Integrated workflow of down-scaled Smart-seq2 and absolute quantitative, single-cell digital PCR validates lack of representability of low-abundant transcripts

Tobias Lange[1,2,§], Tobias Groß[1,2], Ábris Jeney[3], Julia Scherzinger[4], Elly Sinkala[4], Christoph Niemöller[4], Stefan Zimmermann[1], Peter Koltay[1,2], Felix von Stetten[1,5], Roland Zengerle[1,5], Csaba Jeney[1,2,§]

Affiliations

[1]Laboratory for MEMS Applications, Department of Microsystems Engineering (IMTEK), University of Freiburg, Georges-Köhler-Allee 103, Freiburg 79110, Germany

[2]Actome GmbH, Georges-Köhler-Allee 103, Freiburg 79110, Germany

[3]Faculty of Medicine, University of Freiburg, Breisacher Straße 153, Freiburg 79110, Germany

[4]CYTENA GmbH, Zollhallenstraße 5, Freiburg 79106, Germany

[5]Hahn-Schickard, Georges-Köhler-Allee 103, Freiburg 79110, Germany

[§]To whom correspondence should be addressed: csaba.jeney@actome.de, tobias.lange@imtek.uni-freiburg.de

Abbreviations

CI: confidence interval, cl: crude lysate, CV: coefficient of variance, DE: differential expression, DEG: differentially expressed gene, dPCR: digital PCR, FACS: fluorescence-activated cell sorting, log2FC: log2 of fold change between two conditions, RT-qPCR: reverse transcription quantitative PCR, sc: single-cell, scRNA-seq: single-cell RNA sequencing, scRT-ddPCR: single-cell reverse transcription droplet digital PCR, TPM: Transcripts per kilobase million, UMAP: uniform manifold approximation and projection,

Keywords

Down-scaled, dPCR, scRNA-seq, scRT-ddPCR, single cells, Smart-seq2

Abstract

Single-cell RNA sequencing (scRNA-seq) can unmask transcriptional heterogeneity facilitating the detection of rare subpopulations at unprecedented resolution. In response to challenges related to coverage and quantity of transcriptome analysis, the lack of unbiased and quantitative validation methods hampers further improvements. Digital PCR (dPCR) represents such a method as the inherent partitioning could enhance detection by increasing the effective concentration of nucleic acids. Thus, we have developed a novel, integrated workflow combining down-scaled, single-cell Smart-seq2 and absolute quantitative, single-cell digital PCR. Our workflow reduces biological and technical variability by analyzing single cells from the same population using identical steps for liquid and single-cell handling. Using two breast cancer cell lines, MCF7 and BT-474, we validated the performance of our down-scaled Smart-seq2 by comparative clustering with published data and our scRT-ddPCR by comparison of absolute gene mRNA counts to bulk methods. We observed significant differences in signal distributions of low-abundant *ErbB2* in MCF7 between scRNA-seq and scRT-ddPCR. The differences are mirrored in the calculated fold changes. We assume that low-abundant transcripts suffer from dropouts in scRNA-seq while high abundant transcripts (such as *ACTB* or *ErbB2* in BT-474 cells) suffer from minimal dropouts in the down-scaled Smart-seq2. In scRT-ddPCR however, the inherent partitioning of the reaction volume increases the effective concentration of mRNAs and thus improves sensitivity. As a conclusion, our workflow combines transcriptome-wide gene expression profiling by scRNA-seq with the precise and reliable determination of absolute gene mRNA per cell counts and fold changes by scRT-ddPCR. This can correct biased conclusions made solely on the basis of scRNA-seq. We think this workflow is a valuable addition to the single-cell transcriptomic research toolbox and could even become a new standard in fold change validation because of its reliability, ease of use and increased sensitivity.

1 Introduction

RNA-seq is the gold standard for gene expression analysis. Herein, differential expression (DE) analysis between two conditions is pivotal to answer challenging questions in research and clinical applications. In bulk RNA-seq population heterogeneity remains covert, whereas scRNA-seq can capture delicate differences between cells [1]. The growing interest in scRNA-seq is expressed by the development of new platforms and their applications [2–6]. Precise determination of the transcriptional landscape is mandatory to allow applications such as: unmasking transcriptional heterogeneity in healthy and cancerous tissues by functional clustering [7–9], discovering uncharacterized cell types [10], and identifying phylogenetic relationships between cells [11]. Unbiased transcriptional data is also essential to facilitate the detection of rare subpopulations at unprecedented resolution, which enables researches to understand underlying mechanisms of drug resistance development and relapse in disease treatment [5,8,12]. However, the inherent low sample input in scRNA-seq introduces a significant amount of noise and increases the propensity for dropouts and thus cell-to-cell variability [13,14]. This is especially dramatic with respect to low-abundant transcripts. Several imputation tools were developed to compensate for dropouts [13,15], which can further bias the evaluation by over- or underestimating dropout rates. Furthermore, the tremendous variety of platforms and bioinformatic tools has not yet solidified into a consistent pipeline [13,16,17]. For instance, the two most frequently used methods, plate-based Smart-seq2 [18,19] and the droplet-based Chromium system from 10X Genomics [20] showed striking differences in a direct comparison [21]. Smart-seq2 proved to be more sensitive, especially for low-abundant transcripts, hence more genes per cell and less noise were detected. It is hypothesized that through down-scaling of reaction volumes the sensitivity could be improved [22] while conserving data quality [23,24]. This idea follows the concept of dPCR [25] as down-scaling (by partitioning) is an inherent feature and enhances molecular detections by increasing the effective concentration of nucleic acids. To confirm conclusions solely made on the basis of scRNA-seq, it is common practice to validate expression data by scRT-qPCR [26] primarily to control the elevated amount of technical noise and thus dropouts. Several scRT-qPCR workflows have been described [27–31] as well as a few scRT-ddPCR workflows [32–34]. The majority of these workflows use fluorescence-activated cell sorting (FACS) for single-cell isolation [28,29,31,32,35], while other studies use microfluidic devices [33,34], micromanipulators [27] or manual cell picking [30]. Efficient single cell isolation is necessary to improve throughput. Despite its widespread use, single-cell isolation with FACS requires high sample input and inherent shear forces can damage the cells and impair RNA integrity [35,36]. Neither of these workflows use integrated single-cell handling for sequencing and PCR to minimize biological and technical variability. The F.SIGHT™ single-cell dispenser (CYTENA GmbH) [37] is a gentle method and requires minimal sample input (down to 5000 cells in 5 µl). Single-cell isolation is ensured by image-based analysis delivering an image proof of each dispensation event, which can unambiguously be assigned

to the addressed well of the microplate. Therefore, the F.SIGHT™ is an ideal platform to integrate scRNA-seq and gene expression validation by orthogonal PCR means. For such purposes qPCR has been widely used [26] but demands an external calibration for absolute quantification [38], is less sensitive and more susceptible to inhibitors compared to dPCR [25,39]. Especially the lower sensitivity hampers its use for challenging single-cell mRNA quantification with a focus on low-abundant transcripts. Through partitioning, dPCR can reliably detect single molecules in the regular operating mode or as one-step reverse transcription dPCR [25,33,39].

Here we propose a novel workflow, which integrates transcriptome-wide gene expression profiling by down-scaled Smart-seq2 with the precise and reliable determination of absolute gene mRNA per cell counts by scRT-ddPCR. To illustrate the functionality of our workflow, we have chosen two HER2-positive breast cancer cell lines, BT-474 and MCF7, whereof BT-474 cells overexpress *ErbB2* [40]. Single cells are gently isolated exploiting the F.SIGHT™ enabling full control over morphological criteria and ensuring high cell viability [37]. Down-scaled volumes of reagents are dispensed with an immediate drop on-demand technology (I.DOT) [41,42], and subsequent digital PCR enables absolute quantification of gene mRNA per cell counts [25]. We confirmed the down-scaled Smart-seq2 sequencing by comparing with published data and by using three common alignment tools (salmon, kallisto and STAR). Furthermore, we validated the scRT-ddPCR by total RNA extraction methods and found high concordance. Taken together, this workflow enabled us to verify dropouts of low-abundant transcripts in scRNA-seq to potentially correct resulting, biased conclusions.

2 Materials and methods

2.1 Cells and cell culture

MCF7 (ATCC® HTB-22™) and BT-474 (ATCC® HTB-20™) cells were obtained from the BIOSS Centre for Biological Signaling Studies (Freiburg, Germany). MCF7 cells were cultured in DMEM, GlutaMAX™ Supplement (31966021, Gibco™) and BT-474 cells were cultured in DMEM/F12, GlutaMAX™ Supplement (31331028, Gibco™) in Nunc™ EasYFlask™ Cell Culture Flasks (156340, Thermo Scientific™). Both media were supplemented with 10 % FBS (10270106, Gibco™) and 1 % Pen/Strep (15140122, Gibco™). Cells were cultured until ~90 % confluency in a cell culture incubator (Heracell™ 150i CO_2 Incubator, 50116048, Thermo Scientific™) under a 5 % CO_2 atmosphere at 37 °C. Cells were harvested with 1X TrypLE™ Express Enzyme (12604021, Gibco™). Trypsin activity was quenched by addition of medium. The cells were washed twice with DPBS (14040133, Gibco™) and counted (Countess® II Automated Cell Counter, Invitrogen™) including live/dead staining with trypan blue (T10282, Invitrogen™).

2.2 Total RNA isolation and bulk cell lysis

Total RNA was isolated from $1x10^6$ MCF7 and $1x10^6$ BT-474 cells ('bulk') using the RNeasy Mini Kit (74104, Qiagen) in combination with the QIAshredder (79654, Qiagen) for lysate homogenization according to manufacturer's instructions. Simultaneously, total RNA was isolated using the *Quick*-DNA/RNA Microprep Plus Kit (D7005, Zymo Research) with an upfront proteinase K digest and on-column DNase I digest. RNA concentration (**Tab S5**) was measured with the NanoDrop™ One (Thermo Scientific™). RNA integrity was checked on a 1.2 % native agarose gel (2267.1, Roth) using 1X TBE buffer (3061.1, Roth) (**Fig S2c**). 1 µg total RNA was combined with 1X DNA Orange Loading Dye (R0631, Thermo Scientific™) and 60 to 75 % formamide (6749.3, Roth) and heated to 65 °C for 5 min before loading. RNA was visualized with 1X GelRed® Nucleic Acid Stain (SCT123, Milipore). Absolute gene mRNA per cell counts were determined in a dilution series using dPCR (2.5 Droplet digital PCR).

