applied biosystems® by life technologies™

Analysis of cellular heterogeneity

Single-cell profiling on the OpenArray® Real-Time PCR System

Abstract

Gene expression profiling using real-time PCR is a pivotal tool for molecular genomics. Typically, RNA or DNA samples originate from tissue composed of heterogeneous cell populations. When analyzing gene expression profiles from large numbers of cells, the average profile may not be a true representation of the many different profiles that could exist in the cell population (e.g., in different states of growth, differentiation, or activation). As a result, the transcriptional variability of individual cells and any insight into the relationship between specific genes in single cells gets lost (Figure 1). In addition, some cell types such as undifferentiated stem cells or circulating tumor cells may be present at low numbers in tissue, so the gene signature for these cells could be masked by the gene expression profiles of the dominant cell type in a population. Observed heterogeneity in gene expression may indicate the presence of specialized cell types, or result from the stochastic nature of transcription [1]. To fully understand the complexity of tissue and cellular heterogeneity, it is necessary to measure molecular signatures at single-cell resolution. Here we discuss the use of the OpenArray® Real-Time PCR System for gene expression profiling at the single-cell level.

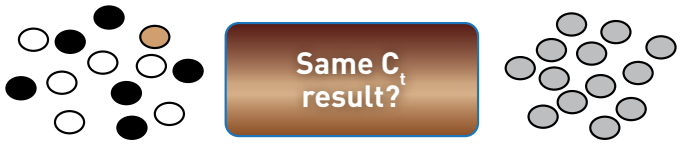


Figure 1. Cellular heterogeneity may be masked in standard gene expression analysis.

Introduction

A typical cell contains ~1–2 pg of mRNA, which translates to a few hundred thousand molecules transcribed from gDNA. The high sensitivity of reverse transcription (RT) combined with quantitative real-time PCR (qPCR) makes it possible to reproducibly and reliably detect gene expression signatures at the level of a single cell. Although there has been increased interest in single-cell profiling, there are no standard tools for this emerging application. To overcome the relatively high level of biological noise in the system and to enable better statistical analysis of cell-to-cell variation, the typical experiment for single-cell profiling requires the collection of 50–100 individual cells. Current qPCR technologies using 384-well plate formats may be both time- and cost-ineffective to achieve research goals for this application. The OpenArray® platform provides an effective solution for quickly profiling a large number of cells across a panel of 56 to 244 genes, and can be used to evaluate a greater number of single-cell events in order to achieve statistical significance.

In addition, Life Technologies offers a validated workflow protocol for single-cell profiling, starting from sample collection and preparation of cellular content from single cells all the way to either absolute transcript quantification using digital PCR or global gene expression profiling using qPCR. Both measurement approaches include an optimized workflow and robust protocol with reagent kits for cell lysis, RT, preamplification (if needed), and real-time PCR performed on the OpenArray® platform.

Workflow for single-cell profiling on the OpenArray® Real-Time PCR System

Single-cell collection

The majority of published single-cell studies have involved three major methods for cell collection: flow cytometry, manual cell harvesting, and laser capture microdissection (LCM). Based on the type of downstream application, one of these methods should be used for the single-cell collection process. The main objective of cell collection in a single-cell profiling experiment is to collect viable functional cells without disturbing expression levels of mRNA transcripts.

Flow cytometry is a technique used to count and sort microscopic particles such as cells. This is accomplished by suspending the cells in a liquid and then streaming them in front of a detector,

wherein a measurement is performed. Sorting and collection can be accomplished by using a variant of flow cytometry called fluorescence-activated cell sorting (FACS) (Figure 2). This technique uses fluorescence or light scatter to sort cells into containers or plates based on a physical characteristic of the cell, such as the expression of a specific cell-surface biomarker. With this technique, cells can be collected one at a time and then used for further analysis such as gene expression.

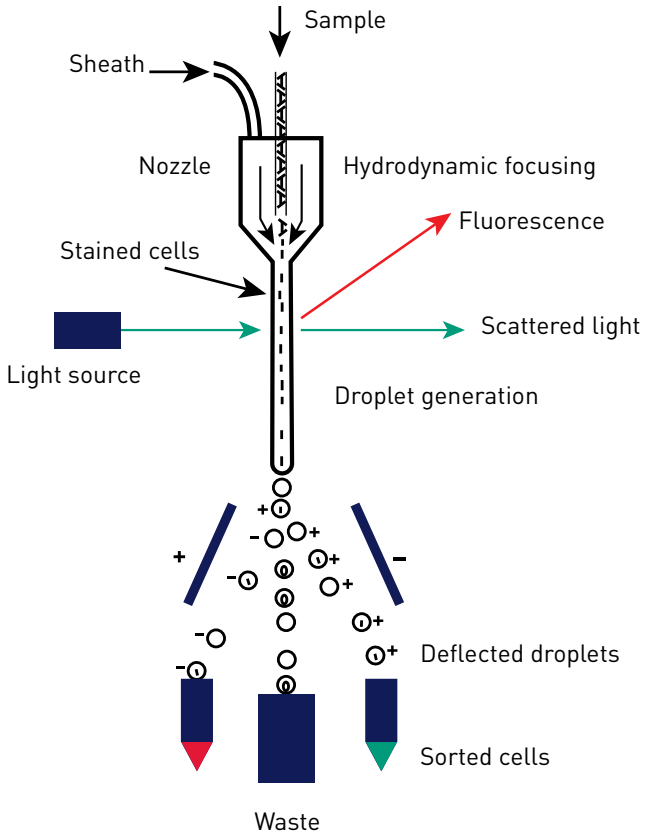


Figure 2. How fluorescence-activated cell sorting works.

