

A microfluidic processor for gene expression profiling of single human embryonic stem cells

Jiang F. Zhong,^{*abc} Yan Chen,^c Joshua S. Marcus,^c Axel Scherer,^c Stephen R. Quake,^d Clive R. Taylor^b and Leslie P. Weiner^a

Received 17th August 2007, Accepted 18th October 2007

First published as an Advance Article on the web 2nd November 2007

DOI: 10.1039/b712116d

The gene expression of human embryonic stem cells (hESC) is a critical aspect for understanding the normal and pathological development of human cells and tissues. Current bulk gene expression assays rely on RNA extracted from cell and tissue samples with various degree of cellular heterogeneity. These ‘cell population averaging’ data are difficult to interpret, especially for the purpose of understanding the regulatory relationship of genes in the earliest phases of development and differentiation of individual cells. Here, we report a microfluidic approach that can extract total mRNA from individual single-cells and synthesize cDNA on the same device with high mRNA-to-cDNA efficiency. This feature makes large-scale single-cell gene expression profiling possible. Using this microfluidic device, we measured the absolute numbers of mRNA molecules of three genes (*B2M*, *Nodal* and *Fzd4*) in a single hESC. Our results indicate that gene expression data measured from cDNA of a cell population is not a good representation of the expression levels in individual single cells. Within the G0/G1 phase pluripotent hESC population, some individual cells did not express all of the 3 interrogated genes in detectable levels.

Consequently, the relative expression levels, which are broadly used in gene expression studies, are very different between measurements from population cDNA and single-cell cDNA. The results underscore the importance of discrete single-cell analysis, and the advantages of a microfluidic approach in stem cell gene expression studies.

Introduction

The fundamental challenge for stem cell gene expression studies is that stem cells are extremely rare in the context of numerous differentiated and mature cells, and there is no definitive marker for the isolation of a homogenous pure stem cell population. There is especially a lack of markers to distinguish closely related stem cells and progenitor cells. Many stem cell gene expression profiling studies have by default used the heterogeneous populations of stem cells and progenitor cells. Data obtained from these population-averaging expression profiles reflect the sum of all the sub-populations, the population-averaging expression. Without knowing the relative percentages of stem cells and progenitor cells in the interrogated population, these gene expressions profiles, which are the sum of all expression profiles from various cell types, are very difficult to interpret and not sufficiently informative. Even if the percentages are known, variation in the expression profiles among individual stem cells in different phases of the cell cycle would not be detectable by these current methods. Single-cell transcriptome analysis can

overcome this hurdle and provide precise information of stem cell gene regulation as revealed in individual separately analyzed cells.

Single-cell gene expression profiling from early embryos has suggested the transient expression of critical regulatory genes, again underscoring the importance of systematical single-cell expression profiling.^{1,2} Currently, multiple color fluorescent-activated cell sorting (FACS) has a limited utility for isolation of pure stem cell populations due to a lack of discriminating cell markers. Single-cell gene expression profiling studies with laser capture microdissection (LCM),^{3–12} patch-clamp analysis,^{2,13,14} as well as *in situ* mRNA amplification^{13,15} have been reported, but analysis of a large number of cells has proven to be very difficult with these methods. Single-cell whole genome microarray gene expression screening^{4,5,16} and single-cell cDNA library construction^{1,17,18} also have been conducted on a limited number of cells. While these studies demonstrate the potential value of single-cell gene expression profiling, they also showed the limitation of these methods in processing a large number of samples. Material loss and low biochemical reaction efficiency (mRNA capture and RT reaction) are other major challenges for single-cell analysis. A single mammalian cell contains 20–40 pg of total RNA,^{19,20} but only 0.5–1.0 pg of mRNA (10^5 to 10^6 mRNA molecules).²¹ Therefore, detecting single-cell mRNA is difficult with current methods.

The microfluidic device described here significantly increases the mRNA-to-cDNA processing efficiency ~5 fold to 54% compared to bulk reactions (~12%).^{22,23} With this device, we measured the absolute copy number of three genes in

^aDepartment of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA.
E-mail: jzhong@usc.edu

^bDepartment of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

^cDepartment of Bioengineering and Electrical Engineering, California Institute of Technology, Pasadena, CA 91125, USA

^dDepartment of Bioengineering, Stanford University, Stanford, CA 94305, USA

individual single hESCs and compared these data to those obtained from the cDNA of FACS sorted hESCs in the G0/G1 phase. Our results indicate that the hESC colony is a heterogeneous cell population and many single cells do not express all three of the interrogated genes. This result suggests that pluripotent hESC colonies are not homogeneous cell populations, rather they are highly heterogeneous cell populations regulated by different gene networks. Besides stochastic factors of individual cells, heterogeneity of cells with respect to cell cycle and other factors may be a major contributor to observed variations in mammalian single-cell gene expression.

