Highly Parallel Genome-Wide Expression Analysis of Single Mammalian Cells

Jian-Bing Fan¹, Jing Chen¹, Craig S. April¹, Jeffrey S. Fisher¹, Brandy Klotzle¹, Marina Bibikova¹, Fiona Kaper¹, Mostafa Ronaghi², Sten Linnarsson³, Takayo Ota³, Jeremy Chien³, Louise C. Laurent⁴,⁵, Sean V. Nisperos⁶, Gina Y. Chen⁶, Jiang F. Zhong⁶

Introduction

Recently, there has been growing interest in obtaining gene expression profiles from single cells, as it has become increasingly evident that the heterogeneity present in cell populations is such that population-based transcriptional profiles may not reflect the regulatory networks functional at the individual cell level [1,2]. Applications for single cell gene expression profiling include lineage determination in early development and organogenesis, including embryogenesis [3,4], neuronal [5–8] and glial [9] cell differentiation, hematopoietic [10,11], bone marrow stromal [12], epidermal [13], heart [14,15] and pancreatic [16] stem cell biology. Apart from facilitating cell lineage mapping an additional key utility of single cell transcriptomics is in clinical diagnostics, particularly the identification of gene expression signatures in circulating tumor cells for use as prognostic markers for metastatic tumors [17] and treatment response [18].

The analysis of single cancer cells can potentially overcome the shortcomings of tumor heterogeneity and help pinpoint driver mutations that spur the initial development of tumors, and identify which mutations lead to metastasis, cancer progression and resistance to therapy. However, a key technological challenge in the transcriptional profiling of single cells is that most whole-genome amplification protocols suffer from significant amplification bias. While there have been several recent advancements in the capture and isolation of single cells, such as cell picking [19,20] and micromfluiddic [1,17,21] devices, there remains a need for the development of high-throughput, whole-genome gene expression assays for single cells. Example of previously reported assays aimed at attempting to overcome the limitation of single cell or near single cell quantities of starting material [for reviews see [4,22,23]] include terminal continuation [24], homomeric tailing [3,10], Ribo-SPIA technology (Ovation Pico WTA and WT-Ovation One-Direct Amplification Systems) [25,26], TransPlex Whole Transcriptome Amplification technology (Pico Profiling) [27], template switching [28,29], multiple displacement amplification (total transcript amplification [30]) and linear antisense RNA amplification [6,8].
The underlying RNA or cDNA amplification strategies employed in most of these studies include either linear antisense RNA amplification or homomeric/TdT tailing followed by exponential amplification. While the former approach has been a mainstay for amplifying nanogram amounts of total RNA, there have been relatively few studies in which single cell quantities have been assayed [6,8]. Reported disadvantages to this approach include inefficiencies during second strand cDNA synthesis and purification [31], a multi-day workflow [32], time-dependent RNA degradation [33], as well as transcript representation bias [34] all of which are associated with successive rounds of amplification. Variations of the latter approach include A- [3,4] or G-tailing [10] in order to tag the 3’ end of the first cDNA strand for global PCR amplification. A third strategy by which cDNA may be tagged makes use of a reverse transcriptase with terminal transferase activity facilitating template-switching [20,28,35,36]. Other options by which the 3’ termini of cDNAs may be tagged, include linker/adaptor ligation [37] or the use of a terminal-tagging oligo (TTO) [38]. The linker/adaptor ligation protocol generally requires several additional enzymatic and washing steps, and is therefore not only prone to loss of material, but also cross-contamination. Because both of these methods currently require nanogram quantities of total RNA as inputs, it is likely that the efficiency with which mRNAs are converted into tagged and amplifiable cDNA templates is lower than either the template-switch or homomeric/TdT tailing methods. Recently, a q29 DNA polymerase-based multiple displacement amplification method was described in which the transcriptomes of single bacterial cells were profiled, yielding assay reproducibilities of $R^2 \approx 0.80$ [30]. While, currently this isothermal technology is adapted for prokaryotes, the authors suggest that it may be modified for use within a eukaryotic context.