Manual cell harvesting uses a variety of technologies to manually select and collect cells. This is usually accomplished using microscopy and a collection apparatus such as a glass micropipette or capillary.

Laser capture microdissection is a technology whereby a laser is coupled to a microscope and focused onto a tissue section on a slide. The tissue section is viewed, and individual cells or clusters of cells are identified either manually or semi-automatically, allowing the selection of targets for isolation. The identified cells are then cut out and separated from the adjacent tissue.

Preparation of cellular cDNA content for qPCR

Complete cell lysis and inhibition of intracellular RNases are critical to the success of the downstream RT reaction. High-efficiency RT is an absolute requirement for single-cell analysis in order to detect low-level transcripts by qPCR. Life Technologies has developed the Single Cell-to-C_T™ Kit to perform all steps of sample preparation, including cell lysis and genomic DNA removal, RT, and preamplification (optional)—all in a single tube. This helps to avoid loss of material that can result from sample transfer steps.

The principle of the TaqMan® preamplification is to amplify target cDNA prior to qPCR analysis. Briefly, cDNA is synthesized from total RNA by random priming. The cDNA for the specific target assays is then preamplified using pooled gene-specific primers to increase the number of targeted copies. The preamplification product is diluted and finally analyzed by qPCR. Previous studies suggest that this technique introduces little to no bias to the final reaction [2].

If absolute quantification of the transcript is required, the preamplification step can be omitted from a protocol and cDNA samples applied directly to TaqMan® OpenArray® Digital PCR Plates.

Results

Internal study

The experimental validation of the Single Cell-to-C_T™ Kit on the OpenArray® platform was designed to:

1. Assess the compatibility of the workflow and reagent performance for nanoscale qPCR reactions;
2. Perform a cross-platform comparison using the same set of cDNAs generated by the Single Cell-to-C_T™ Kit.

Overview

We used the CellSensor® AP-1-*bla* ME-180 cell line, the Single Cell-to-C_T™ Kit protocol (Figure 3), and the OpenArray® system (Figure 4) to analyze up to 960 samples across 56 TaqMan® Assays. The CellSensor® AP-1-*bla* ME-180 cell line contains a beta-lactamase reporter gene under control of the Activator Protein-1 (AP-1) response element stably integrated into ME-180 cells. Epidermal growth factor (EGF) is a peptide that induces cellular proliferation through the EGF receptor. Proliferative effects of EGF signaling occur through several pathways, namely the activation of the Ras and MAP kinase (MAPK) pathways. Activation of MAPK pathways can promote phosphorylation of transcription factors such as c-Fos to generate AP-1 and ELK-1 that contribute to cell proliferation. Therefore, the effects of EGF stimulation on the gene expression profiles of pooled 100-, 10-, and single-cell populations were investigated. The cells were sorted by FACSAria and collected directly into Single Cell-to-C_T™ lysis buffer followed by RT and preamplification with 56 gene-specific primer pools. Preamplified biological replicate samples were tested on the 7900HT Fast and OpenArray® Real-Time PCR Systems (Table 1).

Analysis of single-cell expression profiles

Because of the inherent variability of the expression of any given gene within a certain cell population, transcript levels cannot be normalized to reference genes as is typically done in conventional qPCR. One of the acceptable tools to show variability of transcript levels among analyzed cells is the C_t density plot, where C_t values are binned and plotted against cell number. Figure 5 shows histograms of C_t distribution for 480 single-cell samples, and for 96 samples of pooled 100-cell samples. *FOS*, *ABCA1*, and *EGR1* were detected as the most down-regulated genes affected by EGF treatment, while *SERPINE*, *FGF1*, and *CEACAM1* were the most up-regulated genes.

The analysis of single-cell profiles for many detected genes clearly shows the presence of cell subpopulations within a