Materials and methods

Mold fabrication

All photomasks were designed with AutoCAD software, and printed at a resolution of 20 000 dots per inch on transparency films (CAD/Art services). We used both the control mold and flow mold to define device features. The control mold with 24 μm high features was fabricated with a single step using SU8-2025 (Microchem, USA). The flow mold was fabricated with three lithographic steps. First we defined the 10 μm high column construction flow channels with SU8-2010 (Microchem, USA). Then the 12 μm high output, bead and buffer delivery channels were fabricated using SPR220-7 (Shibley, USA). In addition, a hard bake process at 200 $^{\circ}\text{C}$ facilitating channel rounding was necessary for valve closure. The last step was to construct the 40 μm high cell loading channels with AZ-50 (Clariant). In all optical lithography processes, mold exposures were under UV light on a MJB mask aligner (7 mW cm^{-2}).

Device fabrication

The microfluidic devices were fabricated by multilayer soft lithography with the silicone elastomer polydimethylsiloxane (PDMS, RTV615, General Electric, USA). Each device employs push-up valve geometry and consists of a three layer elastomeric structure.

The molds were first exposed to chlorotrimethylsilane (TMCS, Aldrich) vapor for 2 min to promote elastomer release after the baking steps. For the flow layer of the device, a mixture of PDMS (5 parts RTV615A : 1 part RTV615B) was poured onto the flow mold. After degassing, the flow molds were baked for 45 min at 80 $^{\circ}\text{C}$. For the control layer of the device, a mixture of PDMS (20 parts RTV615A : 1 part RTV615B) was spun on the control mold at 1800 rpm and baked for 30 min at 80 $^{\circ}\text{C}$. The flow layer was separated from the flow mold and flow channel access holes were then punched. Next, the flow and control layers were aligned and baked for 45 min at 80 $^{\circ}\text{C}$. The two layer structure was peeled from the control mold with control channel access holes punched, and was mounted to another thin PDMS layer made by spinning 20 : 1 PDMS mixture on a blank wafer. After baking for 3 hours at 80 $^{\circ}\text{C}$, the three layer structure was bonded to a clean microscope slide and baked overnight at 80 $^{\circ}\text{C}$.

Microfluidic station

Fig. 1 shows the setting of the microfluidic system and the hESC colonies for microfluidic analysis. The microfluidic

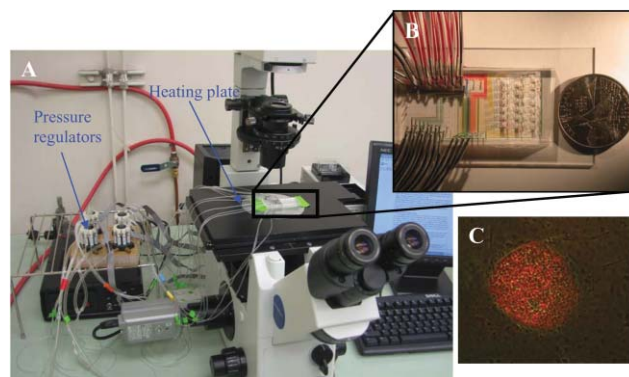


Fig. 1 The setting of the microfluidic device for single hESC mRNA extraction. A. The system includes a microscope, a computer to control air pressure with pressure regulators, and a heating stage to heat the microfluidic chip to desire temperatures. B. A typical microfluidic chip. C. Merged image of immunofluorescent stained (Oct-3/4) and light microscope images from a pluripotent hESC colony. The hESC colony was labeled with mouse α human Oct-3/4 IgG and PE-conjugated rabbit anti-mouse IgG antibodies. Only cells in the center of the hESC colony expressed Oct-3/4. The intensity of the labeling indicates the Oct-3/4 positive cells expressed Oct-3/4 at different levels. The spontaneously differentiated cells around the colony do not express Oct-3/4.

valves within the device are controlled by individual pressure regulators (Fluidigm, USA) and are interfaced *via* 23 gauge stainless steel tubing (New England Small Tube) and tygon tubing (VWR). An NI-DAQ card through a Labview interface (National Instruments) was used to control the pressure regulators.

hESC culture

The hESC line (H9) was obtained from the WiCell Research Institute (Madison, WI) and maintained as instructed.²⁴ Undifferentiated hESCs were cultured on an irradiated (5500 rads) layer of mouse embryonic fibroblast (MEF) feeder cells in 6-well plates or in matrigel-coated plates with MEF conditioned medium. To avoid contamination of mouse cells from the MEF, hESC colonies were dissected and transferred to matrigel-coated plates before using for microfluidic processing.

MEF were prepared from the embryos of 13–14 day pregnant CF-1 mice (Charles River Labs) and stocks were cryopreserved until required for culture of hESCs. hESCs were consistently observed as large clumps that appeared on the matrigel-coated surface, consistent with the published observations of others. Periodic karyotyping confirmed their human diploid chromosomal character. After injection into severe combined immunodeficient (SCID) mice, these cells produced teratomas which include tissues of ectodermal, mesodermal and endodermal origin. Immunostaining of the hESC colonies for alkaline phosphatase and with antibodies specific for SSEA-4, TRA-1-60, TRA-1-81 or Oct-3/4 confirmed the pluripotency of these cells.

