

Highly sensitive and specific microRNA expression profiling using BeadArray technology

Jing Chen, Jean Lozach, Eliza Wickham Garcia, Bret Barnes, Shujun Luo, Ivan Mikoulitch, Lixin Zhou, Gary Schroth and Jian-Bing Fan*

Illumina, Inc. 9885 Towne Centre Drive, San Diego, CA 92121, USA

Received February 25, 2008; Revised June 2, 2008; Accepted June 3, 2008

ABSTRACT

We have developed a highly sensitive, specific and reproducible method for microRNA (miRNA) expression profiling, using the BeadArray™ technology. This method incorporates an enzyme-assisted specificity step, a solid-phase primer extension to distinguish between members of miRNA families. In addition, a universal PCR is used to amplify all targets prior to array hybridization. Currently, assay probes are designed to simultaneously analyse 735 well-annotated human miRNAs. Using this method, highly reproducible miRNA expression profiles were generated with 100–200 ng total RNA input. Furthermore, very similar expression profiles were obtained with total RNA and enriched small RNA species ($R^2 \geq 0.97$). The method has a 3.5–4 log (10^5 – 10^9 molecules) dynamic range and is able to detect 1.2- to 1.3-fold-differences between samples. Expression profiles generated by this method are highly comparable to those obtained with RT-PCR ($R^2 = 0.85$ – 0.90) and direct sequencing ($R = 0.87$ – 0.89). This method, in conjunction with the 96-sample array matrix should prove useful for high-throughput expression profiling of miRNAs in large numbers of tissue samples.

INTRODUCTION

MicroRNAs (miRNAs) are small (~21 nt) endogenous non-coding RNAs that have been shown to influence the abundance and translational efficiency of cognate mRNAs (1,2). Since the discovery of the miRNA lin-4 in *C. elegans* (3), many miRNAs have been identified in a wide variety of plants and metazoans (4). According to the most recent miRBase release (<http://microrna.sanger.ac.uk/>; Release 10.0: August 2007), there are over 5000 validated miRNAs, including 528 human miRNAs (5). There are many more predicted miRNAs that have not been

validated. It has been estimated that there are a total of at least 800 human miRNAs (6).

miRNAs are transcribed as long precursors (pri-miRNAs) that are processed by Drosha, resulting in an ~70-nt stem-loop structure (pre-miRNAs). The pre-miRNAs are transported to the cytoplasm, and are further processed by the Dicer-containing complex, resulting in 17- to 27-nt mature miRNAs. The mature miRNAs are loaded in the RNA-induced silencing complex (RISC) that can effect gene silencing through sequence-specific base pairing with target messenger RNAs (mRNAs), resulting in either transcriptional/translational repression or target breakdown. It has been shown that each miRNA can regulate the expression level of hundreds of different mRNAs, and between 20% and 30% of all transcripts are regulated by miRNAs in mammalian genomes (7). Many developmental and cellular processes have now been found to be under critical regulation by miRNAs; miRNA dis-regulation has been implicated in the aetiology of diseases such as cancer (8,9), heart diseases (10) and Parkinson's disease (11). To further facilitate this type of study, a tool is needed that is sensitive enough to measure the expression levels of miRNAs specifically in small tissue samples.

Several unique attributes of miRNAs, including their small size, lack of polyadenylated tails, tendency to cross-hybridize to their mRNA targets with imperfect sequence homology, significant sequence homology among family members, have made them challenging to quantify. Many methods have been developed for miRNA profiling, including quantitative PCR (12), sequencing (13–17), northern blotting (18,19) and microarray analyses based on either direct hybridization (8,20–25) or hybridization coupled with enzymatic extension (26). While these methods have been used successfully in a variety of studies, they still have some technical limitations. For example, some of these methods need large amounts of starting materials (e.g. >10 µg of total RNA), while some require enrichment of small RNA species in order to lower cross-hybridization from mRNA, even though the enrichment procedure itself adds variation to the measurement.

*To whom correspondence should be addressed. Tel: 858 202 4588; Fax: 858 202 4680; Email: jfan@illumina.com

To overcome these technical difficulties, we have developed a highly sensitive and specific miRNA assay. It offers unique advantages for specificity over other sequence hybridization-based expression profiling platforms. It can use as little as 2 ng total RNA as the starting material, significantly lower than some existing methods. The method has a 3.5–4 log dynamic range and is able to detect 1.2- to 1.3-fold-difference between samples. Furthermore, we demonstrate that robust expression profiles can be generated with miRNAs isolated from formalin-fixed, paraffin-embedded (FFPE) tissues, which opens up new opportunities for analyses of small RNAs from archival human tissues.

