RESEARCH ARTICLE

Targeted transcript quantification in single disseminated cancer cells after whole transcriptome amplification

Franziska C. Durst[1]☯‡, Ana Grujovic[1]☯‡, Iris Ganser[1], Martin Hoffmann[2], Peter Ugocsai[3], Christoph A. Klein[1,2‡]*, Zbigniew T. Czyż[2‡]

1 Experimental Medicine and Therapy Research, University of Regensburg, Regensburg, Germany, **2** Fraunhofer Institute for Toxicology and Experimental Medicine, Project Group Personalized Tumor Therapy, Regensburg, Germany, **3** Caritas Hospital St. Josef, Department of Obstetrics and Gynecology, Regensburg, Germany

☯ These authors contributed equally to this work.
‡ FCD and AG are joint first authors on this work. ZTC and CAK contributed equally as joint senior authors on this work.
* Christoph.Klein@ukr.de

Abstract

Gene expression analysis of rare or heterogeneous cell populations such as disseminated cancer cells (DCCs) requires a sensitive method allowing reliable analysis of single cells. Therefore, we developed and explored the feasibility of a quantitative PCR (qPCR) assay to analyze single-cell cDNA pre-amplified using a previously established whole transcriptome amplification (WTA) protocol. We carefully selected and optimized multiple steps of the protocol, e.g. re-amplification of WTA products, quantification of amplified cDNA yields and final qPCR quantification, to identify the most reliable and accurate workflow for quantitation of gene expression of the *ERBB2* gene in DCCs. We found that absolute quantification outperforms relative quantification. We then validated the performance of our method on single cells of established breast cancer cell lines displaying distinct levels of HER2 protein. The different protein levels were faithfully reflected by transcript expression across the tested cell lines thereby proving the accuracy of our approach. Finally, we applied our method to breast cancer DCCs of a patient undergoing anti-HER2-directed therapy. Here, we were able to measure *ERBB2* expression levels in all HER2-protein-positive DCCs. In summary, we developed a reliable single-cell qPCR assay applicable to measure distinct levels of *ERBB2* in DCCs.

Introduction

The analysis of systemically spread cancer via detection of disseminated cancer cells (DCCs) or circulating tumor cells (CTCs) in distant organs or blood, respectively, faces several technical challenges. First, the frequency of DCCs or CTCs is very low, e.g. ~two DCCs and ~one CTC can be found among 10^6 nucleated cells in bone marrow and peripheral blood, respectively [1, 2], in breast cancer depending on the clinical stage. Second, micrometastatic cancer

cells exhibit phenotypical and genetic heterogeneity affecting their malignant potential and susceptibility to therapy [3]. Therefore, the analysis of metastasis necessitates highly reliable methods enabling the investigation of single cells specifically at its early stages. Single-cell transcriptomes underlie dynamic changes that reflect functional and differentiation processes occurring in individual cells. Therefore, the analysis of individual cell transcriptomes provides a first insight into cell functions relevant for disease progression or therapy resistance.

A single cell is calculated to contain 1 pg of mRNA comprising transcripts expressed over several orders of magnitude, with the majority of genes being represented by less than 100 mRNA copies per cell [4]. For the accurate assessment of heterogeneity among single cells, the applied workflows have to fulfill several specific requirements. First, a method dedicated for the analysis of rare and unique cells should optimally provide sufficient amount of material to run all required downstream analyses. Second, the amplification of single-cell mRNA must be as accurate and comprehensive as possible to essentially preserve the qualitative and quantitative complexity of the sample. False-negative (technical drop-outs) and false-positive results should be reduced to a minimum. Third, an optimal workflow should be highly sensitive allowing the detection of genes expressed at low levels. Various single-cell whole transcriptome amplification (WTA) methods have been developed [4, 5] permitting different types of downstream analyses utilizing qPCR [6–8], microarrays [9–11] or next generation sequencing (NGS) [12–15] as read-outs. Each of the available WTA technologies displays unique strengths and weaknesses reflected by differences in the detection sensitivity [13, 14]. Notably, available WTA strategies do not always provide an overview over the full-length transcripts of a single cell but instead are biased towards specific loci [7, 16] or the ends of the RNA molecules [17–19] limiting downstream analyses to either quantification of selected genes only (targeted analysis) or counting of mRNA molecules, thus limiting the ability to assess transcript diversity or mutational state analysis. The selection of an optimal technique for gene expression analysis in single cells depends on the specific research question, number of samples and genes to be analyzed, as well as the required sensitivity and costs. In the last years, powerful technological advances have been made in the area of single-cell RNA sequencing (scRNA-Seq), resulting in the development of technologies allowing the quantification of gene expression levels in single cells at much higher throughput [20]. Progress in NGS and droplet microfluidics resulted in the development of methods enabling the analysis of hundreds to thousands of individual cells in a parallelized fashion [15] and the detection of thousands of transcripts in a single experiment [21–23]. However, high-throughput methodologies exhibit relative low mRNA capture rates (seizing only ~10–15% of transcripts expressed by each cell [18, 24]) that render them unsuitable for comprehensive analysis of rare cells. Moreover, NGS-based technologies require the utilization of expensive hardware and generate high running costs. Importantly, scRNA-Seq workflows necessitate the use of advanced bioinformatic workflows and scripts to analyze the output data. Thus, users of such methods must have the knowledge and experience how to use (and in some cases establish) bioinformatic pipelines. In contrast, data analysis for qPCR such as relative and absolute quantification methods [25, 26], are well established, standardized and do not require a high level of expertise [27]. Moreover, protocols for preparing qPCR samples are simpler and result in higher sensitivity and reproducibility as compared to NGS-based approaches [7, 27]. Importantly, single-cell qPCR workflows exhibit high levels of reliability and wide and dynamic detection ranges, making them exceptionally well-suited for targeted gene expression analyses in single cells, where sensitivity is essential and the amount of target genes is low [7, 27]. Therefore, the present study aimed to develop a single-cell qPCR assay to quantify gene expression changes in single cells, specifically in patient-derived DCCs. We established a workflow comprised of single-cell WTA, re-amplification of single-cell cDNA, post-WTA normalization of cDNA quantities and qPCR-based data analysis. The new assay

provides means for measuring expression levels of individual pre-selected genes in WTA products generated from single cells in an accurate and reliable fashion.

Materials and methods

Cell lines

BT-474 (ACC 64) and MCF-7 (ACC 115) breast cancer cell lines were obtained from the German Collection of Microorganisms and Cell Cultures (DSMZ). MCF-10A (CRL-10317), a non-tumorigenic mammary epithelial cell line, was obtained from the American Type Culture Collection (ATCC). ZR-75-1 (CRL1500, ATCC) and MDA-MB-453 (ACC 65, DSMZ) cells were purchased from DSMZ. The identity of all cell lines was confirmed by DNA finger printing analysis utilizing the GenePrint 10 System (Promega). BT-474 and MDA-MB-453 cells were cultivated in DMEM medium (Pan-Biotech) supplemented with 10% FCS (Sigma-Aldrich), 2 mM L-Glutamine (Pan-Biotech) and 1% Penicillin/Streptomycin (Pan-Biotech). ZR-75-1 and MCF-7 cells were propagated in RPMI 1640 medium (Pan-Biotech) supplemented with 10% FCS, 2 mM L-Glutamine and 1% Penicillin/Streptomycin. In addition to other components, the medium for MCF-7 cells contained 1 mM Sodium-Pyruvate (Sigma-Aldrich). MCF-10A cells were cultured in DMEM/F12 medium (Pan-Biotech) supplemented with 5% Horse serum (Sigma-Aldrich), 1% Penicillin/Streptomycin, 20 ng/ml EGF (Sigma-Aldrich), 0.5 μg/ml Hydrocortisone (Sigma-Aldrich), 10 μg/ml Insulin (Sigma-Aldrich) and 0.1 μg/ml Cholera toxin (Sigma-Aldrich).

Flow cytometry

Cell line cells were harvested using 0.05% Trypsin/EDTA (Pan-Biotech) for 3 min and stained using a mouse anti-human HER2 antibody conjugated to FITC (20 μg/ml, clone 24D2, BioLegend). Cells were incubated with anti-HER2 antibody for 20 min at 4˚C in the dark and washed once with PBS supplemented with 1% FCS prior to the FACS analysis. A corresponding isotype control (Mouse IgG1, κ-FITC labeled, eBioscience) was used in every staining experiment to determine the background level of fluorescence and set the threshold for specific staining signals. The cells were analyzed on a BD FACS Canto II instrument (BD Bioscience) equipped with FACS DIVA v7.0 software (BD Bioscience). The sorting of single cells was performed with a FACSAria cell sorter (BD Bioscience). FlowJo v10 (Treestar) software was used for the analysis of the obtained data.

Patient sample

A pleural effusion from a patient with HER2-positive metastatic breast cancer was obtained from the Caritas Hospital St. Josef in Regensburg. The ethics committees of the University of Regensburg approved the sampling and genetic analysis of the isolated cells (ethics vote number 17-672-101). The donor provided written informed consent for the processing of the clinical material. The patient received her regular dose of trastuzumab / pertuzumab early in the morning and the pleural effusion was sampled in the late afternoon. The sample was centrifuged for 10 min at 300×g and the cells were washed once using Hank's solution (Biochrom). $2×10^6$ cells were double-stained using anti-EpCAM (1:50, clone HEA-125, Miltenyi) and anti-HER2 antibodies directly conjugated to PE and FITC, respectively. After 15 min incubation at 4˚C on a roller mixer in the dark, the cells were washed once with 1× PBS. EpCAM staining was used to detect DCCs. The stained sample was screened on an Olympus IX-81 inverted fluorescent microscope (Olympus) for the presence of marker positive cells. Single cells were manually isolated with a micromanipulator (Eppendorf, PatchMan NP2) as previously

described [9]. The isolated cells were picked in 1 μl of 1× PBS and transferred into 4.4 μl of lysing buffer containing 4 μl mTRAP™ Lysis Buffer (Active Motif) and 0.4 μl (10 ng) tRNA from E. coli MRE 600 (Roche). The picked cells suspended in the lysis were stored at -80˚C until further processing, i.e. WTA.