Synthesis of cDNA from hESCs

The centers (100 to 200 cells) of hESC colonies from feeder-free cultures (matrigel) were mechanically picked up with a 25

gauge needle. The cells were disassociated into single-cell suspension with trypsin. After being labeled with a DNA-selective dye for living cells, Vybrant DyeCycle Green (Invitrogen, USA), FACS was performed to isolate cells in the G0/G1 phase based on DNA content. These isolated cells were pooled to extract total RNA with Trizol LS (Invitrogen, USA), or loaded into the microfluidic device for processing. Typically, 150 000 FACS sorted cells were pooled for Trizol RNA extraction, and 2000 to 5000 cells were used for microfluidic experiments. To compare the efficiency of bench-top bulk assays to microfluidic assay, the same biochemical reagents were used for the bulk assay and the microfluidic device.

Single-cell lysis, mRNA capture and RT were performed in the same microfluidic device to convert mRNA into cDNA. Dynabeads with oligo (dT)₂₅ (Invitrogen, USA) were used to capture mRNA. Sensiscript RT kit (Qiagen, USA) was used for converting captured mRNA to cDNA. The oligo (dT)₂₅ sequence in Dynabeads serves as both mRNA capture sequence and primer for cDNA synthesis. After oligo (dT)₂₅ beads with attached cDNA were flushed to the collection wells, the wells were cut off from the chip. Centrifugation was used to transfer the beads from individual wells into PCR tubes. These beads with attached cDNA from individual cells were subjected to realtime qPCR for measurement of the number of molecules of interrogated mRNA in IQ5 (Bio-Rad, USA).

Microfluidic device and single-cell process

The basic components of a microfluidic processor are shown in Fig. 2. The processor captures mRNA from 20 single-cells separately and simultaneously, then converts it into individual cDNA. The lysis buffer was loaded into the flow channels until it reached the waste outlets, so as to leave no air bubbles in the channel. Oligo (dT)₂₅ beads (Invitrogen, USA) were then loaded and columns were built serially by addressing flow lines individually with the multiplexer control channels, while keeping the sieve valve actuated. Once columns were built, excess beads still present in the flow channels were flushed with the lysis buffer to the constructed columns. A single-cell suspension was then loaded. By adjusting cell concentration and flow rate to obtain optimal distance between two floating cells, single hESC was captured in individual cell lysis module (approximately 80% of modules captured only one cell). Microscopic examination was performed to ensure that only cDNA from modules that captured one cell will be used for PCR analysis. Cells were then lysed chemically by mixing cells with the lysis buffer in the ~10 nL ring (Fig. 2 insert 1). Mixing occurred by executing a peristaltic pump sequence^{25,26} with control channels. Cell lysates were then pushed *via* pneumatic pressure over the affinity columns to captured mRNA with oligo (dT)₂₅ beads. After washing the columns with a first strand synthesis buffer, the reverse transcription (RT) master mix was introduced. Once the RT master mix filled the flow channels, first strand synthesis was then carried out by heating the processor to 40 °C on a thermal microscope stage. The oligo (dT)₂₅ beads served as both primers (oligo (dT)₂₅ sequences and a solid phase support). The RT reaction mixture (99 µL) was flown over the columns for 45 min until the RT reaction was completed at a flow rate of ~20 µm s⁻¹.

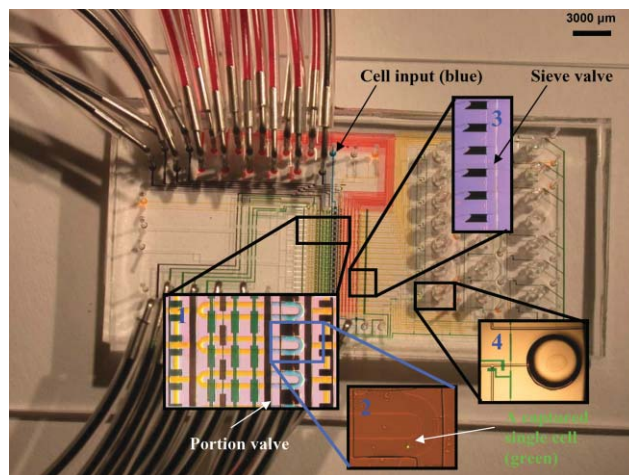


Fig. 2 Single-cell mRNA extraction microfluidic device filled with food dye for illustration. All flow channels are filled with yellow food dye, multiplexer control channels are filled with red dye, collection and waste channels are in blue. The inserts show enlargements of four important areas of the chip. After loading cell suspension from the cell input inlet, single-cells are captured in cell lysis module (Insert 1) within the flow channels (blue). The pump valves are green. The separation valve is black. The lysis buffer is yellow. A captured single hESC is labeled with a fluorescent dye (green) and shown in Insert 2. Cell lysis is performed by opening the portion valve and pumping to mix lysis buffer (yellow) with the captured cell. The resulting cell lysate is pushed through oligo-dT bead columns for mRNA capture. Oligo-dT beads are stacked into columns by closing the sieve valve while loading bead suspension. Insert 3 shows six stacked oligo-dT bead columns next to the sieve valve. After washing beads with buffers, RT reaction master mix is flown through the bead columns to synthesize cDNA from the captured mRNA at 40 °C. After RT reaction, beads with attached cDNA are pushed to collection wells (Insert 4) by opening the sieve valve. The beads are recovered by cutting the wells off the chips and centrifuging a flipped-well in a microcentrifuge tube.