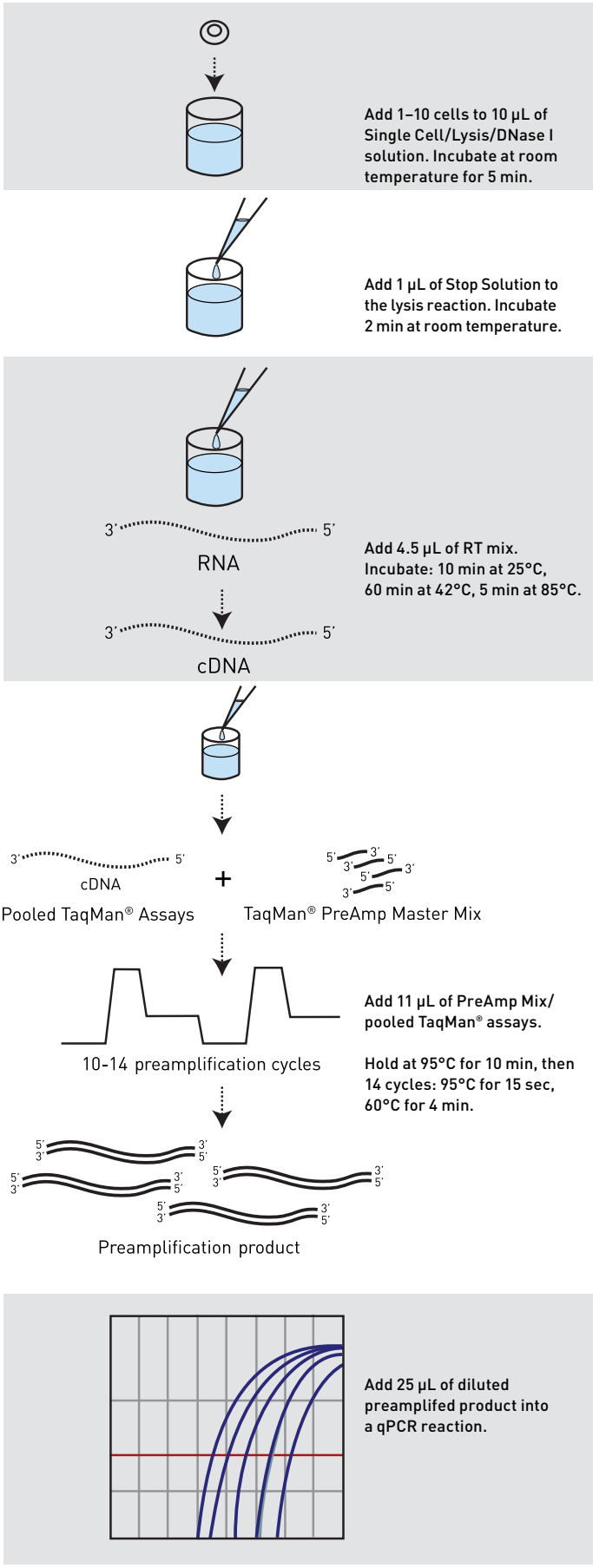


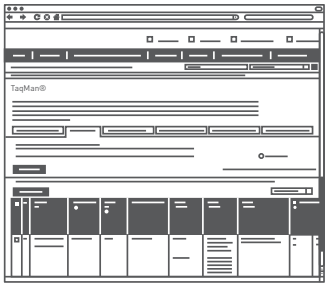
Figure 3. Single Cell-to-C_T™ qPCR workflow.

Table 1. Comparison of Applied Biosystems® 7900HT Real-Time PCR System and OpenArray® Real-Time PCR System workflows for profiling 960 samples.

	7900HT platform	OpenArray® platform
Number of plates needed to profile 56 assays	140 384-well plates	20 OpenArray® plates
Time-to-results	350 hr	16 hr

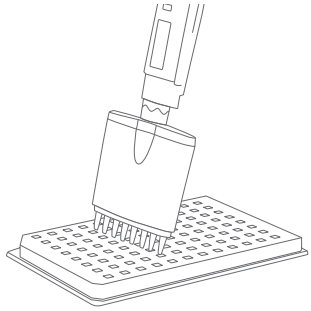
collected cell population. For example, based on the frequency of C_t distribution among all tested 480 single cells, the existence of two subpopulations of cells expressing the *EGR1* gene that coordinates the expression of EGF receptor is evident in treated and untreated cells (Figure 5). Such heterogeneity is masked when expression of *EGR1* is analyzed in pooled 100-cell samples. Moreover, because cellular content for 100 individual cells is pooled together and averaged, the C_t histogram typically follows a normal distribution. However, the C_t histogram for a single-cell population usually has an asymmetric profile with several shoulders, indicating cell heterogeneity within the population. Different cell treatments or changes in the cell environment may "synchronize" transcript levels so that the transcript profile approaches a normal distribution. Nonparametric tests can be used for statistical analysis of asymmetric C_t distribution within single-cell populations.

1. Select assays or panels, and order products.



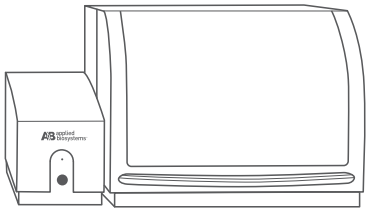
Visit www.appliedbiosystems.com/openarray to select assays and OpenArray® Plate formats for custom plates, or to select from predesigned OpenArray® Pathway Panels or OpenArray® Digital PCR Plates. Digital PCR plates are pre-treated to accept your assays and samples in your lab. All other OpenArray® Plates are delivered with assays dried down in the plate through-holes.