Many of these approaches have not been widely adopted either because they suffer from amplification bias, are not sufficiently scalable or robust for high-throughput applications, are not suitable in eukaryotic contexts, or a combination of these factors. Here we describe a template-switch-based high-throughput method that is capable of generating robust whole-genome gene expression profiles at the single cell level.

Results

The pre-amplification method described here exploits the template switching ability of some reverse transcriptases which allows the 3’ tagging of cDNA, thereby facilitating the incorporation of universal PCR primer sites at both ends of the synthesized cDNAs (Fig. 1). Here we report on the comprehensive characterization of the performance of our single cell gene expression assay, termed Whole-Genome Gene Expression in Single Cells (WG-XSC), using picogram quantities of total RNAs, as well as a variety of different single cell types. We describe the utility of the WG-XSC assay in the transcriptional profiling of single cells and low input material, for which existing conventional methods are not sufficiently sensitive.

Pre-Amplification Assay Optimization

Previous template-switching-based amplification protocols utilized oligo-dT-based priming for cDNA synthesis followed by a single-phase PCR amplification reaction [20,28,35,36]. We made several modifications to these specific steps that led to substantial improvements in both the cDNA yield and representation of single cell quantities of starting material. These improvements are summarized in Table 1.

We first assessed the impact of different cDNA priming methods, during the reverse transcriptase step, on the performance of our assay. Here we evaluated three conditions, namely: oligo-dT (T30), oligo-dT + random hexamer (T30+N6) or oligo-dT + random nonamer (T30+N9). Replicate inputs of 50 pg H9 cell total RNA were used for all tested priming conditions after which pre-amplified products were used as inputs for the 24 K WG-DASL Assay. While typical assay reproducibilities of $R^2 \approx 0.37$ were obtained for the T30 condition, improved self-correlations of $R^2 \approx 0.63$ were observed for both the T30+N6 and the T30+N9 priming conditions (Table 1). We also obtained a concomitant 58% increase in the

![Figure 1. Pre-amplification scheme.](image-url)

(1) First strand cDNA synthesis is primed with tagged oligo-dT and random 9-mer primers. The tagged oligo-dT primer contains a VN anchor followed by a T-30 stretch with a 5’ PCR tag. The tagged random 9-mer consists of a 9-mer followed by the identical 5’ PCR tag. (2) Upon reaching the 5’ terminus of the mRNAs, the reverse transcriptase, via its terminal transferase activity, adds a few nucleotides (predominantly deoxyadenosine) to the 3’ end of the newly synthesized cDNAs. (3) The template-switch primer, which consists of the same 5’ PCR tag as well as a 3’ ribotubulin stretch, anneals via GC complimentary base-pairing to the 3’ end of the cDNAs, thereby serving as a new template for the reverse transcriptase. (4) After cDNA synthesis, both ends of the cDNAs now contain the identical PCR tag, allowing exponential amplification of the entire cDNA population through single primer PCR (5).

doi:10.1371/journal.pone.0030794.g001
assay sensitivity with the T30 + randomer priming conditions yielding approximately 10449 and 10332 probes detected (p<0.01) for the T30+N6 and T30+N9 conditions, respectively, compared with 6595 detected probes (p>0.01) for the T30 condition (Table 1). Moreover, raw intensity correlations of the lower 50 pg with the higher 1 ng input yielded R² = 0.69 for both the T30+N6 and the T30+N9 priming conditions, while yielding R² = 0.47 for the T30 condition (Table 1). Our results clearly demonstrate an improvement in the assay performance with the oligo-dT + randomer compared to the oligo-dT priming.

Previous experiments performed with different numbers of PCR cycles (15, 18, 21, 24 and 27 cycles) using different RNA inputs (50 pg and 1 ng) demonstrated that the assay performance (reproducibility, sensitivity and correlation with higher inputs), was poorest at the extremes of our chosen cycle ranges (15 and 27), but optimal at 21 PCR cycles (data not shown). To reduce the impact of stochastic effects on low copy numbers during the early cycles, we sought to improve the efficiency and fidelity of amplification by applying an altered thermal profile for the first few PCR cycles. We next therefore assessed the effect of two different PCR cycling profiles on our assay performance, namely a single-phase profile with an annealing temperature of 65°C, and a 24 cycle, two-phase profile consisting of an initial five PCR cycles carried out at a lower annealing temperature (58°C), followed by 19 cycles at a higher (65°C) annealing temperature (see Materials and Methods for details). For this experiment, replicate inputs of 50 pg UHR total RNA were processed using the 24 K WG-DASL Assay. RNA quality was assessed using the Bioanalyzer 2100 and yielded RIN values of 9.6 and 9.2 for the UHR and BR samples, respectively (data not shown).