Upon completion of the RT reaction, the waste valves were closed, and collection valves were opened. The beads were sent to collection wells by opening the sieve valves and flowing columns of the processor in a serial manner with a PCR buffer. The fluid multiplexer was used to push beads in each of the 20 reaction channels individually. Beads were collected by cutting the collection wells off the device and centrifuging the beads into PCR tubes.

Standard curves

Human fetal brain cDNA (Invitrogen, USA) and cDNA generated from poly-A RNA control kit (Affymetrix, USA) were used as templates to obtain quantitative PCR amplicons. Quantification of the PCR amplicons was performed by gel densitometry with DNA ladders (Norgen, Canada). These amplicons contain the majority of the cDNA sequence of the respective genes and were used for generating standard curves for qPCR. The *lys* standard curve was generated with a known amount of a 846 bp fragment of the *lys* cDNA. The primers for the amplification were: cagtcaaccttaccgcatt (forward) and acatggacaggaggcatttc (reverse).

Three sets of primers were used to amplify 905 bp, 982 bp and 890 bp fragments of B2M, Nodal and Fzd4 respectively

from human brain cDNA (Invitrogen, USA). The primer sets were: B2M: ggcatctcgaagctgaca (forward) and ccagattaaccaacacatgc (reverse); Nodal: cttcctgagccaacaagagg (forward) and cagactccactgagcccttc (reverse); Fzd4: gggagctctaaatccaca (forward) and ggcatggagatgaaacaca (reverse)

Quantitative PCR

Four sets of multiplex Taqman primers and probes were designed with Beacon Designer (Premier Biosoft International, USA). B2M: aattgctatgtctgggttcatcc (forward), gcttcatgtctcgtatcccacttaac (reverse) and acaaagtcacatggttcacacggcaggca (probe-FAM); Nodal: catacatccagatgtctgctgaaagc (forward), atcagaggcaccacattcttc (reverse) and ccaccgagtcctctcactgtgtgccc (probe-Cy5); Fzd4: cgacccatccgcatctcc (forward), acattggcacataaacagaacaagg (reverse) and ccgaaacctcctccaagattcagcact (probe-FAM). Multiplex quantitative PCR was performed with IQ-5 (bio-rad, USA) and Quantitect Taqman PCR kit (Qiagen, USA) for B2M, Nodal and Fzd4. The qPCR of *lys* was performed independently from the multiplex qPCR.

Results and discussion

Device efficiency

Single-cell measurement of mRNA is difficult. One difficulty results from the loss of material during the steps of single-cell capture, lysis, mRNA isolation and cDNA synthesis. Unlike the DNA molecule, mRNA is very susceptible to degradation by widely existing RNase. Therefore, it is essential to carry out the cDNA first strand synthesis on the same device immediately after mRNA capture. Another difficulty is the low mRNA-to-cDNA efficiency, mRNA capture (~40–50%) and RT reaction (~20%)^{22,23} in bulk assays.

In order to measure the absolute number of molecules of mRNA with standard curves, the mRNA-to-cDNA efficiency of the processor must be obtained. An artificial Poly-A RNA standard (Affymetrix, USA) was used to compare the input mRNA copy number to the cDNA copy number which was measured by qPCR at the end of the processes. This RNA standard contains known amounts of artificial RNAs with poly-A tails from 4 *B. subtilis* genes, *lys*, *phe*, *thr* and *dap*. The *lys* gene was used as RNA standard for calculation of mRNA-to-cDNA efficiency. A standard curve was generated with a known amount of *lys* DNA at 10 fold dilutions. The mRNA capture and RT reaction were carried out at both conventional micro-litre level and at nano-litre scale with the microfluidic processor.

In the conventional reaction, a 50 μL RT reaction was carried out after capturing 4 μL of artificial mRNA (7.3 nM) with 30 μL oligo (dT)₂₅ beads as instructed by the manufacturer. One micro litre of the RT product (corresponding to 2.16×10^8 molecules of *lys* artificial mRNA) was used for realtime quantitative PCR (qPCR) detection. Based on the standard curve, the final detected number of molecules was $2.5 \pm 0.5 \times 10^7$ molecules (Table 1). The processing efficiency of conventional micro-litre scale reaction is 12%.

To compare the efficiency of conventional and microfluidic approaches, the same artificial RNA sample was introduced into the cell capturing chamber, processed into cDNA and detected with realtime qPCR. Each of the 20 cell-capture chambers in the microfluidic device has a volume of 3.4 nL resulting in 1.84×10^6 molecules of input *lys* artificial mRNA. The absolute *lys* cDNA molecule number detected by qPCR with the standard curve is $1.0 \pm 0.3 \times 10^6$. The total processing efficiency was 54% which is approximately 5 times the efficiency of the conventional approach. A series of 10 fold dilutions of the artificial RNA were used and they verified that the processing efficiencies of both approaches do not vary significantly in different concentrations of the input artificial RNA. The standard deviations of both bench-top and microfluidic reactions were also calculated and are similar (Table 1). The small standard deviation of the 20 reactions in the same device indicates consistent processing efficiency among the 20 reaction chambers.

Gene expression of single hESCs

Conventional gene expression studies were carried out with RNA extracted from cell populations. To investigate whether the gene expression of cell population is a reasonable sum of the expressions of individual cells, the expression levels of three genes were measured by two approaches: with pooled cDNA from the hESC population and with cDNA from individual hESCs.