2. Add samples and master mix.



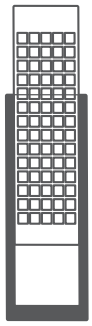
For gene expression and genotyping applications, mix cDNA or DNA samples with master mix in 384-well sample plates. For digital PCR applications, mix assays, samples, and master mix in 384-well sample plates.

3. Load samples.



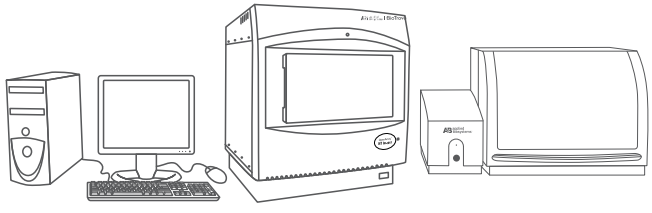
Load sample mixes onto an OpenArray® Plate with the OpenArray® AccuFill™ System.

4. Insert into case, and seal.



Insert the OpenArray® Plate into the OpenArray® case filled with immersion fluid, and seal with glue.

5. Cycle, image, and analyze data.



For genotyping applications, cycle in the Dual Flat Block GeneAmp® PCR System 9700 and transfer to the OpenArray® Real-Time PCR Instrument for imaging. For real-time and digital PCR applications, cycle and image on the OpenArray® Real-Time PCR Instrument, and analyze data.

Figure 4. OpenArray® Real-Time PCR System workflow.

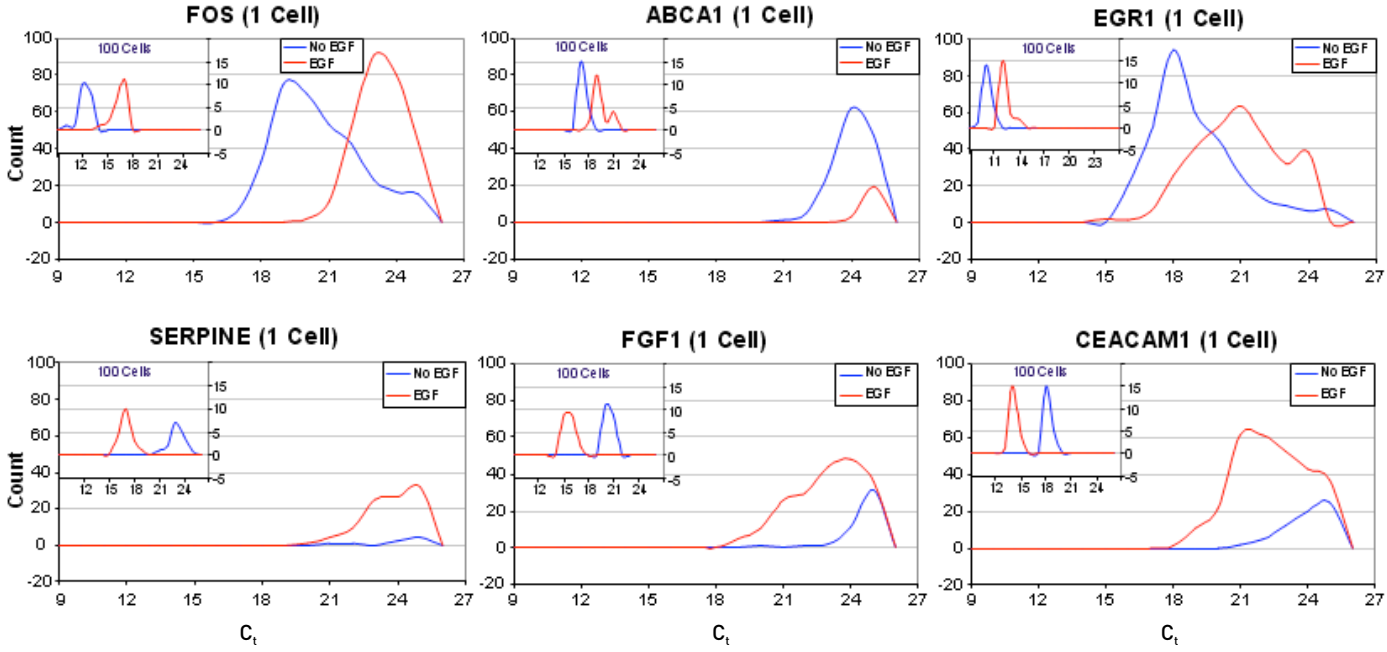


Figure 5. The average gene expression profile of a cell population may not reflect the variability exhibited by single cells. Main graphs show distribution of C_t among 840 single-cell samples, and insets show distribution among 96 samples of pooled 100-cell samples.