Whole Transcriptome Amplification (WTA) of single cells and quality control

Single cells were subjected to WTA using a previously described protocol [9]. Lot numbers of the TdT enzyme used in a period of work on this publication are following: 4217219, 4202212, 4216713, 4222131, 4236381, 4241072, 4247609, 4227719, 4263188, 4270799, 4256884, 4304640, 4307658, 4321970, 4324377, 4344851 and 4324377. Primer sequences for WTA and re-amplification are provided in S2 Table. The quality of isolated primary cDNA amplification products and re-amplified WTA products was examined utilizing endpoint PCR controlling for the expression of three housekeeping genes: *EEF1A1*, *ACTB* and *GAPDH*. Each PCR reaction was composed of the following ingredients: 1 μl of the 10× FastStart PCR Buffer comprising 20 mM MgCl$_2$ (Roche Diagnostics), 0.2 μl dNTPs (10 mM each), 1 μl of the primer-mix (8 μM of each forward and reverse primer for each tested gene–*EEF1A1*, *ACTB* and *GAPDH*; S1 Table), 0.2 μl BSA (20 mg/ml, Roche Diagnostics), 6.5 μl HPLC Gradient Grade H$_2$O, 0.1 μl FastStart Taq Polymerase (5 U/μl, Roche Diagnostics) and 1 μl of the fivefold diluted primary WTA product. The PCR was performed on a MJ Research Peltier Thermal Cycler Tetrad (Bio-Rad) using the following program: initial denaturation at 95˚C for 4 min was followed by 32 cycles of 30 s at 95˚C, 30 s at 58˚C, 90 s at 72˚C and a final elongation step of 7 min at 72˚C. Positive and negative controls were included in every PCR run.

Gene expression analysis using endpoint PCR

Endpoint PCR was conducted to verify the presence of selected genes of interest (S1 Table) in WTA products. The primary or re-amplified WTA product was diluted five times with water and used as a template in each test PCR (1 μl per sample). The templates were mixed with 1 μl of the 10× PCR buffer containing PCR Grade dNTP Mix (10 mM each), 0.5 μl of the forward and reverse primer (8 μM each; S1 Table), 0.25 μl BSA (10 mg/ml, Roche Diagnostics), 0.1 μl PANScript DNA Polymerase (5 U/μl, Pan Biotech) and 6.65 μl HPLC Grade H$_2$O to a final volume of 10 μl. The PCR was run using the following program: initial denaturation of 4 min at 95˚C was followed by 42 cycles of 30 s at 95˚C, 30 s at 58˚C, 90 s at 72˚C and a final elongation step for 7 min at 72˚C. Positive and negative control was included in every PCR run.

Re-amplification of primary WTA single-cell products

Re-amplification of primary WTA products was performed in a reaction volume of 50 μl comprising 5 μl Expand Long Template Buffer 1 (Expand Long Template PCR System, Roche Diagnostics), 6 μl of the CP2-15C or CP2-9C primer (2.88 μM; S2 Table), 1.75 μl dNTPs (10 mM each), 7.5 μl 20% Formamide (Sigma Aldrich), 1.5 μl DNA Pol Mix (5 U/μl, Expand Long Template PCR System, Roche Diagnostics), 27.25 μl PCR HPLC Gradient Grade H$_2$O and 1 μl template (primary WTA product). The re-amplification was run on a PTC DNA Engine 2 Tetrad Thermocycler using the following program: 1 min at 95˚C, 5 cycles comprising 15 s at 94˚C, 1 min at 60˚C and 3 min 30 s at 65˚C, 3 cycles of 15 s at 94˚C, 1 min at 60˚C and 3 min 30 s at 65˚C (the elongation step was extended by 10 s per cycle) and a final elongation step of 7 min at 65˚C. A negative control was included in every run. The quality of the re-amplified product was examined using the same endpoint PCR assay as the one used for primary WTA products (see above).

Quantitative Real-Time PCR and statistical methods

Quantitative PCR (qPCR) was performed for selected genes of interest (*ERBB2* and reference genes) using a LightCycler 480 instrument (Roche). Each qPCR comprised 5 µl of the template cDNA, 10 µl iQTM SYBR Green Supermix (Bio-Rad), 1 µl of each forward and reverse primer (8 µM each, S1 Table) and 3 µl PCR HPLC Gradient Grade H$_2$O. The quantitative PCR was run using the following program: 1 cycle for 5 min at 95˚C (temp. ramp of 4.4˚C/s), 38 cycles for 20 s at 95˚C (ramp 4.4˚C/s), 15 s at 58˚C (ramp 2.2˚C/s) and 15 s at 72˚C (ramp 4.4˚C/s; fluorescence signal was measured at the end of each elongation step). Subsequently, melting curves were generated using the following procedure: 1 cycle for 5 s at 95˚C (ramp 4.4˚C/s), 1 cycle for 1 min at 50˚C (ramp 2.2˚C/s), 1 cycle of DNA melting, wherein the temperature was continuously increased to 95˚C with a ramp of 0.11˚C/s with continuous fluorescence measurement 5 times/s) followed by a final cooling to 40˚C for 30 s (2.2˚C/s). The melting plots were examined to validate the specificity of the PCR amplification. Samples with an appearance of melting curves different than expected were excluded from downstream analyses. The crossing point (Cp) values were determined with the LightCycler 480 Software using the second derivative maximum method applying the high sensitivity algorithm. All single-cell WTA products were analyzed in technical triplicates. Cp-values were averaged across the technical replicates before further data processing. Samples with average Cp-values >33 were considered as negative. Mock controls were included in all qPCR runs.

Relative qPCR quantification analysis

Template cDNA (i.e. WTA or re-amplified WTA) was diluted 50 times and 5 µl of the diluted cDNA was used per qPCR reaction. Primer efficiencies were calculated based on the analysis of standard curves generated using serially diluted WTA samples. For each measured sample, the mean Cp-value of the gene of interest was first normalized to the expression of a calibrator sample (which allows the normalization of technical errors; for this a reference cDNA was used) and second to a single reference gene (to account for biological variation) by calculating ΔΔCp values according to the following formula as described before [25]:

$$\Delta\Delta Cp = \frac{(E_{Target})^{\Delta Cp_{Target}(Calibrator-Sample)}}{(E_{Reference})^{\Delta Cp_{Reference}(Calibrator-Sample)}}$$

in which E represents the primer efficiency.

Visualization of the relative qPCR quantification data and downstream analyses were conducted utilizing log2-transformed ratios resulting in negative ΔΔCp values.

Absolute qPCR quantification analysis

The absolute qPCR quantification analysis of *ERBB2* expression levels necessitates the normalization of the template input. Quantification of cDNA yields in the individual samples was spectrophotometrically conducted using either the NanoDrop 2000 instrument or the Qubit dsDNA BR kit and the Qubit fluorometer. To allow an accurate measurement of cDNA, yields of WTA products were either purified or subjected to double-stranded DNA reconstitution (see below). The DNA input for each qPCR was normalized to 5 ng and run as described above. The obtained Cp-values generated by the LightCycler 480 Software (Roche) were converted into log10 copy numbers. For this, a standard curve measurement was conducted with serially diluted cDNA standards comprising amplicons of the *ERBB2* transcript. Concentrations of the standards ranged from 1.00E-08 to 1.00E-04 ng/µl covering copy numbers from 3.00E02 to 3.00E06 molecules/µl.

Purification of WTA products

Purification of WTA samples was conducted to remove buffer and remaining WTA reagents (dNTPs, proteins and primers) that may negatively influence downstream processes (i.e. cDNA yield quantification and qPCR). 10–15 μl of the primary or re-amplified WTA product was purified using QIAquick PCR Purification Kit (QIAgen) according to the manufacturer's instruction with several changes: (i) No pH-indicator was added to the PB buffer. (ii) Purified cDNA was eluted from the purification column using HPLC Gradient Grade H_2O instead of EB buffer provided by the manufacturer of the kit. (iii) PCR HPLC Gradient Grade H_2O (typically 15–20 μl) used for the elusion was pipetted on the silica membrane of the column, followed by 5 min incubation at room temperature prior to the final centrifugation (elution) step. To allow a more optimal distribution of the elution liquid in the silica membrane on the purification column, spin assembly was centrifuged at 500 rpm for 30 s followed by the final centrifugation at 13,000 rpm for 60 s. The concentration of each purified sample was measured using a NanoDrop 2000c (Thermo Fisher Scientific) utilizing 1 μl of the purified cDNA.

Double stranded DNA reconstitution of WTA products

Double strand cDNA (dscDNA) reconstitution was conducted to convert single stranded cDNA amplicons present in the WTA products into double stranded molecules allowing a fluorometric quantification of cDNA yields without the need for prior purification of the WTA samples. The dscDNA-synthesis was performed with 10 μl of the template DNA (i.e. primary or re-amplified WTA products) added to 1 μl Expand Long Template Buffer 2 (Expand Long Template PCR System, Roche Diagnostics), 1 μl dNTPs (10 mM each), 1 μl CP2-15C primer (2 μM; S2 Table), 0.5 μl DNA Pol Mix (Roche Diagnostics) and 6.5 μl HPLC Gradient Grade H_2O. The reactions were run using a PTC DNA Engine 2 Tetrad Thermocycler for 2 h at 68˚C. The concentration of the resulting dscDNA was measured using the Qubit BR Kit and the Qubit 2.0 Fluorometer (Thermo Fisher Scientific) according to the manufacturer's instruction using the Broad Range Assay. 2 μl of the dscDNA were used for each measurement.

Bioanalyzer assay

To analyze the fragment size distribution of the WTA samples, the Agilent High Sensitivity DNA Kit (Agilent Technologies) was used according to the manufacturer's instructions. WTA samples underwent dscDNA reconstitution to determine the concentration of the dscDNA. The samples were further diluted to 1 ng/μl and 1 μl was used for analysis in the Agilent 2100 Bioanalyzer.