Based on the mRNA-to-cDNA efficiency calculated above, the numbers of three mRNAs in a single hESC were measured. Standard curves were constructed with known amounts of full length *B2M*, *Fzd 4* and *Nodal* cDNA. The three standard curves overlap each other, indicating similar qPCR efficiency for each set of the primers (Fig. 3A). These standard curves have a range of 2 to 2×10^6 molecules that allow the direct comparison of the expression levels of these three genes from their amplification curves (Fig. 3B). The minimal detectable level is 4 molecules (2/54%) for our microfluidic processor, and 17 molecules (2/12%) for bulk assays with these standard curves.

Table 1 Microfluidic device increasing mRNA capture and reverse transcription (RT) efficiency. Known amounts of artificial *Lys* mRNAs in a mixture with 3 other poly-A artificial RNAs were processed with traditional microliter scale reaction or with microfluidic device in nanoliter scale. A standard curve was generated with known amount of *lys* DNA for the detection of the number of molecules of cDNA ($y = -3.5547x + 39.538$ $R^2 = 0.99$). The molecule number of *lys* cDNA was calculated from the standard curve with real-time PCR threshold cycles. Processing efficiencies were calculated by dividing the measured number of cDNA molecules with the number of input mRNA molecules

	Input <i>Lys</i> RNA (molecule number)	Threshold Cycle	Detected cDNA (molecule number)	Efficiency (%)
Microlitre Reaction (Bench-top)	2.16×10^8	13.22 ± 0.37 ($n = 20$)	$2.6 \pm 0.6 \times 10^7$	12
Nanolitre Reaction (Microfluidic)	1.84×10^6	18.15 ± 0.46 ($n = 20$)	$1.0 \pm 0.3 \times 10^6$	54

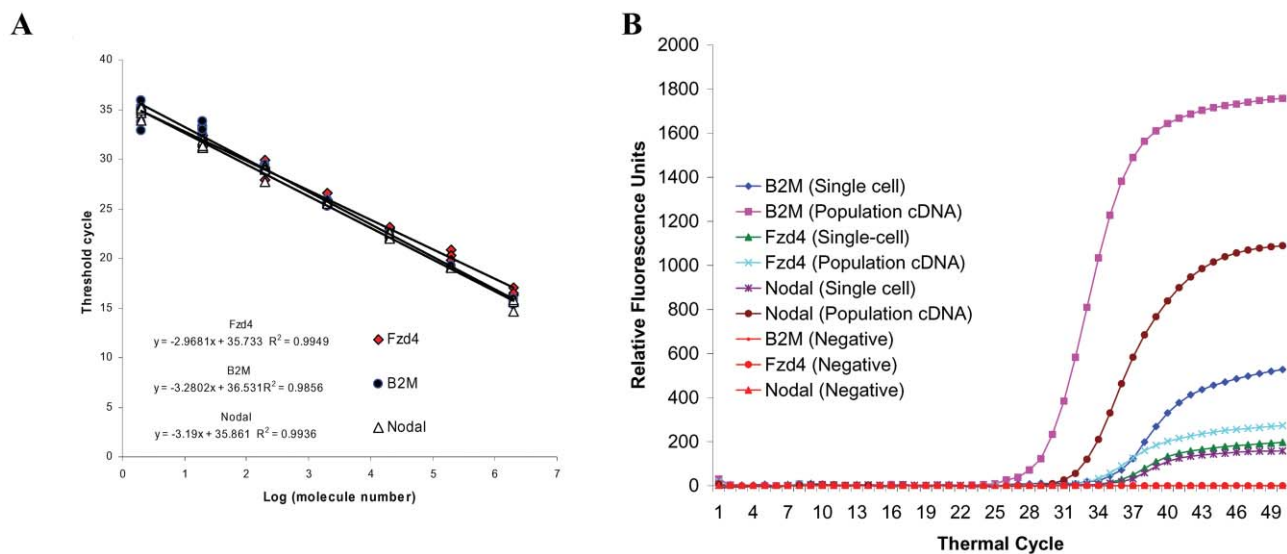


Fig. 3 Measuring the absolute molecule number of the three genes in a single hESC with multiplex quantitative PCR. A. Standard curves are generated with known amounts of plasmid DNA containing the full sequence of the genes. The curves cover from 2 to 2×10^6 copies of the respective genes. With our primer design, all the curves overlap each other, and indicate similar PCR efficiency. B. The multiplex quantitative PCR amplification curves obtained from cDNA of hESC colonies are plotted with curves obtained from cDNA of a representative hESC. Because the standard curves of the three genes are very similar, these amplification curves show that the expression ratio of *B2M* and *Nodal* is similar in population cDNA and this single-cell cDNA. However, the expression of *Fzd4* and *Nodal* is very similar in this particular single hESC, but very different in the hESC population. Unlike this single hESC, some single hESC do not express all three genes. This result suggests the heterogeneity of hESC and underscores the importance of single-cell analysis.