Crude lysates ('cl') from $1x10^6$ MCF7 and $1x10^6$ BT-474 cells were prepared using 500 µl LBTW (proprietary buffer of Actome GmbH). The lysate was then diluted to 10 cells/µl with DPBS and 100 nl (single-cell volume equivalent) were dispensed into the dPCR master mix using I.DOT (2.3 Liquid dispensation using I.DOT).

2.3 Liquid dispensation using I.DOT

The Immediate Drop-on-demand Technology (I.DOT One; Dispendix, Stuttgart, Germany) [41,42] with I.DOT PURE plates 90 µm orifice (Dispendix, Stuttgart, Germany) was used to

154 dispense 0.5 µl LBTW into a 384-well V-bottom plate (0030623304, Eppendorf®). Prior to
155 dispensation, the I.DOT was calibrated for the applied liquids to ensure reliable dispensation.
156 The reduced volumes for down-scaled Smart-seq2 (2.6.1 cDNA synthesis using Smart-seq2)
157 and down-scaled library preparation (2.6.2 Library preparation using Nextera XT and
158 sequencing) were dispensed using the I.DOT.
159

160 ### 2.4 Single-cell dispensation using F.SIGHT™

161 The single cell dispensing procedure was performed as described earlier [37,43,44]. The
162 F.SIGHT™ single-cell dispenser (CYTENA GmbH, Freiburg, Germany) is an improved
163 version of the single-cell printer (SCP) [37]. Both MCF7 and BT-474 cell concentrations were
164 adjusted to $1x10^6$ cells/ml and loaded into a Dispensing Cartridge (CYTENA GmbH,
165 Freiburg, Germany). The settings for MCF7 cells were 10 to 25 µm cells size (BT-474: 10 to
166 30 µm) and 0.5 to 1 roundness (same for BT-474) (**Fig 1a**, **S3a** and **S3b**). The F.SIGHT™
167 can reliably dispense single cells in minimal liquid volumes [44]. The single-cell dispensation
168 efficiency (fraction of successful single-cell isolation events from targeted single-cell
169 isolation events) is usually around 90 % [37], and is additionally controlled by cell images
170 unambiguously assigned to each dispensation event (**Fig 1a**). Thus, other than single-cell
171 dispensation events like multiple cells per droplet or empty droplets can be excluded.

172

173 ### 2.5 Droplet digital PCR

174 *ErbB2* and *ACTB* mRNAs were analyzed using the naica® Crystal Digital PCR System (Stilla
175 Technologies, Villejuif, France) [45]. Master mix was prepared as follows: 11.5 µl qScript
176 XLT 1-Step RT-qPCR ToughMix (2X) (95132, Quantabio), 1.15 µl TaqMan Assay
177 Hs01001580_m1 (20X) (ErbB2) (4331182, Applied Biosystems™), 1.15 µl TaqMan Assay
178 Hs01060665_g1 (20X) (ACTB) (4448489, Applied Biosystems™), 0.23 µl fluorescein
179 (100X, prepared according to "Fluorescein preparation for naica® system" from Stilla
180 Technologies) (0681-100G, VWR Chemicals), 1 µl of diluted RNA sample (or a single cell),
181 ad 23 µl H_2O. After thorough mixing, 20 µl of the reaction mix were transferred bubble-free
182 to the Sapphire Chips (Stilla Technologies, Villejuif, France). The dPCR conditions of the
183 Geode cycler were: partitioning of the reaction mix, cDNA synthesis (50 °C, 10 min), initial
184 denaturation (95 °C, 1 min); followed by 45 cycles of denaturation (95 °C, 6 s), annealing and
185 extension (60 °C, 45 s) and finally the pressure was released. The chips were transferred to
186 the Prism3 reader and imaged using exposure times: 65 ms and 150 ms for FAM and HEX
187 channel (82 mm focus). Afterwards, droplet quality was manually controlled and in case of
188 poor quality, e.g. coalescence or air bubbles, the respective areas were excluded from further
189 analysis. All NTCs were negative (data not shown). The average droplet volume using 1X
190 qScript XLT 1-Step RT-qPCR ToughMix is 0.548 nl. Hence, the corresponding analysis

191 configuration file was used for quantification (User Manual v2.1 of the Crystal Miner
192 Software, Stilla Technologies, 2018). We comply with the dMIQE guidelines [46,47] and
193 report all essential information (**Tab S7**).

194

195 ### 2.6 Down-scaled single-cell RNA sequencing

196 #### 2.6.1 cDNA synthesis using Smart-seq2

197 For cDNA synthesis the SMART-Seq Single Cell Kit (634472, Takara BIO) at 1/10 of the
198 original reaction volume was used (down-scaled Smart-seq2 workflow). Briefly, 1.15 µl of
199 the lysis buffer (0.1 µl Reaction Buffer, 0.1 µl 3' SMART-Seq CDS Primer II A and 0.95 µl
200 dH_2O) were dispensed in skirted 384-well PCR plates (4ti-0384/X, 4titude) using the I.DOT.
201 Subsequently, single cells (MCF7 and BT-474) and NTC (empty droplets from the
202 F.SIGHT™) were dispensed into the lysis buffer. The plates were sealed (AB0558, Thermo
203 Scientific™) and snap-frozen at -80 °C until further processing. After thawing, primers were
204 annealed in a C1000 Touch™ Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA) at
205 72 °C for 3 min. Afterwards, 0.75 µl of RT Master Mix (0.4 µl SMART-Seq sc First Strand
206 Buffer, 0.1 µl SMART-Seq sc TSO, 0.05 µl RNase Inhibitor and 0.2 µl SMARTScribe II
207 Reverse Transcriptase) were added to each well using the I.DOT. The plate was processed in
208 the C1000 Thermal Cycle for cDNA synthesis at 42 °C for 180 min, 70 °C for 10 min and
209 4 °C hold. Next, 3 µl PCR Master Mix (2.5 µl SeqAmpCB PCR Buffer (2X), 0.1 µl PCR
210 Primer, 0.1 µl SeqAmp DNA Polymerase and 0.3 µl dH_2O) were added to each well and
211 cDNA was amplified (95 °C for 1 min, 19 cycles: 98 °C for 10 sec, 65 °C for 30 sec, and
212 68 °C for 3 min; 72 °C for 10 min and 4 °C hold). Purification of cDNA was performed using
213 9 µl of AMPure XP bead suspension (A63880, Beckman Coulter) per well according to
214 manufacturer's instructions. In brief, beads and cDNA were incubated for 8 min at room
215 temperature. The beads were separated using magnetic separation for 5 min and beads were
216 washed with 30 µl 80 % ethanol for 30 sec. Afterwards, the beads were resuspended in 17 µl
217 10 mM Tris-HCl and incubated for 8 min at room temperature. The beads were separated
218 using magnetic separation for 5 min. 15 µl supernatant of each well were transferred to a fresh
219 384-well plate. The cDNA quantity was determined with the Quant-iT™ PicoGreen™
220 dsDNA Assay Kit (P7589, Invitrogen™) in 384-well plates (4ti-0203, 4titude). Fluorescent
221 intensities were measured using the Spark 10M Microplate Reader (Tecan, Männedorf,
222 Switzerland). cDNA quality was determined with Agilent's 2100 Bioanalyzer using the High
223 Sensitivity DNA Kit (5067-4626, Agilent) according to manufacturer's instructions (**Fig 1b**).

224

225 #### 2.6.2 Library preparation using Nextera XT and sequencing

226 Prior to tagmentation cDNA was normalized to 0.2 ng/µl. Tagmentation was performed using
227 the Nextera XT DNA Library Preparation Kit (FC-131-1024, Illumina) at 10-fold
228 down-scaled reaction volumes. 1 µl of Tagment DNA Buffer and 0.5 µl Amplicon Tagment
229 Mix were added to each well using the I.DOT. The amplified and purified cDNA was
230 tagmented for 8 min at 55 °C in a C1000 Thermal Cycler. Transposome activity was
231 quenched by addition of 0.5 µl Neutralize Tagment Buffer and incubation at room
232 temperature for 5 min. MCF7 and BT-474 libraries were independently amplified with index
233 primers N7xx and S5xx of Nextera XT Index Kit v2 Set A (FC-131-2001, Illumina) and
234 Nextera XT Index Kit v2 Set B (FC-131-2002, Illumina). 1.5 µl Nextera PCR Master Mix and
235 0.5 µl of a unique combination of primers were added to each well using the I.DOT and
236 tagmented cDNA was amplified in a C1000 Thermal Cycler (72 °C for 3 min, 95 °C for
237 30 sec, 12 cycles: 95 °C for 10 sec, 55 °C for 30 sec, and 72 °C for 30 sec; 72 °C for 5 min
238 and 10 °C hold). The libraries were pooled, and purified according to the previously
239 mentioned bead clean-up procedure (AMPure XP bead suspension, A63880, Beckman
240 Coulter), except that after removal of ethanol, 75 µl of resuspension buffer were added and
241 incubated for 3 min. 73 µl of eluted library were transferred to a fresh tube. The quality of
242 tagmented libraries was determined using the High Sensitivity DNA Kit on Agilent's 2100
243 Bioanalyzer (**Fig 1c**). The pooled library was sequenced on the NextSeq 500 System
244 (Illumina, San Diego, CA, USA) using the High-Output v2.5 Kit (20024906, Illumina) with
245 75 bp single-end reads.

246

247 2.6.3 Bioinformatics data analysis

248 FASTQ files were generated with bcl2fastq v2.20 ("--no-lane-splitting" flag). Sample quality
249 was assessed with FASTQC v0.11.9 (exemplary images: **Fig S1a** to **S1d**). The samples were
250 aligned to the GRCh38 cDNA reference transcriptome from Ensembl using salmon v1.3.0
251 ("--validateMappings" flag) [48] or kallisto v0.46.1 [49], as well as to GRCh38.p13 genome
252 with STAR 2.7.5c in solo mode [50]. The outputs were analyzed in Jupyter Notebooks [51]
253 (jupyter core v4.7.0, jupyter-notebook v6.1.6) using R version 4.1.2. Transcript abundances
254 (TPM) and count estimates were imported with the tximeta [52] package for salmon and
255 tximport [53] for kallisto and summarized to genes using the summarizeToGene() statement.
256 STAR alignment files were counted with featureCounts v2.0.3 [54] and the count matrices
257 were directly imported. Cells with an alignment efficiency below 80 % were filtered out
258 (salmon and kallisto) (**Fig 1d** and **Tab S1**). Cells with less than 60 % of uniquely mapped
259 reads were filtered out (STAR) (**Fig 1e** and **Tab S1**). In both cases only cells with at least
260 1E+5 detected reads were considered. Pseudo-bulk differential expression analysis was
261 performed with DESeq2 v1.32.0 [55]. In order to evaluate transcriptional similarity between
262 cells analyzed on the down-scaled Smart-seq2 platform to published data (GSE151334) [4],
263 the datasets were combined, batch-corrected with BBKNN [56] and projected into a two-

dimensional space with SCANPY's (v1.8.1) [57] built-in UMAP-function (uniform manifold approximation and projection).