Having obtained robust data using picogram quantities of RNA, we next repeated the experiment, using individual cells as inputs. Here we used single HeLa and primary brain tumor (BT) cells. As before, all samples were processed in triplicate. We observed a similar trend to that obtained for the RNA equivalent inputs, with intra-sample self-reproductivities of R²=0.42, R²=0.69 and R²=0.96 for the 10 pg, 50 pg and 1 ng UHR and R²=0.34, R²=0.61 and R²=0.95 for the 10 pg, 50 pg and 1 ng BR RNA inputs, respectively (Fig. 2A, Table 1). By contrast, comparisons between the UHR and BR RNA samples, at the 50 pg input level, yielded inter-sample correlations of R²=0.39 (Fig. 2A), whereas UHR vs. BR inter-sample correlations at the 1 ng input level yielded R²=0.61, suggesting that, based on gene expression profiles, our assay can reliably differentiate between different low RNA inputs.

Table 1. Values are averages for at least two technical replicates.

| Table 1. Optimization of Preamplification Assay |
|-----------------------------------------------|
| **Input** | **Condition** | **Self-Reproducibility (R²)** | **Correlation with 1 ng (R²)** | **Sensitivity** | **Probe Concordance (%)** |
|-------------------------|----------------|-------------------------------|-------------------------------|----------------|-------------------------|
| 50 pg UHR total RNA⁴    | T30 + one-phase PCR | 0.374                         | 0.473                         | 6595           | 95.9                    |
| 50 pg UHR total RNA⁴    | T30 + two-phase PCR | 0.481                         | 0.585                         | 8019           | 95.1                    |
| 50 pg H9 total RNA⁴     | T30 + N6 + one-phase PCR | 0.626                        | 0.695                         | 10449          | 92.4                    |
| 50 pg H9 total RNA⁴     | T30 + N9 + one-phase PCR | 0.627                        | 0.688                         | 10332          | 92.2                    |
| 50 pg UHR total RNA⁴    | T30 + N9 + two-phase PCR | 0.698                        | 0.806                         | 13443          | 96.6                    |
| Single HeLa cells⁴      | T30 + N9 + two-phase PCR | 0.757                        | 0.801                         | 11083          | 97.4                    |

Values shown for the self-reproducibility and correlation are derived from all probes.

⁴Sensitivity is calculated as the number of probes detected at p-value<0.01.

⁵Probe concordance is calculated as a percentage of the number of probes with matching detected calls at p-value<0.01 between the low (50 pg or single cell) and standard (1 ng) inputs divided by the total number of probes detected in the lower input.

doi:10.1371/journal.pone.0030794.t001

in input levels. In order to further characterize our assay we used T30+N9 priming together with the two-phase PCR profile described earlier to assay two different RNA inputs. Triplicate aliquots of UHR and BR, each at 10 pg, 50 pg and 1 ng total RNA were used in conjunction with the 29 K WG-DASL HT Assay. RNA quality was assessed using the Bioanalyzer 2100 and yielded RIN values of 9.6 and 9.2 for the UHR and BR samples, respectively (data not shown). On average, our intra-sample self-reproductivities were R²=0.42, R²=0.69 and R²=0.96 for the 10 pg, 50 pg and 1 ng UHR and R²=0.34, R²=0.61 and R²=0.95 for the 10 pg, 50 pg and 1 ng BR RNA inputs, respectively (Fig. 2A, Table 1). By contrast, comparisons between the UHR and BR RNA samples, at the 50 pg input level, yielded inter-sample correlations of R²=0.39 (Fig. 2A), whereas UHR vs. BR inter-sample correlations at the 1 ng input level yielded R²=0.61, suggesting that, based on gene expression profiles, our assay can reliably differentiate between different low RNA inputs.

