Analytical Biochemistry 452 (2014) 103–113

Contents lists available at ScienceDirect

Analytical Biochemistry

journal homepage: www.elsevier.com/locate/yabio

Validation of high-throughput single cell analysis methodology

Alison S. Devonshire *, Marc-Olivier Baradez, Gary Morley, Damian Marshall, Carole A. Foy

LGC, Teddington, Middlesex TW11 0LY, UK

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form 27 February 2014
Accepted 1 March 2014
Available online 11 March 2014

Keywords:
Single cell
RT–qPCR
mRNA
Digital PCR
Gene expression
Normalization

ABSTRACT

High-throughput quantitative polymerase chain reaction (qPCR) approaches enable profiling of multiple genes in single cells, bringing new insights to complex biological processes and offering opportunities for single cell-based monitoring of cancer cells and stem cell-based therapies. However, workflows with well-defined sources of variation are required for clinical diagnostics and testing of tissue-engineered products. In a study of neural stem cell lines, we investigated the performance of lysis, reverse transcription (RT), preamplification (PA), and nanofluidic qPCR steps at the single cell level in terms of efficiency, precision, and limit of detection. We compared protocols using a separate lysis buffer with cell capture directly in RT–PA reagent. The two methods were found to have similar lysis efficiencies, whereas the direct RT–PA approach showed improved precision. Digital PCR was used to relate preamplified template copy numbers to C_q values and reveal where low-quality signals may affect the analysis. We investigated the impact of calibration and data normalization strategies as a means of minimizing the impact of inter-experimental variation on gene expression values and found that both approaches can improve data comparability. This study provides validation and guidance for the application of high-throughput qPCR workflows for gene expression profiling of single cells.

Introduction

Single cell analysis has yielded new insights into biological phenomena that are driven by cell populations composed of different cell types, such as in colon cancer [1], or cells responding heterogeneously to a given stimulus, such as T-cell responses to vaccines [2]. In addition to messenger RNA (mRNA)[1] expression profiling using reverse transcription (RT)–quantitative polymerase chain reaction (qPCR), microarray profiling, and RNA–Seq [3–5], methodologies to study genetic variation [6,7], DNA methylation [8], and protein and metabolite levels [9] have also been developed. As single cell analysis moves from being a research tool closer to applications in clinical diagnostics [10,11] and regenerative medicine [12,13], the molecular assays used to screen for disease- and tissue-related traits require stringent validation in order to meet criteria for approval as in vitro diagnostic tests [14] as well as means for ongoing standardization and quality control (QC). Measurement performance characteristics such as linearity, precision, and limit of detection, as well as appropriate controls, are all important aspects of clinical diagnostic tests or assays to assess the quality and consistency of tissue-engineered and stem cell products [15].

RT–qPCR is a key methodology for accurate and precise measurement of mRNA biomarkers and is already used for clinical monitoring and disease stratification [16,17]. A key development enabling high-throughput expression analysis of single cells is the use of microfluidic qPCR arrays that combine screening of hundreds of gene targets with the quantitative accuracy and dynamic range offered by RT–qPCR [18]; this has been applied to research in a wide range of fields, including cancer development, neurology, and stem cell biology [1,19,20]. The impact of RT, preamplification (PA), and qPCR steps on assay precision has been characterized for single cell RT–qPCR using standard microliter volume qPCR

* Corresponding author. Fax: +44 (0) 208 9432767.
E-mail address: alison.devonshire@lgcgroup.com (A.S. Devonshire).

[1] *Abbreviations used:* mRNA, messenger RNA; RT, reverse transcription; qPCR, quantitative polymerase chain reaction; QC, quality control; PA, preamplification; dPCR, digital PCR; LOD, limit of detection; 4-OHT, 4-hydroxy-tamoxifen; DMEM, Dulbecco's modified Eagle's medium; PBS, phosphate-buffered saline; CE, cell equivalents; EDTA, ethylenediaminetetraacetic acid; B2M, β-2-microglobulin; PPIA, peptidylprolyl isomerase A; RPLP0, large ribosomal protein 0; C_q, quantification cycle; NCC, no cell control; NTC, no template control; RT–minus, reverse transcriptase–negative; cDNA, complementary DNA; RQ, relative quantity; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; TBP, TATA-binding protein; YWHAZ, tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide; DCX, doublecortin; GATA6, GATA binding protein 6; GFAP, glial fibrillary acidic protein; MAPT, microtubule-associated protein tau; MASH1, mammalian achaete–scute complex homolog 1; NCAM1, neural cell adhesion molecule 1; NEFL, neurofilament, light polypeptide; NES, nestin; NGN1, neurogenin 1; SOX2, SRY (sex determining region Y)–box 2; GOI, gene of interest; RG, reference gene; SD, standard deviation; CV, coefficient of variation; LCM, laser capture microscopy; gDNA, genomic DNA; MIQE, minimum information for publication of quantitative real-time PCR experiments; FD, fold difference.

http://dx.doi.org/10.1016/j.ab.2014.03.001
0003-2697/

instruments [21,22]; however, to our knowledge, sources of technical variability within the entire workflow of high-throughput single cell analysis have not been investigated.

We previously investigated the accuracy, linearity, and precision of high-throughput nanofluidic qPCR platforms for gene expression biomarker analysis with conventional real-time platforms [23] as well as the performance of different RT–qPCR and PA protocols for single cell analysis [24]. Following on from this work, in the current study we investigate the precision of RT–PA and nanofluidic qPCR for a high-throughput single cell analysis approach. We investigate differences between alternative protocols with a separate lysis and DNase step compared with capture of the single cell directly in RT–PA buffer. We also demonstrate how digital PCR (dPCR) can be used for validation of assay performance characteristics such as limit of detection (LOD).

We draw on a model relevant to the fields of stem cell biology and regenerative medicine by applying the validation approaches to the measurement of single cells from two human neural stem cell lines: CTX0E03 and CTX0E16. Both cell lines were generated from stem cells from the same donor; however, CTX0E03 cells have been shown to be effective in a rat model of ischemic stroke [25] and are now in clinical trials for treatment of stroke-related disability [26], whereas CTX0E16 cells do not show clinical efficacy. We use data from a comparison study of the two cell lines at the single cell level (further results of which will be published elsewhere) to assess the impact of different mRNA quantification strategies, including calibration by a standard curve, reference gene, and global normalization.

Materials and methods

Cell culture

Sister neural stem cell lines CTX0E03 and CTX0E16 were provided by ReNeuron (Guildford, UK). The cell lines were established from somatic stem cells in the cortical neuroepithelium of the same donor and immortalized with the c-mycER^{TAM} gene, which is conditionally expressed in the presence of 4-hydroxy-tamoxifen (4-OHT) and enables cell expansion. In the absence of 4-OHT and growth factors (see below), cells undergo differentiation into neural cell types [25].

CTX0E03 cells (passage 9) and CTX0E16 cells (passage 16) were maintained in laminin-coated flasks (Sigma) in RMM medium [Dulbecco's modified Eagle's medium (DMEM)/F12 medium (Gibco) containing 0.03% human serum albumin (VWR), 100 μg/ml human apo-transferrin (Sigma), 16.2 μg/ml putrescine dihydrochloride (Sigma), 5 μg/ml insulin (Sigma), 60 ng/ml progesterone (Sigma), 2 mM L-glutamine (Sigma), 40 ng/ml sodium selenite (Sigma), 10 ng/ml basic fibroblast growth factor (Peprotech), 20 ng/ml epidermal growth factor (Peprotech), and 100 mM 4-OHT (Sigma)] at 37 °C and 5% CO_2.

To prepare bulk RNA from CTX cells under proliferating and differentiated conditions, a time course experiment was performed. CTX cells were revived from liquid nitrogen into six T75 flasks and maintained in RMM medium (20 ml of medium per flask with growth factors and 4-OHT as above) with medium renewal every 2 to 3 days. When the cells reached 80% confluence, one flask of cells (representing the basal condition) was rinsed with 10 ml of phosphate-buffered saline (PBS) without calcium or magnesium (PAA, part no. H15-002) and cells were dissociated with 4 ml of TrypZean, a recombinant form of trypsin that is free from animal source contaminants (Lonza), for 5 min at 37 °C. TrypZean was neutralized by the addition of 8 ml of Trit inhibitor (DMEM/F12 medium containing 0.044% human serum albumin (VWR), 0.55 mg/ml trypsin inhibitor (Sigma), and 0.25 units of Benzonase (Merck)), and the cells were counted using an automated cell counter (Vi-Cell XL, Beckman Coulter). The cells were centrifuged at 1500 rpm in an Eppendorf 5702 benchtop centrifuge, and the pooled cells were adjusted to a concentration of 1×10^6 cells/ml with PBS. The cells were recounted as before to ensure that the correct cell density had been achieved. The cell suspension was processed according to Section 'RNA isolation and preparation of reference RNA' in order to prepare cell pellets and cell lysates. The remaining flasks were maintained in culture as before until the cells reached confluency. At this point, an additional flask was harvested as above, and this represented "day 0" cells. The remaining flasks were washed with 10 ml of PBS per flask, and the medium was changed to differentiation medium (RMM medium without growth factors or 4-OHT) with medium renewal every 2 to 3 days. Further flasks were harvested for RNA isolation every 7 days (7, 14, 21, and 28 days post-confluence and growth factor withdrawal).