2.7 Statistical analysis

Groups were initially tested for normal distribution (Shapiro-Wilk test) and for homoscedasticity (F-test) upon which information a suitable test was chosen (if not stated differently): Mann-Whitney test (normal distribution rejected), Welch's t-test (normal distribution, heteroscedasticity) or Student's t-test (normal distribution, homoscedasticity). Distributions were compared using the Kolmogorov-Smirnov test. Bonferroni correction was applied for multiple testing corrections. Significance levels are indicated as follows: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns $p > 0.05$. Boxplots indicate the inner quartiles of the data (25 % to 75 %). Whiskers show 1.5xIQR (interquartile range). The median is drawn as a horizontal line. The mean is represented by a square. Individual data points are shown as dots. All plots and statistical analyses were performed with OriginPro 2021 (OriginLab Corporation).

2.8 Bootstrapping comparison

To compare fold changes between two methods (*a*: scRNA-seq and *b*: scRT-ddPCR), we used a bootstrapping comparison because regular statistical tests suffer from p-value inflation after repetitive bootstrapping. The algorithm is described in **Fig S4** based on the ratio r_g (**Eq 1**). In brief, four arrays of expression values are needed (one per cell line and per method). A subset of each initial expression array is randomly subsampled with replacement (same length as initial array). The log2FCs of the means of these expression arrays are calculated per method. Then the ratio r_g of these log2FCs is determined (**Eq 1**). Subsampling and ratio calculation is repeated 1,000 times. Finally, mean and 95 % confidence intervals (CI) of the new array r_g is determined. If the 95 % CI overlaps with 1, the methods are assumed to yield the same fold change.

$$r_g = \frac{log_2\left(\frac{\overline{gene\ x}_{MCF7}}{\overline{gene\ x}_{BT-474}}\right)_a}{log_2\left(\frac{\overline{gene\ x}_{MCF7}}{\overline{gene\ x}_{BT-474}}\right)_b} \quad (1)$$

3 Results

3.1 Down-scaling of Smart-seq2

Unlike other scRNA-seq workflows, single-cell isolation was performed with the F.SIGHT™ single-cell dispensation technology that uses a microfluidic chip generating free-flying, picoliter-sized droplets in which single cells are encapsulated and delivered to the microplate [35–37]. Image-based analysis intercepts a permanent vacuum suction when single cells with of the desired morphological criteria are detected in the nozzle. High precision is ensured by automatic dispenser offset compensation (AOC) enabling single-cell deposition into few hundred nanoliters in 384-well plates [44]. Simultaneously, the F.SIGHT™ records an image series for each dispensation event, which can be unambiguously assigned to the addressed well of the microplate [37]. Based on the images, the cells can be qualitatively stratified according to roundness and size from a heterogeneous population of particles (**Fig 1a**, only colored dots are cells, the grey dots are artifacts out of specification) resulting in a homogeneous cell population (**Fig 1a**, boxplots at the edges). As demonstrated previously, neural networks can be applied on these images to judge each single cell's integrity [58]. For MCF7 and BT-474 cell, the single-cell isolation efficiency was 93 % and 96 %, respectively (**Tab S1**). The cells were directly dispensed in to the lysis buffer and processed down-scaled Smart-seq2. Representative electropherograms of cDNA and tagmented cDNA are shown in **Fig 1b** and **1c**. The average fragment length for tagmented cDNA was 459 bp and 432 bp for MCF7 and BT-474 cells, respectively. FastQC analysis revealed an average Phred score of above 30 for both cell lines (**Fig S1a** to **S1d**). We analyzed sequencing data using three common aligners: salmon [48], kallisto [49] and STAR [50]. Based on alignment efficiency ≥ 80 % (salmon and kallisto) or fraction of uniquely mapped reads ≥ 60 % (STAR), cells of poor quality were excluded from downstream analysis (**Fig 1d** and **1e** and **Tab S1**). Further on, cells with less than 1E+5 reads were also excluded from analysis (**Fig 1f** and **Tab S1**). The median of genes is indistinguishable between aligners and cell lines. 11676 genes per single MCF7 cell and 11682 genes per single BT-474 cell were detected (**Fig 1g**). Furthermore, data was comparatively clustered by UMAP against external data from Isakova, Neff and Quake [4] to validate our workflow against published data. The cluster of MCF7 cells from our data exactly overlapped with published Smart-seq2 data of MCF7 cells, while other cells form independent clusters including the reference cell lines HEK293T and fibroblasts (**Fig 2a**). Upon different input to DESeq2 (salmon, kallisto and STAR aligner count matrices) we performed pseudo-bulk DE analysis (**Fig 2c**). Interestingly, the number of DEGs varies with the alignment tool used (**Fig 2d** and **Tab S2**). As expected [40], *ErbB2* expression was found to be significantly higher in BT-474, while ACTB as a housekeeping gene did not exhibit significant differential expression between cell lines (**Fig 2b**, **2c** and **S1e**). Additionally, we evaluated the expression of two marker genes for MCF7 cells, *KRT8*

329 and *TFF1* [4]. Both were not significantly overexpressed in MCF7 compared to BT-474,
330 presumably because of close relatedness (**Fig S1e** and **2a**).

3.2 Validation of scRT-ddPCR using bulk methods

For the integrated scRT-ddPCR, the cells were treated as in the case of the down-scaled Smart-seq2 using F.SIGHT™ (**Fig S3a** and **S3b**) and I.DOT for cell and liquid handling except that the cells were dispensed into a customized lysis buffer (LBTW) ensuring complete dispersion of cell material. After lysis, the gene mRNA per cell counts were determined directly form the lysate using digital PCR. First, we investigated varied volumes of LBTW lysis buffer, as the carry-over of detergents may impact droplet formation or reverse transcription and thus PCR efficiency [32,59]. The results indicate that as the volume of lysis buffer increases, the number of formed droplets decreases (**Fig 3a**) due increased areas of coalescence (**Fig S2a**). The use of optimal 0.5 µl LBTW produces no areas of coalescence and the number of droplets is not significantly reduced, despite a drop of ~18 % in total droplet number (**Fig 3a**). Similarly, we analyzed the effect of the lysis buffer on the measured absolute concentration of *ErbB2* and *ACTB* mRNAs. We could only detected a non-significant overestimation of mRNA concentrations (**Fig S2b**). Thus, we used 0.5 µl lysis buffer in subsequent experiments. Of note, at low target concentrations the subsampling error becomes significant [25] due to the loss of mRNAs in the non-partitioned (dead) volume, which is ~34 % according to manufacturer's information. We reduced to the loaded master mix volume (~18 % dead volume) without performance effects (data not shown) to minimize the loss of transcripts. The well-documented differential expression of *ErbB2* in MCF7 and BT-474 cells [40] was taken advantage to demonstrate the ability of our scRT-ddPCR for absolute quantification. We could observe an expression of 9 *ErbB2* mRNA/MCF7 cell (79 % CV), while BT-474 cells expressed a ~50-fold higher amount (453 *ErbB2* mRNA/BT-474 cell (42 % CV)) (**Fig 3b** and **Tab S3**). Absolute *ACTB* mRNA counts were found to be similar in MCF7 and BT-474 cells using our scRT-ddPCR method (MCF7: 66 *ACTB* mRNA/cell, 44 % CV; BT-474: 114 *ErbB2* mRNA/cell, 70 % CV; $p > 0.05$, Mann-Whitney test with Bonferroni correction; **Fig 3b** and **Tab S3**). We could not detect any correlation between cell size and number of mRNAs per singe cell (**Fig S3c** and **S3d**).To verify the ability of the lysis buffer to exert full dispersion of cell material prior to compartmentalization and full capture of all mRNAs by dPCR, we used unbiased bulk methods. Two established total bulk RNA isolation methods (assuming 100 % isolation efficiency) were used to determine the amount of *ErbB2* and *ACTB* mRNAs in bulk upon which the absolute amount of mRNAs per single cell were calculated. We could not detect quantification differences between the isolation methods despite obvious RNA integrity differences (**Fig S2c**, **S2d** and **Tab S5**). Similarly, a single-cell volume equivalent from a crude lysate ('cl') of $1x10^6$ MCF7 or BT-474 cell was dispensed into the dPCR mix and *ErbB2* and *ACTB* counts per single cell were determined to validate that the lysis conditions have no effect on dPCR or the detectability of the transcripts. Comparing these extraction methods ('sc', 'bulk', 'cl') for BT-474 cells yields no significant differences for both genes *ErbB2* or *ACTB* (**Fig 3b**). Also *ErbB2* counts in MCF7 cells are not significantly different

370 between extraction methods. However, *ACTB* counts from 'bulk' are significantly higher than
371 from 'sc' or 'cl' ($p < 0.001$, Mann-Whitney test with Bonferroni correction). Although, *ACTB*
372 is considered to be a housekeeping gene, its variability due to possible uncontrolled
373 conditions is already described [60]. In this given case we assume, the variability might be
374 related to the different passage numbers (**Tab S4**) or to a different confluency of the cell
375 culture. Overall, the results of absolute gene mRNA counts of total bulk RNA isolation
376 methods and similarly the results of the crude lysates confirm unbiased, quantitative transcript
377 detection by our scRT-ddPCR method (**Fig 3b**). The larger CVs of single-cell data compared
378 to 'bulk' and 'cl' (**Tab S3**) are expected and considered to recapitulate expression variability.