Statistical analysis

The statistical analysis was conducted using the GraphPad Prism 6.01 software package (GraphPad Software Inc). -ΔΔCp values and log10 copy number values were tested for Gaussian distribution using the D'Agostino and Pearson omnibus normality tests. The Pearson's (parametric) or Spearman's (non-parametric) correlation tests were used to compare the sample sets. The unpaired t-test (two-tailed, CI = 95%) with Welch's correction was used to determine the statistical significance of differences between gene expression levels measured in BT-474 and MCF-7 cells. The one-way ANOVA test with Tukey's multiple comparisons test was used to assess the statistical significance between gene expression in more than two tested cell populations. Values of $p < 0.05$ (marked with a * in the figures) were considered statistically significant, whereas $p < 0.01$ (**), $p < 0.001$ (***) and $p < 0.0001$ (****) were recognized as highly significant.

Results

Relative quantification of gene expression in single cells

Relative transcript quantification by qPCR, for which the expression level of a gene of interest is calculated relative to the expression of one or multiple reference gene(s), so called house-keeping genes, is widely used to measure mRNA levels in multiple sample types [28]. However, to assure maximal accuracy of measurements one has to select reference genes that (i) are stably and consistently expressed at detectable levels across the studied sample collective and (ii) remain unaltered in the experimental set-up used. To identify the most suitable reference genes, we assembled a shortlist of putative housekeeping genes using three different approaches. First, we selected eight genes (*POLR2A*, *G6PD*, *HPRT1*, *ABCF1*, *GPS1*, *VPS72*, *CANX* and *TBP*) involved in basic metabolic processes as we hypothesized that these need to be expressed in every cell. Second, we re-evaluated a series of single-cell microarray-based gene expression profiling experiments to find the most consistently and stably expressed genes. For this, we used samples stemming from various experimental conditions including *ex vivo* isolated cell types that had been amplified by our previously reported WTA method prior to microarray hybridization [9, 29–31]. Utilizing this approach, we identified 1,544 candidate genes. From this group we selected seven candidate genes (*WNT10A*, *IRX3*, *MYOM1*, *NUDT13*, *ASCL2*, *EHMT2* and *DUSP15*) exhibiting strong and constant expression for further testing. Lastly, we picked additional five candidates (*EMC7*, *RAB7A*, *REEP5*, *VCP* and *PSMB4*) from a previously reported RNA-Seq-derived list of highly uniformly expressed housekeeping genes [32]. We designed and tested one primer pair for every candidate gene (S1 Table). Following initial tests, eight genes (*ABCF1*, *GPS1*, *VPS72*, *WNT10A*, *IRX3*, *MYOM1*, *NUDT13*, *ASCL2*) were excluded either (i) due to the lack of amplification of any gene product; (ii) multiple amplification products or (iii) other than expected RFLP patterns in endpoint PCR (S1 Fig). The remaining candidate genes (*POLR2A*, *G6PD*, *HPRT1*, *CANX*, *TBP*, *EHMT2*, *DUSP15*, *EMC7*, *RAB7A*, *REEP5*, *VCP* and *PSMB4*) were then examined for the stability and uniformity of expression in a cohort of WTA products obtained from cell pools and single cells isolated from various tissues: (i) fourteen cell pools (consisting of approximately 40 cells each) from patients with acute lymphocytic leukemia, (ii) sixteen single EpCAM+ DCCs from patients with prostate cancer and (iii) fifteen cell pools (consisting of a few hundred cells) from primary cultures of patients with melanoma, bronchial or prostate carcinoma (five cell pools each). First screenings were conducted using endpoint PCRs. Surprisingly, none of the twelve genes under investigation was consistently expressed in all tested samples. The transcript detection rate varied considerably (11–93%) between different candidate genes (Fig 1A, S3 Table). To exclude low-abundantly expressed genes, we set a detection rate cut-off to 60% (i.e. the tested gene must be expressed in more than 60% of samples) resulting in the omission of three rarely expressed genes (*CANX*, *DUSP15* and *TBP*) from further analyses (Fig 1A). Next, we re-analyzed the same sample set by qPCR to assess the level and uniformity of the remaining nine reference genes. The melt curve analyses revealed an unspecific amplification (a diverging appearance of melt curves across samples) of the *PSBM4* gene resulting in its exclusion from further analyses. *RAB7A*, *EMC7* and *REEP5* genes showed the most consistent expression across all tested samples (Fig 1B). To assess the uniformity of expression, we calculated log2-transformed ratios (-ΔΔCp values) (see Materials and methods) for each putative reference gene against all remaining ones. We hypothesized that stably expressed genes exhibit low variances of -ΔΔCp. Among the genes tested, *POLR2A* showed the most stable expression (Fig 1C, S4 Table). Notably, for all tested genes except *RAB7A* the variance of measured -ΔΔCp was higher in single-cell WTA products as compared to those generated from cell pools (Fig 1C), indicating that the expression of individual genes is noisier at the single-cell level as
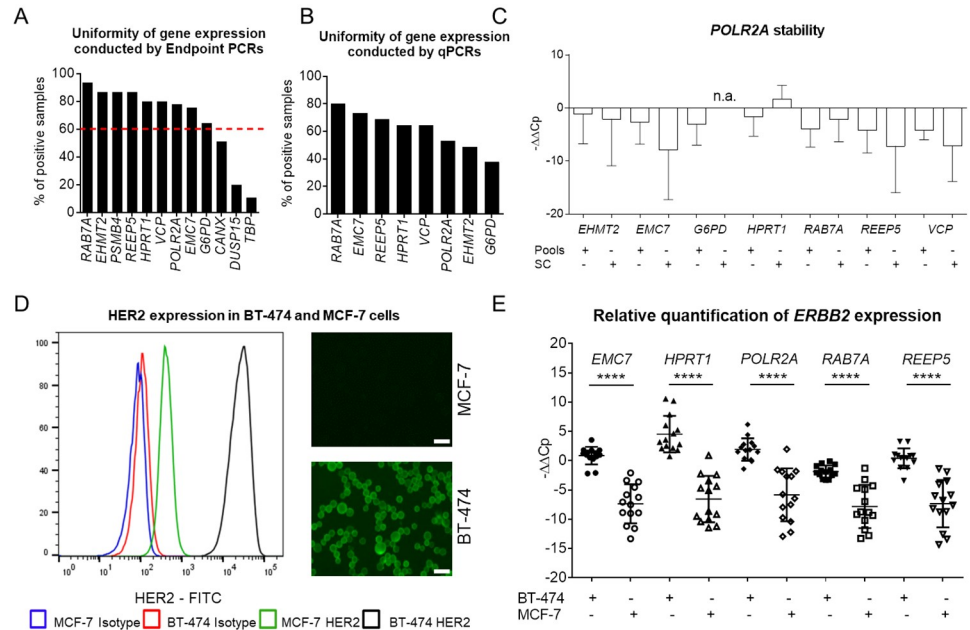
Fig 1. Relative qPCR quantification of single cell transcripts. (A) Stable and uniform gene expression (depicted as the percentage of samples tested positive for selected transcripts) of twelve genes examined by endpoint PCRs of WTA products obtained from three sample sets including cell pools and single cells. The red dashed line indicates the threshold for further testing (60%). (B) Uniformity of gene expression (%) of the remaining eight candidate reference genes as measured by qPCR. (C) Stability of gene expression depicted as -ΔΔCp (log2-transformed ratios) calculated for *POLR2A* vs the remaining candidate genes. The expression of *G6PD* could not be measured in single cells (n.a. = not assessed). SC = single-cell WTA products, Pools = WTA products generated from a pool of cells. (D) HER2 expression in BT-474 and MCF-7 breast cancer cells analyzed by FACS (left panel) and microscopically (right panels). HER2 expression was 62-fold higher in BT-474 as compared to MCF-7 cells. Median FITC (BT-474) = 24,659; Median FITC (MCF-7) = 399. Microscope settings corrections: Brightness: +20%, contrast: -40%; Scale bar 20 μm. (E) Relative quantification of *ERBB2* expression at the single-cell level in BT-474 and MCF-7 cells calculated separately using single reference genes as indicated. -ΔΔCp were calculated and plotted for every single-cell qPCR measurement. Mean ± SD of -ΔΔCp values; Unpaired t-test with Welch's correction; *** p<0.001, **** p<0.0001.

compared to bulk measurements. Finally, taking into account the gene expression consistency and uniformity, five genes (*RAB7A*, *EMC7*, *REEP5*, *POLR2A* and *HPRT1*) were chosen as potential reference genes for further testing.

Next, we examined whether the normalization to each of the selected reference genes enables a robust quantitative measurement of mRNA levels of a selected target gene in single-cell WTA products. To this end, we investigated expression levels of the *ERBB2* gene in two breast cancer cell lines—BT-474 and MCF-7—known to differ in the expression levels of HER2, the protein product of the *ERBB2* gene [33, 34]. FACS analysis revealed a 62-fold higher HER2 expression in BT-474 as compared to MCF-7 cells (Median FITC [BT-474] = 24,659; Median FITC [MCF-7] = 399; Fig 1D). From each cell line twenty-two single cells were picked by micromanipulation and subsequently subjected to WTA. All forty-four cells were tested for the presence of transcripts of three housekeeping genes (*ACTB*, *GAPDH* and *EEF1A1*; WTA quality control assay–see Material and methods), the target gene (*ERBB2*) and the five reference genes (*RAB7A*, *EMC7*, *REEP5*, *POLR2A* and *HPRT1*) using endpoint PCRs. The drop-out rates (i.e. the failure to detect the expression of a tested transcript) for each PCR assay are documented in Table 1 and S5 Table. Of note, drop-outs were detected at a considerable frequency for transcripts of the tested reference genes, which hindered a measurement of gene expression in single cells when using the relative quantification method. One MCF-7 (6.7%)

Table 1. Stepwise drop-out rates during sample selection for qPCR.