Having obtained robust data using picogram quantities of RNA, we next repeated the experiment, using individual cells as inputs. Here we used single HeLa and primary brain tumor (BT) cells. As before, all samples were processed in triplicate. We observed a similar trend to that obtained for the RNA equivalent inputs, with intra-sample self-reproductivities of R²=0.42, R²=0.69 and R²=0.96 for the 10 pg, 50 pg and 1 ng UHR and R²=0.34, R²=0.61 and R²=0.95 for the 10 pg, 50 pg and 1 ng BR RNA inputs, respectively (Fig. 2A, Table 1). By contrast, comparisons between the UHR and BR RNA samples, at the 50 pg input level, yielded inter-sample correlations of R²=0.39 (Fig. 2A), whereas UHR vs. BR inter-sample correlations at the 1 ng input level yielded R²=0.61, suggesting that, based on gene expression profiles, our assay can reliably differentiate between different low RNA inputs.

Performance with RNA Inputs and Single Cells

A key performance characteristic of any single cell genomics assay is its ability to discriminate among different samples at low input levels. In order to further characterize our assay we used T30+N9 priming together with the two-phase PCR profile described earlier to assay two different RNA inputs. Triplicate aliquots of UHR and BR, each at 10 pg, 50 pg and 1 ng total RNA were used in conjunction with the 29 K WG-DASL HT Assay. RNA quality was assessed using the Bioanalyzer 2100 and yielded RIN values of 9.6 and 9.2 for the UHR and BR samples, respectively (data not shown). On average, our intra-sample self-reproductivities were R²=0.42, R²=0.69 and R²=0.96 for the 10 pg, 50 pg and 1 ng UHR and R²=0.34, R²=0.61 and R²=0.95 for the 10 pg, 50 pg and 1 ng BR RNA inputs, respectively (Fig. 2A, Table 1). By contrast, comparisons between the UHR and BR RNA samples, at the 50 pg input level, yielded inter-sample correlations of R²=0.39 (Fig. 2A), whereas UHR vs. BR inter-sample correlations at the 1 ng input level yielded R²=0.61, suggesting that, based on gene expression profiles, our assay can reliably differentiate between different low RNA inputs.

Having obtained robust data using picogram quantities of RNA, we next repeated the experiment, using individual cells as inputs. Here we used single HeLa and primary brain tumor (BT) cells. As before, all samples were processed in triplicate. We observed a similar trend to that obtained for the RNA equivalent inputs, with intra-sample self-reproductivities of R²=0.42, R²=0.69 and R²=0.96 for the 10 pg, 50 pg and 1 ng UHR and R²=0.34, R²=0.61 and R²=0.95 for the 10 pg, 50 pg and 1 ng BR RNA inputs, respectively (Fig. 2A, Table 1). By contrast, comparisons between the UHR and BR RNA samples, at the 50 pg input level, yielded inter-sample correlations of R²=0.39 (Fig. 2A), whereas UHR vs. BR inter-sample correlations at the 1 ng input level yielded R²=0.61, suggesting that, based on gene expression profiles, our assay can reliably differentiate between different low RNA inputs.