MATERIALS AND METHODS

RNA samples

Small RNA-containing total RNAs extracted from four cell lines [prostate adenocarcinoma (PC-3), breast adenocarcinoma (MCF-7), embryonic kidney (293) and cervical Adenocarcinoma (HeLa)] and human tissues were purchased from Ambion, Inc. In general, total RNAs were used for miRNA profiling experiments if not otherwise stated; for some experiments, small RNA molecules were enriched using Invitrogen's PureLink miRNA Isolation Kit. FFPE samples were purchased from Biochain Institute, Inc., and RNA was extracted using Qiagen's RNeasy FFPE Kit.

miRNA profiling on universal BeadArray platform

The method is a modification of an assay that we developed previously for high-throughput gene expression profiling, the DASL[®] Assay (cDNA-mediated annealing, selection, extension and ligation) (27). The miRNA method similarly targets specific sequences with sets of oligonucleotides that are extended, and then labelled during PCR amplification. As shown in Figure 1, miRNAs were first polyadenylated using Poly-A Polymerase [5 μ l RNA plus 5 μ l polyadenylation reaction reagent (PAS, Illumina), incubated at 37°C for 60 min, then heat inactivated at 70°C for 10 min]. The standard miRNA expression profiling assay protocol requires an input of 100–200 ng of total RNA, although amounts as low as 2 ng have shown good reproducibility.

The introduced poly A tail was then used as a priming site for cDNA synthesis [8 μ l polyadenylation reaction plus 8 μ l cDNA synthesis reagent (CSS, Illumina), incubated at 42°C for 60 min, then heat inactivated at 70°C for 10 min]. As shown in Figure 1, the primer used for cDNA synthesis was biotinylated and contained a universal PCR primer sequence that was used later in the assay. After cDNA synthesis, miRNAs were individually interrogated using specific oligonucleotides. A single miRNA-specific Oligo (MSO) was designed against each mature miRNA sequence, which consists of three parts: at the 5'-end is another universal PCR priming site; in the middle is an address sequence used for capturing the product on the array; and at the 3'-end is a miRNA-specific sequence. The second universal PCR priming site is shared among all