The pluripotent hESC colonies are often considered and treated as a homogenous population. However, our data indicate that this is not the case. As Fig. 3B shows, the absolute expression levels (mRNA molecule number) measured from cDNA equivalents to 1000 cells are not even close to 1000 fold of the levels measured from a typical single-cell. With cDNA equivalent to 1000 G1/G0 phase hESC, the absolute molecule numbers of *B2M*, *Nodal* and *Fzd4* are $6,034 \pm 660$, 402 ± 55 and 69 ± 19 respectively. However, the distributions of the molecule numbers of these 3 genes are diverse in single hESCs (Fig. 4). Among the 54 interrogated single hESCs, 14.8%, 37% and 37% of hESC did not express detectable levels (less than 4 copies) of *B2M*, *Fzd4* and *Nodal* respectively. The expression of *B2M* ranged from 4 to 76 copies, with the majority between 30 and 50 copies. The expression of *Fzd4* and *Nodal* ranged from 6 to 548 copies, and 22 to 504 copies respectively. All cells expressed at least one of the 3 genes.

Many gene expression studies, such as microarray or quantitative PCR, are based on the relative expression level of genes with normalization to a house keeping gene which is assumed to have a similar expression in all samples. Our data show that even the relative expression level of the three genes are very different between measurements derived from cDNA of a cell population and separately from individual cells. In the measurement from population cDNA, the expression level of *Fzd4* is approximately 8 fold lower than the expression of *Nodal* after being normalized to *B2M*. However, in a representative single hESC, the expression levels of *Nodal* and *Fzd4* are very similar after normalization with *B2M* (Fig. 4).

Conclusions

Microfluidic devices that are designed to manipulate nanolitre amounts of reagents provide a desirable platform for single-cell gene expression processing. As our data indicate, capturing mRNA and converting it to cDNA in a microfluidic device not only consumes significantly less reagents, but most importantly increases the efficiency of biochemical reactions. The microfluidic device described has a mRNA-to-cDNA efficiency 5 times better than the corresponding bench top reactions. Previous studies showed a 40–55% mRNA capture efficiency of oligo (dT)₂₅,²² and ~20% RT efficiency²³ in conventional bench top reactions. The 12% overall efficiency of our bulk assay is consistent with these previous studies (55% \times 20%). With our microfluidic device, the mRNA-to-cDNA processing efficiency is 54% (5 fold better than bulk assay). This finding indicates a very efficient mRNA-to-cDNA processing that may be due to the nanolitre scale reaction. This mRNA-to-cDNA efficiency is similar to the previously reported efficiency from other microfluidic devices.²⁷

Due to the nanolitre scale nature of the microfluidic device, the input number of mRNA in conventional reaction and microfluidic reaction is ~100 fold different when same RNA standard is used. To rule out the possibility of RNA concentration in processing efficiency, serial 10-fold dilutions of the RNA standard were used for the same experiments. The mRNA-to-cDNA efficiencies are not significantly different for both bench-top and microfluidic reactions. Therefore, the high mRNA-to-cDNA efficiency is a desirable characteristic of microfluidic devices. With this 54% processing efficiency, the microfluidic device described potentially can detect 2 mRNA copies of interrogated genes.