Laser capture microscopy

CTX0E03 or CTX0E16 cells were revived from liquid nitrogen into laminin-coated T75 flasks containing RMM medium (above) and maintained in culture for 3 days. After this time, the cells were rinsed 1× with PBS (PAA) and dissociated with TrypZean (as in Section 'Cell culture'). Cell density was assayed (as in Section 'Cell culture'), and 2×10^5 cells were transferred to laminin-coated PET laser dissection microscope slides (Zeiss) and allowed to adhere for 1 h at 37 °C and 5% CO_2. After this time, the medium was replaced with 1 ml of fresh RMM medium and the cells were placed at 37 °C and 5% CO_2 for 18 h. The medium was removed and replaced with 1 ml of RMM medium containing 4 μM calcein AM (Invitrogen). Cells were incubated at 37 °C and 5% CO_2 for 1 h to allow fluorescence to develop. The medium was removed from the cells, and the cells were rinsed once with 1 ml of PBS before fixing in 95% ethanol (Fisher)/5% acetic acid (Sigma) at −20 °C. Following 10 min of incubation at room temperature, the fixative was removed and the slides were air-dried before laser dissection using a Zeiss PALM Laser Capture Microscope. Cells were selected on the basis of displaying calcein fluorescence (excitation 488 nm/emission 516 nm) as an indicator of viability. (Only live cells convert nonfluorescent calcein AM to fluorescent calcein through the action of intracellular esterases.)

After cell selection, but prior to cutting, 15 μl of Cells Direct lysis buffer or RT–PA mix (see Sections 'Single cell processing and RT–PA (single cells captured in lysis buffer) and Single cell processing and RT–PA (single cells captured in RT–PA buffer)') was added to wells of a 96-well capture plate (Zeiss) and individual cells were dissected and catapulted into wells containing the mix. The collection plate was placed on top of a 96-well plate, and the samples were transferred to the plate by centrifuging for 1 min at 1000 rpm in a Jouan CR412l centrifuge. Plates were then sealed using a silicone sealing mat (Web Scientific) and placed on ice until heating (lysis buffer) or RT–PA. Cells captured in lysis buffer were heated at 75 °C for 15 min (PTC 225 Tetrad PCR System, MJ Research) prior to freezing at −80 °C. Cells captured in RT–PA buffer were processed according to Section 'Single cell processing and RT–PA (single cells captured in RT–PA buffer)'.

RNA isolation and preparation of reference RNA

For isolation of purified RNA from CTX0E03 or CTX0E16 cell lines, cells were cultured as in Section 'Cell culture'. Following cell detachment, cells were resuspended in PBS at a concentration of 10^6 cells/ml. Aliquots (1 ml) were prepared, and the cells were pelleted by centrifugation at 13,500 rpm in an MSE MicroCentaur benchtop centrifuge for 2 min. The supernatant was removed, and cell pellets were stored at −80 °C until required. Alternatively,

40 μl of the cell suspension was added to 440 μl of Cells Direct lysis buffer (Invitrogen) consisting of 400 μl of Resuspension Buffer and 40 μl of Lysis Enhancer (final concentration 83 cells/μl).

RNA was isolated from cell pellets using an RNeasy Minikit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Purified RNA was quantified by ultraviolet (UV) spectroscopy (Nanodrop, Thermo Scientific). A_{260}/A_{280} ratios of more than 2.0 were obtained for all samples. RNA integrity was analyzed using a 2100 Bioanalyzer (Agilent). RNA integrity number (RIN) values of 10.0 were obtained for all samples. For the comparison of alternative workflows (see Section 'Impact of a separate lysis step on measurement precision and efficiency' in Results), purified CTX0E03 total RNA was diluted in Cells Direct lysis buffer or nuclease-free water (workflow A or B, respectively) and 450 pg of RNA was added per RT–PA reaction (Sections 'Single cell processing and RT–PA (single cells captured in lysis buffer) and Single cell processing and RT–PA (single cells captured in RT–PA buffer)').

For the preparation of reference RNA, equal quantities (500 ng) of purified RNA from CTX0E03 and CTX0E16 under basal growth conditions or following withdrawal of growth factors (1, 2, 3, and 4 weeks) were mixed and stored in aliquots at −80 °C. The reference RNA was diluted to 7500, 750, 75, and 7.5 pg total RNA/μl (nominal values 10^4, 10^3, 10^2, and 10^1 cell equivalents (CE)) with nuclease-free water, and 5 μl of each dilution was added to RT–PA reactions (Section 'Single cell processing and RT–PA (single cells captured in RT–PA buffer)').

qPCR assays

Details of qPCR assays used in this study are provided in Supplementary Table 1 (see online supplementary material). Final concentration of each primer and probe for all assays was 900 and 250 nM, respectively (defined as 1× concentration). For target PA, a multiplex mix of all assays was prepared by mixing equal volume of all assays (final concentration 0.05×). Assay performance encompassing the entire analytical workflow, including cell lysis, RT–PA, and qPCR (Fig. 1), was defined using a dilution series of CTX0E16 cell lysate (see Section 'Contribution of RT–PA and qPCR to assay linearity and precision' in Results). A dilution series was prepared in Cells Direct lysis buffer containing the equivalent of 1000, 100, 10, 1, and 0.1 cells (CE) per RT–PA reaction ($n = 3$). RT–PA–qPCR efficiency was calculated based on the slope of the RNA dilution series, and LOD was defined for 18 assays that were expressed in CTX0E03 and CTX0E16 cells (Supplementary Table 2).

Single cell processing and RT–PA (single cells captured in lysis buffer)

Single cells captured in lysis buffer (Fig. 1, workflow A) were processed as follows based on the Cells Direct One-Step RT–qPCR manual (revision date: 11 June 2010, part no. 25-0870). A total of 52 single CTX0E03 cells (excluding negative controls) (see Section 'Single cell negative controls') were collected and processed for the comparison of different workflows. Samples were DNase treated using DNase supplied with the Cells Direct kit by the addition of 1.0 μl DNase and 1.8 μl 10× DNase buffer to a final volume of 17.8 μl and incubation of the samples for 5 min at room temperature. The reaction was stopped by the addition of 4 μl of 25 mM ethylenediaminetetraacetic acid (EDTA) and heated for 10 min at 70 °C (GeneAmp PCR System 9700 PCR, Applied Biosystems). The entire DNase-treated lysate was used for RT–PA in a final reaction volume of 52 μl. RT–PA reactions consisted of 2× Cells Direct reaction buffer, SuperScript III RT/Platinum *Taq* (1 μl), and the multiplex assay mix such that each primer pair was present at a final concentration of 0.05× (45 nM). RT–PA thermal cycling conditions were as follows: 50 °C for 15 min (RT), followed by 95 °C for 2 min, followed by 18 cycles of 95 °C for 15 s and 60 °C for 4 min (PTC 225 Tetrad PCR System, MJ Research). Reactions were diluted 1:5 in Tris–EDTA (pH 8.0) (Fluka) and stored at −20 °C.

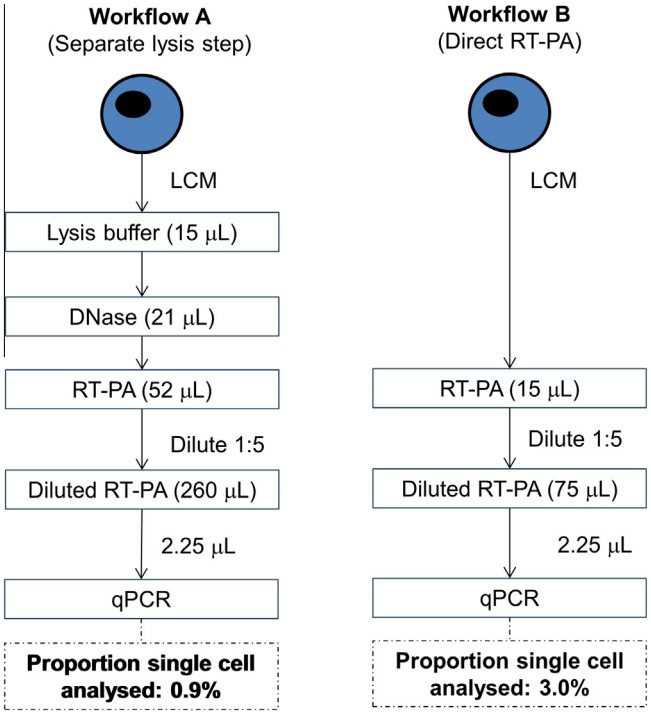

Fig.1. Alternative single cell analytical workflows. Workflow A: single cells are captured in a separate lysis reagent and undergo cell lysis followed by DNase treatment prior to RT–PA and qPCR (BioMark). Workflow B: single cells are captured and processed directly in RT–PA reagent prior to qPCR. Volume restrictions associated with each stage mean that a smaller percentage of single cell material can be analyzed by qPCR for workflow A compared with workflow B.

Single cell processing and RT–PA (single cells captured in RT–PA buffer)

Single cells captured in RT–PA buffer (Fig. 1, workflow B) were processed according to BioMark Advanced Development Protocol Number 5 (Fluidigm) [27] with an increased reaction volume of 15 μl. Final primer concentrations and thermal cycling conditions were as specified for single cells captured in lysis buffer (Section 'Single cell processing and RT–PA (single cells captured in lysis buffer)'). Reactions were diluted 1:5 in Tris–EDTA (pH 8.0) (Fluka) following RT–PA and stored at −20 °C. A total of 42 single CTX0E03 cells (excluding negative controls) (Section 'Single cell negative controls') were collected for the comparison of different workflows. In a separate study comparing the expression profiles of the two cell lines, 117 CTX0E03 single cells and 117 CTX0E16 single cells were collected in three independent experiments. Following exclusion of samples due to no gene expression being detected for any target or reference gene expression (β-2-microglobulin (B2M), peptidylprolyl isomerase A (PPIA), and large ribosomal protein 0 (RPLP0)) being absent or more than 5 quantification cycle (C_q) units higher than the experimental mean, 15 CTX0E03 samples and 20 CTX0E16 samples were excluded from the dataset (Supplementary Table 3).