379

3.3 Comparison of Smart-seq2 and scRT-ddPCR

Conclusions made solely on the basis of scRNA-seq might be biased because of noise and dropouts and thus need confirmation by orthogonal PCR means [26]. Because of increased sensitivity, absolute quantification and higher tolerance towards inhibitors [25,33,38,39], we chose scRT-ddPCR for unbiased validation of scRNA-seq data. Assessing the normalized signal distributions upon the individual *ErbB2* expression values per singe MCF7 cell (based on TPM normalized values (salmon, kallisto), raw counts (STAR) or absolute gene mRNA counts per cell (RT-ddPCR); **Fig 2b** and **3b**), we found for scRNA-seq, independent of the alignment tool, that around 80 % of the cells accumulated in the first bin (**Fig 4a**) – a typical indication of a dropout event [13–15]. For the *ErbB2* signal distribution from MCF7 scRT-ddPCR data, we observed a significantly different shape (Kolmogorov-Smirnov test with Bonferroni multiple testing correction; **Tab S6**). However, 'bulk' and 'cl' signal distributions from *ErbB2* expression in MCF7 cells were also significantly different from scRT-ddPCR (**Tab S6**), which we assume to be consequence of the bulk approach. For high-abundant transcripts such as *ErbB2* in BT-474 cells, we could detect differences between the alignment tools especially between salmon and STAR and kallisto and STAR (**Tab S6**), which could reflect missing normalization of counts. And again, we see the difference between 'bulk-cl' and scRT-ddPCR. However, *ACTB* signal distributions show no such behavior with one exception (**Fig 4b** and **Tab S6**). Based on the expression values (**Fig 2b**) log2FCs were calculated for all conditions (**Fig 4c**). As expected, *ACTB* log2FCs from scRNA-seq fluctuate within 0 ± 1, which is the null hypothesis of DESeq2 [55] meaning there is no differential expression between the cell lines. These calculated log2FCs do not differ from the log2FCs of DESeq2 (blue and green arrows; **Fig 4c**) and could be entirely confirmed by scRT-ddPCR underlining the suitability of *ACTB* as a housekeeping gene in the light of fold change calculations. The fluctuation probably depicts statistical noise. Log2FCs from both, scRNA-seq and RT-ddPCR, confirm the overexpression of *ErbB2* in BT-474 cells, although to a different extent (**Fig 4c**). To compare log2FCs we bootstrapped their ratio (**Eq 1** and **Fig S4**) and calculated 95 % CIs. If these 95 % CI overlaps with 1, the methods are assumed to determine the same log2FC. Within the scRNA-seq or the RT-ddPCR group no significant differences could be observed, whereas the differences between the groups is significant. RT-ddPCR methods predict stronger overexpression of *ErbB2* in BT-474 compared to MCF7 (**Fig 4c**), which can potentially be explained by the biased detection of *ErbB2* expression in MCF7 cells with Smart-seq2 (**Fig 4a**).

4 Discussion

In this study we presented an integrated workflow combing down-scaled Smart-seq2 and single-cell reverse transcription digital PCR. We demonstrate the need for validation of scRNA-seq data by orthogonal scRT-ddPCR. The key feature of dPCR is the detection of absolute mRNA counts – a ground truth. Especially for low-abundant transcripts such as *ErbB2* in MCF7 cells, this is pivotal as we show that scRNA-seq could not reliably detect this gene (**Fig 4a** and **4c**). Down-scaling of scRNA-seq platforms is frequently applied to increase throughput, reduce costs and is assumed to increase sensitivity [22]. Here we showed that the novel combination of instruments delivers high quality data with the down-scaled Smart-seq2 protocol (**Fig 1**). Validation by comparative UMAP-clustering against published data revealed excellent conformity (**Fig 2a**). We anticipated from the dPCR concept [25] that down-scaling of Smart-seq2 could enhance molecular detections. However, down-scaling did not enable us to reliably detect the expression of *ErbB2* in MCF7 cells (**Fig 4a**) presumably because of simultaneous reverse transcription of multiple poly(A)-mRNAs into cDNAs. Compared to 3'-counting methods such as the Chromium system, full-length protocols like Smart-seq2 demonstrated better coverage of low-abundant transcripts [3,21], but are still not sensitive enough for the detection of *ErbB2* mRNA in MCF7 cells as we showed. We assume that our workflow is compatible with Smart-seq2-relatives such as FLASH-seq or Smart-seq3xpress [23,24]. Additionally, we successfully validated our scRT-ddPCR method by bulk methods (**Fig 3b**) underlining that scRT-ddPCR can serve as a ground truth for the validation of conclusions made on the basis of scRNA-seq data. We found that *ErbB2* could be detected reliably by scRT-ddPCR not suffering from dropout (**Fig 4a**) potentially because of the inherent partitioning step in dPCR, which increases the effective mRNA concentration. At sufficiently low sample concentration, no competing reactions bias cDNA synthesis. Currently, the amount of genes analyzed in by scRT-ddPCR is limited to the capability of the dPCR instrument. At the time we conducted the experiments, the highest degree of multiplexing was three colors [45]. Recent developments in dPCR instrumentation allow five (QIAcuity, Qiagen) or six (Prims6, Stilla Technlogies) color detection. However, approaches of monochrome multiplexing, such as photo bleaching, could double the degree of multiplexing [61,62]. Comparing quantitative results from scRNA-seq and scRT-ddPCR, we found that log2FCs for high-abundant transcripts are not different between methods, while we observed significant differences in *ErbB2* log2FCs between methods (**Fig 4c**). Because of dropouts, *ErbB2* log2FCs from scRNA-seq underestimate the fold differences between the cell lines, which could have severe implications for large-scale DE analyses and conclusions made on this basis. High-abundant transcripts can be detected in both methods of our workflow (**Fig 4b**). We sought to minimize biological and technical variability by using cells from the same population and the same high precision instrumentation for both methods regarding single-cell isolation and liquid handling. However, both methods use different priming strategies for cDNA synthesis [63] and in sequencing protocols more PCR steps are

453 required (cDNA amplification, tagmented library amplification, bridge amplification), which
454 potentiate biases and potentially favor dropouts. Bias in scRNA-seq becomes already obvious
455 when using different bioinformatic tools as we observed that salmon and kallisto aligner
456 predicted a significant overexpression of *OLFML3* and *RAMP3* in MCF7 cells, which was not
457 visible using STAR aligner (**Fig 2c** and **S1e**) and to our knowledge there is no supporting
458 evidence in literature. This underlines the necessity for an independent validation method that
459 allows the reliable detection of absolute mRNA counts such as scRT-ddPCR. To further
460 refine our integrated workflow, in a next iteration we could include external RNA samples
461 from the External RNA Controls Consortium (ERCC) in scRNA-seq and the scRT-ddPCR.
462 Furthermore, to reduce noise [6], we could think of using a protocol with unique molecular
463 identifiers (UMIs) such as Smart-seq3, Smart-seq3xpress or FLASH-seq [23,24,64], which
464 also allow molecular counting. Our here presented workflow can also serve as a platform for
465 assays assessing other entities of cells such as proteins [32].
466 In summary, we have established an integrated workflow combing transcriptome-wide gene
467 expression profiling by scRNA-seq and highly reliable detection of absolute gene mRNA
468 counts per single cell by scRT-ddPCR (ground truth method). Our workflow uses the same
469 instrumentation for single-cell isolation and liquid handling reducing technical variability.
470 Additionally, biological variability is reduced to the heterogeneity of the cell population itself
471 as single cells are isolated from the same bulk. Furthermore, we showed that potentially
472 through partitioning digital PCR enables the detection of low-abundant transcripts, which is
473 still an obstacle in scRNA-seq. In conclusion, we think this workflow is a valuable addition to
474 the toolbox of researchers interested in single-cell transcriptomics because of its reliability,
475 ease of use, reduced costs and increased sensitivity.

476

Author contribution statement

Conceptualization: Csaba Jeney, Tobias Lange
Investigation: Tobias Lange, Tobias Groß
Methodology: Tobias Lange, Julia Scherzinger, Elly Sinkala, Ábris Jeney, Csaba Jeney, Tobias Groß
Formal analysis: Tobias Lange, Ábris Jeney
Supervision and funding acquisition: Csaba Jeney, Stefan Zimmermann, Peter Koltay, Felix von Stetten, Roland Zengerle
Writing – original draft: Tobias Lange
Writing – review and editing: Csaba Jeney, Stefan Zimmermann, Peter Koltay, Felix von Stetten, Roland Zengerle, Christoph Niemöller, Tobias Lange, Tobias Groß, Ábris Jeney

All authors read and approved the final version of the manuscript.

491 Declaration of competing interest

492 J.S., E.S. and C.N. are employees of CYTENA GmbH, which produces the F.SIGHT™
493 single-cell dispenser used in this study. T.L., T.G., P.K, and C.J. are employees of Actome
494 GmbH, which develops the LBT lysis buffer used in this study and P.K., R.Z. and C.J. are
495 shareholders of Actome GmbH. The remaining authors declare no competing interest.

496

Data and code availability

scRNA-seq data is available from GEO under accession number: GSE201443. The code for bootstrapping comparison is available upon reasonable request.

Acknowledgments

We would like to thank the Actome GmbH for providing research licenses, for help with laboratory work and for advices on the manuscript. We would like to acknowledge the Lighthouse Core Facility of the Universitätsklinkium Freiburg for assistance with dPCR. We appreciate the RNA isolation kit gift form Zymo Research Europe (Freiburg). We acknowledge Martin Adam Stoffel (University of Edinburgh) for ideas on bootstrapping. This work was supported by grants from the Baden-Württemberg Stiftung in the framework of "Methoden in den Lebenswissenschaften" (MONOGRAM – MET-ID55) and the Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg (DINAMIK – 7533-7-11.10-6).

511 References

512 [1] H.C. Fan, G.K. Fu, S.P.A. Fodor, Combinatorial labeling of single cells for gene
513 expression cytometry, Science (80-.). 347 (2015).
514 https://doi.org/10.1126/science.1258367.

515 [2] G. Chen, B. Ning, T. Shi, Single-cell RNA-seq technologies and related computational
516 data analysis, Front. Genet. 10 (2019) 1–13. https://doi.org/10.3389/fgene.2019.00317.

517 [3] C. Ziegenhain, B. Vieth, S. Parekh, B. Reinius, A. Guillaumet-Adkins, M. Smets, H.
518 Leonhardt, H. Heyn, I. Hellmann, W. Enard, Comparative Analysis of Single-Cell
519 RNA Sequencing Methods, Mol. Cell. 65 (2017) 631-643.e4.
520 https://doi.org/10.1016/j.molcel.2017.01.023.

521 [4] A. Isakova, N. Neff, S.R. Quake, Single-cell quantification of a broad RNA spectrum
522 reveals unique noncoding patterns associated with cell types and states, Proc. Natl.
523 Acad. Sci. U. S. A. 118 (2021) 1–9. https://doi.org/10.1073/pnas.2113568118.

524 [5] S.M. Shaffer, M.C. Dunagin, S.R. Torborg, E.A. Torre, B. Emert, C. Krepler, M.
525 Beqiri, K. Sproesser, P.A. Brafford, M. Xiao, E. Eggan, I.N. Anastopoulos, C.A.
526 Vargas-Garcia, A. Singh, K.L. Nathanson, M. Herlyn, A. Raj, Rare cell variability and
527 drug-induced reprogramming as a mode of cancer drug resistance, Nature. 546 (2017)
528 431–435. https://doi.org/10.1038/nature22794.