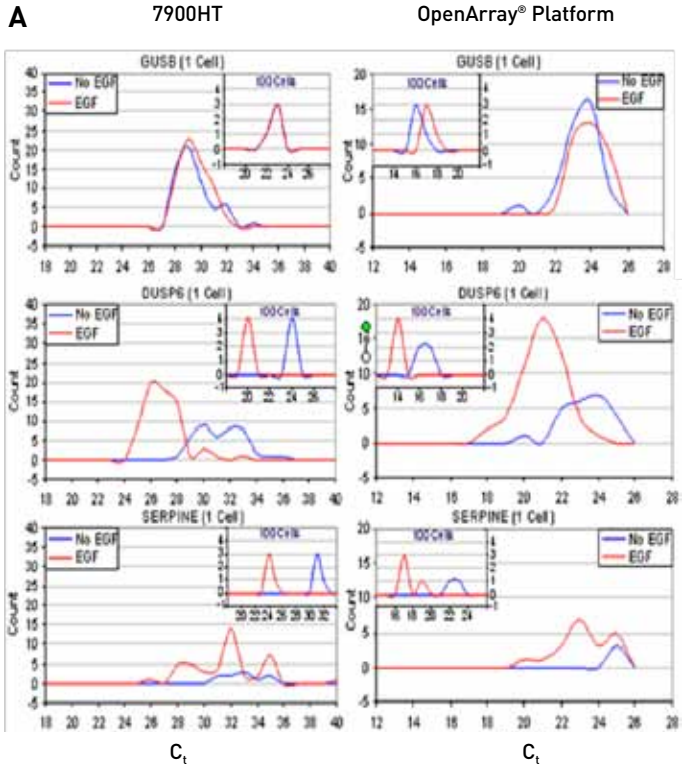
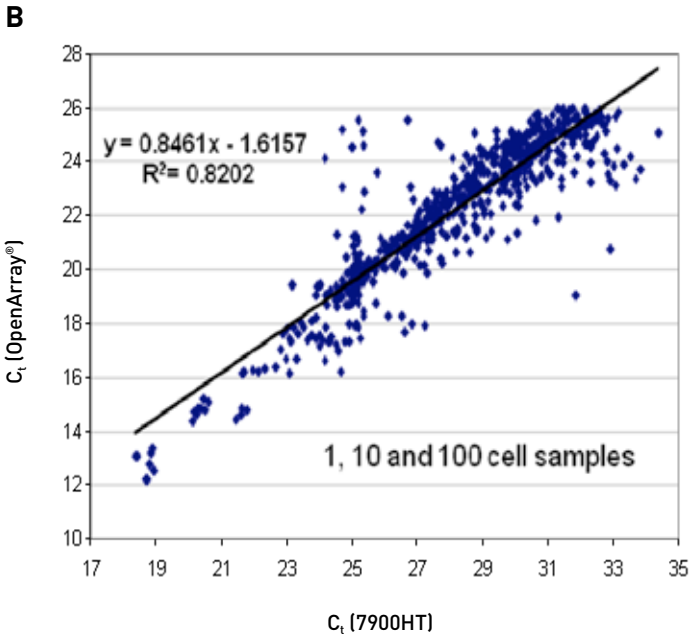


Figure 6. Cross-platform comparison (7900HT Fast Real-Time PCR System vs. OpenArray® system). Data were obtained on the two platforms for single- and 100-cell samples. **(A)** Biological replicates were used for comparison. **(B)** Scatter plot of C_t values from single-, 10-, and 100-cell samples from the 7900HT system vs. the OpenArray® system.



Cross-platform comparison

Eighty-four single-, 4-, 10-, and 100-cell samples (±EGF) were profiled for 10 genes in duplicate using the 7900HT Fast Real-Time PCR System. The same TaqMan® Assays were used on the OpenArray® platform. The cross-platform comparisons (Figure 6) are presented as C_t distribution plots for single- and 100-cell samples (insert). Both platforms produced similar expression profiles: no change in *GUSB* expression, up-regulation of *DUSP6* (C_t shift to the left), and an increase in the number of cells expressing *SERPINE* with EGF treatment. Figure 6B shows a scatter plot of C_t values generated on the 7900HT system vs. the OpenArray® system; the data indicate good correlation between the platforms (R^2 = 0.82).

Profiling of single embryonic stem cells

Overview

There is increasing evidence that pluripotent embryonic stem cells (ESCs), once thought to be a uniform cell type, are in fact highly heterogeneous, especially in terms of RNA expression. In recent studies, researchers at the University of Connecticut confirmed that the expression of key developmental control genes in ESCs is highly heterogeneous and that profiling of single stem cell populations is required to understand the cellular complexity of stem cell biology. Until recently, understanding the dynamics of both miRNA and mRNA expression within a cell population has been severely limited by the lack of an efficient and effective means to assay gene expression within individual cells.

The goal of this study was to demonstrate that TaqMan® OpenArray® Panels containing assays for detection of genes involved in pluripotency and differentiation can be used for analysis of cell heterogeneity in a single embryonic stem cell population. The results obtained on the OpenArray® platform were consistent with results generated earlier using a 384-well plate on the 7900HT platform. This study confirms the robust reproducibility of the OpenArray® system, and establishes this real-time PCR platform as a reliable method for high-throughput single-cell profiling experiments.