Single cell negative controls

No cell control (NCC) and no template control (NTC) reactions containing 15 μl of lysis buffer (Section 'Single cell processing and RT–PA (single cells captured in lysis buffer)') or 15 μl of Cells Direct RT–PA buffer (Section 'Single cell processing and RT–PA

(single cells captured in RT–PA buffer)') were included in each 96-well single cell collection plate. The NCC consisted of a section of the microscope slide surface that did not contain a fluorescently labeled single cell (Section 'Laser capture microscopy'). No template material was added to the NTC well as a control for contamination at the cell capture and RT–PA stages. A reverse transcriptase-negative (RT–minus) control reaction consisted of a single cell sample processed using an RT–PA reaction containing 2 units of Platinum *Taq* DNA polymerase (Invitrogen) in place of SuperScript III RT/Platinum *Taq* mix.

qPCR analysis (BioMark Dynamic Arrays)

Preamplified complementary DNA (cDNA) samples were analyzed by qPCR using BioMark 48.48 Dynamic Arrays and the BioMark System according to the manufacturer's instructions as described previously [24]. Preamplified cDNA (2.25 μl) was added per sample inlet of the Dynamic Array. Duplicate qPCR reactions were performed consisting of replicate assay inlets. A further NTC (Section 'Single cell negative controls') was included in each qPCR experiment consisting of nuclease-free water added instead of preamplified template in one sample inlet of the Dynamic Array. Thermal cycling conditions consisted of 50 °C for 2 min, 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. C_q data were generated using BioMark Real-Time PCR Analysis software (version 3.0.2) with detector-specific automatic threshold (see Supplementary Table 4) and linear baseline correction settings and a default quality threshold of 0.65. Data were exported for processing using Microsoft Excel 2003/2007.

dPCR analysis (BioMark Digital PCR Arrays)

BioMark 12.765 Digital Array Integrated Fluidic Circuits (IFCs) (Fluidigm) used for dPCR analysis consisted of 12 panels each containing 765 partitions of 6 nl volume. For the analysis of preamplified cDNA samples from single cells, 8-μl digital PCR assays were prepared according to the manufacturer's instructions and consisted of 2× Gene Expression Mastermix (Applied Biosystems), 1× BioMark Sample Loading Reagent (Fluidigm), 1× qPCR assay, and 2.25 μl of sample. Depending on the C_q of the target gene based on the Dynamic Array, the sample was diluted in to an optimal range for dPCR analysis [28]. Thermal cycling conditions consisted of a 10-min heat activation step (at 95 °C), followed by 40 cycles of 15 s at 95 °C, followed by 60 s at 60 °C. dPCR data were processed using BioMark Digital PCR Analysis software (version 3.0.2) with automatic threshold and linear baseline correction settings and a default quality threshold of 0.65. Positive PCR amplifications were counted between cycles 20 and 35.

Data analysis

Statistical analysis of the effect of alternative workflows on target C_q values and RT–minus C_q values was performed by two-way analysis of variance (ANOVA) with Bonferroni post hoc tests using GraphPad Prism (version 5.04).

C_q data from a study of single cells from the CTX0E03 and CTX0E16 cell lines (Fig. S2A in Supplementary material) were processed using Genex Enterprise (version 5.3.6, MultiD Analyses, Göteborg, Sweden) based on guidelines from Stahlberg and et al. [29]; C_q values above 26 were replaced with a C_q cutoff value of 28 (C_q values using BioMark arrays tend to be ~10 units lower than those using the standard qPCR instrument due to the 1000-fold ($\log_2 = 9.96$) higher concentration associated with 10-nl chambers [23]). Values were transformed to relative quantities (RQs) relative to minimum expression values (i.e., for $C_q = 28$, RQ = 1.0), and missing data cells were given the value RQ = 0.5. RQ expression values were $\log_2$ transformed because single cell expression data have been observed to be log-normally distributed [3] and denoted as $\log_2$(Ex).

Calibrated gene expression values based on a reference RNA standard curve (Section 'RNA isolation and preparation of reference RNA') were exported from the BioMark Real-Time PCR Analysis software (version 3.0.2) and further analyzed in Microsoft Excel (2007). Calibrated data were treated in the same way as C_q data; missing values were replaced with a value of 0.5× minimum observed value (resulting in the same relative difference between minimum values and missing data of minus 1 as C_q data on the $\log_2$Ex scale). Calibrated values were $\log_2$ transformed to the same scale as the C_q processed/normalized data.

To test the robustness of different data normalization strategies, processed C_q values ($\log_2$(Ex)) were normalized in Genex using (i) a reference gene or (ii) a global normalization approach by subtracting the mean expression value of (i) six reference gene $\log_2$(Ex) values (B2M, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), PPIA, RPLP0, TATA-binding protein (TBP), and tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide (YWHAZ)) or (ii) all 16 gene targets (B2M, doublecortin (DCX), GAPDH, GATA binding protein 6 (GATA6), glial fibrillary acidic protein (GFAP), microtubule-associated protein tau (MAPT), mammalian achaete–scute complex homolog 1 (MASH1), neural cell adhesion molecule 1 (NCAM1), neurofilament light polypeptide (NEFL), nestin (NES), neurogenin 1 (NGN1), PPIA, RPLP0, SRY (sex determining region Y)–box 2 (SOX2), TBP, and YWHAZ) (for each cell) from the expression from the value for each gene target according to Eq. (1):

$$C_q(\text{norm}) = C_q(\text{GOI}) - \frac{1}{n}\sum_{i=1}^{n} Cq_{\text{RGI}} \quad (1)$$

where C_q(norm) is normalized C_q value, C_q(GOI) is C_q of gene of interest, C_{qRG} is C_q of reference gene (or all gene targets for global normalization), and n is number of reference genes (reference gene normalization) or all gene targets (global normalization).

Expression in CTX0E03 single cells was calculated relative to CTX0E16 single cells (Supplementary Table 7) by subtraction of the mean $\log_2$(Ex) value for CTX0E16 single cells from all single cell data for both cell lines, such that mean $\log_2$(Ex) for CTX0E16 = 0.0.

Results

Contribution of RT–PA and qPCR to assay linearity and precision

To investigate sources of quantification bias and imprecision in single cell analysis reflecting all stages of the high-throughput analytical workflow (Fig. 1), we performed a dilution series of cell lysate to 1000, 100, 10, 1, and 0.1 CE ($n = 3$) based on cell count. This dilution series was used to define the performance of each assay in terms of linearity encompassing both RT–PA and qPCR steps (an example for B2M is given in Fig. 2A, with efficiency information for other assays provided in Supplementary Table 2).

Data from the dilution series corresponding to 1 CE (or LOD if 1 CE was not detected (Supplementary Table 2)) were used to assess measurement precision. Variation between replicate RNA measurements expressed as standard deviation of C_q values (SD (C_q)) was compared with mean C_q value in order to investigate whether target abundance influenced assay variability. Variation between replicate RNA samples for the majority of assays was less than 0.45 C_q unit (equivalent to a coefficient of variation (CV) < 35%) (Supplementary Table 2). The contributions of the two reaction stages (RT–PA and qPCR) to the overall technical error were compared (Fig. 2B). We found that most assays demonstrated a qPCR error of ⩽0.4 C_q SD, although this increased for some assays for