Having obtained robust data using picogram quantities of RNA, we next repeated the experiment, using individual cells as inputs. Here we used single HeLa and primary brain tumor (BT) cells. As before, all samples were processed in triplicate. We observed a similar trend to that obtained for the RNA equivalent inputs, with intra-sample self-reproductivities of R²=0.42, R²=0.69 and R²=0.96 for the 10 pg, 50 pg and 1 ng UHR and R²=0.34, R²=0.61 and R²=0.95 for the 10 pg, 50 pg and 1 ng BR RNA inputs, respectively (Fig. 2A, Table 1). By contrast, comparisons between the UHR and BR RNA samples, at the 50 pg input level, yielded inter-sample correlations of R²=0.39 (Fig. 2A), whereas UHR vs. BR inter-sample correlations at the 1 ng input level yielded R²=0.61, suggesting that, based on gene expression profiles, our assay can reliably differentiate between different low RNA inputs.
Figure 2. Raw signal intensity correlations between replicates of low input RNAs and whole cells. (A) 50 pg UHR and BR total RNA and (B) single HeLa and brain tumor (BT) cells; 50 cell tumorsphere (TS) and adherent cells (AC). Pair-wise scatterplots of at least two replicates for each input type are shown for all 29 K probes across the full range of raw signal intensities. Correlations are the square of Pearson’s correlation coefficient. doi:10.1371/journal.pone.0030794.g002
and 100 under-expressed genes in the tumorspheres relative to their attached counterpart. Using DAVID [39] we extracted GO terms and compared the relative frequencies of these terms in each list to that in the human genome to identify statistically significant enriched GO terms. Within the over-expressed gene list several transcription factors such as SMAD2, SHOC2 and KLHL7 were included under the enriched GO term “nucleus” (p = 0.71E-3) consistent with a stem cell (Fig. S1) versus a more differentiated cellular phenotype. Conversely, for the under-expressed genes over-represented GO terms included “cell/biological adhesion” (p = 5.5E-5) as well as “extracellular space” (p = 5.3E-3) suggestive of a cell-matrix adhesion pathway for an attached (Fig. S1) as opposed to a suspended cell culture.

In order to determine the extent to which the gene expression profiles obtained at low input levels correlated with those obtained with higher inputs, we directly compared raw signal intensities between the lower and higher inputs. Correlations between 50 pg and 1 ng total RNA typically yielded R^2 = 0.80 (Fig. 3A, Table 1), whereas correlations between 10 pg and 1 ng total RNA typically yielded R^2 = 0.59 (Fig. 3B). Single cell correlations with 1 ng total RNA, derived from a corresponding bulk cell culture, yielded R^2 = 0.80 (Fig. 3C). At p < 0.01, we detected ~13,443 and 10,180 probes for the 50 and 10 pg RNA inputs, respectively, whilst detecting ~14,156 and 11,083 probes for the 5-cell and single cell inputs (Table 1), respectively. This level of sensitivity represents approximately 77% (50 pg), 58% (10 pg), 80% (5-cell) and 63% (single cell) of the total number of probes detected in the higher, standard inputs. Furthermore, when the lists of probes detected (p < 0.01) in the lower inputs were intersected with those detected in the higher 1 ng inputs, we obtained probe concordance values of ~96.6%, 97.1% and 97.4% for the 50 pg (Fig. 3D, Table 1), 10 pg (Fig. 3E) and single HeLa cells (Fig. 3F, Table 1), respectively. The percentage of false positive probes detected in the lower inputs was <2% of the total number of probes detected in the higher standard inputs. Taken together these results demonstrate that our assay is sufficiently sensitive to reliably detect, in low inputs, most of the genes that are detected at standard higher inputs, and that the expression profiles derived from these lower inputs accurately recapitulate those obtained in higher inputs.

**Discussion**

Over the last few years there have been several reported studies on either single cell gene expression profiling using low gene density (1–100) assays [1,2,16,40–42], or inputs of small populations of cells [10–150] [27,43] or 100 pg–1 ng total RNA [44–46] at the whole transcriptome level. Very few genome-wide studies have been reported in which the assay performance has been rigorously characterized using whole single cells or RNA equivalents as inputs [3,4,10,47].

Our WG-XSC assay is highly reproducible, typically yielding R^2 = 0.76 and ~0.69 for single cell and 50 pg RNA inputs, respectively. The transcript representation as assessed by the correlation between lower inputs and larger standard inputs yielded R^2 = 0.80 and ~0.81 for single cell and 50 pg RNA inputs, respectively. Of the few microarray-based single cell transcriptional studies with self-correlation metrics the reported R values range between 0.73–0.91 [3,10,48]. Exact comparisons between these studies and the current study is challenging as either the underlying experimental designs differ [10] or the analysis methods are different [3,10,48]. A recent mRNA-Seq study in which single mouse oocytes were assayed, reported assay reproducibilities of R^2 = 0.97 [4]. It should be noted however that these cells are atypical in size and therefore are also not directly comparable with the current study.