	BT-474		MCF-7	
	Input cells	**Drop-outs (%)**	**Input cells**	**Drop-outs (%)**
Isolated single cells	22	-	22	-
WTA quality control	22	0/22 (0)	22	3/22 (13.6)
Endpoint PCR of the target gene	22	2/22 (9.1)	19	4/19 (21.1)
Endpoint PCRs of the reference gene(s)	20	6/20 (30)	15	1/15 (6.7)
qPCR	14	0	14	0

https://doi.org/10.1371/journal.pone.0216442.t001

and six BT-474 single cells (30%) had to be excluded from qPCR analysis due to non-detect-able expression of reference transcripts in the endpoint PCR analysis (Table 1). Still, we tested the remaining fourteen single cells from each BT-474 and MCF-7 cell line by qPCR (summing up to 28 single cells). Subsequent statistical analysis of single-cell qPCR data showed a signifi-cant difference in *ERBB2* transcript level between BT-474 (HER2^{hi}) and MCF-7 (HER2^{lo}) cal-culated for all target-reference gene pairs (Fig 1E). As a consequence of the pre-selection by endpoint PCR, no further drop-outs were observed for qPCR. In summary, we identified five reference genes suitable for the relative quantification of mRNA expression levels in single-cell WTA products.
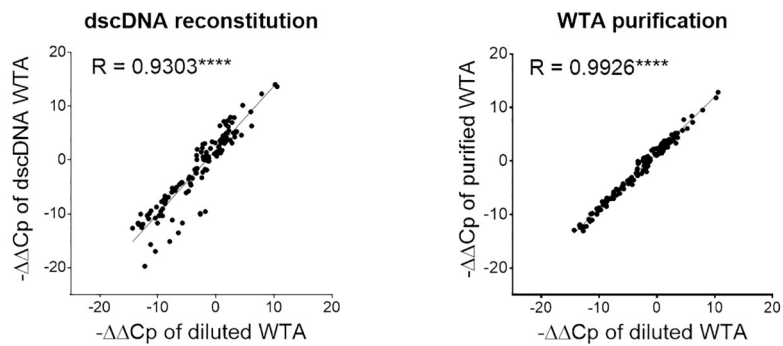
Absolute quantification of gene expression in single cells

In contrast to relative qPCR quantification, absolute quantification does not require measure-ments of reference genes in samples tested but a careful normalization of cDNA inputs. For this, the concentration of template cDNA needs to be determined and equalized across all sam-ples. However, since single-cell WTA products are mixtures of single stranded and double stranded cDNA amplicons (sscDNA and dscDNA, respectively) as well as primers, dNTPs and other reagents used in the WTA procedure, direct measurement of cDNA yields utilizing spectrophotometric or fluorometric methods is prone to error. We therefore compared two approaches to assess DNA concentrations in single-cell WTA products. In the first approach, the samples were purified using the QIAquick PCR Purification Kit (to remove residual buffer, dNTPs and primers) prior to a spectrophotometric measurement. In the second procedure, the samples were subjected to dscDNA reconstitution of the WTA products enabling a subse-quent fluorometric measurement using dsDNA-binding dyes. Both protocols were applied to the WTA products of single BT-474 and MCF-7 cells mentioned above. Subsequently, 5 ng of the obtained cDNA were used in every qPCR to measure expression levels of the *ERBB2* gene. In line with the results obtained by relative quantification, the absolute quantification allowed us to detect distinct expression levels of the *ERBB2* gene in HER2^{hi} and HER2^{lo} cells using both cDNA processing strategies (Fig 2A). Of note, only one BT-474 single cell that underwent dscDNA reconstitution had to be excluded from qPCR analysis due to discrepant amplifica-tion plots (S5 Table). Accordingly, we can report a drop-out rate of only 2% (1/44) for single-cell samples subjected to absolute quantification (i.e. relative to both MCF-7 and BT-474 cells after WTA purification or dscDNA reconstitution). As an advantage of absolute quantification over relative quantification, the calculated differences in *ERBB2* copy numbers reflected more faithfully the 62-fold difference in HER2 protein expression level between HER2^{hi} and HER2^{lo} cells (50-fold or 51-fold more *ERBB2* copy numbers in BT-474 cells as compared to MCF-7 cells using dscDNA or purified samples, respectively) (Fig 1D, Fig 2A). To better define the impact of the sample pre-treatment (i.e. dscDNA synthesis or WTA purification) on the accu-racy of qPCR measurements, we generated more data points not restricted to *ERBB2* alone by
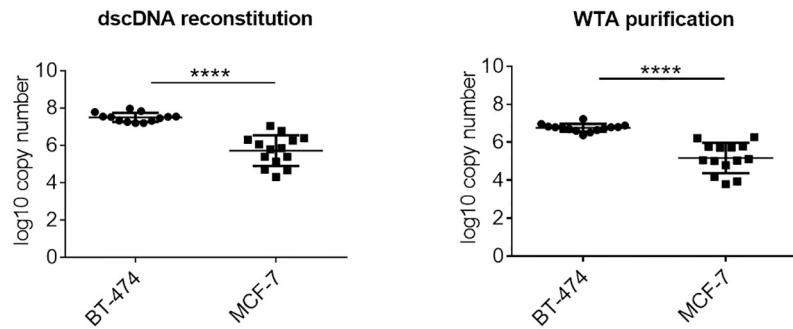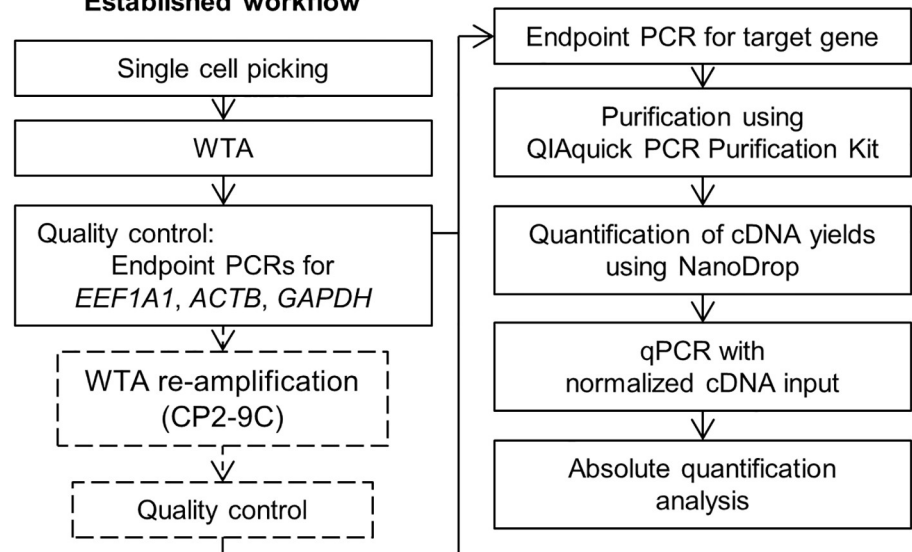
Fig 2. Novel qPCR-based workflow for highly accurate gene expression analysis in single cells. (A) Quantification of the *ERBB2* gene expression levels in MCF-7 and BT-474 single cells using absolute qPCR quantification. WTA products underwent either double-stranded cDNA (dscDNA) reconstitution (left panel) or purification using the QIAquick PCR Purification Kit (right panel) prior to the normalization of cDNA input and qPCR. Cp values were converted to log10 copy numbers using an external standard curve. Mean ± SD; Unpaired t-test with Welch's correction; **** p<0.0001. (B) Correlation of *ERBB2* expression levels obtained using the relative quantification strategy for the diluted WTA samples as compared to the WTA products subjected to dscDNA reconstitution (left panel) or purification (right panel). Each point represents one -ΔΔCp value calculated for one *ERBB2*-reference gene pair (-ΔΔCp for all target/reference gene combinations were plotted). Pearson's correlation coefficient R. **** p<0.0001. (C) Workflow of the established qPCR assay for profiling gene expression levels in single cells using the absolute quantification strategy.

measuring the expression levels of all five previously selected reference genes (*RAB7A*, *EMC7*, *REEP5*, *POLR2A* and *HPRT1*) (S5 Table). Careful analysis of the qPCR data generated for *RAB7A* gene revealed that two samples (one generated from MCF-7 and one from BT-474) subjected to dscDNA exhibited discrepant amplification curves as compared to other samples included in the collective and were therefore excluded from further analyses (S5 Table). Direct comparison of calculated -ΔΔCp values obtained for all possible gene pairs showed a high level of correlation between the results generated using unprocessed primary WTA products and the results obtained after dscDNA reconstitution or WTA purification (Fig 2B). Here, -ΔΔCp values obtained from purified samples correlated better with the corresponding values generated from untreated specimens as compared to the samples subjected to dscDNA reconstitution (Fig 2B). Based on this correlation analysis and the observed drop-outs of samples subjected to dscDNA reconstitution, we concluded that WTA purification is more suited as a cDNA pre-treatment method than dscDNA reconstitution. Summarizing, we established a workflow consisting of i) single-cell WTA of cells isolated by e.g. micromanipulation including an optional WTA re-amplification step (see below), ii) a WTA quality control assay (endpoint PCR targeting the three housekeeping genes *ACTB*, *GAPDH* and *EEF1A1*), iii) an optional endpoint PCR for the selected gene of interest (acting as a screening to select samples suitable for downstream analyses), iv) a purification of amplified cDNA, v) a spectroscopic cDNA quantification, vi) a normalization of cDNA input and vii) a qPCR measurement with an absolute quantification used for the data analysis (Fig 2C). This workflow was utilized in the subsequent experiments.