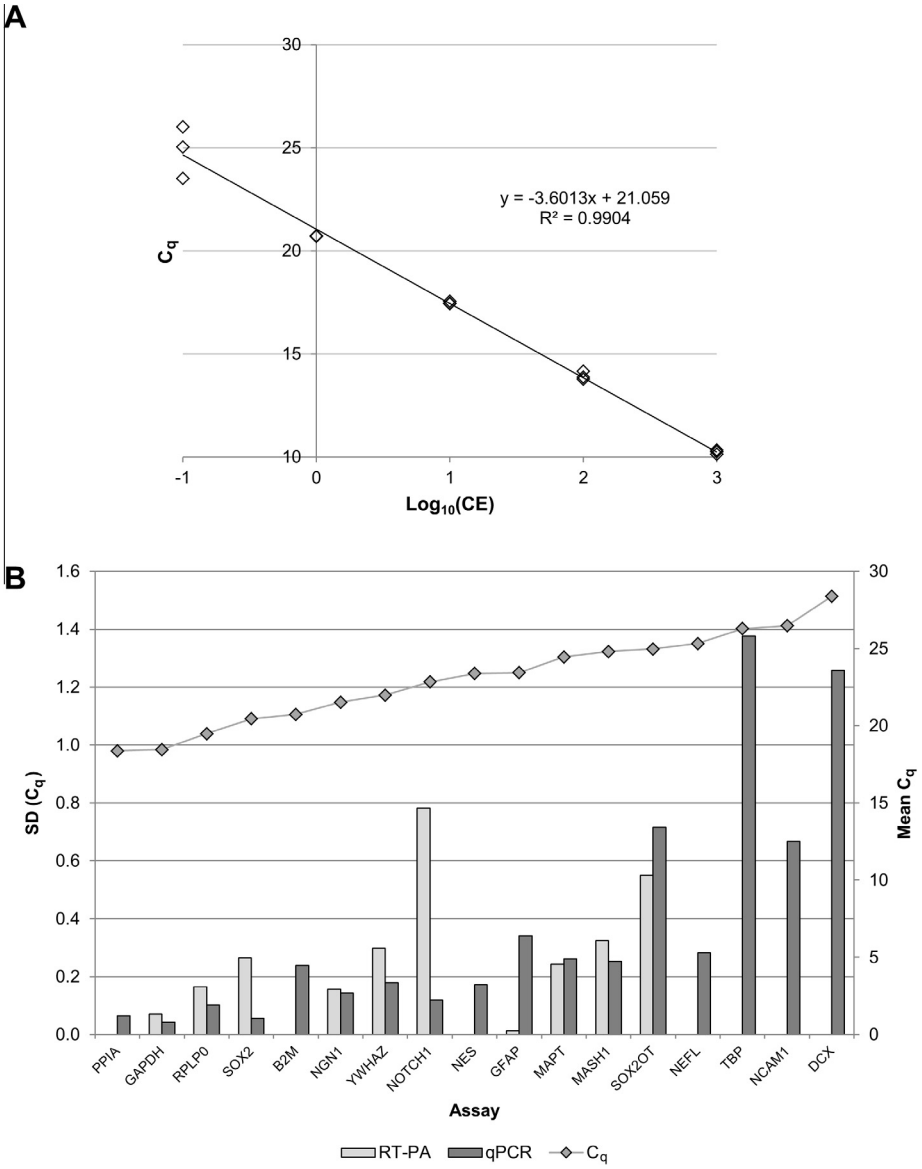

Fig. 2. Evaluation of assay performance using a single cell analytical workflow: Reaction linearity and variability of RT–PA and qPCR reactions. (A) Example of characterization of reaction linearity for target gene (B2M) using a dilution series of CTX0E16 cell lysate to 0.1, 1, 10, 100, and 1000 CE. A slope of –3.60 indicates an efficiency of 90%. C_q values are shown as individual datapoints. (B) Variability of RT–PA and qPCR reactions at the single cell level. RT–PA and qPCR error are displayed for 16 assays based on replicate RT–PA ($n = 3$) and qPCR ($n = 2$ for each RT–PA replicate) reactions. For assays where qPCR error is greater than the overall error, RT–PA error could not be calculated; therefore, only qPCR error is displayed.

target abundances of a mean C_q value of 25 or above. For some assays, qPCR error exceeded overall variation (PPIA, B2M, NES, NEFL, TBP, NCAM1, and DCX), meaning that it was not possible to calculate the error associated specifically with the RT–PA stage [30]. For other assays where qPCR error was less than overall variation, the error attributable to the RT–PA stage was also calculated; in most cases, RT–PA error was $\leqslant$0.3 SD C_q (equivalent to a CV of $\leqslant$23%), and it was found that this component tended to be the dominant component of the overall error for more highly expressed genes ($C_q < 23$).

Impact of a separate lysis step on measurement precision and efficiency

During previous work developing laser capture microscopy (LCM) methodology for single cell analysis, we implemented a protocol whereby the single cells were captured in 15 µl of Cells Direct lysis buffer and samples were subsequently analyzed by one-step RT–qPCR or preamplified using the Cells Direct RT–PCR reagents [24]. The volume of capture buffer used for LCM was previously optimized to 15 µl in order to ensure even coverage of the "wells" of the 96-sample capture lid and avoid significant evaporation during the procedure. Therefore, it was effectively fixed for our cell capture method. The workflow using a separate lysis step (Fig. 1, workflow A) gave the option of DNase treatment of the single cell lysate prior to RT and PA. In contrast, in the standard protocol of single cell analysis using BioMark Dynamic Arrays [27], the single cell was captured directly in the Cells Direct RT–PCR buffer used for RT and PA (Fig. 1, workflow B). It was assumed that cell lysis occurred during the initial incubation at 50 °C (RT phase). The latter protocol (workflow B) was more economical because it used smaller volumes of RT–PCR reagents and offered in theory a greater LOD because a greater percentage of single cell mRNA (3%) was loaded per inlet of a Dynamic Array compared with workflow A (1%)

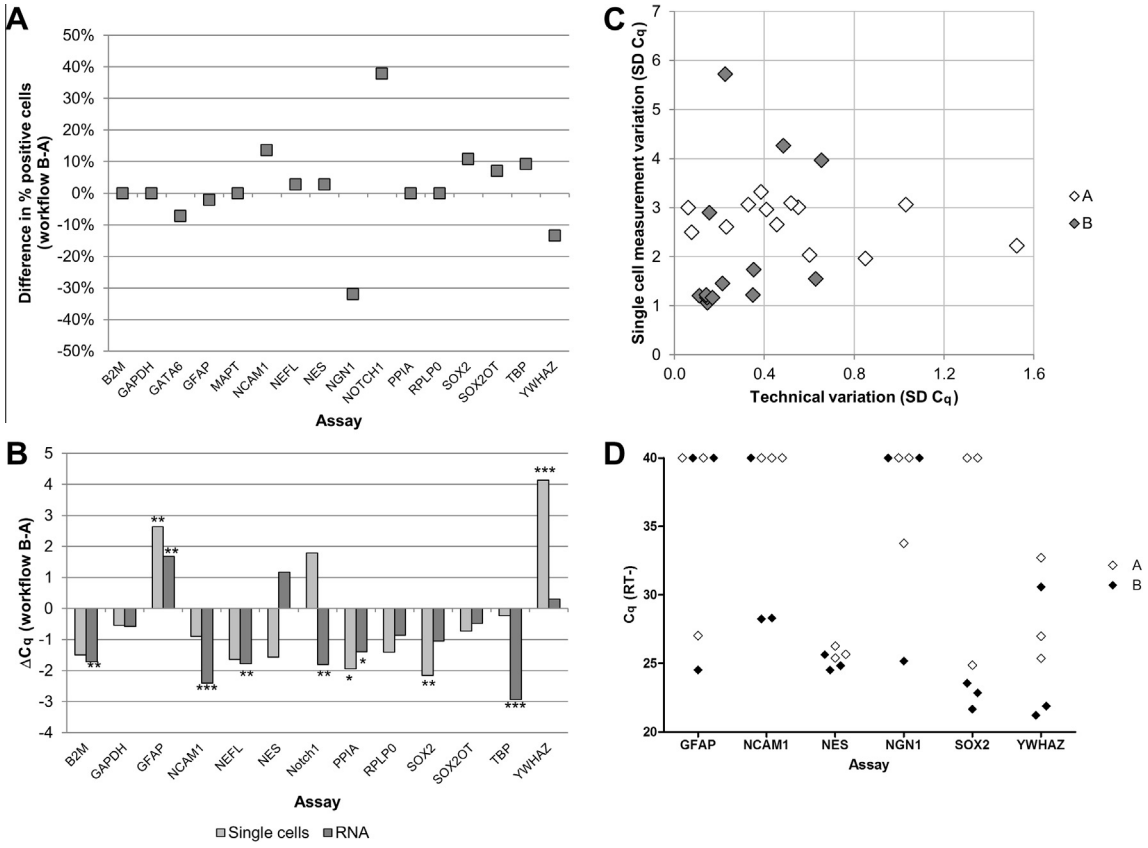

Fig. 3. Comparison of alternative workflows for single cell analysis. Workflows A and B (Fig. 1) were compared between CTX0E03 single cells ($n \sim 50$) and purified total RNA ($n = 3$) was processed using the two workflows. (A) Difference in assay detection rate (percentage of positive single cell samples) between workflows A and B. Negative and positive values indicate a higher percentage of positive cells for workflows A and B, respectively. (B) Difference in mean C_q values between single cell and RNA samples processed using workflow B versus workflow A. Significant differences between the C_q values for the two workflows are indicated by asterisks: $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. (C) Comparison of technical variation (purified RNA reactions) and biological variation (single cell measurements) for the two workflows expressed as SD (C_q). Each assay is depicted as an individual datapoint. (D) RT–minus reactions were performed with single CTX0E03 cells ($n = 3$), and C_q values for each reaction are displayed as individual data points. (Negative reactions were assigned a C_q value of the maximum number of cycles, i.e., 40.)

(Fig. 1). However, the potential disadvantages of workflow B may be that lysis of the cells and release of cytoplasmic mRNA in RT–PA buffer could be less efficient compared with a specific lysis reagent and that a lack of DNase treatment leads to genomic DNA (gDNA) interference in mRNA measurements. Therefore, we investigated the impact of the alternative workflows on lysis efficacy, LOD, precision, and gDNA contamination by comparing results from CTX0E03 single cells and purified RNA samples processed using both protocols.

Initially, we compared the number of failed single cell samples, where no mRNA signal was detected in any of the target genes, between the two protocols as an indicator of differential lysis failure. For both approaches, 10% of cells processed using workflow A (separate lysis) and 12% of cells processed using workflow B (direct RT–PA) were classed as failed samples on this basis (data not shown), which is within the range of the typical LCM collection failure rate (found to be 0–12%) (data not shown). This result suggests that the overall lysis efficacy of both protocols was similar for the cell line under investigation.

We next investigated whether the differences in input between the two workflows resulted in any qualitative differences in the LOD of individual gene targets in single cells (Fig. 3A). The percentages of single cells expressing 16 different target genes were compared for the two approaches. mRNA targets were detected in a similar proportion (±10%) of single cells using both approaches (Fig. 3A), with differences in expression between the two workflows not correlating with C_q of the target gene (Fig. S1A), suggesting that workflow A was not associated with a reduction in LOD compared with workflow B.