Two obvious, but critical steps that could impact levels of reproducibility and representation include the extent of cell lysis as well as the efficiency with which low abundance mRNA molecules are converted to cDNA. In order to minimize the loss of material, and maximize the synthesis of cDNA in an unbiased fashion, our protocol specifically incorporates the use of a

---

**Figure 3. Intensity and detected probe concordance comparisons between low and higher 1 ng inputs.** Raw signal intensity correlations between (A) 50 pg (x-axis) and 1 ng (y-axis) UHR total RNA; (B) 10 pg (x-axis) and 1 ng (y-axis) UHR total RNA; (C) single HeLa cell (x-axis) and 1 ng (y-axis) HeLa total RNA. The overlapping sets of detected probes between the low and higher inputs are shown for both the RNA equivalent (D, E) and single cell (F) inputs. All probe values shown are at a threshold of p < 0.01.

doi:10.1371/journal.pone.0030794.g003
were used for each pre-amplification reaction. CA, USA). Different inputs, as indicated in the Results section, from Life Technologies/Ambion, Austin, TX, USA) and Univer-

(HeLa-S3) and FirstChoice Human Brain Reference (BR) (both following vendors, FirstChoice Human Cervical Adenocarcinoma RNase-free water. Commercial RNAs were purchased from the Lysis Buffer (SLB, Illumina, Inc.).

biomarker discovery.
molecular lineage tracing studies and clinical studies aimed at being extremely valuable in a variety of contexts, including detailed development and tumor stem cell biology.

Our high-throughput assay generates whole-genome gene expression profiles with single cell or low RNA inputs. This robust and scalable method for profiling a variety of cell types at the single cell level can be applied to critical questions in a broad range of areas, including developmental biology and cancer biology. We have used the technology for gene expression profiling in circulating tumor cells isolated from prostate cancer and biology. We have used the technology for gene expression profiling in circulating tumor cells isolated from prostate cancer and ovarian cancer patients’ blood, as well as molecular and functional characterization of early lineage commitment of human hematopoietic stem cells (data not shown). The ability to obtain genome-wide gene expression data on many individual cells in parallel will be extremely valuable in a variety of contexts, including detailed molecular lineage tracing studies and clinical studies aimed at biomarker discovery.

Materials and Methods

RNA Extraction
RNA from the WA09 (H9) [49] human embryonic stem cell line was extracted using TR1zl (Life Technologies/Invitrogen, Carlsbad, CA, USA) according to the manufacturer’s instructions after which the precipitated RNA pellet was resuspended in 10 μl RNase-free water. Commercial RNAs were purchased from the following vendors, FirstChoice Human Cervical Adenocarcinoma (HeLa-S3) and FirstChoice Human Brain Reference (BR) (both from Life Technologies/Ambion, Austin, TX, USA) and Universal Human Reference (UHR) (Agilent/Stratagene, Santa Clara, CA, USA). Different inputs, as indicated in the Results section, were used for each pre-amplification reaction.

Sorting of Cultured Cells with a Phase-Switch Microfluidics Device
A microfluidics device with a phase-switch feature was used for isolating individual cells. Briefly, cultured cells were harvested with trypsinization and washed with PBS, whereafter a single cell suspension in PBS was loaded into a phase-switch microfluidics device for encapsulation of individual cells into droplets. Cells were encapsulated from the aqueous phase (PBS) into droplets in the oil phase by either laser-cavitation or T-junction break-up of immiscible threads as previously described [50]. With optimized parameters, each droplet contained only one single HeLa-S-Fucci (RIKEN BioResource Center Cell Bank, Ibaraki, Japan) [51] or brain tumor (BT) cell. This was visually confirmed under a fluorescent microscope. This approach minimizes the stress on sorted cells and facilitates the manipulation of single cells in droplets. Individual sorted cells were aliquoted in ≤1 μl Single cell Lysis Buffer (SLB, Illumina, Inc.).