529 [6] S. Islam, A. Zeisel, S. Joost, G. La Manno, P. Zajac, M. Kasper, P. Lönnerberg, S.
530 Linnarsson, Quantitative single-cell RNA-seq with unique molecular identifiers, Nat.
531 Methods. 11 (2014) 163–166. https://doi.org/10.1038/nmeth.2772.

532 [7] P. van Galen, V. Hovestadt, M.H. Wadsworth II, T.K. Hughes, G.K. Griffin, S.
533 Battaglia, J.A. Verga, J. Stephansky, T.J. Pastika, J. Lombardi Story, G.S. Pinkus, O.
534 Pozdnyakova, I. Galinsky, R.M. Stone, T.A. Graubert, A.K. Shalek, J.C. Aster, A.A.
535 Lane, B.E. Bernstein, Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to
536 Disease Progression and Immunity, Cell. 176 (2019) 1265-1281.e24.
537 https://doi.org/10.1016/j.cell.2019.01.031.

538 [8] K.T. Kim, H.W. Lee, H.O. Lee, S.C. Kim, Y.J. Seo, W. Chung, H.H. Eum, D.H. Nam,
539 J. Kim, K.M. Joo, W.Y. Park, Single-cell mRNA sequencing identifies subclonal
540 heterogeneity in anti-cancer drug responses of lung adenocarcinoma cells, Genome
541 Biol. 16 (2015) 1–15. https://doi.org/10.1186/s13059-015-0692-3.

542 [9] I. Tirosh, B. Izar, S.M. Prakadan, M.H. Wadsworth, D. Treacy, J.J. Trombetta, A.
543 Rotem, C. Rodman, C. Lian, G. Murphy, M. Fallahi-Sichani, K. Dutton-Regester, J.R.
544 Lin, O. Cohen, P. Shah, D. Lu, A.S. Genshaft, T.K. Hughes, C.G.K. Ziegler, S.W.
545 Kazer, A. Gaillard, K.E. Kolb, A.C. Villani, C.M. Johannessen, A.Y. Andreev, E.M.
546 Van Allen, M. Bertagnolli, P.K. Sorger, R.J. Sullivan, K.T. Flaherty, D.T. Frederick, J.
547 Jané-Valbuena, C.H. Yoon, O. Rozenblatt-Rosen, A.K. Shalek, A. Regev, L.A.
548 Garraway, Dissecting the multicellular ecosystem of metastatic melanoma by single-
549 cell RNA-seq, Science (80-.). 352 (2016) 189–196.
550 https://doi.org/10.1126/science.aad0501.

551 [10] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson,
552 M.W. Szcześniak, D.J. Gaffney, L.L. Elo, X. Zhang, A. Mortazavi, A survey of best
553 practices for RNA-seq data analysis, Genome Biol. 17 (2016) 1–19.
554 https://doi.org/10.1186/s13059-016-0881-8.

[11] S. Müller, S.J. Liu, E. Di Lullo, M. Malatesta, A.A. Pollen, T.J. Nowakowski, G. Kohanbash, M. Aghi, A.R. Kriegstein, D.A. Lim, A. Diaz, Single-cell sequencing maps gene expression to mutational phylogenies in PDGF - and EGF -driven gliomas , Mol. Syst. Biol. 12 (2016) 889. https://doi.org/10.15252/msb.20166969.

[12] D.T. Miyamoto, Y. Zheng, B.S. Wittner, R.J. Lee, H. Zhu, K.T. Broderick, R. Desai, D.B. Fox, B.W. Brannigan, J. Trautwein, K.S. Arora, N. Desai, D.M. Dahl, L. V. Sequist, M.R. Smith, R. Kapur, C.L. Wu, T. Shioda, S. Ramaswamy, D.T. Ting, M. Toner, S. Maheswaran, D.A. Haber, RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance, Science (80-.). 349 (2015) 1351–1356. https://doi.org/10.1126/science.aab0917.

[13] M.D. Luecken, F.J. Theis, Current best practices in single-cell RNA-seq analysis: a tutorial, Mol. Syst. Biol. 15 (2019). https://doi.org/10.15252/msb.20188746.

[14] A. Haque, J. Engel, S.A. Teichmann, T. Lönnberg, A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications, Genome Med. 9 (2017) 1–12. https://doi.org/10.1186/s13073-017-0467-4.

[15] L. Zhang, S. Zhang, Comparison of Computational Methods for Imputing Single-Cell RNA-Sequencing Data, IEEE/ACM Trans. Comput. Biol. Bioinforma. 17 (2020) 376–389. https://doi.org/10.1109/TCBB.2018.2848633.

[16] B. Vieth, S. Parekh, C. Ziegenhain, W. Enard, I. Hellmann, A systematic evaluation of single cell RNA-seq analysis pipelines, Nat. Commun. 10 (2019) 1–11. https://doi.org/10.1038/s41467-019-12266-7.

[17] L.A. Corchete, E.A. Rojas, D. Alonso-López, J. De Las Rivas, N.C. Gutiérrez, F.J. Burguillo, Systematic comparison and assessment of RNA-seq procedures for gene expression quantitative analysis, Sci. Rep. 10 (2020) 1–15. https://doi.org/10.1038/s41598-020-76881-x.

[18] S. Picelli, Full-length single-cell RNA sequencing with smart-seq2, Methods Mol. Biol. 1979 (2019) 25–44. https://doi.org/10.1007/978-1-4939-9240-9_3.

[19] S. Picelli, O.R. Faridani, Å.K. Björklund, G. Winberg, S. Sagasser, R. Sandberg, Full-length RNA-seq from single cells using Smart-seq2, Nat. Protoc. 9 (2014) 171–181. https://doi.org/10.1038/nprot.2014.006.

[20] G.X.Y. Zheng, J.M. Terry, P. Belgrader, P. Ryvkin, Z.W. Bent, R. Wilson, S.B. Ziraldo, T.D. Wheeler, G.P. McDermott, J. Zhu, M.T. Gregory, J. Shuga, L. Montesclaros, J.G. Underwood, D.A. Masquelier, S.Y. Nishimura, M. Schnall-Levin, P.W. Wyatt, C.M. Hindson, R. Bharadwaj, A. Wong, K.D. Ness, L.W. Beppu, H.J. Deeg, C. McFarland, K.R. Loeb, W.J. Valente, N.G. Ericson, E.A. Stevens, J.P. Radich, T.S. Mikkelsen, B.J. Hindson, J.H. Bielas, Massively parallel digital transcriptional profiling of single cells, Nat. Commun. 8 (2017). https://doi.org/10.1038/ncomms14049.

[21] X. Wang, Y. He, Q. Zhang, X. Ren, Z. Zhang, Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2, Genomics, Proteomics Bioinforma. 19 (2021) 253–266. https://doi.org/10.1016/j.gpb.2020.02.005.

[22] S. Mora-Castilla, C. To, S. Vaezeslami, R. Morey, S. Srinivasan, J.N. Dumdie, H. Cook-Andersen, J. Jenkins, L.C. Laurent, Miniaturization Technologies for Efficient Single-Cell Library Preparation for Next-Generation Sequencing, J. Lab. Autom. 21

599 (2016) 557–567. https://doi.org/10.1177/2211068216630741.

600 [23] V. Hahaut, D. Pavlinic, C. Cowan, S. Picelli, Lightning Fast and Highly Sensitive Full-
601 Length Single-cell sequencing using FLASH-Seq, BioRxiv. 4 (2021)
602 2021.07.14.452217.

603 [24] M. Hagemann-Jensen, C. Ziegenhain, R. Sandberg, Scalable full-transcript coverage
604 single cell RNA sequencing with Smart-seq3xpress, BioRxiv. (2021)
605 2021.07.10.451889.

606 [25] A.S. Basu, Digital Assays Part I: Partitioning Statistics and Digital PCR, SLAS
607 Technol. 22 (2017) 369–386. https://doi.org/10.1177/2472630317705680.

608 [26] D. Zucha, M. Kubista, L. Valihrach, Tutorial: Guidelines for Single-Cell RT-qPCR,
609 Cells. 10 (2021) 2607. https://doi.org/10.3390/cells10102607.

610 [27] M. Bengtsson, M. Hemberg, P. Rorsman, A. Ståhlberg, Quantification of mRNA in
611 single cells and modelling of RT-qPCR induced noise, BMC Mol. Biol. 9 (2008) 1–11.
612 https://doi.org/10.1186/1471-2199-9-63.

613 [28] S. Darmanis, C.J. Gallant, V.D. Marinescu, M. Niklasson, A. Segerman, G.
614 Flamourakis, S. Fredriksson, E. Assarsson, M. Lundberg, S. Nelander, B. Westermark,
615 U. Landegren, Simultaneous Multiplexed Measurement of RNA and Proteins in Single
616 Cells, Cell Rep. 14 (2016) 380–389. https://doi.org/10.1016/j.celrep.2015.12.021.

617 [29] C. Brion, S.M. Lutz, F.W. Albert, Simultaneous quantification of mrna and protein in
618 single cells reveals post-transcriptional effects of genetic variation, Elife. 9 (2020) 1–
619 34. https://doi.org/10.7554/eLife.60645.

620 [30] I. Kays, B.E. Chen, Protein and RNA quantification of multiple genes in single cells,
621 Biotechniques. 66 (2019) 15–21. https://doi.org/10.2144/btn-2018-0130.

622 [31] E. Jonasson, L. Andersson, S. Dolatabadi, S. Ghannoum, P. Åman, A. Ståhlberg, Total
623 mRNA Quantification in Single Cells: Sarcoma Cell Heterogeneity, Cells. 9 (2020).
624 https://doi.org/10.3390/cells9030759.

625 [32] C. Albayrak, C.A. Jordi, C. Zechner, J. Lin, C.A. Bichsel, M. Khammash, S. Tay,
626 Digital Quantification of Proteins and mRNA in Single Mammalian Cells, Mol. Cell.
627 61 (2016) 914–924. https://doi.org/10.1016/j.molcel.2016.02.030.

628 [33] J. Lin, C. Jordi, M. Son, H. Van Phan, N. Drayman, M.F. Abasiyanik, L. Vistain, H.L.
629 Tu, S. Tay, Ultra-sensitive digital quantification of proteins and mRNA in single cells,
630 Nat. Commun. 10 (2019) 1–10. https://doi.org/10.1038/s41467-019-11531-z.