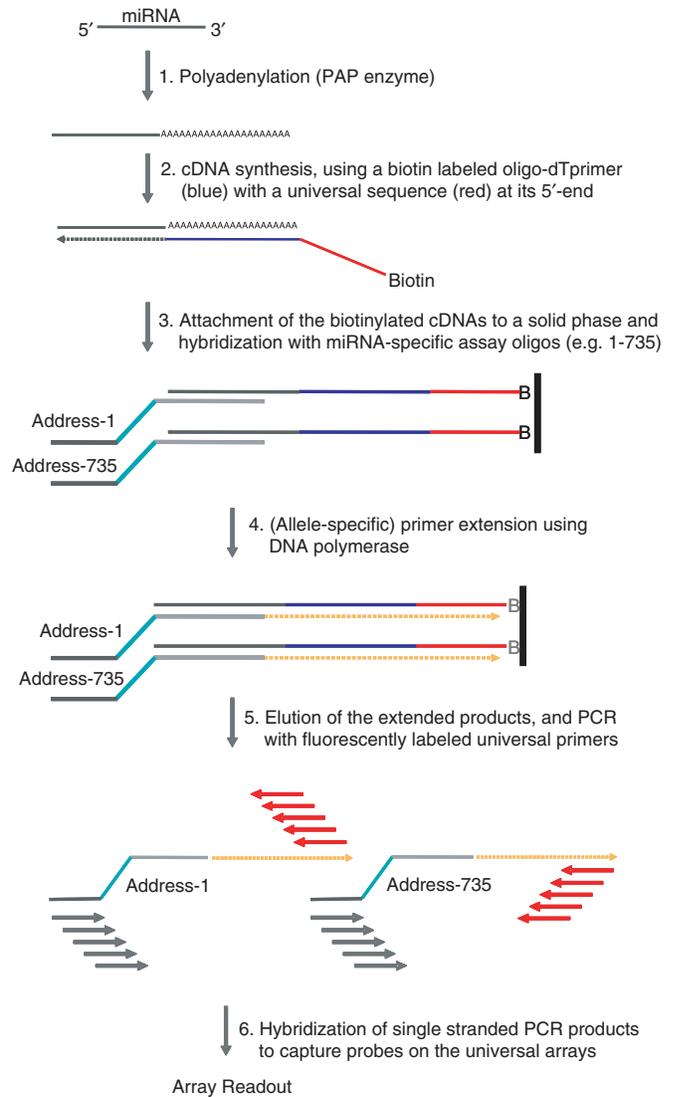


Figure 1. miRNA assay scheme. (1) Polyadenylate RNA: add multiple A (>18 bases) to 3'-ends of total RNA or purified short RNA species, including microRNAs. (2) cDNA synthesis of microRNA: synthesize cDNA using a biotin-labelled oligo-dT primer (blue) with a universal sequence (red) at its 5'-end. (3) Hybridize assay oligos to cDNA: attach biotinylated cDNA to a solid phase and hybridize with a pool of microRNA-specific oligos. (4) microRNA-specific primer extension: extend primers using DNA polymerase. (5) Universal PCR: elute the extended products and perform PCR with fluorescently labelled universal primers. Bind double-stranded PCR products to a solid phase and prepare the labelled, single-stranded PCR products for hybridization. (6) Hybridize ssDNA to arrays: hybridize PCR product to capture probes on the universal arrays.

MSO's, and each address sequence is associated uniquely with each of the 735 miRNA targets (see later).

The miRNA assay probes correspond to 470 well-annotated human miRNA sequences (miRBase: <http://microrna.sanger.ac.uk/>, version 9.1, February 2007 Release) and 265 miRNAs identified recently (28,29). Assay probes were designed with a T_m of $60 \pm 8.6^\circ\text{C}$ and a length of 17–21 nt (average 18 nt). To maximize assay specificity, candidate probes were examined collectively to minimize sequence similarity between probes, particularly

Table 1. Internal single base mismatch controls and 3'-end mismatch controls

Internal single base mismatch control	Assay probe sequence	Intensity ratio of perfect match/mismatch probes
hsa-let-7a	TGAGGTAGTAGGTTGTATAG	
let-7a_mis1	TGAGGTAGTAGGTTCTATAG	60
hsa-let-7c	TGAGGTAGTAGGTTGTATG	
let-7c_mis1	TGAGGTAGTAGGTTCTATG	37
let-7c_mis2	TGAGGTAGTAGCTTGTATG	44
hsa-let-7f	TGAGGTAGTAGATTGTATAGT	
let-7f_mis1	TGAGGTAGTAGATTCTATAGT	65
hsa-miR-152	TCAGTGCATGACAGAACTT	
miR-152_mis1	TCAGTGCATGACACAACCTT	19
hsa-miR-182	TTTGGCAATGGTAGAACTC	
miR-182_mis1	TTTGGCAATGGTACAACCTC	52
3'-end mismatch control		
RNU24	ACATTTTAAACCACCAAG	
RNU24-C	ACATTTTAAACCACCAAC	37
RNU24-A	ACATTTTAAACCACCAA	15
RNU24-T	ACATTTTAAACCACCAAT	38
RNU66	TGAGGTGGTTCTTTCTATCC	
RNU66-G	TGAGGTGGTTCTTTCTATCG	12
RNU66-T	TGAGGTGGTTCTTTCTATCT	21

at their 3'-ends. The sequence information for all the 735 miRNA-specific probes is included in Supplementary Table 1. As controls, we also designed central mismatch probes for miRNAs hsa-let-7a, let-7c, let-7f, miR-152 and miR-182, and 3'-end mismatch probes for small nucleolar RNAs RNU24 and RNU66 (Table 1).

The subsequent DASL assay process and array hybridization were performed as described previously (27). Briefly, 15 µl of the cDNA synthesis reaction was added to 5 µl of the multiplexed MSO pool (MAP, Illumina) and 30 µl of a reagent containing streptavidin paramagnetic particles (OB1, Illumina), heated to 70°C, and allowed to anneal to 40°C. All 735 human miRNAs were assayed simultaneously. After binding and washing, the annealed MSOs were extended through the cDNA primer, forming an amplifiable product. The extended oligos were eluted from the streptavidin beads and added to a PCR reaction, in which one of the universal primers was fluorescently labelled and the other universal primer was biotinylated. The PCR products were captured on streptavidin paramagnetic beads, washed and denatured to yield single-stranded fluorescent molecules to hybridize to the arrays.