Culture of Tumorspheres and Adherent Cells
An ovarian cancer cell line, RMG1 [52] was cultured in three T-150 flasks until 80–90% confluency in M199/MCDB105 medium (Sigma, St. Louis, MO, USA) supplemented with 5% FBS (HyClone Laboratories Ltd, Logan, UT, USA), penicillin (100 U/ml), and streptomycin (100 μg/ml). To culture under stem cell conditions [53] cells were trypsinized and then resuspended in DMEM/F12 medium supplemented with 5 μg/ml insulin (Novo Nordisk Inc., Princeton, NJ, USA), 20 ng/ml epidermal growth factor (R&D Systems, Inc., Minneapolis, MN, USA), 10 ng/ml basic fibroblast growth factor (Invitrogen, Carlsbad, CA, USA), 0.4% bovine serum albumin (BD Falcon; Bedford, MA, USA), penicillin (100 U/ml), and streptomycin (100 μg/ml) in two 100 mm Ultra Low Attachment plates (Corning, Lowell, MA, USA) over three weeks. Spheroids were selected using a 40 μm cell strainer (BD Falcon, Bedford, MA, USA), after which half of the spheroids were cultured in one 35 mm dish (adherent cells, AC) and the other half were grown in one well of one 6 well low attachment plate in stem cell medium (tumorspheres, TS) for two days. Cells were grown at 37°C in a 5% CO₂ air atmosphere. The brain tumor (BT) cells, derived from U118 human glioblastoma cells (ATCC HTB-15), were purchased from the ATCC (Rockville, MD, USA) and maintained in DMEM supplemented with 10% fetal bovine serum, glutamine, and penicillin/streptomycin as recommended by the ATCC.

Cell Lysis, cDNA Synthesis and Pre-Amplification
All cell lysis and cDNA reactions were performed using 0.2 ml Maxymum Recovery PCR tubes (Axogen, Union City, CA, USA). The cell lysis, reverse transcription, template switching and pre-amplification reactions were all performed in a single tube. Briefly, for cell lysis, 1.8 μl SLB was added directly to the isolated single cell. Tubes were placed in a thermocycler and heated to 72°C for 3 min, followed by five min at 4°C. After cell lysis, 3.2 μl Single cell cDNA Synthesis Buffer (SCB, Illumina, Inc.) was added to the lysed single cell. The reverse transcription and template switching reactions were performed at 42°C for 60 min, followed by a 10 min 70°C inactivation step. After cDNA synthesis 32 μl of Single cell PCR Mix (SPM, Illumina, Inc.) was added directly to the unpurified products followed by amplification using a PCR cycling profile which consisted of an initial denaturation of 95°C for 1 min, followed by 5 cycles of (95°C for 20 sec 58°C for 30 sec and 68°C for 3 min), 9 cycles of (95°C for 20 sec, 65°C for 30 sec and 68°C for 3 min), 10 cycles of (95°C for 30 sec, 65°C for 30 sec and 68°C for 3 min+6 sec/cycle) and 1 cycle of 72°C for 10 min. For cell-equivalent RNA inputs, the SLB and SCB were added directly to the RNA (the cell lysis step was omitted) and were reverse-transcribed and pre-amplified in the identical manner to that described for the cell lysates.

Whole-Genome DASL® Assay
For whole-genome gene expression analysis, we used either the Whole-Genome DASL Assay or the Whole-Genome DASL HT Assay, an updated version of the original Whole-Genome DASL Assay [54]. Briefly, the WG-DASL assay probes ~24 K targets (~18 K unique genes) and uses the HumanRef-8 v3 BeadChip, while the WG-DASL HT Assay interrogates ~29 K targets (~21 K unique genes), based on content derived from the National Center for Biotechnology Information Reference Sequence Database (Release 98, November 2009) and uses the Human HT-12 v4 BeadChip. For both 24 K and 29 K assays we used 10 μl of the 37 μl (27%) pre-amplified cDNA product which we annealed directly to the 24 K or 29 K oligo pool and then proceeded exactly as previously described [54].
Whole-Genome Gene Expression of Single Cells

Microarray Data Analysis

Unless otherwise stated, all data were analyzed in an un-normalized, raw state. All individual samples were assayed a minimum of two times. After scanning, intensity data were imported into GenomeStudio® v2.0 where the data quality was assessed using several assay controls. Detection p-values were computed using several hundred negative controls to determine gene expression detection limits. Assay performance metrics are described further in the Results section. All of the microarray data are MIAME-compliant (http://www.mged.org/Workgroups/MIAME/miame.html) and have been submitted to GEO (Accession Number: GSE34365). Over-representation analysis of differentially expressed genes was performed using DAVID [39], which reports functional gene categories as statistically significant gene ontology (GO) terms.