631 [34] J. Ma, G. Tran, A.M.D. Wan, E.W.K. Young, E. Kumacheva, N.N. Iscove, P.W.
632 Zandstra, Microdroplet-based one-step RT-PCR for ultrahigh throughput single-cell
633 multiplex gene expression analysis and rare cell detection, Sci. Rep. 11 (2021) 1–18.
634 https://doi.org/10.1038/s41598-021-86087-4.

635 [35] A. Gross, J. Schoendube, S. Zimmermann, M. Steeb, R. Zengerle, P. Koltay,
636 Technologies for Single-Cell Isolation, Int. J. Mol. Sci. 16 (2015) 16897–16919.
637 https://doi.org/10.3390/ijms160816897.

638 [36] J. Riba, S. Zimmermann, P. Koltay, Technologies for Automated Single Cell Isolation,
639 Handb. Single Cell Technol. (2018) 1–28. https://doi.org/10.1007/978-981-10-4857-
640 9_9-1.

[37] A. Gross, J. Schöndube, S. Niekrawitz, W. Streule, L. Riegger, R. Zengerle, P. Koltay, Single-Cell Printer: Automated, On Demand, and Label Free, J. Lab. Autom. 18 (2013) 504–518. https://doi.org/10.1177/2211068213497204.

[38] M. Alikian, A.S. Whale, S. Akiki, K. Piechocki, C. Torrado, T. Myint, S. Cowen, M. Griffiths, A.G. Reid, J. Apperley, H. White, J.F. Huggett, L. Foroni, RT-qPCR and RT-digital PCR: A comparison of different platforms for the evaluation of residual disease in chronic myeloid leukemia, Clin. Chem. 63 (2017) 525–531. https://doi.org/10.1373/clinchem.2016.262824.

[39] S.C. Taylor, G. Laperriere, H. Germain, Droplet Digital PCR versus qPCR for gene expression analysis with low abundant targets: From variable nonsense to publication quality data, Sci. Rep. 7 (2017) 1–8. https://doi.org/10.1038/s41598-017-02217-x.

[40] F.C. Durst, A. Grujovic, I. Ganser, M. Hoffmann, P. Ugocsai, C.A. Klein, Z.T. Czyż, Targeted transcript quantification in single disseminated cancer cells after whole transcriptome amplification, PLoS One. 14 (2019) 1–21. https://doi.org/10.1371/journal.pone.0216442.

[41] A. Traube, T. Brode, Vorrichtung zur Aufnahme einer Flüssigkeit sowie Vorrichtung zur Aufbringung von Flüssigkeiten auf Probenträger und Verfahren hierzu, DE102007041071, 2010.

[42] M. Klinger, C. Laske, M. Graeve, M. Thoma, A. Traube, A Sensor for the In-Flight Detection of Single Fluorescent Microbodies in Nanoliter Droplets, IEEE Sens. J. 20 (2020) 5809–5817. https://doi.org/10.1109/JSEN.2020.2972268.

[43] T. Groß, C. Jeney, D. Halm, G. Finkenzeller, G.B. Stark, R. Zengerle, P. Koltay, S. Zimmermann, Characterization of CRISPR/Cas9 RANKL knockout mesenchymal stem cell clones based on single-cell printing technology and emulsion coupling assay as a low-cellularity workflow for single-cell cloning, BioRxiv. (2020) 1–19. https://doi.org/10.1101/2020.08.17.253559.

[44] J. Riba, N. Renz, C. Niemöller, S. Bleul, D. Pfeifer, J.M. Stosch, K.H. Metzeler, B. Hackanson, M. Lübbert, J. Duyster, P. Koltay, R. Zengerle, R. Claus, S. Zimmermann, H. Becker, Molecular genetic characterization of individual cancer cells isolated via single-cell printing, PLoS One. 11 (2016) 1–15. https://doi.org/10.1371/journal.pone.0163455.

[45] J. Madic, A. Zocevic, V. Senlis, E. Fradet, B. Andre, S. Muller, R. Dangla, M.E. Droniou, Three-color crystal digital PCR, Biomol. Detect. Quantif. 10 (2016) 34–46. https://doi.org/10.1016/j.bdq.2016.10.002.

[46] A.S. Whale, W. De Spiegelaere, W. Trypsteen, A.A. Nour, Y.-K. Bae, V. Benes, D. Burke, M. Cleveland, P. Corbisier, A.S. Devonshire, L. Dong, D. Drandi, C.A. Foy, J.A. Garson, H.-J. He, J. Hellemans, M. Kubista, A. Lievens, M.G. Makrigiorgos, M. Milavec, R.D. Mueller, T. Nolan, D.M. O'Sullivan, M.W. Pfaffl, S. Rödiger, E.L. Romsos, G.L. Shipley, V. Taly, A. Untergasser, C.T. Wittwer, S.A. Bustin, J. Vandesompele, J.F. Huggett, The Digital MIQE Guidelines Update: Minimum Information for Publication of Quantitative Digital PCR Experiments for 2020, Clin. Chem. 66 (2020) 1012–1029. https://doi.org/10.1093/clinchem/hvaa125.

[47] J.F. Huggett, C.A. Foy, V. Benes, K. Emslie, J.A. Garson, R. Haynes, J. Hellemans, M. Kubista, R.D. Mueller, T. Nolan, M.W. Pfaffl, G.L. Shipley, J. Vandesompele, C.T. Wittwer, S.A. Bustin, The digital MIQE guidelines: Minimum information for

686 publication of quantitative digital PCR experiments, Clin. Chem. 59 (2013) 892–902.
687 https://doi.org/10.1373/clinchem.2013.206375.

688 [48] R. Patro, G. Duggal, M.I. Love, R.A. Irizarry, C. Kingsford, Salmon provides fast and
689 bias-aware quantification of transcript expression, Nat. Methods. 14 (2017) 417–419.
690 https://doi.org/10.1038/nmeth.4197.

691 [49] N.L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq
692 quantification, Nat. Biotechnol. 34 (2016) 525–527. https://doi.org/10.1038/nbt.3519.

693 [50] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M.
694 Chaisson, T.R. Gingeras, STAR: Ultrafast universal RNA-seq aligner, Bioinformatics.
695 29 (2013) 15–21. https://doi.org/10.1093/bioinformatics/bts635.

696 [51] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K.
697 Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing,
698 Jupyter Notebooks—a publishing format for reproducible computational workflows,
699 Position. Power Acad. Publ. Play. Agents Agendas - Proc. 20th Int. Conf. Electron.
700 Publ. ELPUB 2016. (2016) 87–90. https://doi.org/10.3233/978-1-61499-649-1-87.

701 [52] M.I. Love, C. Soneson, P.F. Hickey, L.K. Johnson, N. Tessa Pierce, L. Shepherd, M.
702 Morgan, R. Patro, Tximeta: Reference sequence checksums for provenance
703 identification in RNA-seq, PLoS Comput. Biol. 16 (2020) 1–13.
704 https://doi.org/10.1371/journal.pcbi.1007664.

705 [53] C. Soneson, M.I. Love, M.D. Robinson, Differential analyses for RNA-seq: Transcript-
706 level estimates improve gene-level inferences, F1000Research. 4 (2016) 1–23.
707 https://doi.org/10.12688/F1000RESEARCH.7563.2.

708 [54] Y. Liao, G.K. Smyth, W. Shi, FeatureCounts: An efficient general purpose program for
709 assigning sequence reads to genomic features, Bioinformatics. 30 (2014) 923–930.
710 https://doi.org/10.1093/bioinformatics/btt656.

711 [55] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion
712 for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 1–21.
713 https://doi.org/10.1186/s13059-014-0550-8.

714 [56] K. Polański, M.D. Young, Z. Miao, K.B. Meyer, S.A. Teichmann, J.E. Park, BBKNN:
715 Fast batch alignment of single cell transcriptomes, Bioinformatics. 36 (2020) 964–965.
716 https://doi.org/10.1093/bioinformatics/btz625.

717 [57] F.A. Wolf, P. Angerer, F.J. Theis, SCANPY: large-scale single-cell gene expression
718 data analysis, Genome Biol. 19 (2018) 15. https://doi.org/10.1186/s13059-017-1382-0.

719 [58] J. Riba, J. Schoendube, S. Zimmermann, P. Koltay, R. Zengerle, Single-cell dispensing
720 and 'real-time' cell classification using convolutional neural networks for higher
721 efficiency in single-cell cloning, Sci. Rep. 10 (2020) 1–9.
722 https://doi.org/10.1038/s41598-020-57900-3.

723 [59] I. Smyrlaki, M. Ekman, A. Lentini, N. Rufino de Sousa, N. Papanicolaou, M.
724 Vondracek, J. Aarum, H. Safari, S. Muradrasoli, A.G. Rothfuchs, J. Albert, B.
725 Högberg, B. Reinius, Massive and rapid COVID-19 testing is feasible by extraction-
726 free SARS-CoV-2 RT-PCR, Nat. Commun. 11 (2020) 1–12.
727 https://doi.org/10.1038/s41467-020-18611-5.

728 [60] B. Kozera, M. Rapacz, Reference genes in real-time PCR, J. Appl. Genet. 54 (2013)

729 391–406. https://doi.org/10.1007/s13353-013-0173-x.