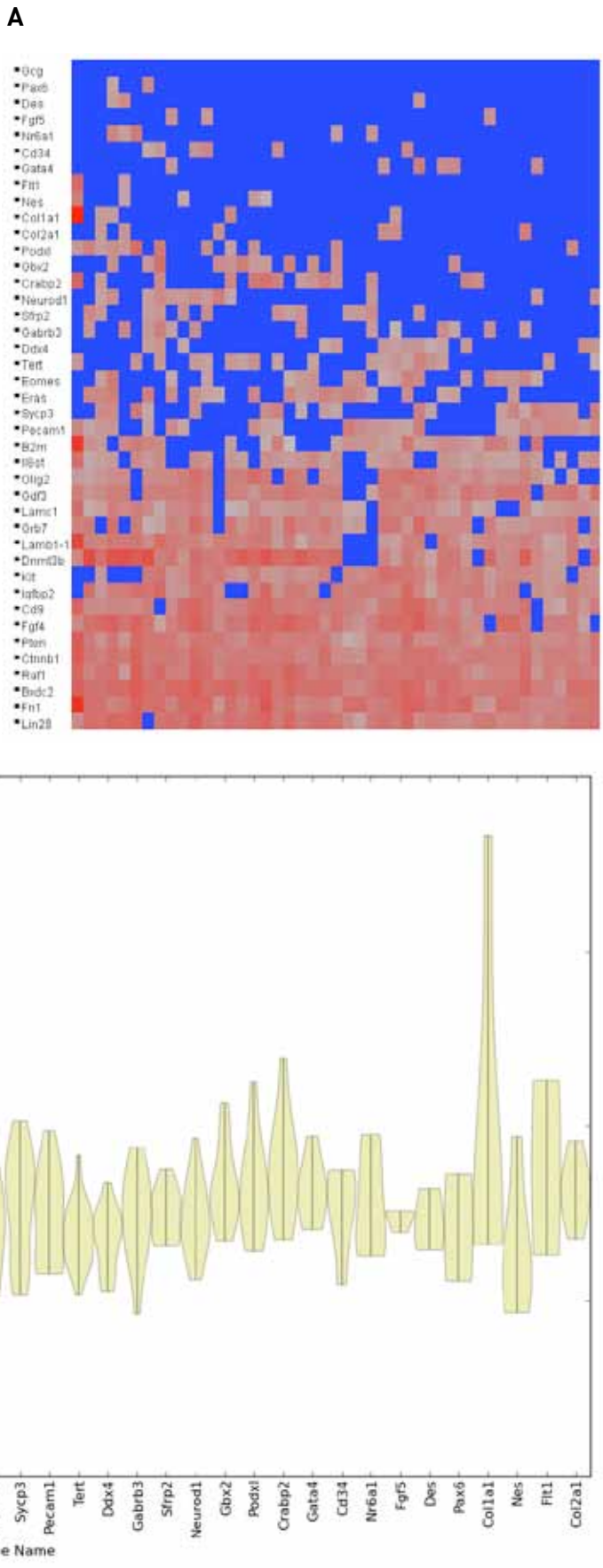


Figure 7. Analysis of gene expression in 47 single mESCs. (A) Hierarchical clustering. Expression levels range from blue (not detected) to red (high detection). Housekeeping genes as well as regulators of pluripotency are expressed in all cells at high levels as expected, demonstrating the ability of the OpenArray® platform to reliably detect transcripts at the single cell level. **(B)** Violin plots depicting the expression level (vertical axis) and probability density (width of each plot) of the single-cell expression data shown in (A).

Analysis of expression profiles in single mouse ESCs

Single mouse embryonic stem cells (mESCs) were collected via FACS into 96-well plates containing Single Cell-to-C_T™ Lysis Buffer followed by DNase treatment, RT, and preamplification performed using either gene-specific or miRNA Megaplex™ Primer Pools from Life Technologies. The expression of 64 mRNA transcripts in 48 single cells and 61 miRNAs in 27 cells using the OpenArray® platform was detected. For dPCR, the Single Cell-to-C_T™ Kit was used without the preamplification step. Cells were lysed following FACS, RNA was reverse transcribed, and the entire RT reaction was applied to the OpenArray® plate using Gapdh and Nanog TaqMan® Assays. The results are presented by hierarchial clustering analysis and violin plots. Hierarchial clustering revealed two major classes of gene expression among tested cells (Figure 7A): uniform and heterogeneous. Housekeeping genes as well as regulators of pluripotency (Figure 7B) are expressed in all cells at high levels as expected, demonstrating the ability of the OpenArray® platform to reliably detect transcripts at the single-cell level. Violin plots depicting the expression level (vertical axis) and probability density (width of each plot) of the single-cell expression data are shown in Figure 7B. Genes were sorted by expression frequency from high (left) to low (right), and all zero values have been removed to demonstrate that both uniformly and heterogeneously expressed genes are detected at the same high levels. This indicates that heterogeneous expression is not the result of sporadic detection of low-level transcripts.