The impact of differences in the proportion of input material (Fig. 1) was also assessed by comparing the difference in C_q values (ΔC_q) using the two workflows for both single cells and purified RNA samples (Fig. 3B) for 13 target genes that were expressed in the majority of single cells. For both sample sources, a reduction in mean C_q was apparent for the majority (10/13) of the assays analyzed using workflow B, with four assays (GFAP, YWHAZ, NES (purified RNA only), and Notch1 (single cells only)) demonstrating the opposite pattern. Of the 10 assays demonstrating a reduction in C_q with workflow B, a mean ΔC_q of −1.21 was observed for purified RNA and −1.26 for single cells. This is broadly in agreement with the predicted ΔC_q difference of −1.58 based on the 3-fold higher input quantity in workflow B compared with workflow A (Fig. 1).

Differences between the two workflows in terms of precision were also compared based on replicate measurements of purified RNA and single cells in order to investigate different potential sources of variability, with variation between replicate reactions with purified RNA reflecting technical variation and variation between replicate single cell measurements being composed of both biological differences between cells and technical error. For workflow A the majority of assays (10/13) displayed a technical variation in terms of SD (C_q) < 0.6 for replicate measurements of purified RNA, whereas for workflow B the majority of assays showed an SD (C_q) of <0.4 C_q (Fig. 3C, x axis). The differences in precision appeared to arise from differences in the processing steps

upstream of qPCR because the qPCR error correlated well between the two approaches (slope of 1.00, intercept of −0.02, $R^2 > 0.95$) (Fig. S1B) with the exception of GFAP and YWHAZ, which also showed a decrease in performance based on ΔC_q (Fig. 3B).

The overall variation between single cells, which is composed of both biological and technical sources of variation, was compared between the two methods (Fig. 3C, *y* axis). For the majority of targets (9/13), the variation between single cell measurements was lower using workflow B (SD $C_q \leqslant 2.0$) compared with workflow A, with a median reduction in SD (C_q) of 1.3 for all 13 assays. Because single cells were processed using the two workflows and were collected from the same population, they should not be inherently more variable in one sample set compared with another; it was concluded that the additional variation observed between single cell measurements in workflow A versus workflow B had arisen from differences in technical factors, adding further evidence to the hypothesis that there may be an improvement in technical precision for processing of single cells using workflow B compared with workflow A.

Finally, the influence of absence of DNase treatment in workflow B compared with workflow A was assessed by performing triplicate RT–minus reactions for each method consisting of a single cell sample that was processed using Cells Direct reaction buffer with only *Taq* polymerase instead of both Superscript III and *Taq* polymerase. Of the 16 target assays, 6 were susceptible to gDNA amplification, as evidenced by amplification of the RT–minus controls (GFAP, NCAM1, NES, NGN1, SOX2, and YWHAZ) (Fig. 3D). NCAM1 gDNA amplification was absent in workflow A and present in workflow B, whereas all three NES and YWHAZ replicates showed evidence of gDNA amplification in both workflows. For other targets, a sporadic pattern of RT–minus amplification was observed for both workflows. These findings suggest that gDNA is not completely removed by DNase treatment in workflow A, although the impact of this on RT–minus controls is assay dependent.

Validation of assay LOD using dPCR

Typically, single cell gene expression data are composed of both dichotomous and continuous components [31], that is, whether a cell is positive or negative in its expression of a given marker as well as the expression level of mRNA per cell, reflected in the C_q value in the case of RT–qPCR measurement. Therefore, the LOD of the RT–qPCR method used to profile single cells is particularly important because it determines whether a cell is referred to as positive or negative for a particular mRNA marker, which may form the basis for discriminating between different cell types. We sought to validate the performance of four assays for gene targets that were expressed in less than 100% of single cells in sample sets of approximately 100 CTX0E03 and CTX0E16 single cells (Supplementary Table 3): GATA6, MASH1, NCAM1, and NGN1 (Supplementary Table 5). We hypothesized that, due to the small volume of preamplified sample (4 nl) analyzed per 10-nl reaction chamber of the BioMark 48.48 Dynamic Arrays, low copy numbers of preamplified template from single cell samples may be below the LOD of this workflow. For example, an LOD of a single cDNA molecule per single cell requires efficient PA in order to ensure that the preamplified template is reliably detected by qPCR (see example 1 in Supplementary Table 6). A minimum of 5 copies per Dynamic Array reaction chamber has been calculated as a concentration where Poisson statistics predict that 99.3% of reactions are positive [32]. For a 96.96 Dynamic Array, a minimum of 5 template copies per chamber corresponds to a preamplified sample containing approximately 730 molecules/μl [32], whereas approximately 555 molecules/μl would be required post-PA due to the larger reaction chamber of the 48.48 Dynamic Arrays used here. Therefore, we analyzed selected samples using dPCR that had been denoted as positive or negative using the Dynamic Array qPCR assays, reasoning that the 300-fold larger volume that could be analyzed by dPCR (1.25 μl/panel) would enable the positive/negative nature of the sample to be verified and an absolute copy number value to the template preamplified cDNA molecules to be assigned.

We compared the concurrence between the qPCR and dPCR expression patterns (Table 1). GATA6, MASH1, and NGN1 showed 100% concurrence in terms of positive or negative cell from both qPCR and dPCR analyses. However, positive dPCR amplifications for NCAM1 were observed for four samples that were denoted as negative by qPCR. We also compared the Dynamic Array C_q value of the sample with the predicted number of template copies per Dynamic Array 10-nl reaction chamber based on the dPCR measurement of sample concentration. Samples that were predicted to have a copy number of less than 5 molecules per reaction chamber had a sporadic pattern of expression on the Dynamic Arrays, with only one of the two qPCR replicates being positive and C_q values greater than 27 (Table 1).

We also predicted how many cDNA copies may have been present in the original single cell samples based on the number of preamplified PCR copies obtained (Table 1), modeled on exponential amplification for 17 cycles with the target-specific efficiencies calculated in Section 'Contribution of RT–PA and qPCR to assay linearity and precision' (Supplementary Table 2). cDNA copy numbers per cell of less than 1 copy were predicted for eight samples for NCAM1 and for one sample for NGN1 (Table 1), suggesting that the efficiency of PCR-based PA may have been lower than predicted in these cases. Given the increased detection of NCAM1 in the absence of DNase treatment (Fig. 3D), the NCAM1 signal in single cell samples may be attributable to gDNA contamination, suggesting that for this target a robust strategy in terms of applying C_q data cutoff or RT–minus background subtraction should be employed (discussed further in Discussion).

Quantification approaches for single cell mRNA analysis

Different data analysis approaches for quantification of gene expression in single cells have been applied, including the use of a standard curve composed of RNA or DNA [21,22] and the comparison of relative expression levels based on C_q values [1,31]. In addition, different normalization approaches may be applied to single cell expression data, including normalization to reference genes [18] and to an average of all assayed transcripts (global normalization) [29].

We sought to compare the effects of these different quantification and normalization approaches on single cell data collected from these two neural stem cell lines: CTX0E03 and CTX0E16. The ability of the different data analysis approaches to handle sources of inter-experimental variation was investigated in a series of three independent experiments, each consisting of the LCM collection of 39 CTX0E03 and 39 CTX0E16 single cells (Fig. S2A). A separate 96-well plate was used to collect cells from each cell line and perform RT–PA (Fig. 1, workflow B). Preamplified cDNA samples from each 96-well capture/RT–PA plate were analyzed on a separate 48.48 Dynamic Array. Data from a total of approximately 100 single cells from each cell line were analyzed following sample QC (Supplementary Table 3).

We prepared a reference RNA to act as a calibration material composed of purified total RNA from both proliferating and differentiating CTX0E03 and CTX0E16 cells in order to maximize representation of all transcripts of interest (Section 'RNA isolation and preparation of reference RNA'). A dilution series composed of this reference RNA containing equivalent to approximately 10^4, 10^3, 10^2, and 10^1 cells ($n = 1$) was included for each 96-well plate used

for cell collection and RT–PA. Calibrated expression values were calculated for each Dynamic Array. The calibration curve also provided a useful means to monitor RT–PA–qPCR efficiency between independent RT–PA reactions and qPCR arrays ($n = 6$). RT–PA–qPCR efficiency varied within acceptable parameters of between 90 and 110% for the majority of markers (Fig. S2B).

We measured the expression of three target genes (NEFL, NES, and SOX2), which were expressed at different (NEFL) or similar (NES and SOX2) levels in single cells from the two lines, and calculated their relative expression levels following alternative data processing approaches: (i) non-normalized C_q values, (ii) calibrated values, (iii) C_q values normalized to the mean of the six reference genes (B2M, GAPDH, PPIA, RPLP0, TBP, and YWHAZ), and (iv) globally normalized C_q values (Fig. 4). We compared the impact of the different quantification approaches in terms of the fold difference in expression between the two cell lines and the comparability of single cell data between separate experiments, reflected in SD of expression within single cells from the same cell line (Supplementary Table 7).

Non-normalized and calibrated expression values gave a similar estimation of the fold difference (FD) in NEFL expression in CTX0E03 versus CTX0E16 single cells of $\log_2(\text{Ex}) \sim +4.7$ (~26-fold), whereas reference gene and globally normalized data gave a lower mean of $\log_2(\text{Ex})$ of +3.2 (9.4-fold) and +4.0 (15.6-fold), respectively (Fig. 4A). Variation in NEFL expression in CTX0E16 single cells was similar for all quantification approaches with SD ($\log_2$(Ex)) of approximately 2.0, whereas calculated variation in CTX0E03 NEFL expression was reduced in calibrated data and normalized data (both reference gene and global methods), with a reduction in SD of 0.3 and approximately 0.5 $\log_2(\text{Ex})$, respectively (Supplementary Table 7).