Supporting Information

Figure S1 Ovarian cancer cells (RMG1) were cultured in stem cell media. (A) a spheroid in a ultra-low attachment plate, (B) a spheroid in a tissue-culture treated plate, after two days in a regular plate. Scale bar = 90 μm.

(PDF)

Acknowledgments

We would like to thank Shujin Luo and Jerry Wang at Illumina, Inc. for helpful discussions.

Author Contributions

Conceived and designed the experiments: CSA MB J Chen JBF JSF FK MR. Performed the experiments: CSA J Chen BK. Analyzed the data: CSA J Chen JBF. Contributed reagents/materials/analysis tools: GYC J Chen LCL SL SVN TO JFZ. Wrote the paper: CSA JBF.
41. Gibson JD, Jakuba CM, Boucher N, Holbrook KA, Carter MG, et al. (2009) Single-cell transcript analysis of human embryonic stem cells. Integr Biol (Camb) 1: 540–551.

42. Sta˚hlberg A, Bengtsson M (2010) Single-cell gene expression profiling using reverse transcription quantitative real-time PCR. Methods 50: 292–298.

43. Irimia D, Mindrinos M, Russo M, Xiao W, Wilhelm J, et al. (2009) Genome-wide transcriptome analysis of 150 cell samples. Integr Biol (Camb) 1: 99–107.

44. Dafforn A, Chen P, Deng G, Herrler M, Iglesias D, et al. (2004) Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis. Biotechniques 37: 854–857.

45. Singh R, Maganti RJ, Jabba SV, Wang M, Deng G, et al. (2005) Microarray-based comparison of three amplification methods for nanogram amounts of total RNA. Am J Physiol Cell Physiol 288: C1179–1189.

46. Lang JE, Magbanua MJ, Scott JH, Makrigiorgos GM, Wang G, et al. (2009) A comparison of RNA amplification techniques at sub-nanogram input concentration. BMC Genomics 10: 328.

47. Ramos CA, Bowman TA, Boles NC, Merchant AA, Zheng Y, et al. (2006) Evidence for diversity in transcriptional profiles of single hematopoietic stem cells. PLoS Genet 2: e159.

48. Appay V, Bosio A, Jabba SV, Wang M, Deng G, et al. (2007) Sensitive gene expression profiling of human T cell subsets reveals parallel post-thymic differentiation for CD4+ and CD8+ lineages. J Immunol 179: 7406–7414.

49. Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, et al. (1998) Embryonic stem cell lines derived from human blastocysts. Science 282: 1145–1147.

50. Zhong JF, Chen Y, Marcus JS, Scherer A, Quake SR, et al. (2006) A microfluidic processor for gene expression profiling of single human embryonic stem cells. Lab Chip 6: 68–74.

51. Sakaue-Sawano A, Kurokawa H, Morimura T, Hanyu A, Hama H, et al. (2006) Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. Cell 132: 487–498.

52. Kiguchi K, Iwamori M, Mochizuki Y, Kishikawa T, Tsukazaki K, et al. (1998) Selection of human ovarian carcinoma cells with high dissemination potential by repeated passage of the cells in vivo into nude mice, and involvement of Le(x)-determinant in the dissemination potential. Jpn J Cancer Res 89: 923–932.

53. Zhang S, Balch C, Chau MW, Lai HC, Matei D, et al. (2008) Identification and characterization of ovarian cancer-initiating cells from primary human tumors. Cancer Res 68: 4311–4320.

54. April C, Klotzle B, Royce T, Wickham-Garcia E, Boyanovisky T, et al. (2009) Whole-genome gene expression profiling of formalin-fixed, paraffin-embedded tissue samples. PLoS One 4: e6162.