Sensitivity of the new single-cell qPCR workflow

The dependency of the relative quantification on a simultaneous expression of multiple genes (i.e. target and reference gene(s)) hinders the quantification if the expression of one of the tested transcripts cannot be detected. Missing expression values were prominent in single-cell samples before precluding gene expression analysis of these samples when using the relative quantification method (Table 1, S5 Table). In contrast, absolute quantification provided the means for a more robust and reliable gene expression profiling in single-cell WTA products. Therefore, we decided to study the accuracy of the absolute quantification method in greater detail. We attempted to measure finer differences in gene expression levels than those quantified when analyzing BT-474 and MCF-7 cells (62-fold difference). Accordingly, we chose three well-established breast cancer cell lines (MCF-10A, ZR-75-1 and MDA-MB-453) exhibiting similar, partially overlapping, levels of HER2 protein expression as determined by flow cytometry (Fig 3A and 3B) [34]. FACS analysis of ZR-75-1 showed a 3.5-fold higher expression level of HER2 protein compared to MCF-10A, and a 2.3-fold lower amount of HER2 protein compared to MDA-MB-453 cells (Fig 3B). We utilized FACS to isolate cell populations exhibiting distinct levels of HER2 (similar within a given cell population but distinct from other groups) from each cell line (Median FITC [MCF-10A] = 436; Median FITC [ZR-75-1] = 1,529; Median FITC [MDA-MB-453] = 3,575; Fig 3B). Subsequently, ten single cells from each sorted subpopulation were isolated by micromanipulation and subjected to our qPCR workflow (Fig 2C). Based on our QC criteria, we excluded two single cells from further examination leaving twenty-eight single-cell WTA products for qPCR (S6 Table). In line with the results of previous experiments, FACS and qPCR-based analyses provided concordant results. Quantitative PCR analysis enabled the detection of distinct mRNA levels of the *ERBB2* gene in MCF-10A, ZR-75-1 and MDA-MB-453 cells. In all but one comparison, i.e. MCF-10A vs ZR-75-1 (the cell lines exhibiting medium and low expression levels of the *ERBB2* gene), we detected significantly different expression levels of *ERBB2* (Fig 3C). We concluded that the quantification of
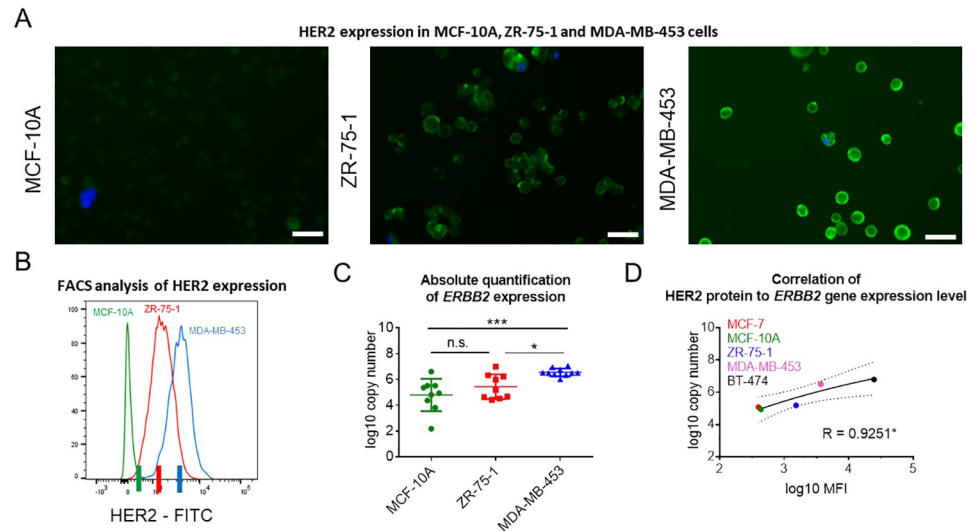
Fig 3. Highly accurate analysis of differential expression facilitated by the new qPCR assay. HER2 expression analysis in three breast cancer cell lines (MCF-10A, ZR-75-1 and MDA-MB-453) assessed microscopically (A) and by flow cytometry (B). (A) Brightness: +20%, contrast: -40%. Scale bar indicates 20 μm. (B) Median FITC (MCF-10A) = 436; Median FITC (ZR-75-1) = 1,529; Median FITC (MDA-MB-453) = 3,575; 3.5-fold (MCF-10A vs. ZR-75-1) and 2.3-fold (ZR-75-1 vs. MDA-MB-453) increase in HER2 expression levels. Colored lines on the *x*-axis indicate the signal intensities used as thresholds for sorting of cell populations prior to single-cell isolation and analysis. (C) Quantification of *ERBB2* expression at the single-cell level was conducted following the newly established protocol (Fig 2C). Cp values were converted to log10 copy numbers using an external standard curve. Mean ± SD; Tukey's multiple comparisons test was applied, n.s. = not significant, * $p<0.05$, ** $p<0.001$. (D) Correlation between measured HER2 protein and *ERBB2* gene expression levels. Log10 converted median fluorescent intensity (MFI) values derived from FACS analysis and log10 converted copy numbers calculated using the absolute quantification method are plotted. Pearson's correlation coefficient R, * $p<0.05$.

moderate to high expression levels is feasible with our approach, while measurements are less accurate in samples exhibiting extremely low expression levels of the target gene. This may be due to noisy expression profiles (i.e. varying expression levels across different cells), suboptimal conversion rates of transcripts to WTA amplicons, technical issues related to the quantification of low-abundance transcripts [24, 35, 36] or it reflects biological differences associated with low and high abundant transcripts during the translation. Nonetheless, our method enabled to detect distinct levels in transcript expression in single-cell WTA products of cell lines exhibiting a 2.3-fold difference in expression of the corresponding protein. Notably, when considering all five tested breast cancer cell lines (MCF-7, MCF-10A, ZR-75-1, MDA-MB-453 and BT-474), MFI values obtained by FACS analysis correlated highly with transcript copy numbers (Pearson R = 0.925; p = 0.0243; Fig 3D).

Transcript quantification using re-amplified WTA products

Since the amount of cDNA contained in a single-cell WTA product is limited, thus restricting the amount of downstream analyses, we sought for ways to increase the amount of the available material without introducing any considerable bias into the sample's representation. For this reason, we developed two re-amplification protocols for WTA products differing in the sequence of used primers. Both WTA re-amplification methods employ PCR primers consisting of an universal adapter sequence identical to the one used in the primary WTA [9] at the 5'-end and a run of either fifteen or nine cytosines at the 3'-end (denoted CP2-15C and CP2-9C, respectively; S2 Table). Poly-C tails enable primer binding to oligo-G motifs present in the flanks of every WTA amplicon allowing a re-amplification of primary WTA products. First,
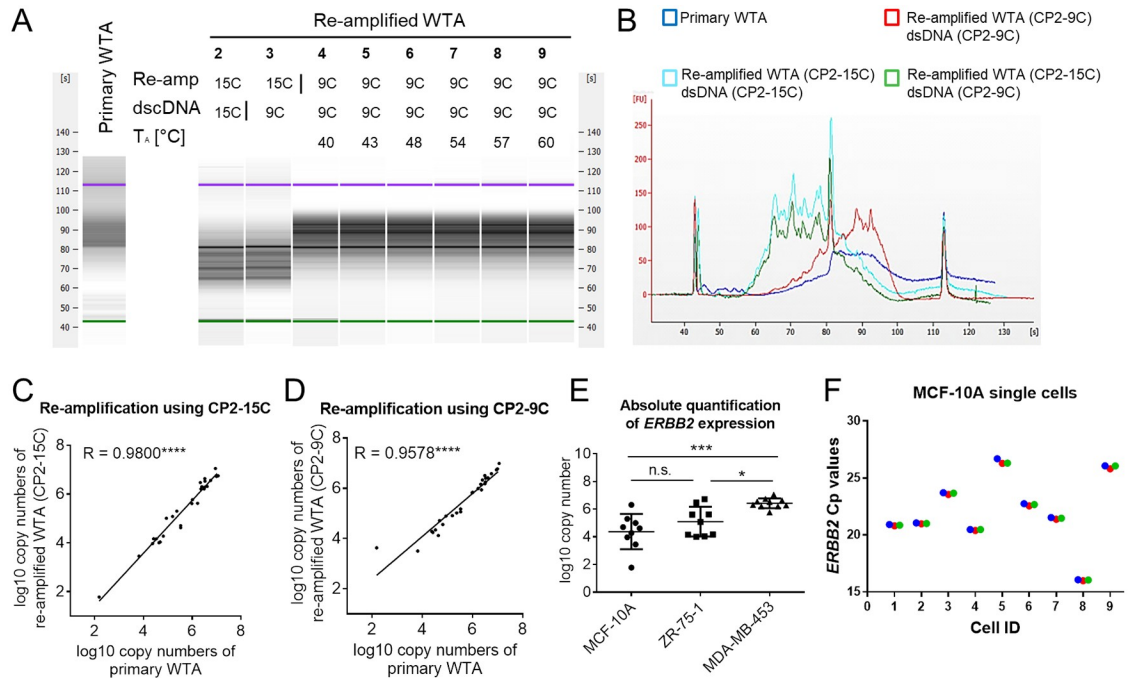
Fig 4. Single-cell qPCR analysis utilizing re-amplified WTA products. (A,B) Fragment size distribution of primary WTA and corresponding WTA re-amplification products (generated with either the CP2-15C or the CP2-9C primer) of one representative BT-474 single cell assessed by Bioanalyzer assay. (A) 2–9: Re-amplified WTA products generated using the indicated primer sequence for re-amplification and dscDNA-synthesis. 4–9: Samples re-amplified with different annealing temperatures (T$_A$). (B) Electropherogram of selected samples. (C,D) Correlation between qPCR results conducted with primary and re-amplified WTA using CP2-15C (C) and CP2-9C (D) primers for re-amplification. (E) Absolute quantification of *ERBB2* transcript levels in MCF-10A, ZR-75-1 and MDA-MB-453 cells using re-amplified single-cell WTA products (generated using the CP2-9C primer). (F) Cp values of *ERBB2* expression generated for MCF-10A single cells after re-amplification using CP2-9C primer. Red, green and blue represent three different replicates.