Non-normalized and calibrated expression values were similar for NES and SOX2 expression, with mean expression levels calculated to be slightly higher in CTX0E03 versus CTX0E16 single cells (Supplementary Table 7). Reference gene normalization led to CTX0E03 NES and SOX2 expression being determined to be lower than that of CTX0E16, which is likely due to the higher expression of reference genes in the former cell line. Following global

Table 1
dPCR analysis of single cell preamplified cDNA samples.

Assay	C_q (detection pattern)	PCR copies/µl preamplified cDNA	Predicted PCR copies per 10-nl reaction	Predicted cDNA copies per single cell
GATA6	25.5 (+/+)	3657	15	2
GATA6	25.3 (+/+)	5104	21	3
GATA6	N/D	0	N/A	N/A
GATA6	N/D	0	N/A	N/A
GATA6	N/D	0	N/A	N/A
GATA6	N/D	0	N/A	N/A
GATA6	N/D	0	N/A	N/A
GATA6	N/D	0	N/A	N/A
GATA6	N/D	0	N/A	N/A
GATA6	N/D	0	N/A	N/A
MASH1	27.1 (+/+)	4206	17	12
MASH1	27.3 (+/+)	3347	14	10
MASH1	28.0 (+/−)	1203	5	3
MASH1	26.8 (+/+)	1054	**4**	3
MASH1	N/D	0	N/A	N/A
MASH1	N/D	0	N/A	N/A
MASH1	N/D	0	N/A	N/A
MASH1	N/D	0	N/A	N/A
MASH1	N/D	0	N/A	N/A
MASH1	N/D	0	N/A	N/A
MASH1	N/D	0	N/A	N/A
NCAM1	26.3 (+/+)	5023	20	5
NCAM1	26.6 (+/+)	3325	13	3
NCAM1	27.5 (+/+)	1301	5	1
NCAM1*	N/D	424	**2**	**<1**
NCAM1	27.6 (+/−)	398	**2**	**<1**
NCAM1	29.6 (+/−)	210	**1**	**<1**
NCAM1*	N/D	207	**1**	**<1**
NCAM1	28.5 (+/−)	193	**1**	**<1**
NCAM1	28.7 (+/−)	117	**<1**	**<1**
NCAM1*	N/D	30	**<1**	**<1**
NCAM1*	N/D	3	**<1**	**<1**
NGN1	24.8 (+/+)	3376	14	3
NGN1	25.0 (+/+)	4377	18	4
NGN1	27.4 (+/−)	235	**1**	**<1**
NGN1	N/D	0	0	N/A
NGN1	N/D	0	0	N/A
NGN1	N/D	0	0	N/A
NGN1	N/D	0	0	N/A
NGN1	N/D	0	0	N/A
NGN1	N/D	0	0	N/A
NGN1	N/D	0	0	N/A
NGN1	N/D	0	0	N/A

Note: Preamplified cDNA samples from selected CTX0E03 and CTX0E16 single cells were analyzed by 12.765 Digital PCR arrays, and the concentrations were compared with C_q values measured by Dynamic Arrays ($n = 2$ reactions). Samples that displayed discordant results in detection pattern between Dynamic Arrays are indicated by an asterisk (*). The concentration of the preamplified cDNA samples was used to predict PCR product copy number per 10-nl Dynamic Array reaction chamber and cDNA copy number per sample, based on 17 cycles of exponential amplification with gene-specific efficiencies (Supplementary Table 2). Samples with predicted copy number of less than 5 copies/reaction or less than 1 cDNA copy/cell are highlighted in bold. C_q values: (+/+), both qPCR reactions were positive (mean C_q value given); (+/−) 1 of 2 qPCR reactions was positive (single C_q value given); N/D, not detected (both qPCR replicates were negative); N/A, not applicable (no copies detected by dPCR).

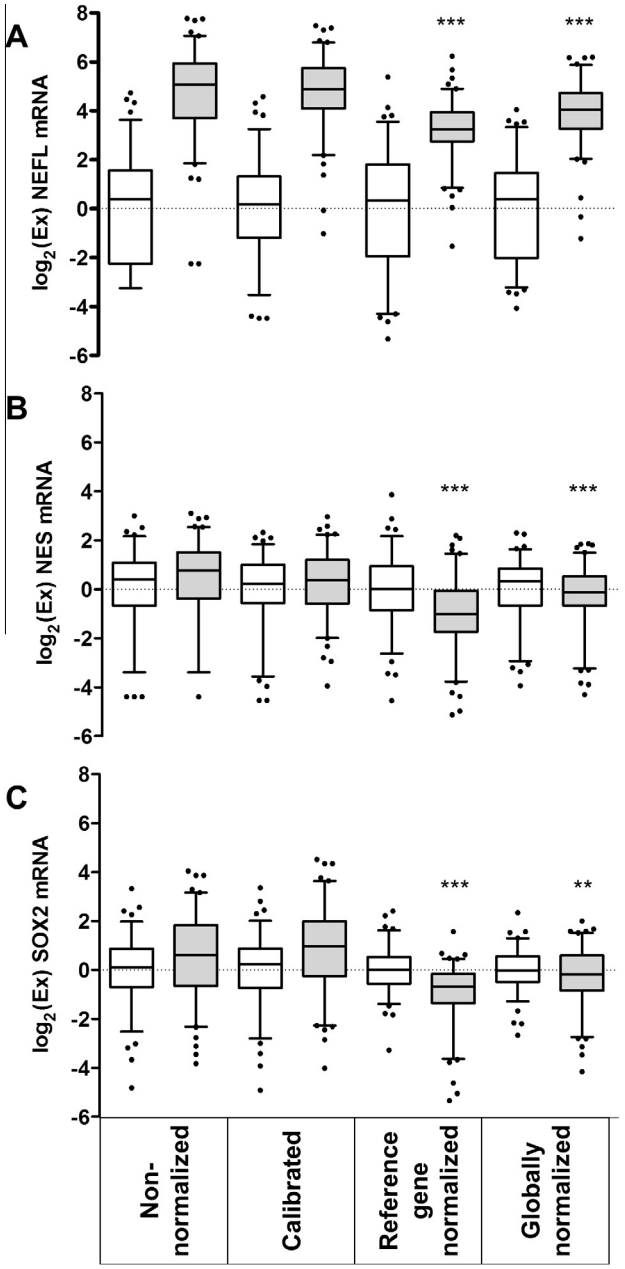

Fig. 4. Comparison of approaches for single cell data quantification and normalization. Expression profiles of three GOIs (A: NEFL; B: NES; C: SOX2) in approximately 100 CTX0E16 and CTX0E03 single cells are compared using four different analysis approaches: (i) non-normalized C_q data, (ii) calibration of C_q values to a reference RNA standard curve, (iii) normalization to the mean C_q of six reference genes, and (iv) global normalization to the average expression value of all gene targets per cell. Expression values are expressed as $\log_2$-transformed values relative to the mean value in CTX0E16 cells with box plots displaying median, inter-quartile range, and 5–95% confidence intervals. Single cell data outside of the 5–95% confidence intervals are shown as individual points. Significant differences in FD in expression (CTX0E03 vs. CTX0E16) compared with the non-normalized C_q values are indicated by asterisks: $^{**}P < 0.01$; $^{***}P < 0.001$).

normalization, mean NES and SOX2 expression was calculated to be the same in single cells from the two sister cell lines. Reference gene and global normalization strategies were also associated with a reduction of the variation of the data for SOX2 expression of 0.4 and 0.5 $\log_2$(Ex) units, respectively, with a more modest reduction in SD being demonstrated for globally normalized NES data of 0.3 units (Supplementary Table 7).

Discussion

Nanofluidic high-throughput qPCR platforms such as the BioMark system differ from other qPCR platforms such as that evaluated by Reiter and coworkers [22] in that the reaction volume is nanoliter scale (10 nl in the case of 48.48 arrays) compared with standard microliter scale qPCR volumes. PA of template is required for expression analysis using the BioMark platform if less than 10^4 mRNA copies/μl are present in a sample [23]; therefore, validation of assay performance for the full analytical workflow, including lysis, RT, PA, and qPCR stages, is fundamental (Fig. 2A and Supplementary Table 2). Our results build on those of others, illustrating that the combined PA and qPCR efficiency of the majority (93%) of assays was more than 90% [32], by demonstrating that the entire workflow (including the RT step that is known to be a major source of variability [33]) can reproducibly produce good linearity with efficiencies between 90 and 110% (Fig. S2B). Our approach of assay validation including the RT step is also in keeping with a primer/probe qualification strategy proposed by Dominguez and coworkers that advocates evaluation of assay performance criteria using an mRNA dilution series [34].

We characterized sources of technical variability using this high-throughput workflow (Fig. 2B) and found that variation in the qPCR tended to become the dominant contributor to overall technical variability with increasing C_q values of measured target. This is in keeping with the findings of others who have noted an increasing proportion of noise as C_q values approach the LOD of the instrument [29,32]. This may originate due to sampling variation at low copy numbers [32] or the stochastic nature of PCR amplification at low copy numbers [35].

The overall precision of single cell measurements was compared between alternative workflows involving cell capture in lysis buffer prior to RT–PA (workflow A) or direct capture in RT–PA reagent (workflow B) (Fig. 1). As demonstrated by a reduction in variation between purified RNA measurements for workflow B (Fig. 3C), it was concluded that more direct processing of single cells by cell capture in RT–PA reagent improved precision, which may be due to fewer sample handling and pipetting stages between cell capture and qPCR analysis. Alternatively, the higher template concentration in workflow B and associated lower C_q values (Fig. 3B) may have led to reduced stochastic variation during RT–PA or qPCR. The higher technical precision observed with workflow B also translated to a decrease in the variability in single cell measurements with this approach (Fig.3C).