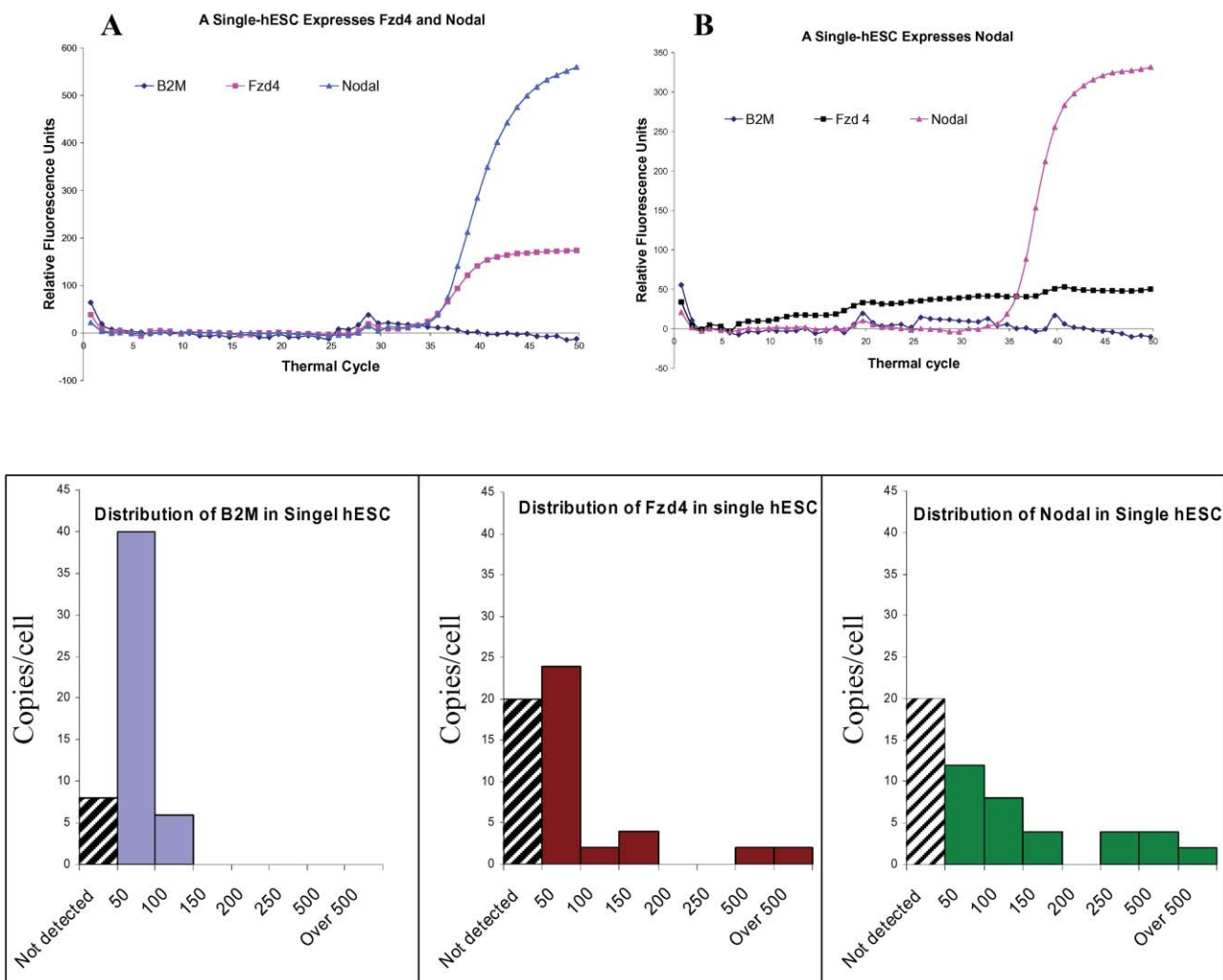


Fig. 4 Expression of *B2M*, *Nodal* and *Fzd4* in single hESCs. A. A single-hESC expresses both *Nodal* and *Fzd4*, but does not express *B2M* in a detectable level. B. A single hESC expresses only *Nodal*. The mRNAs of *B2M* and *Fzd4* are undetectable. C, D and E shows the distribution of *B2M*, *Fzd4* and *Nodal* in single hESCs. Among the 54 interrogated single hESCs, 14.8%, 37% and 37% of cells (indicated by black strait pattern) do not express detectable levels (less than 4 copies) of *B2M*, *Fzd4* and *Nodal* respectively. The distribution pattern of expressions is narrower for *B2M* compared to the other 2 genes. The discontinuous distribution of *Nodal* and *Fzd4* suggest a high heterogeneity of the 54 cells.

In our device, there are 20 individual single-cell processing reactors for simultaneously processing. The standard deviation of threshold cycles of the 20 reactors is small (0.37) and similar to those from conventional reactions (~ 0.5). These data indicate that the processing efficiencies among the 20 reactors are very similar. Similar results are also obtained in multiple lots of devices. This aspect is critical for comparing results from individual cells intra- and inter-devices.

With this device, we profiled the expression of three genes in hESC. Our data indicated that even for FACS-sorted G1/G0 phase hESC, the expression data from population cDNA can not be used to estimate the gene expression level in individual cells. Therefore, the population averaging effect distorts the expression levels of the three genes in bulk assays conducted with population cDNA from hESC colonies. Because the population averaging effects are not the same for all three genes, the relative expression levels are also distorted. The qPCR with cDNA from 1000 hESC indicates that *Nodal* is expressed approximately 8 times more than *Fzd4* after

normalization to *B2M*. However, in individual cells the expressions of these genes are very diverse. The difference observed in pooled cDNA may be due to the different number of cells expressing the respective genes.

It has been reported that gene expression is highly variable in the single-cell levels.^{28,29} A speculation is that the high variability is due to experimental variation. In the present study, we showed that experimental variation may not be the main factor. The stochastic gene expression behavior of single-cells (biological noise) has been reported in various studies of prokaryote cells.^{30–34} Biochemical processes such as transcription, translation, RNA and protein degradation have been thought of as the primary contributors. However, heterogeneity of the mammalian cell populations may be a factor related to the observed expression variations in single-cell analysis of mammalian cells. When gene expression studies are conducted at the single-cell level, we must recognize that no two cells are identical. In a particular mammalian cell population, two cells always are different in cell cycle,

differentiation stages, and environmental stimulation. The variation of gene expression at the single-cell levels is expected, and the reason may be due to stochastic expression fluctuation, or due to heterogeneity of the cell populations. In mammalian cells which are regulated by more complex gene networks than yeast and bacteria, cell heterogeneity may be the major contributor of variation of gene expression at single-cell levels.

Acknowledgements

This work was supported by grants from the National Institutes of Health (NIH 1R01 HG002644-01A1) and the Nancy Davis Foundation for Multiple Sclerosis. The authors thank Dr Richard Vestewig for his suggestions and proof reading of the manuscript. The authors also thank the laboratory of Dr Melvin Simon and the laboratory of Dr Vijay Kalra for their assistance in real-time PCR.

