian Anal.* **1**, 121–143 (2006).
60. Lawson, C. L. & Hanson, R. J. *Solving Least Squares Problems* (SIAM, 1995).
61. Lin, Y. *et al.* Evaluating stably expressed genes in single cells. doi: https://doi.org/10.1101/229815.
62. Tabula Muris Consortium *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).
63. Chubb, J. R. & Liverpool, T. B. Bursts and pulses: Insights from single cell studies into transcriptional mechanisms. *Curr. Opin. Genet. Dev.* **20**, 478–484 (2010).
64. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
65. Liu, S. & Trapnell, C. Single-cell transcriptome sequencing: Recent advances and remaining challenges. *F1000Res.* **5**, 2 (2016).
66. Zhang, H., Jenkins, G., Zou, Y., Zhu, Z. & Yang, C. J. Massively parallel single-molecule and single-cell emulsion reverse transcription polymerase chain reaction using agarose droplet microfluidics. *Anal. Chem.* **84**, 3599–3606 (2012).
67. Gong, Y., Ogunniyi, A. O. & Love, J. C. Massively parallel detection of gene expression in single cells using subnanolitre wells. *Lab Chip* **10**, 2334–2337 (2010).
68. Thompson, A. M. *et al.* Self-digitization microfluidic chip for absolute quantification of mRNA in single cells. *Anal. Chem.* **86**, 12308–12314 (2014).
69. Sanchez-Freire, V., Ebert, A. D., Kalisky, T., Quake, S. R. & Wu, J. C. Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns. *Nat. Protoc.* **7**, 829–838 (2012).
70. Lee, J. H. *et al.* Fluorescent in situ sequencing (FISSEQ) of RNA for gene expression profiling in intact cells and tissues. *Nat. Protoc.* **10**, 442–458 (2015).
71. Reuben, R. C. & Gefter, M. L. A DNA-binding protein induced by bacteriophage T7. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 1846–1850 (1973).
72. Scherzinger, E., Litfin, F. & Jost, E. Stimulation of T7 DNA polymerase by a new phage-coded protein. *Mol. Gen. Genet.* **123**, 247–262 (1973).
73. Araki, H. & Ogawa, H. A T7 amber mutant defective in DNA-binding protein. *Mol. Gen. Genet.* **183**, 66–73 (1981).

74. Nakai, H. & Richardson, C. C. The effect of the T7 and *Escherichia coli* DNA-binding proteins at the replication fork of bacteriophage T7. *J. Biol. Chem.* **263**, 9831–9839 (1988).
75. Kubu, C. J. HotStart-IT®: A novel hot start PCR method based on primer sequestration. *Biotechniques* **44**, 275–277 (2008).
76. Wilson, I. G. Inhibition and facilitation of nucleic acid amplification. *Appl. Environ. Microbiol.* **63**, 3741–3751 (1997).
77. Schrader, C., Schielke, A., Ellerbroek, L. & Johne, R. PCR inhibitors—Occurrence, properties and removal. *J. Appl. Microbiol.* **113**, 1014–1026 (2012).
78. Opel, K. L., Chung, D. & McCord, B. R. A study of PCR inhibition mechanisms using real time PCR. *J. Forensic Sci.* **55**, 25–33 (2010).
79. Kim, Y. T., Tabor, S., Bortner, C., Griffith, J. D. & Richardson, C. C. Purification and characterization of the bacteriophage T7 gene 2.5 protein. A single-stranded DNA-binding protein. *J. Biol. Chem.* **267**, 15022–15031 (1992).
80. Bryant, F. R. & Lehman, I. R. On the mechanism of renaturation of complementary DNA strands by the recA protein of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 297–301 (1985).
81. Zou, Z. *et al.* ssDNA hybridization facilitated by T7 ssDNA binding protein (gp2.5) rapidly initiates from the strand terminus or internally followed by a slow zippering step. *Biochimie* **147**, 1–12 (2018).
82. Shokri, L., Rouzina, I. & Williams, M. C. Interaction of bacteriophage T4 and T7 single-stranded DNA-binding proteins with DNA. *Phys. Biol.* **6**, 025002 (2009).
83. Rajagopal, A. *et al.* Significant expansion of real-time PCR multiplexing with traditional chemistries using amplitude modulation. *Sci. Rep.* **9**, 1053 (2019).
84. Zhong, Q. *et al.* Multiplex digital PCR: Breaking the one target per color barrier of quantitative PCR. *Lab Chip* **11**, 2167–2174 (2011).
85. Chisti, Y. Hydrodynamic damage to animal cells. *Crit. Rev. Biotechnol.* **21**, 67–110 (2001).
86. Csaszar, E., Cohen, S. & Zandstra, P. W. Blood stem cell products: Toward sustainable benchmarks for clinical translation. *BioEssays* **35**, 201–210 (2013).
87. Quon, G. & Morris, Q. ISOLATE: a computational strategy for identifying the primary origin of cancers using high-throughput sequencing. *Bioinformatics* **25**(21), 2882–2889 (2009).
88. Qiao, W. *et al.* PERT: A method for expression deconvolution of human blood samples from varied microenvironmental and developmental conditions. *PLoS Comput. Biol.* **8**(12), e1002838 (2012).
89. Guckenberger, D. J., de Groot, T. E., Wan, A. M. D., Beebe, D. J. & Young, E. W. K. Micromilling: A method for ultra-rapid prototyping of plastic microfluidic devices. *Lab Chip* **15**, 2364–2378 (2015).
90. Wan, A. M. D., Sadri, A. & Young, E. W. K. Liquid phase solvent bonding of plastic microfluidic devices assisted by retention grooves. *Lab Chip* **15**, 3785–3792 (2015).
91. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Acknowledgements

The authors thank Céline Bauwens, Joel Östblom, and Daniel Aguilar-Hidalgo for critical reading and editing of the manuscript. This work was supported by the McEwen Centre for Regenerative Medicine, the Canadian Institutes of Health Research, the Natural Sciences and Engineering Research Council of Canada, Medicine by Design, Terry Fox Foundation, the Canadian Cancer Research Institute, the Stem Cell Network, the Princess Margaret Hospital Foundation, the Campbell Family Institute for Cancer Research, and the Ontario Ministry of Health and Long Term Care.

Author contributions

J.M. and G.T. designed and performed the experiments. A.M.D.W and E.W.K.Y aided in the development and fabrication of the imaging chips. J.M. developed the analysis pipeline and simulation. J.M. and G.T. processed the experimental data and performed the analysis. J.M. drafted the manuscript and designed the figures with inputs from all authors. E.K., N.N.I., and P.W.Z. conceived the study. N.N.I. and P.W.Z. were in charge of overall direction and planning.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-86087-4.

Correspondence and requests for materials should be addressed to P.W.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.