we investigated the qualitative representation of the transcriptome after re-amplification and compared the fragment size distributions of PCR amplicons in the corresponding primary and re-amplified WTA products. The analysis revealed that the utilization of the CP2-15C primer during WTA re-amplification resulted in an enrichment of short WTA amplicons while the re-amplified WTA products generated using the CP2-9C primer showed a more similar fragment size distribution to the corresponding primary WTA products (Fig 4A and 4B). Next, we tested the impact of the annealing temperature used in the WTA re-amplification protocol (using the CP2-9C primer) on the fragment size distribution of the resulting products. Across the entire range of annealing temperatures tested (40–60˚C), the fragment size distribution of the re-amplified WTA products remained unchanged and therefore comparable to the primary products (Fig 4A). To examine the application of re-amplified WTA samples in qPCR, we tested both WTA re-amplification protocols (utilizing either the CP2-15C or the CP2-9C primer) using the previously generated primary WTA samples from all five breast cancer cell lines. We analyzed the resulting WTA re-amplification products using the same qPCR workflow we had applied to the corresponding primary WTA samples (see above, S7–S9 Tables). First, we assessed the correlation between gene expression measurements, which were obtained for both primary and re-amplified WTA samples. Direct comparison of the results obtained by both relative and absolute quantification methods showed a high correlation between the sample types (Fig 4C and 4D, S2 Fig). Quantitative PCR results obtained using both WTA re-amplification protocols (CP2-9C or CP2-15C primer) showed similar levels of

correlation to the data generated with primary WTA products (Pearson R for CP2-15C: 0.98, p<0.0001 and CP2-9C: 0.96, p<0.0001; Fig 4C and 4D). Relative and absolute quantification enabled us to distinguish different expression levels of the *ERBB2* gene in re-amplified WTA products of single breast cancer cells with a comparable level of confidence as in the original WTA samples (Fig 4E, S2 Fig). The technical replicates of Cp-values generated after re-amplification of WTA samples using CP2-9C primer did not vary considerably while the biological variation between single cells became clearly evident (Fig 4F). Moreover, in line with the results obtained for primary WTA products, absolute qPCR quantification using re-amplified WTA samples was more accurate when conducted with purified samples as compared to specimens subjected to dscDNA synthesis (S2B and S2C Fig). In summary, qPCR analysis of single-cell primary and re-amplified WTA products provided highly concordant results. Since the re-amplification protocol utilizing the CP2-9C primer was less prone to introduce a representation bias into the re-amplified WTA products, we decided to further use this variant of the protocol.

ERBB2 expression in single disseminated breast cancer cells

Having established a workflow for qPCR-based gene expression analysis using re-amplified single-cell WTA products (Fig 2C), we tested it using patient-derived DCCs. For this, we isolated DCCs from a pleural effusion of a metastatic breast cancer patient, who had received her regular dose of the HER2-targeting antibodies trastuzumab / pertuzumab about six hours before sampling. Following a double-staining against EpCAM and HER2, single and double-positive single cells were detected and isolated by micromanipulation (Fig 5A). In total, we isolated fifteen EpCAM+/HER2- cells (i.e. EpCAM-positive DCCs lacking HER2 expression), four EpCAM$^{lo/-}$HER2+ cells (putative DCCs, low positive or negative for EpCAM and positive for HER2). The ratio of HER2$^+$ / HER2$^-$ cells was fully consistent with the administration of very high doses of HER2-targeting antibodies a few hours before *ex vivo* staining that successfully blocked the microscopic HER2 protein detection in the majority of cells. In addition, three double-negative cells (presumably non-epithelial, non-tumor cells) and two pools of BM cells were isolated. The isolated single cells were subjected to WTA and quality assessment as described above (Fig 2C). All but one single-cell WTA product displayed an expression of at least two out of three tested housekeeping genes, meeting our QC criteria, and were included in further analyses. All re-amplified single-cell WTA products were tested positive for at least two housekeeping genes in a subsequent endpoint control PCR (S10 Table). Thereafter, all re-amplified single-cell WTA products were further processed and tested for the expression of the *ERBB2* gene by endpoint PCR (Fig 5B) and qPCR (Fig 5C). Quantitative PCR analysis revealed that all HER2+ cells expressed high levels of *ERBB2* transcripts, confirming the reliability of our measurement. Interestingly, HER2- cells displayed two levels of transcript abundance: one group with a similar expression level as the HER2+ cells and another group with highly reduced or absent expression (p<0.0001, Fig 5D, S10 Table). While transcript detection in cells for which no HER2-staining could be achieved is expected after blocking the antibody-antigen binding by excessively high doses given for therapy, the detection of a population without protein and transcript detection may indicate the formation of a treatment resistant subclone.

Discussion

The analyses of disseminated cancer cells (DCCs) and other rare cell populations face several unique challenges. First, DCCs are extremely rare and genetically heterogeneous making every sample precious and unique. Therefore, the analysis of such cells necessitates highly reliable
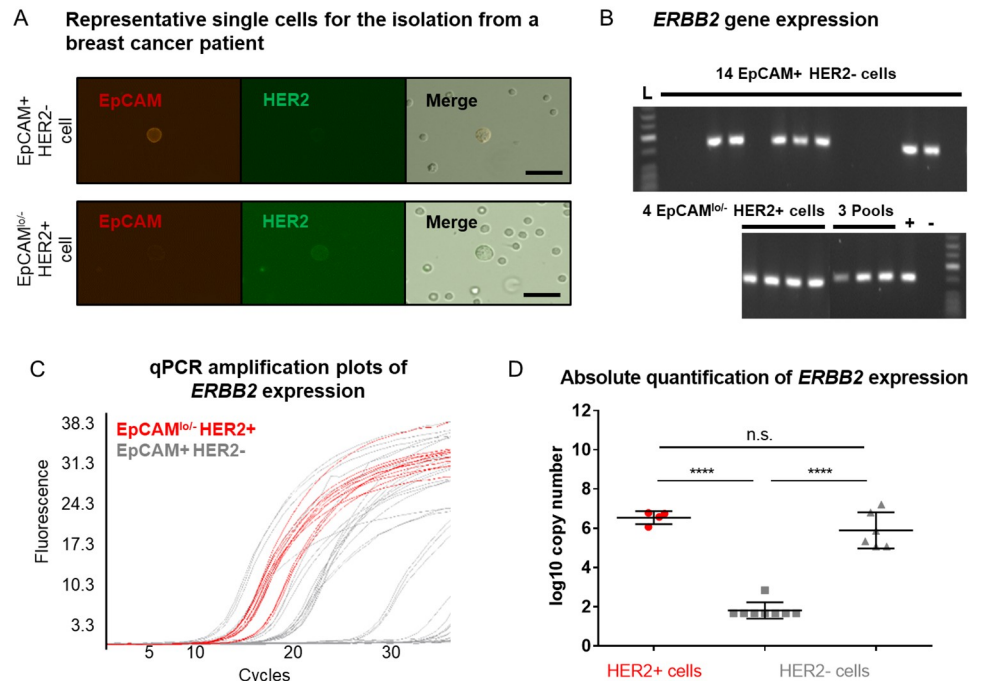
Fig 5. HER2/*ERBB2* expression in single cells derived from a clinical patient. (A) Cells from a pleural effusion of a metastatic breast cancer patient with a HER2-positive tumor were stained using anti-EPCAM-Cy3 and anti-HER2-FITC antibodies. Cells showing a clear membranous staining pattern were picked. Scale bar indicates 20 μm. (B) *ERBB2* expression was examined in primary WTA products by endpoint PCR. (C,D) Quantitative PCR analysis of *ERBB2* expression was performed using re-amplified cDNA. (C) Amplification curves obtained for HER2+ and HER2- cells were labeled with red and gray colors, respectively. (D) Quantification of *ERBB2* expression at the single-cell level was conducted following the newly established protocol (Fig 2C). Cp values were converted to log10 copy numbers using an external standard curve. Mean ± SD; one-way ANOVA testing with Tukey's multiple comparisons test. Color code as in C, **** $p < 0.0001$.

https://doi.org/10.1371/journal.pone.0216442.g005

single-cell technologies enabling to minimize rates of technical drop-outs and allowing highly sensitive measurements. For this reason, we established a novel workflow to reliably quantify transcript levels of selected genes using single-cell cDNA libraries generated using a previously described WTA protocol [9]. The new protocol has been established to measure expression levels of the *ERBB2* gene that is of high clinical relevance. The new method enabled to measure distinct levels of *ERBB2* gene expression in cancer cell lines and patient-derived DCCs and allows addressing specific research questions with little resources.

Optimized single-cell qPCR-based protocols are very reliable, rarely suffer from technical failures [27] and exhibit very low technical noise thus allowing an accurate quantification of transcripts down to ~100 copies per cell [36]. For these reasons, we decided to use qPCR as the read-out in our approach. To maximize the reliability of the new protocol, we carefully selected and optimized all steps of our workflow.

A key consideration in the context of qPCR-based gene expression analysis in single cells is the data processing and normalization. Among many proposed strategies such as normalization to total RNA [37], cell size [38] or introduced RNA/DNA spikes [39], the most common way of normalization is to quantify the expression ratio of the target gene relative to the mRNA levels of reference genes–herein referred to as relative quantification method [40]. However, due to the ubiquitous cell-to-cell variability in the expression of traditional reference genes, relative quantification of single cells is prone to introduce random error into single-cell analyses [41] unlike relative quantitation using bulk material [41]. Various studies showed that

gene expression occurs in temporal bursts also affecting housekeeping genes used as references. In relative quantification, these bursts result in false variations in transcript levels between individual cells [42–45]. Indeed, our data support these reports showing considerable intra-cellular variability in the gene expression of the candidate reference genes rendering most of them unsuitable for relative quantification. We noted a considerable variation between the tested candidate reference genes (the detection rates in single cells ranged from 11% to 93%) consistent with the reported dynamic changes in the transcriptomes of single cells [46–48]. To minimize the amount of missing expression data, we decided to use only the most consistently and stably expressed reference genes (*RAB7A*, *EMC7*, *REEP5*, *POLR2A*, *HPRT1*) in our analysis. Employment of these genes enabled a clear discrimination between the two breast cancer cell lines BT-474 and MCF-7 with high and low HER2 expression levels, respectively, of the target gene *ERBB2*. Still, despite our extensive survey for the best references, an exclusion of samples from data analysis due to lack of expression of reference genes could not be avoided, thus challenging the use of the relative expression quantification for the analysis of rare cells such as DCCs.