The overall variation between measurements of single cells from the same cell line was between 1.0 and 2.0 SD (C_q) for the majority of targets, which is estimated to be 1.0 SD (C_q) higher than the technical variation of $\leqslant$0.4 SD (C_q) observed for the majority of assays (Fig. 3C). This finding confirms recommendations by other authors [29,32] that technical imprecision is generally less than biological variation in single cell experiments and provides justification for a lack of technical replication for single cell experiments, especially for studies where hundreds of targets are screened. However, a thorough initial characterization of assay performance is recommended once a subset of markers has been selected because indicators of cell types or subtypes as outliers can demonstrate above average technical variation (Fig. 3C).

Materials suitable for characterization of assay performance at the single cell level could include total RNA extracted from population of the cells of interest (as used in this study) or RNA from cell lines representative of different cell types [36]. However, due to the heterogeneity of the expression of some markers between different single cells within a population, it may sometimes be problematic to relate the average expression per cell from a population-based sample, such as the reference RNA sample prepared

here, to a single cell expression level. dPCR may provide a useful approach for relating the transcript copies present in such reference RNA samples to those present in a single cell sample by absolute quantification of both reference and single cell samples without the requirement of a standard curve for each target gene [37,38].

Our work extends the efforts of previous authors [22] to make single cell analysis compliant with the MIQE (minimum information for publication of quantitative real-time PCR experiments) guidelines [39], which require disclosure of the properties of the nucleic acid extraction, DNase treatment, and assessment of how residual DNA may be a source of contamination. For most single cell analysis workflows, there is not a separate extraction stage; however, it is important to validate the efficacy of cell lysis. We compared the use of a specific lysis reagent with lysis of the cell on heating during the RT step in this study in order to demonstrate that direct processing in RT–PA reagent (Fig. 1, workflow B) provides complete release of cellular mRNAs (Fig. 3A and B). For this study of neural stem cells, similar lysis efficiencies were achieved with the two protocols. However, use of a specific lysis reagent may be required for complete lysis of alternative cell types or cell matrices in the case of isolation of single cells from whole tissues.

For the MIQE requirement for information on DNase treatment and gDNA controls, we presented an alternative method whereby DNase treatment can be included in a single cell workflow (Fig. 1, workflow A) and used this to investigate the magnitude of gDNA contamination in single cell samples (Fig. 3D). Although redesign of an assay is a possibility for assays that detect gDNA, sometimes in the case of single exon genes such as those studied here (SOX2 and NGN1) or targets for which processed pseudogenes are present in the genome, it is not possible to redesign an assay to span exon–exon boundaries. It was found that the contribution of gDNA to RT–minus amplification varied according to assay, with some targets amplifying only in the absence of DNase treatment (e.g., NCAM1) and other targets showing a more sporadic pattern of detection, being positive in some RT–minus reactions and negative in others (e.g., GFAP), suggesting that DNase treatment is not effective at completely removing the trace amounts of gDNA present in single cell samples. This suggests that the sensitivity to gDNA contamination needs to be evaluated on an assay-specific basis with multiple RT–minus single cell samples (as opposed to purified total RNA extracts) such as those used in the methods incorporated into this study (Fig. 3D).

Evaluation of preamplified samples using dPCR (Table 1) also suggests that assays susceptible to amplification of gDNA may lead to false positive signals that can be difficult to distinguish from genuine mRNA expression at higher C_q values. In the case of assays that are susceptible to gDNA contamination, our results suggest that it is not possible to assign an LOD of 1 cDNA copy per cell for such an assay. Given that 2 copies of a target are present in gDNA and these will undergo a doubling of copy number during the first round of PA, whereas single-stranded cDNA will not, an LOD of 4 copies should be the baseline for analysis of differential mRNA expression of targets that are susceptible to interference from gDNA. This would correspond to a copy number of approximately 11 per 48.48 Dynamic Array reaction chamber given an estimated average RT–PA–qPCR efficiency of 90% (Supplementary Table 6, example 2). The dPCR approach used in this study demonstrates how C_q values can be related to the copies of preamplified template per single cell sample and a C_q cutoff assigned as LOD for each assay in order to calculate the relative expression levels as described by Livak and coworkers [32].

C_q is being increasingly used as an endpoint measurement for high-throughput gene expression analysis [1,31]. We sought to evaluate different quantification and normalization methods for minimizing the effect of differences between experiments that may affect C_q and detract from discerning true biological differences between single cells. We investigated how the calibration of C_q values to a RNA-based standard curve or normalization of C_q values using reference gene or global normalization strategies affected the expression profile of selected GOIs (NEFL, NES, and SOX2) in CTX0E03 and CTX0E16 single cells (Fig. 4 and Supplementary Table 7). The mean FD in expression of the GOIs between the CTX0E03 and CTX0E16 cell lines was modulated significantly by reference gene or global data normalization, but not by standard curve-based calibration of C_q values. The normalization approaches may provide a strategy to remove nonspecific aspects linked to differences in total mRNA levels from the analysis of specific target genes.

Our comparison of normalization strategies also provides evidence for the hypothesis that global normalization could be a means to account for cell-specific variations in lysis and efficiencies of the RT, PA, and qPCR steps [29], which would contribute to variability in single cell data. By performing three independent experiments including both cell culture and analytical stages, we tested the robustness of the data analysis protocols employed (Fig. 4). We found that the variation within single cell data from each cell line reflected by SD ($\log_2$(Ex)) was generally lower when the data were normalized globally, which may be due to normalization of technical differences among the three independent experiments performed or differences linked to cell size (e.g., linked to the cell cycle) as noted above. Calibration of C_q values improved data comparability for two of the three targets analyzed, suggesting that the use of a calibration curve can in some cases compensate for technical factors that may differ between experiments such as qPCR threshold setting and RT efficiency. These findings suggest that calibration and/or normalization strategies may improve the precision of single cell assays for the expression of specific targets in the case of diagnostic or QC tests, where the impact of nonspecific differences (e.g., cell cycle or reagent efficacy) needs to be minimized. Recently, RT–dPCR has emerged as a promising alternative approach for maximizing the precision of single cell measurements of a limited set of mRNA targets, with the possibility of analyzing a greater proportion of a single cell without the requirement for cDNA PA [40,41].

Conclusions

This study has defined sources of technical variability affecting RT, PA, and nanofluidic qPCR using a high-throughput single cell analysis workflow and has shown how an RNA-based standard curve can be used to characterize and monitor reaction efficiency. Investigation of how gDNA may contribute to mRNA signals when cells are analyzed directly in lysis reagents substantiated recommendations to introduce C_q cutoffs into single cell analysis pipelines in order to ensure that nonspecific signals and sampling noise do not obscure differential mRNA expression patterns. We demonstrated that dPCR can be used as a means of validating assay LOD for high-throughput analysis and for relating target copy number to C_q values. With an increasing number of methods available for analysis of the transcriptome [5,42,43], this study contributes toward RT–qPCR-based analytical methods becoming a well-defined "gold standard" within a portfolio of methods for analysis of single cell components.

Acknowledgments

The work described in this article was funded by the U.K. National Measurement System. We thank John Sinden (ReNeuron) for providing the CTX cell lines and Jack Price (King's College London) for sharing information relating to target GOIs from microarray profiling of the CTX cell lines. We thank Fluidigm for

permission to reproduce an image of a 48.48 Dynamic Array Integrated Fluidic Circuit (Fig. S2). We also thank Jim Huggett for detailed assessment of the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ab.2014.03.001.

References

[1] P. Dalerba, T. Kalisky, D. Sahoo, P.S. Rajendran, M.E. Rothenberg, A.A. Leyrat, S. Sim, J. Okamoto, D.M. Johnston, D. Qian, M. Zabala, J. Bueno, N.F. Neff, J. Wang, A.A. Shelton, B. Visser, S. Hisamori, Y. Shimono, M. van de Wetering, H. Clevers, M.F. Clarke, S.R. Quake, Single-cell dissection of transcriptional heterogeneity in human colon tumors, Nat. Biotechnol. 29 (2011) 1120–1127.

[2] L. Flatz, R. Roychoudhuri, M. Honda, A. Filali-Mouhim, J.P. Goulet, N. Kettaf, M. Lin, M. Roederer, E.K. Haddad, R.P. Sekaly, G.J. Nabel, Single-cell gene-expression profiling reveals qualitatively distinct CD8 T cells elicited by different gene-based vaccines, Proc. Natl. Acad. Sci. USA 108 (2011) 5724–5729.

[3] M. Bengtsson, A. Stahlberg, P. Rorsman, M. Kubista, Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels, Genome Res. 15 (2005) 1388–1392.

[4] I. Tietjen, J.M. Rihel, Y. Cao, G. Koentges, L. Zakhary, C. Dulac, Single-cell transcriptional analysis of neuronal progenitors, Neuron 38 (2003) 161–175.

[5] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B.B. Tuch, A. Siddiqui, K. Lao, M.A. Surani, MRNA-Seq whole-transcriptome analysis of a single cell, Nat. Methods 6 (2009) 377–382.

[6] H.C. Fan, J. Wang, A. Potanina, S.R. Quake, Whole-genome molecular haplotyping of single cells, Nat. Biotechnol. 29 (2011) 51–57.

[7] C. Zong, S. Lu, A.R. Chapman, X.S. Xie, Genome-wide detection of single-nucleotide and copy-number variations of a single human cell, Science 338 (2012) 1622–1626.