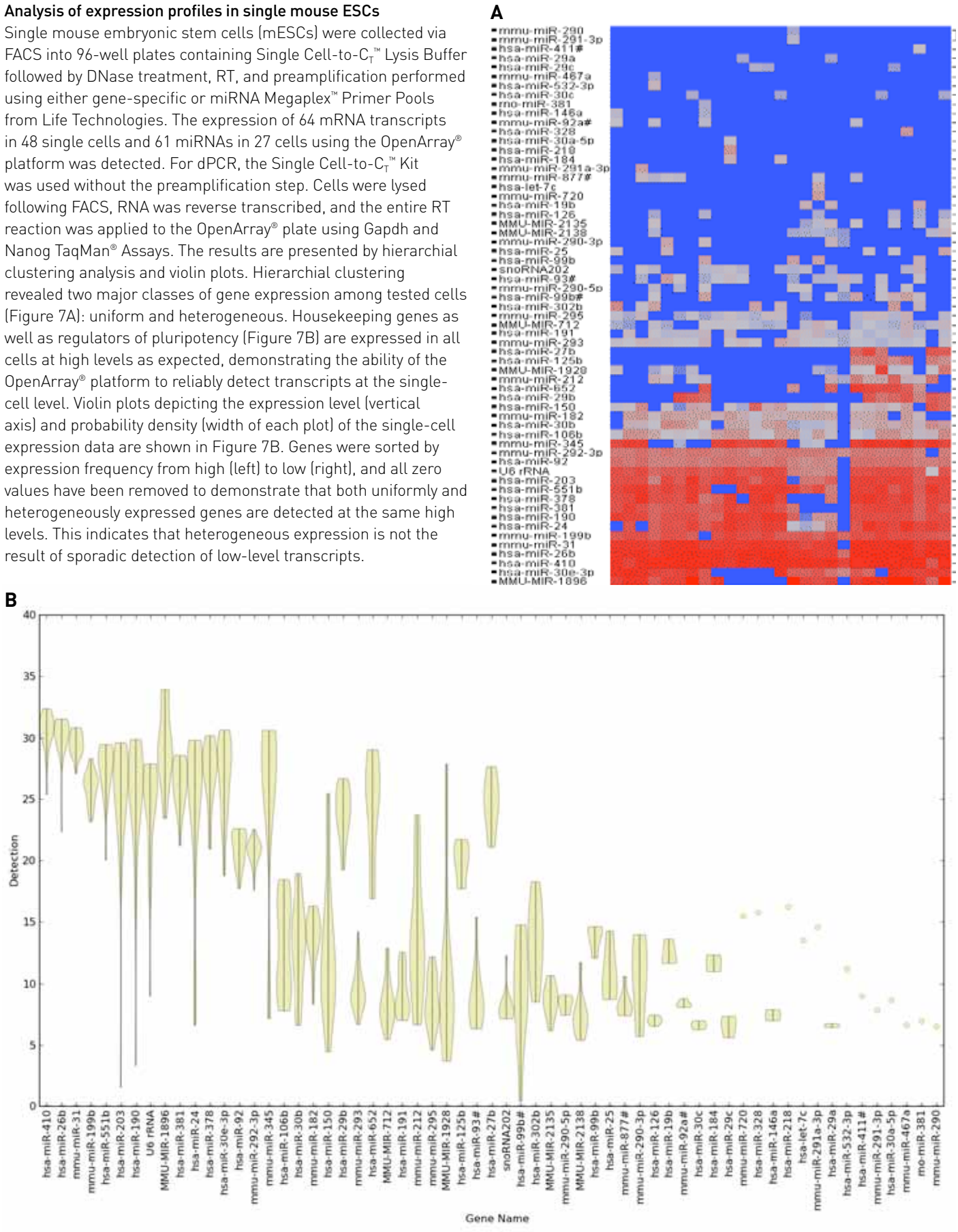


Figure 8. Analysis of the expression of 61 miRNAs in 24 single cells. (A) Hierarchical clustering analysis. **(B)** Violin plot depicting the distribution of miRNA expression levels in single mESCs. C_t values for specific miRNAs were sorted according to expression frequency from high (left) to low (right), and all zero values have been removed, showing a correlation between expression level and expression frequency.

An additional subset of mESCs was collected for miRNA profiling (Figure 8). Cells were lysed using Single Cell-to-C_T™ lysis buffer followed by RT reaction using gene-specific Megaplex™ RT Pools. The resulting cDNA was used in preamplification with both pool A and pool B primers. Hierarchical clustering analysis of the expression of 61 miRNAs in 24 single cells reveals a variety of expression patterns. MicroRNAs involved in stem cell maintenance, such as miR-293, miR-125b, mir-285, and others are expressed in the majority of cells at high levels, while those involved in later differentiation are expressed heterogeneously and at considerably lower per-cell levels.

Application of digital PCR for analysis of transcript levels in single mESCs

Overview

Digital PCR is a new approach to nucleic acid detection and quantification. The key difference between digital PCR and traditional and widely used real-time PCR is sample partitioning into a large number of PCR reaction wells. In digital PCR, quantification relies on the number of positive/negative calls for each real-time PCR reaction within a partitioned sample, and the result is expressed as a number of transcript copies detected per reaction or sample volume. The technique has the sensitivity and linearity of quantification sufficient to address transcript measurements in single cells, without the need to preamplify cDNA.