To overcome the limitations associated with reference genes, we established a workflow utilizing absolute quantification, which requires the quantification of the target gene only but necessitates a careful normalization of cDNA input. For this reason, we sought for methods allowing a reliable quantitation of cDNA yields in single-cell WTA products. Spectrophotometric cDNA quantification preceded by purification of WTA products proved to be the best method in our hands allowing a sensitive quantification of *ERBB2* gene expression. Consequently, we conclude that the most reliable approach for the normalization and quantification of single-cell qPCR data is the absolute quantification method.

The drawback of this method lies in the difficulty to automate. A completely automated system such as Fluidigm C1 is more appropriate for such requirements, but its limitation in the number of analyses on each single-cell sample is unacceptable when working with rare single cells isolated from patients. Another highly advanced system for high-throughput qPCR analysis based on droplet digital PCR (ddPCR) provides highly accurate analyses without great loss of cell material enabling further downstream analyses. Nevertheless, expensive equipment is required, and its application makes only sense if many cells are isolated from one individual sample. The advantage of our approach on the other hand is that very rare cells collected from many patients can be screened for the expression level of selected target genes in a cost effective and rather fast procedure.

To select a method for normalization and analysis of rare single cells, we sought for ways to maximize the amount of material available from each cell. For this, we developed two versions of a protocol for the re-amplification of WTA products. However, we observed that WTA re-amplification is less reliable and generates high amounts of truncated amplicons when the amplification primer comprised long (15 bp long) oligo(C) motifs. Similar observations have been made when using primers containing long oligo(T) motifs during reverse transcription [49]. It is likely that in both cases truncation of transcripts resulted from an internal priming of cDNA amplicons. Nevertheless, a second version of the WTA re-amplification protocol, utilizing primers with shorter (9 bp long) oligo(C) priming motifs, facilitated an efficient amplification of primary WTA products and reliable quantification of *ERBB2* levels by qPCR. Therefore, our new assay allows the quantification of gene expression in primary and re-amplified WTA products with an equal accuracy.

We tested our new workflow for its sensitivity and accuracy. The method allowed the detection of both evident discrepancies in gene expression level of the *ERBB2* gene between MCF-7 and BT-474 cells as well as more subtle differences measured between MCF-10A, ZR-75-1 and MDA-MB-453 cells. In all analyses conducted in breast cancer cells, we found an excellent

correlation between *ERBB2* transcript abundance and HER2 protein levels corresponding to the previously published data [33]. Notably, only six cells (five in the first and one in the second comparison) had to be excluded from the analyses due to undetectable expression of *ERBB2*. This demonstrates the excellent performance and accuracy of our method.

Finally, in a proof of principle experiment, we applied our new workflow to DCCs from a pleural effusion of a metastatic breast cancer patient with a HER2-positive tumor. We categorized the collected DCCs into HER2-expressing and non-expressing cells and assessed the *ERBB2* expression levels by qPCR. In agreement with the HER2 protein expression assessed by immunostaining, the expression of *ERBB2* measured by single-cell quantitative PCR was high in the HER2+ DCCs group. In the HER2- group (i.e. immunofluorescent negative cells for the HER2 channel), we identified single cells either expressing the *ERBB2* transcripts similarly to the HER2+ cells (6/14) or highly reduced expression (8/14). Since the patient had been treated with the HER2-targeting drugs trastuzumab and pertuzumab a few hours before sampling, a blocking of the staining antibody was expected. In fact, the majority of the detected and isolated cells were HER2- upon staining, indicating the efficient binding of the treatment antibodies. The absence of *ERBB2* transcripts in some cells may then reflect either the reported biological variation between single cells resulting from stochastic expression bursts [45] or indicates the formation of a drug-resistant clone that successfully lost the addiction to the oncogene *ERBB2*. However, the faithful capture of *ERBB2* transcripts in all HER2 protein positive cells indicates the reliability of our method in a clinical setting, whereas the detection of a transcript negative cell population may furthermore pave the way for studies addressing mechanisms of therapy resistance and clonal selection directly in patients.

Supporting information

S1 Fig. Exclusion of reference genes based on gene expression profiling. Gradient PCRs using different annealing temperatures (1–12: 53°C; 53.3°C; 54°C; 54.8°C; 56.1°C; 57.7°C; 59.6°C; 61.2°C; 62.3°C; 63.2°C; 63.8°C; 64°C); NC = Negative Control; L = 2-log DNA ladder. Template: reference cDNA.
(TIF)

S2 Fig. Reproducible qPCR results using re-amplified WTA products. (A) Relative quantification analysis. Correlation between log2-transformed ratios (-ΔΔCp) of re-amplified compared to primary WTA products (left panel). Spearman's correlation coefficient R. The right panel shows the relative quantification of *ERBB2* expression at the single-cell level within BT-474 and MCF-7 using single reference genes as indicated. -ΔΔCp were calculated for every single cell, mean ± SD. (B,C: left panels) Correlation of *ERBB2* qPCR results obtained by the absolute quantification strategy in re-amplified BT-474 and MCF-7 single cells between dscDNA reconstitution (B) or purified WTA (C) and diluted primary WTA products. Pearson's correlation coefficients R. (B,C: right panels) Significant discrimination between BT-474 and MCF-7 cells by *ERBB2* gene expression levels. Cp values were converted to log10 copy numbers using an external standard curve. Mean ± SD; Unpaired t-test with Welch's correction; **** p<0.0001.
(TIF)

S1 Table. Oligonucleotides used for amplification of target DNA sequences.
(XLSX)

S2 Table. Oligonucleotides used for WTA and re-amplification.
(XLSX)

S3 Table. Overall gene expression of reference genes across sample sets obtained by end-point PCRs.
(XLSX)

S4 Table. Stability of reference genes.
(XLSX)

S5 Table. Gene expression analyses of primary WTA derived from BT-474 and MCF-7 single cells.
(XLSX)

S6 Table. Gene expression analyses of primary WTA derived from MCF-10A, ZR-75-1, MDA-MB-453 single cells.
(XLSX)

S7 Table. Gene expression analyses in re-amplified WTA (CP2-15C) of BT-474 and MCF-7 single cells.
(XLSX)

S8 Table. Gene expression in re-amplified WTA (CP2-15C) of MCF-10A, ZR-75-1 and MDA-MB-453 single cells.
(XLSX)

S9 Table. Gene expression in re-amplified WTA (CP2-9C) of MCF-10A, ZR-75-1 and MDA-MB-453 single cells.
(XLSX)

S10 Table. Gene expression analyses of picked single cells from a clinical sample.
(XLSX)

Acknowledgments

Author Contributions

Conceptualization: Christoph A. Klein, Zbigniew T. Czyż.

Data curation: Franziska C. Durst, Ana Grujovic, Iris Ganser, Martin Hoffmann.

Formal analysis: Franziska C. Durst, Ana Grujovic.

Funding acquisition: Ana Grujovic.

Investigation: Franziska C. Durst, Ana Grujovic, Iris Ganser.

Methodology: Franziska C. Durst, Ana Grujovic, Iris Ganser.

Resources: Peter Ugocsai.

Supervision: Christoph A. Klein, Zbigniew T. Czyż.

Validation: Franziska C. Durst, Ana Grujovic, Iris Ganser.

Visualization: Franziska C. Durst, Ana Grujovic.

Writing – original draft: Franziska C. Durst, Ana Grujovic.

Writing – review & editing: Christoph A. Klein, Zbigniew T. Czyż.

References

1. Pierga JY, Bonneton C, Vincent-Salomon A, de Cremoux P, Nos C, Blin N, et al. Clinical significance of immunocytochemical detection of tumor cells using digital microscopy in peripheral blood and bone marrow of breast cancer patients. Clin Cancer Res. 2004; 10(4):1392–400. PMID: 14977842.

2. Miller MC, Doyle GV, Terstappen LW. Significance of Circulating Tumor Cells Detected by the Cell-Search System in Patients with Metastatic Breast Colorectal and Prostate Cancer. J Oncol. 2010; 2010:617421. https://doi.org/10.1155/2010/617421 PMID: 20016752

3. Bulfoni M, Turetta M, Del Ben F, Di Loreto C, Beltrami AP, Cesselli D. Dissecting the Heterogeneity of Circulating Tumor Cells in Metastatic Breast Cancer: Going Far Beyond the Needle in the Haystack. Int J Mol Sci. 2016; 17(10). https://doi.org/10.3390/ijms17101775 PMID: 27783057

4. Macaulay IC, Voet T. Single cell genomics: advances and future perspectives. PLoS Genet. 2014; 10 (1):e1004126. https://doi.org/10.1371/journal.pgen.1004126 PMID: 24497842

5. Kolodziejczyk AA, Kim JK, Svensson V, Marioni JC, Teichmann SA. The technology and biology of single-cell RNA sequencing. Mol Cell. 2015; 58(4):610–20. https://doi.org/10.1016/j.molcel.2015.04.005 PMID: 26000846.