References

- 1 J. Adjaye, R. Daniels, V. Bolton and M. Monk, *Genomics*, 1997, **46**, 337–344.
- 2 L. Rambhatla, B. Patel, N. Dhanasekaran and K. E. Latham, *Mol. Reprod. Dev.*, 1995, **41**, 314–324.
- 3 M. K. Chiang and D. A. Melton, *Dev. Cell*, 2003, **4**, 383–393.
- 4 F. Kamme and M. G. Erlander, *Curr. Opin. Drug Discov. Dev.*, 2003, **6**, 231–236.
- 5 L. Luo, R. C. Salunga, H. Guo, A. Bittner, K. C. Joy, J. E. Galindo, H. Xiao, K. E. Rogers, J. S. Wan, M. R. Jackson and M. G. Erlander, *Nat. Med.*, 1999, **5**, 117–122.
- 6 I. Tietjen, J. M. Rihel, Y. Cao, G. Koentges, L. Zakhary and C. Dulac, *Neuron*, 2003, **38**, 161–175.
- 7 D. A. Rappolee, A. Wang, D. Mark and Z. Werb, *J. Cell. Biochem.*, 1989, **39**, 1–11.
- 8 L. H. Trumper, G. Brady, A. Bagg, D. Gray, S. L. Loke, H. Griesser, R. Wagman, R. Brazier, R. D. Gascoyne and S. Vicini, *Blood*, 1993, **81**, 3097–3115.
- 9 G. Brady, F. Billia, J. Knox, T. Hoang, I. R. Kirsch, E. B. Voura, R. G. Hawley, R. Cumming, M. Buchwald and K. Siminovitch, *Curr. Biol.*, 1995, **5**, 909–922.
- 10 A. K. Dixon, P. J. Richardson, K. Lee, N. P. Carter and T. C. Freeman, *Nucl. Acids Res.*, 1998, **26**, 4426–4431.
- 11 M. R. Emmert-Buck, R. F. Bonner, P. D. Smith, R. F. Chuaqui, Z. Zhuang, S. R. Goldstein, R. A. Weiss and L. A. Liotta, *Science*, 1996, **274**, 998–1001.
- 12 C. A. Suarez-Quian, S. R. Goldstein, T. Pohida, P. D. Smith, J. I. Peterson, E. Wellner, M. Ghany and R. F. Bonner, *Biotechniques*, 1999, **26**, 328–335.
- 13 J. Eberwine, H. Yeh, K. Miyashiro, Y. Cao, S. Nair, R. Finnell, M. Zettel and P. Coleman, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 3010–3014.
- 14 K. M. Schmidt-Ott, S. Tuschick, F. Kirchhoff, A. Verkhatsky, L. Liefeldt, H. Kettenmann and M. Paul, *J. Cardiovasc. Pharmacol.*, 1998, **31**, S364–366.
- 15 P. B. Crino, J. Q. Trojanowski, M. A. Dichter and J. Eberwine, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 14152–14157.
- 16 F. Kamme, R. Salunga, J. Yu, D. T. Tran, J. Zhu, L. Luo, A. Bittner, H. Q. Guo, N. Miller, J. Wan and M. Erlander, *J. Neurosci.*, 2003, **23**, 3607–3615.
- 17 S. Y. Ying, H. M. Lui, S. L. Lin and C. M. Chuong, *Biotechniques*, 1999, **27**, 410–412.
- 18 L. H. Trumper, G. Brady, S. Vicini, J. Cossman and T. W. Mak, *Ann. Oncol.*, 1992, **4**, 25–26.
- 19 R. C. Roozmond, *Histochem. J.*, 1976, **8**, 625–638.
- 20 E. Uemura, *Brain Res. Bull.*, 1980, **5**, 117–119.
- 21 G. Brady, *Yeast*, 2000, **17**, 211–217.
- 22 Y. Hamaguchi, Y. Aso, H. Shimada and M. Mitsushashi, *Clin. Chem.*, 1998, **44**, 2256–2263.
- 23 J. Zhang and C. D. Byrne, *Biochem. J.*, 1999, **337**, 231–241.
- 24 J. A. Thomson, J. Itskovitz-Eldor, S. S. Shapiro, M. A. Waknitz, J. J. Swiergiel, V. S. Marshall and J. M. Jones, *Science*, 1998, **282**, 1145–1147.
- 25 H. P. Chou, M. A. Unger and S. R. Quake, *Biomed. Microdev.*, 2001, **3**, 323–330.
- 26 J. S. Marcus, W. F. Anderson and S. R. Quake, *Analyt. Chem.*, 2006, **78**, 956–958.
- 27 L. Warren, D. Bryder, I. L. Weissman and S. R. Quake, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 17807–17812.
- 28 M. Bengtsson, A. Stahlberg, P. Rorsman and M. Kubista, *Genome Res.*, 2005, **15**, 1388–1392.
- 29 N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain and M. B. Elowitz, *Science*, 2005, **307**, 1962–1965.
- 30 J. R. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi and J. S. Weissman, *Nature*, 2006, **441**, 840–846.
- 31 P. S. Swain and A. Longtin, *Chaos*, 2006, **16**.
- 32 G. M. Suel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo and M. B. Elowitz, *Science*, 2007, **315**, 1716–1719.
- 33 M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain, *Science*, 2002, **297**, 1183–1186.
- 34 J. Ryley and O. M. Pereira-Smith, *Yeast*, 1965, **23**, 1065–1073.