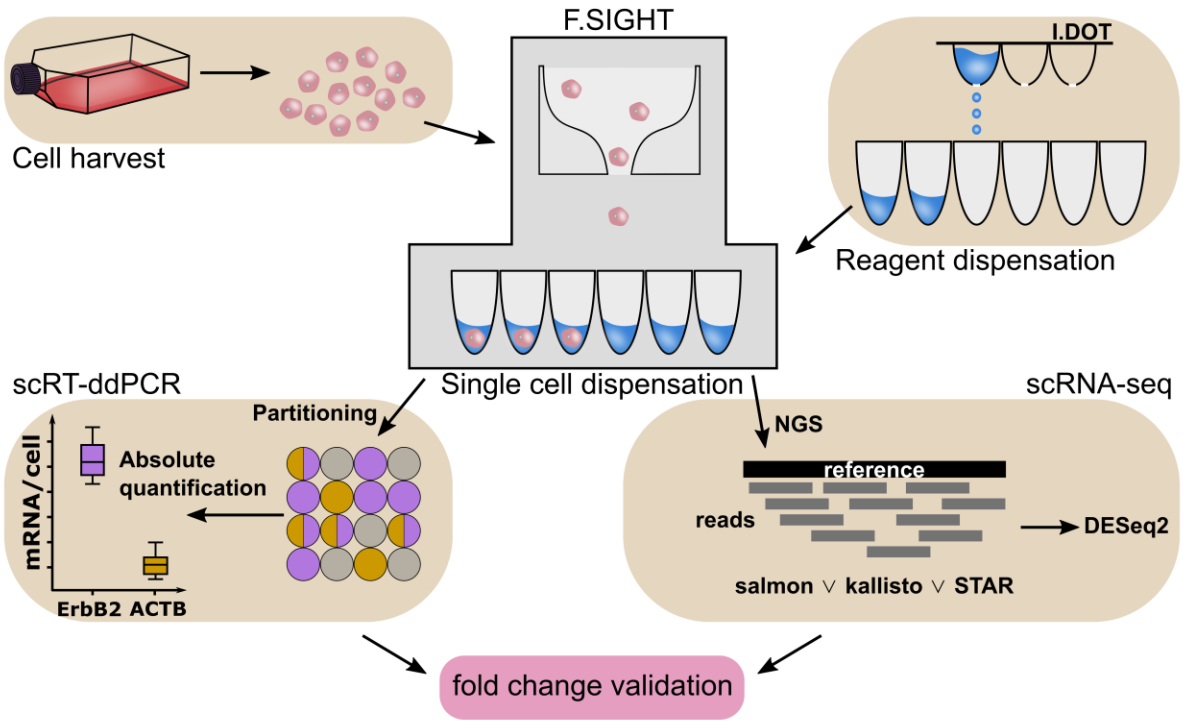
730 [61] F. Schuler, M. Trotter, R. Zengerle, F. Von Stetten, Monochrome Multiplexing in
731 Polymerase Chain Reaction by Photobleaching of Fluorogenic Hydrolysis Probes,
732 Anal. Chem. 88 (2016) 2590–2595. https://doi.org/10.1021/acs.analchem.5b02960.

733 [62] F. Schlenker, E. Kipf, N. Borst, T. Hutzenlaub, R. Zengerle, F. Von Stetten, P. Juelg,
734 Virtual Fluorescence Color Channels by Selective Photobleaching in Digital PCR
735 Applied to the Quantification of KRAS Point Mutations, Anal. Chem. 93 (2021)
736 10538–10545. https://doi.org/10.1021/acs.analchem.1c01488.

737 [63] A. Ståhlberg, J. Håkansson, X. Xian, H. Semb, M. Kubista, Properties of the Reverse
738 Transcription Reaction in mRNA Quantification, Clin. Chem. 50 (2004) 509–515.
739 https://doi.org/10.1373/clinchem.2003.026161.

740 [64] M. Hagemann-Jensen, C. Ziegenhain, P. Chen, D. Ramsköld, G.J. Hendriks, A.J.M.
741 Larsson, O.R. Faridani, R. Sandberg, Single-cell RNA counting at allele and isoform
742 resolution using Smart-seq3, Nat. Biotechnol. 38 (2020) 708–714.
743 https://doi.org/10.1038/s41587-020-0497-0.

744



Graphical abstract: Integrated workflow combining scRT-ddPCR and scRNA-seq with subsequent validation of fold changes.

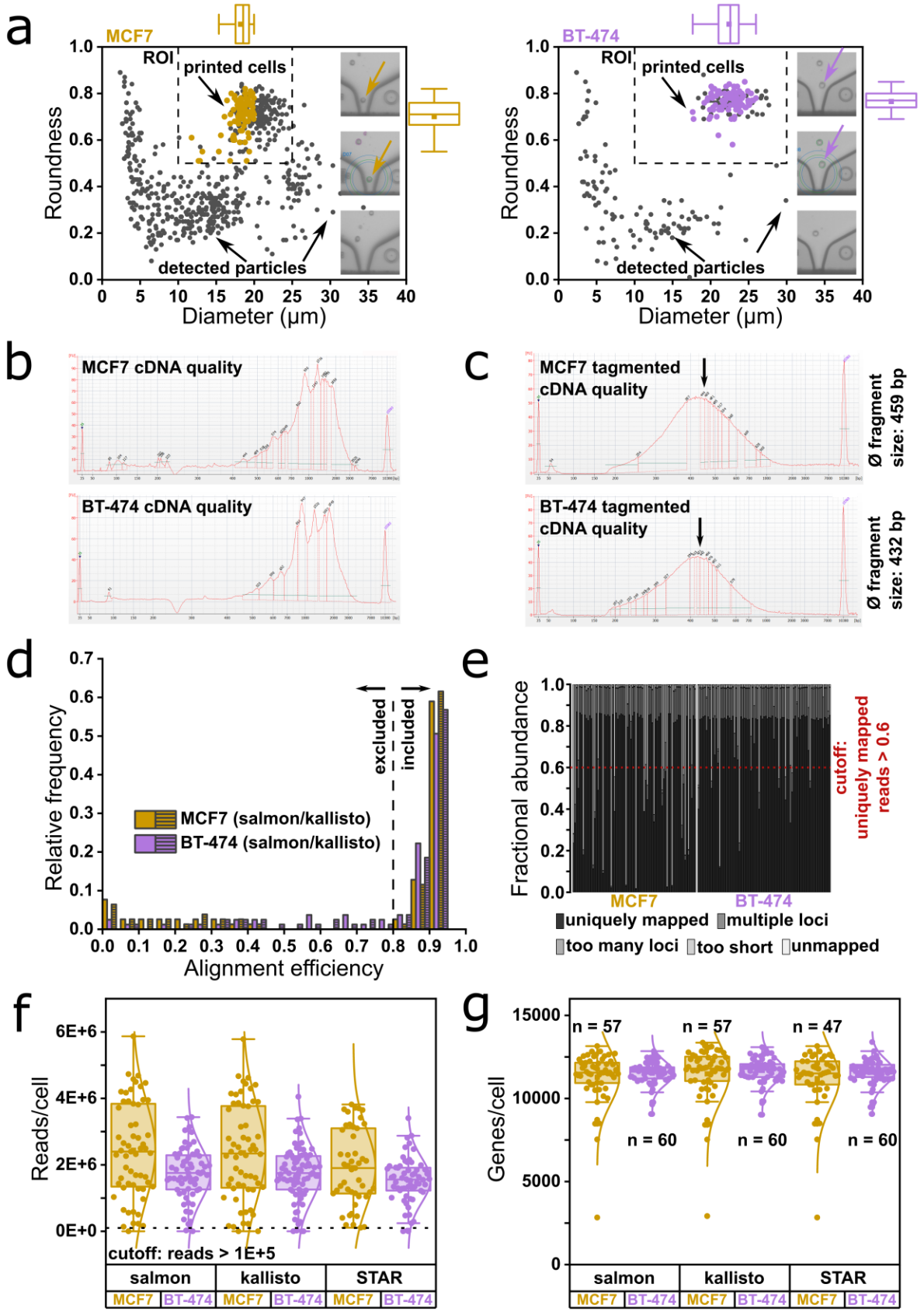


Figure 1: Quality control of down-scaled Smart-seq2 workflow with MCF7 and BT-474 cells. a) Scatter plots (roundness vs. diameter) of printed MCF7 and BT-474 cells (colored dots) or detected particles (grey dots). Boxplots show roundness and diameter distributions of printed cells (n = 84). ROI (region of interest) depicts the defined morphological criteria. Representative images of the printing process enabling image-based exclusion of multiple cells per droplet or empty droplets are shown. These images can be unambiguously assigned to the

addressed wells of the microplate. **b, c)** Representative electropherograms (Agilent's Bioanalyzer) of cDNA and tagmented cDNA size distributions for both cell lines. The average cDNA length after tagmentation was 459 bp for MCF7 and 432 bp for BT-474 cells. **d)** Alignment efficiency of salmon and kallisto aligner. Cells with less than 80 % alignment efficiency were excluded from further analyses. **e)** Alignment statistics (fraction of uniquely mapped reads, fraction of reads mapped to multiple loci, fraction of reads mapped to too many loci, fraction of reads too short for mapping, fraction of unmapped reads) for MCF7 and BT-474 cells using STAR aligner. Cells with less than 60 % of uniquely mapped reads were excluded from further analyses. **f)** Total read counts per cell for MCF7 and BT-474 cells using salmon, kallisto or STAR aligner. Cells with less than 1E+5 transcripts were excluded from further analyses. **g)** Gene per cell counts for MCF7 cells (median across all aligners: 11676 genes per cell) and BT-474 cells (median across all aligners: 11682 genes per cell) after all steps of filtering, listed by the used aligners. The median of genes per cell is the same independent of aligner and cell line ($p > 0.05$, Mann-Whitney test with Bonferroni correction). The amount of cells excluded after each filtering step is shown in **Tab S1**.