6. Huang H, Goto M, Tsunoda H, Sun L, Taniguchi K, Matsunaga H, et al. Non-biased and efficient global amplification of a single-cell cDNA library. Nucleic Acids Res. 2014; 42(2):e12. https://doi.org/10.1093/nar/gkt965 PMID: 24141095

7. Kroneis T, Jonasson E, Andersson D, Dolatabadi S, Stahlberg A. Global preamplification simplifies targeted mRNA quantification. Sci Rep. 2017; 7:45219. https://doi.org/10.1038/srep45219 PMID: 28332609

8. Taniguchi K, Kajiyama T, Kambara H. Quantitative analysis of gene expression in a single cell by qPCR. Nat Methods. 2009; 6(7):503–6. https://doi.org/10.1038/nmeth.1338 PMID: 19525960.

9. Hartmann CH, Klein CA. Gene expression profiling of single cells on large-scale oligonucleotide arrays. Nucleic Acids Res. 2006; 34(21):e143. https://doi.org/10.1093/nar/gkl740 PMID: 17071717

10. Kurimoto K, Yabuta Y, Ohinata Y, Ono Y, Uno KD, Yamada RG, et al. An improved single-cell cDNA amplification method for efficient high-density oligonucleotide microarray analysis. Nucleic Acids Res. 2006; 34(5):e42. https://doi.org/10.1093/nar/gkl050 PMID: 16547197

11. Kamme F, Salunga R, Yu J, Tran DT, Zhu J, Luo L, et al. Single-cell microarray analysis in hippocampus CA1: demonstration and validation of cellular heterogeneity. J Neurosci. 2003; 23(9):3607–15. PMID: 12736331.

12. Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, et al. Quantitative assessment of single-cell RNA-sequencing methods. Nat Methods. 2014; 11(1):41–6. https://doi.org/10.1038/nmeth.2694 PMID: 24141493

13. Svensson V, Natarajan KN, Ly LH, Miragaia RJ, Labalette C, Macaulay IC, et al. Power analysis of single-cell RNA-sequencing experiments. Nat Methods. 2017; 14(4):381–7. https://doi.org/10.1038/nmeth.4220 PMID: 28263961

14. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell. 2017; 65(4):631–43 e4. https://doi.org/10.1016/j.molcel.2017.01.023 PMID: 28212749.

15. Haque A, Engel J, Teichmann SA, Lonnberg T. A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications. Genome Med. 2017; 9(1):75. https://doi.org/10.1186/s13073-017-0467-4 PMID: 28821273

16. Kolodziejczyk AA, Lonnberg T. Global and targeted approaches to single-cell transcriptome characterization. Brief Funct Genomics. 2018; 17(4):209–19. https://doi.org/10.1093/bfgp/elx025 PMID: 29028866

17. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, et al. Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res. 2011; 21(7):1160–7. https://doi.org/10.1101/gr.110882.110 PMID: 21543516

18. Islam S, Kjallquist U, Moliner A, Zajac P, Fan JB, Lonnerberg P, et al. Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. Nat Protoc. 2012; 7(5):813–28. https://doi.org/10.1038/nprot.2012.022 PMID: 22481528.

19. Hashimshony T, Wagner F, Sher N, Yanai I. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep. 2012; 2(3):666–73. https://doi.org/10.1016/j.celrep.2012.08.003 PMID: 22939981.

20. Potter SS. Single-cell RNA sequencing for the study of development, physiology and disease. Nat Rev Nephrol. 2018; 14(8):479–92. https://doi.org/10.1038/s41581-018-0021-7 PMID: 29789704

21. Fan HC, Fu GK, Fodor SP. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. Science. 2015; 347(6222):1258367. https://doi.org/10.1126/science.1258367 PMID: 25657253.

22. Nakamura T, Yabuta Y, Okamoto I, Aramaki S, Yokobayashi S, Kurimoto K, et al. SC3-seq: a method for highly parallel and quantitative measurement of single-cell gene expression. Nucleic Acids Res. 2015; 43(9):e60. https://doi.org/10.1093/nar/gkv134 PMID: 25722368

23. Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, et al. Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science. 2014; 343(6172):776–9. https://doi.org/10.1126/science.1247651 PMID: 24531970

24. AlJanahi AA, Danielsen M, Dunbar CE. An Introduction to the Analysis of Single-Cell RNA-Sequencing Data. Mol Ther Methods Clin Dev. 2018; 10:189–96. https://doi.org/10.1016/j.omtm.2018.07.003 PMID: 30094294

25. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001; 29(9):e45. https://doi.org/10.1093/nar/29.9.e45 PMID: 11328886

26. Rutledge RG, Cote C. Mathematics of quantitative kinetic PCR and the application of standard curves. Nucleic Acids Res. 2003; 31(16):e93. https://doi.org/10.1093/nar/gng093 PMID: 12907745

27. Stahlberg A, Kubista M. Technical aspects and recommendations for single-cell qPCR. Mol Aspects Med. 2018; 59:28–35. https://doi.org/10.1016/j.mam.2017.07.004 PMID: 28756182.

28. Pfaffl MW, Horgan GW, Dempfle L. Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. 2002; 30(9):e36. https://doi.org/10.1093/nar/30.9.e36 PMID: 11972351

29. Klein CA, Seidl S, Petat-Dutter K, Offner S, Geigl JB, Schmidt-Kittler O, et al. Combined transcriptome and genome analysis of single micrometastatic cells. Nat Biotechnol. 2002; 20(4):387–92. https://doi.org/10.1038/nbt0402-387 PMID: 11923846.

30. VerMilyea MD, Maneck M, Yoshida N, Blochberger I, Suzuki E, Suzuki T, et al. Transcriptome asymmetry within mouse zygotes but not between early embryonic sister blastomeres. EMBO J. 2011; 30 (9):1841–51. https://doi.org/10.1038/emboj.2011.92 PMID: 21468028

31. Suzuki T, Asami M, Hoffmann M, Lu X, Guzvic M, Klein CA, et al. Mice produced by mitotic reprogramming of sperm injected into haploid parthenogenotes. Nat Commun. 2016; 7:12676. https://doi.org/10.1038/ncomms12676 PMID: 27623537

32. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends Genet. 2013; 29(10):569–74. https://doi.org/10.1016/j.tig.2013.05.010 PMID: 23810203.

33. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006; 10(6):515–27. https://doi.org/10.1016/j.ccr.2006.10.008 PMID: 17157791

34. Subik K, Lee JF, Baxter L, Strzepek T, Costello D, Crowley P, et al. The Expression Patterns of ER, PR, HER2, CK5/6, EGFR, Ki-67 and AR by Immunohistochemical Analysis in Breast Cancer Cell Lines. Breast Cancer (Auckl). 2010; 4:35–41. PMID: 20697531

35. Devonshire AS, Elaswarapu R, Foy CA. Applicability of RNA standards for evaluating RT-qPCR assays and platforms. BMC Genomics. 2011; 12:118. https://doi.org/10.1186/1471-2164-12-118 PMID: 21332979

36. Bengtsson M, Hemberg M, Rorsman P, Stahlberg A. Quantification of mRNA in single cells and modelling of RT-qPCR induced noise. BMC Mol Biol. 2008; 9:63. https://doi.org/10.1186/1471-2199-9-63 PMID: 18631407

37. Figueredo DS, Barbosa MR, Coimbra DG, Dos Santos JLA, Costa EFL, Koike BDV, et al. Usual normalization strategies for gene expression studies impair the detection and analysis of circadian patterns. Chronobiol Int. 2018; 35(3):378–91. https://doi.org/10.1080/07420528.2017.1410168 PMID: 29219623.

38. Blasi T, Buettner F, Strasser MK, Marr C, Theis FJ. cgCorrect: a method to correct for confounding cell-cell variation due to cell growth in single-cell transcriptomics. Phys Biol. 2017; 14(3):036001. https://doi.org/10.1088/1478-3975/aa609a PMID: 28198357.

39. Bower NI, Moser RJ, Hill JR, Lehnert SA. Universal reference method for real-time PCR gene expression analysis of preimplantation embryos. Biotechniques. 2007; 42(2):199–206. https://doi.org/10.2144/000112314 PMID: 17373485.

40. Maurin M. Real-time PCR as a diagnostic tool for bacterial diseases. Expert Rev Mol Diagn. 2012; 12(7):731–54. https://doi.org/10.1586/erm.12.53 PMID: 23153240.

41. Stahlberg A, Rusnakova V, Kubista M. The added value of single-cell gene expression profiling. Brief Funct Genomics. 2013; 12(2):81–9. https://doi.org/10.1093/bfgp/elt001 PMID: 23393397.

42. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 2006; 4(10):e309. https://doi.org/10.1371/journal.pbio.0040309 PMID: 17048983

43. Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, et al. Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc Natl Acad Sci U S A. 2012; 109(43):17454–9. https://doi.org/10.1073/pnas.1213530109 PMID: 23064634

44. Chubb JR, Trcek T, Shenoy SM, Singer RH. Transcriptional pulsing of a developmental gene. Curr Biol. 2006; 16(10):1018–25. https://doi.org/10.1016/j.cub.2006.03.092 PMID: 16713960

45. Raj A, van Oudenaarden A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell. 2008; 135(2):216–26. https://doi.org/10.1016/j.cell.2008.09.050 PMID: 18957198

46. Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. Imaging individual mRNA molecules using multiple singly labeled probes. Nat Methods. 2008; 5(10):877–9. https://doi.org/10.1038/nmeth.1253 PMID: 18806792

47. Larson DR, Singer RH, Zenklusen D. A single molecule view of gene expression. Trends Cell Biol. 2009; 19(11):630–7. https://doi.org/10.1016/j.tcb.2009.08.008 PMID: 19819144

48. Kaufmann BB, van Oudenaarden A. Stochastic gene expression: from single molecules to the proteome. Curr Opin Genet Dev. 2007; 17(2):107–12. https://doi.org/10.1016/j.gde.2007.02.007 PMID: 17317149.

49. Nam DK, Lee S, Zhou G, Cao X, Wang C, Clark T, et al. Oligo(dT) primer generates a high frequency of truncated cDNAs through internal poly(A) priming during reverse transcription. Proc Natl Acad Sci U S A. 2002; 99(9):6152–6. https://doi.org/10.1073/pnas.092140899 PMID: 11972056