Results

We applied the digital PCR approach to the quantification of Gapdh and Nanog transcripts in single mESCs (Figure 9). Cells were collected directly into the lysis buffer, followed by the RT reaction. In order to show the linear relationship between transcript level and number of cells collected for digital PCR analysis, we collected 20, 10, 5, 3, and single mESCs.

Conclusion

When analyzing gene expression profiles from large numbers of cells, the average profile may not capture the cellular heterogeneity existing in the cell population. The OpenArray® platform provides an effective solution for quickly profiling a large number of cells across a panel of 56 to 244 genes, and can be used to evaluate a greater number of single-cell events in order to achieve statistical significance. Life Technologies offers validated workflows for absolute transcript quantification and global gene expression profiling in single cells using the OpenArray® platform.

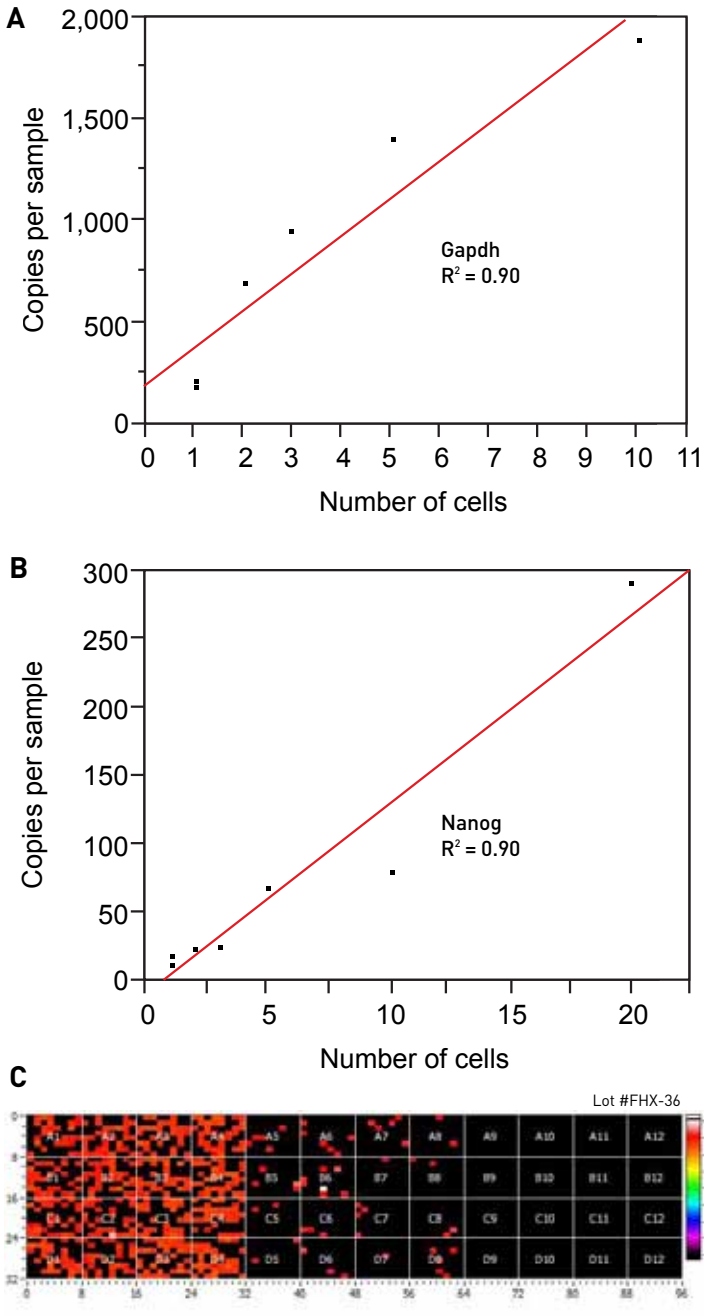


Figure 9. Digital PCR results for Gapdh (above) and Nanog detected on an OpenArray® Plate. Samples contained varying numbers of cells ranging from one to 20 cells. **(A–B)** Scatter plots depicting the strong linear relationship between transcript quantity and the number of cells per sample (r^2 = 0.90) for Gapdh **(A)** and Nanog **(B)**. **(C)** Detection of Gapdh and Nanog transcripts in individual cells. A1–D2: cell 1, Gapdh; A3–D4: cell 2, Gapdh; A5–D6: cell 1, Nanog; A7–D8: cell 2, Nanog; A9–D10: NTC, Gapdh; A11–D12: NTC, Nanog. As expected, Gapdh is expressed at considerably higher levels than Nanog and is consistent in both tested cells.

References

1. Elowitz MB, Levine AJ, Siggia ED et al. (2002) *Science* 297:1183–1186.
2. Li J, Smyth P, Cahill S et al. (2008) *BMC Biotechnology* 8:10.

Headquarters
5791 Van Allen Way | Carlsbad, CA 92008 USA | Phone +1.760.603.7200 | Toll Free in the USA 800.955.6288
www.lifetechnologies.com