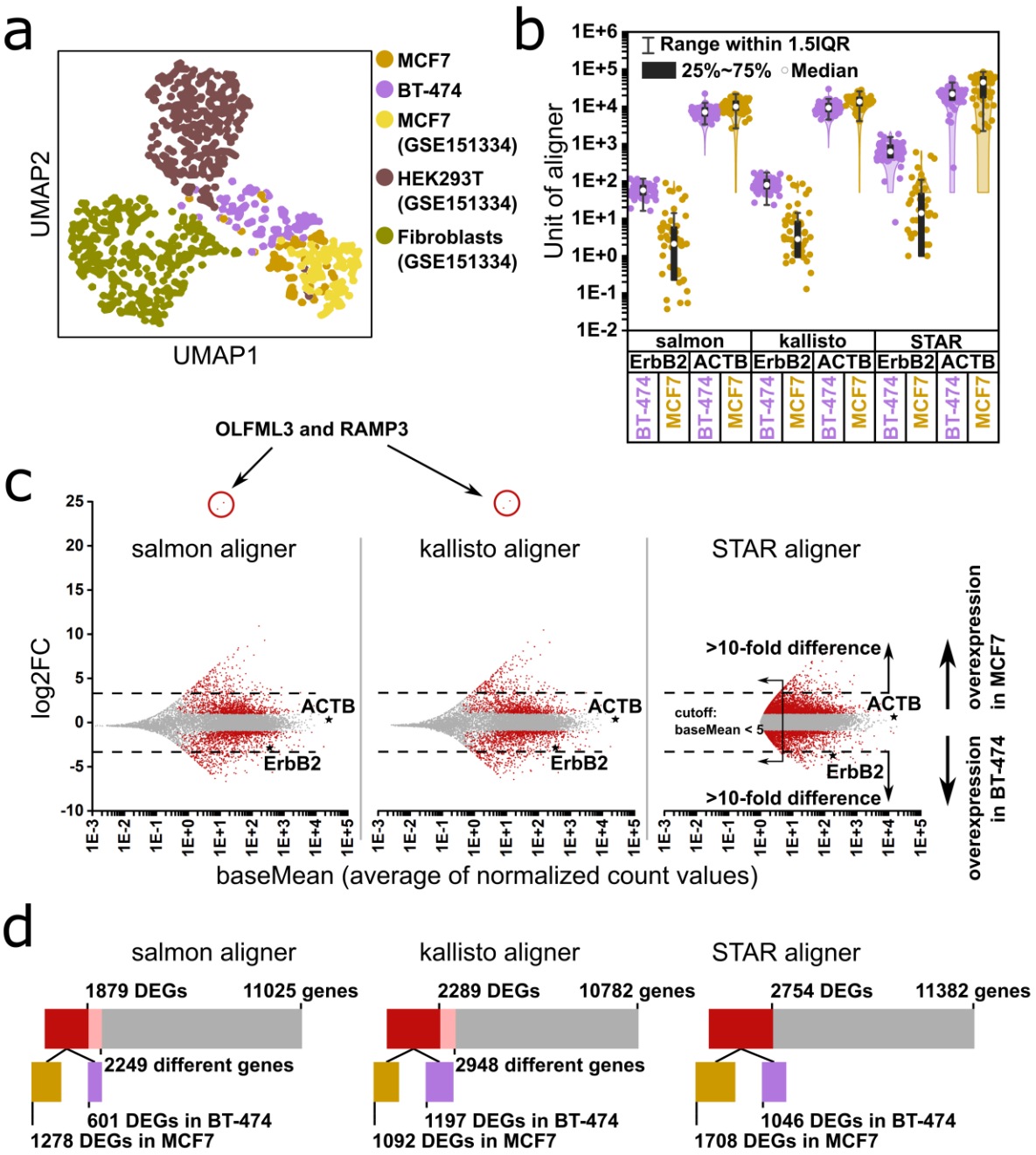


Figure 2: Validation of down-scaled Smart-seq2 by clustering and different bioinformatics pipelines. a) UMAP clustering of MCF7 and BT-474 cells along with MCF7 cells, HEK293T cells and fibroblasts from GSE151334. **b)** Violin plots of *ErbB2* and *ACTB* expression values in MCF7 and BT-474 cells from salmon, kallisto and STAR aligner in the respective units (salmon and kallisto: TPM: transcripts per kilobase million; STAR: raw counts). **c)** Bland-Altman plots of gene expression in MCF7 over BT-474 cells with salmon, kallisto or STAR aligner input to DESeq2. Each dot symbolizes a gene with its average expression value in both cell lines (baseMean; x-axis) and the log2FC (log2 of fold change; y-axis). Dots colored in red are significantly differentially expressed genes (DEGs); log2FC > 1 and $p_{adj} < 0.05$: overexpression in MCF7 cells; log2FC < -1 and $p_{adj} < 0.05$: overexpression in BT-474 cells. *ErbB2* and *ACTB* expression values are highlighted with stars. The dashed line indicates 10-fold overexpression in either cell line. **d)** Total number of genes analyzed in DESeq2 with the number of different genes (only log2FC > 1 or log2FC < -1) and the number of DEGs overexpressed in either cell line (see also **Tab S2**).

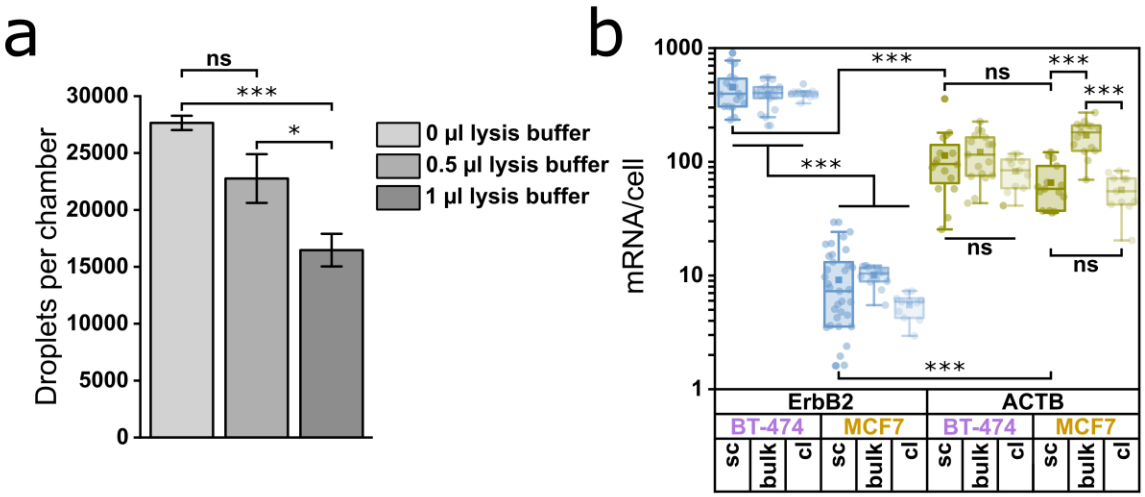


Figure 3: Validation of scRT-ddPCR workflow. a) Impact of lysis buffer volume on the number of droplets generated per reaction chamber of one Sapphire Chip. Bar plots show mean values with standard deviation as error bars. Groups were compared using student's t-test with Bonferroni correction (n = 3). **b)** Absolute gene mRNA per cell counts from different methods ('sc' = scRT-ddPCR; 'bulk' = quantification from bulk isolated RNA; 'cl' = quantification from a crude lysate) according to the genes *ErbB2* and *ACTB* and the cell lines BT-474 and MCF7. Groups were compared using Mann-Whitney test with Bonferroni correction (n ≥ 11). Significance levels not indicated are non-relevant comparisons for this work.

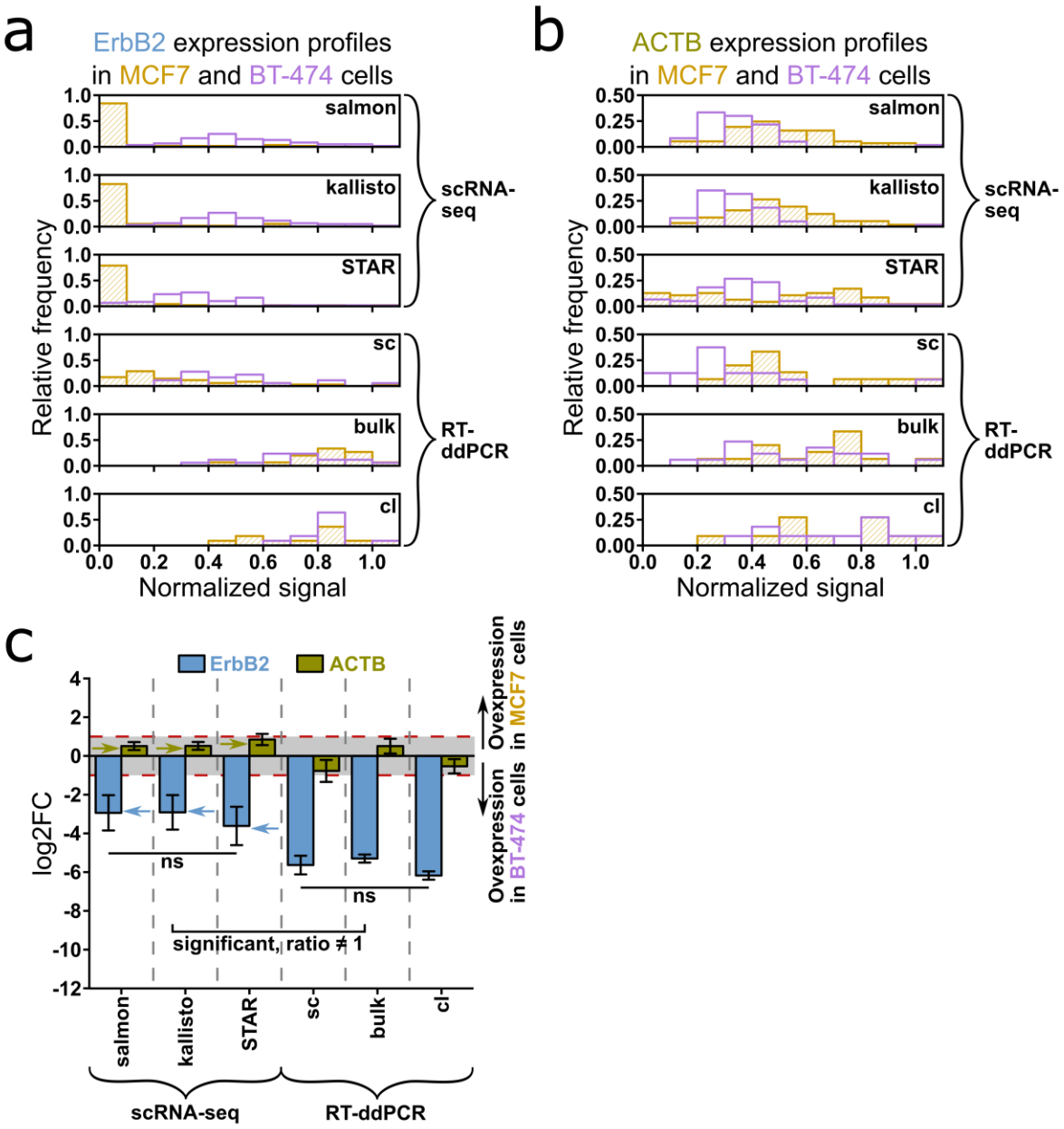


Figure 4: Direct comparison of scRT-ddPCR and scRNA-seq. a) Distribution of normalized *ErbB2* expression signal (normalized to maximum signal) from the three aligners and sc/bulk/clRT-ddPCR in MCF7 and BT-474 cells. The distributions were compared using a Kolmogorov-Smirnov test with Bonferroni correction for multiple testing and the significance levels can be obtained from **Tab S3**. **b)** Distribution of normalized *ACTB* expression signal (normalized to maximum signal) from the three aligners and sc-/bulk-/clRT-ddPCR in MCF7 and BT-474 cells. The distributions were compared using a Kolmogorov-Smirnov test with Bonferroni correction for multiple testing and the significance levels can be obtained from **Tab S3**. **c)** Log2FCs (MCF7 vs. BT-474) for all methods used. The *ACTB* log2FCs fall within the range of 0 ± 1. *ErbB2* log2FCs indicate that *ErbB2* has a higher expression in BT-474 cells but to a different extend between scRNA-seq and RT-ddPCR. The log2FCs were compared using a bootstrap ratio comparison (2.8 Bootstrapping comparison). Arrows highlight the log2FCs calculated by DESeq2. Bars depict mean log2FCs with bootstrapped error bars indicating 95 % CI.