[8] M. Kantlehner, R. Kirchner, P. Hartmann, J.W. Ellwart, M. Alunni-Fabbroni, A. Schumacher, A high-throughput DNA methylation analysis of a single cell, Nucleic Acids Res. 39 (2011) e44.

[9] R. Trouillon, M.K. Passarelli, J. Wang, M.E. Kurczy, A.G. Ewing, Chemical analysis of single cells, Anal. Chem. 85 (2013) 522–542.

[10] N. Varadarajan, B. Julg, Y.J. Yamanaka, H. Chen, A.O. Ogunniyi, E. McAndrew, L.C. Porter, A. Piechocka-Trocha, B.J. Hill, D.C. Douek, F. Pereyra, B.D. Walker, J.C. Love, A high-throughput single-cell analysis of human CD8+ T cell functions reveals discordance for cytokine secretion and cytolysis, J. Clin. Invest. 121 (2011) 4322–4331.

[11] A.A. Powell, A.H. Talasaz, H. Zhang, M.A. Coram, A. Reddy, G. Deng, M.L. Telli, R.H. Advani, R.W. Carlson, J.A. Mollick, S. Sheth, A.W. Kurian, J.M. Ford, F.E. Stockdale, S.R. Quake, R.F. Pease, M.N. Mindrinos, G. Bhanot, S.H. Dairkee, R.W. Davis, S.S. Jeffrey, Single cell profiling of circulating tumor cells: transcriptional heterogeneity and diversity from breast cancer cell lines, PLoS One 7 (2012) e33788.

[12] M. Gu, P.K. Nguyen, A.S. Lee, D. Xu, S. Hu, J.R. Plews, L. Han, B.C. Huber, W.H. Lee, Y. Gong, P.E. de Almeida, J. Lyons, F. Ikeno, C. Pacharinsak, A.J. Connolly, S.S. Gambhir, R.C. Robbins, M.T. Longaker, J.C. Wu, Microfluidic single-cell analysis shows that porcine induced pluripotent stem cell-derived endothelial cells improve myocardial function by paracrine activation, Circ. Res. 111 (2012) 882–893.

[13] Z. Li, C. Zhang, L.P. Weiner, P.-Y. Chiou, Y. Zhang, J.F. Zhong, Molecular characterization of heterogeneous mesenchymal stem cells with single-cell transcriptomes, Biotechnol. Adv. 31 (2013) 312–317.

[14] European Commission, Directive 98/79/EC: in vitro diagnostic medical devices, 1998.

[15] J. Carmen, S.R. Burger, M. McCaman, J.A. Rowley, Developing assays to address identity, potency, purity, and safety: cell characterization in cell therapy process development, Regen. Med. 7 (2012) 85–100.

[16] R.D. Press, S. Kamel-Reid, D. Ang, BCR–ABL1 RT–qPCR for monitoring the molecular response to tyrosine kinase inhibitors in chronic myeloid leukemia, J. Mol. Diagn. 15 (2013) 565–576.

[17] M. Cronin, C. Sangli, M.L. Liu, M. Pho, D. Dutta, A. Nguyen, J. Jeong, J. Wu, K.C. Langone, D. Watson, Analytical validation of the Oncotype DX genomic diagnostic test for recurrence prognosis and therapeutic response prediction in node-negative, estrogen receptor-positive breast cancer, Clin. Chem. 53 (2007) 1084–1091.

[18] V. Sanchez-Freire, A.D. Ebert, T. Kalisky, S.R. Quake, J.C. Wu, Microfluidic single-cell real-time PCR for comparative analysis of gene expression patterns, Nat. Protoc. 7 (2012) 829–838.

[19] K.H. Narsinh, N. Sun, V. Sanchez-Freire, A.S. Lee, P. Almeida, S. Hu, T. Jan, K.D. Wilson, D. Leong, J. Rosenberg, M. Yao, R.C. Robbins, J.C. Wu, Single cell transcriptional profiling reveals heterogeneity of human induced pluripotent stem cells, J. Clin. Invest. 121 (2011) 1217–1221.

[20] A. Citri, Z.P. Pang, T.C. Sudhof, M. Wernig, R.C. Malenka, Comprehensive qPCR profiling of gene expression in single neuronal cells, Nat. Protoc. 7 (2012) 118–127.

[21] M. Bengtsson, M. Hemberg, P. Rorsman, A. Stahlberg, Quantification of mRNA in single cells and modelling of RT–qPCR induced noise, BMC Mol. Biol. 9 (2008) 63.

[22] M. Reiter, B. Kirchner, H. Müller, C. Holzhauer, W. Mann, M.W. Pfaffl, Quantification noise in single cell experiments, Nucleic Acids Res. 39 (2011) e124.

[23] A.S. Devonshire, R. Elaswarapu, C.A. Foy, Applicability of RNA standards for evaluating RT–qPCR assays and platforms, BMC Genomics 12 (2011) 118.

[24] B.C. Fox, A.S. Devonshire, M.O. Baradez, D. Marshall, C.A. Foy, Comparison of reverse transcription–quantitative polymerase chain reaction methods and platforms for single cell gene expression analysis, Anal. Biochem. 427 (2012) 178–186.

[25] K. Pollock, P. Stroemer, S. Patel, L. Stevanato, A. Hope, E. Miljan, Z. Dong, H. Hodges, J. Price, J.D. Sinden, A conditionally immortal clonal stem cell line from human cortical neuroepithelium for the treatment of ischemic stroke, Exp. Neurol. 199 (2006) 143–155.

[26] ReNeuron, PISCES clinical trial in stroke patients, <http://www.reneuron.com/the-pisces-clinical-trial-in-disabled-stroke-patients>.

[27] Fluidigm, BioMark Advanced Development Protocol Number 5: Single-Cell Gene Expression Protocol for the BioMark 48.48 Dynamic Array–Real-Time PCR, part No. 68000107.

[28] Fluidigm, BioMark Advanced Development Protocol Number 9: Calculating Copy Number Using a 12.765 Digital Array, part No. 68000119 Rev. B.

[29] A. Stahlberg, V. Rusnakova, A. Forootan, M. Anderova, M. Kubista, RT–qPCR work-flow for single-cell data analysis, Methods 59 (2013) 80–88.

[30] A. Stahlberg, J. Hakansson, X. Xian, H. Semb, M. Kubista, Properties of the reverse transcription reaction in mRNA quantification, Clin. Chem. 50 (2004) 509–515.

[31] A. McDavid, G. Finak, P.K. Chattopadyay, M. Dominguez, L. Lamoreaux, S.S. Ma, M. Roederer, R. Gottardo, Data exploration, quality control, and testing in single-cell qPCR-based gene expression experiments, Bioinformatics 29 (2013) 461–467.

[32] K.J. Livak, Q.F. Wills, A.J. Tipping, K. Datta, R. Mittal, A.J. Goldson, D.W. Sexton, C.C. Holmes, Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells, Methods 59 (2013) 71–79.

[33] A. Stahlberg, M. Kubista, M. Pfaffl, Comparison of reverse transcriptases in gene expression analysis, Clin. Chem. 50 (2004) 1678–1680.

[34] M.H. Dominguez, P.K. Chattopadhyay, S. Ma, L. Lamoreaux, A. McDavid, G. Finak, R. Gottardo, R.A. Koup, M. Roederer, Highly multiplexed quantitation of gene expression on single cells, J. Immunol. Methods 391 (2013) 133–145.

[35] S.A. Bustin, T. Nolan, Pitfalls of quantitative real-time reverse-transcription polymerase chain reaction, J. Biomol. Tech. 15 (2004) 155–166.

[36] A. Stahlberg, D. Andersson, J. Aurelius, M. Faiz, M. Pekna, M. Kubista, M. Pekny, Defining cell populations with single-cell gene expression profiling: correlations and identification of astrocyte subpopulations, Nucleic Acids Res. 39 (2011) e24.

[37] R. Sanders, D.J. Mason, C.A. Foy, J.F. Huggett, Evaluation of digital PCR for absolute RNA quantification, PLoS One 8 (2013) e75296.

[38] L. Warren, D. Bryder, I.L. Weissman, S.R. Quake, Transcription factor profiling in individual hematopoietic progenitors by digital RT–PCR, Proc. Natl. Acad. Sci. USA 103 (2006) 17807–17812.

[39] S.A. Bustin, V. Benes, J.A. Garson, J. Hellemans, J. Huggett, M. Kubista, R. Mueller, T. Nolan, M.W. Pfaffl, G.L. Shipley, J. Vandesompele, C.T. Wittwer, The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments, Clin. Chem. 55 (2009) 611–622.

[40] N. Farago, A.K. Kocsis, S. Lovas, G. Molnar, E. Boldog, M. Rozsa, V. Szemenyei, E. Vamos, L.I. Nagy, G. Tamas, L.G. Puskas, Digital PCR to determine the number of transcripts from single neurons after patch-clamp recording, Biotechniques 54 (2013) 327–336.

[41] A.K. White, K.A. Heyries, C. Doolin, M. Vaninsberghe, C.L. Hansen, High-throughput microfluidic single-cell digital polymerase chain reaction, Anal. Chem. 85 (2013) 7182–7190.

[42] T. Hashimshony, F. Wagner, N. Sher, I. Yanai, CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification, Cell Rep. 2 (2012) 666–673.

[43] M. Wu, M. Piccini, C.Y. Koh, K.S. Lam, A.K. Singh, Single cell microRNA analysis using microfluidic flow cytometry, PLoS One 8 (2013) e55044.