The universal arrays used for fluorescent reporting consist of capture oligos immobilized on beads and randomly assembled into wells etched in the ends of fibre optic bundles, which are arranged in a matrix to match a 96-well plate (Sentrix® Array Matrix, Illumina) (30). The identity of each bead is determined before hybridization to the miRNA assay product, and the same arrays are used to report the results of similar assays employing the address sequence technique (GoldenGate® Genotyping Assay, DASL Gene Expression Assay, GoldenGate Methylation Assay) (30,31). Arrays were scanned on the BeadArray Reader, and automatic image registration and intensity

extraction software was used to derive intensity data per bead type corresponding to each miRNA (32).

Microarray data analysis

The array intensity data were imported into BeadStudio v3.2 (Illumina), a software package that permits visualization and normalization of the data. We used the 'Average' normalization method for all the analyses reported here except for assay reproducibility, where, given the number of replicates, 'Quantile' normalization appeared to be a better option. The 'Average' normalization method computes a global scaling factor that is applied to all probes and all arrays. The 'Quantile' normalization method was described previously (33). The normalized intensities and detection *P*-values were exported and further analysed using the *R* environment (version 2.6), in combination with Bioconductor packages. Assessment of limit of detection, dynamic range and fold-difference detection were performed using a combination of cell line RNA mixtures and synthetic RNA spikes.

Real-time quantitative RT-PCR (qPCR)

qPCR analyses were performed on the ABI Prism 7900HT sequence detection system (Applied Biosystems). RT-PCR primers for 12 miRNAs (miR-100, 125a, 125b, 135a, 146a, 150, 17-3p, 221, 26a, 31, 93 and 328) were purchased from ABL. Reverse transcription with miRNA-specific primer was performed using ABI's TaqMan MicroRNA Reverse Transcription kit, followed by real-time PCR protocol using miRNA-specific TaqMan primers as suggested by the manufacturer.

Digital gene expression (DGE) profiling

Digital gene expression profiling using the Genome Analyzer sequencing platform (Illumina) was performed for comparison and validation of miRNA assay results. Small RNAs (size ranging from 18 to 30 nt) were isolated from 10 µg total RNA on a 15% PAGE-Urea gel, and ligated to RNA adapter-1 (5'-UCGUAUGCCGUCU UCUGCUU-3'), and the ligated material was then ligated to adapter-2 (5'-GUUCAGAGUUCUACAGUCCGAC GAUC-3'). The final ligated materials were reverse transcribed with a RT-primer (5'-CAAGCAGAAGACGG CATA CGA-3'), and PCR amplified 15 cycles with primer-1 (5'-CAAGCAGAAGACGGCATA CGA-3') and primer-2 (5'-AATGATACGGCGACCACCGACA GGTTCAGAGTTCTACAGTCCGA-3'). Please note that primer-2 does not match adaptor-2 across the full length. Instead, extra bases were added to the 5'-end of the adaptor sequence for a tailed PCR. The amplified products, i.e. the small RNA libraries were loaded onto a Cluster Station (Illumina) where they bound to complementary adapter oligos grafted onto a proprietary flow cell substrate. The Cluster Station isothermally amplified these cDNA constructs to create clonal clusters of ~1000 copies each. The resulting high-density array of template clusters on the flow cell was directly sequenced by a fully automated Genome Analyzer (Illumina). Using a sequencing-by-synthesis approach, four proprietary fluorescently labelled, reversibly terminated nucleotides were used to

sequence the millions of clusters base by base in parallel with an accuracy rate >99.6% per cycle. A 32-bp read length was obtained for each cluster, and the adapter sequence immediately flanking the small RNA sequence was removed. The resulting tag sequences were blasted against the mature sequences of the 735 miRNAs present on the Illumina microarray panel using Eland, an alignment programme within the Genome Analyzer software suite that allows up to two mismatches. Tag sequences matched to the 735 miRNA sequences by the Eland algorithm were counted and used for cross-platform comparison.

RESULTS

The miRNA assay scheme

A diagram illustrating the miRNA assay scheme is shown in Figure 1. As discussed earlier, one MSO is designed to assay each miRNA on the panel. A given address sequence is uniquely associated with a miRNA sequence, which is complementary to a capture sequence immobilized on the universal array (see later).

The input RNA (total RNA) is first polyadenylated and converted to cDNA. It is worthwhile to point out that the method gets faithful measurements of all miRNAs only if all of them are equally well-polyadenylated. All MSOs corresponding to the 735 miRNAs are hybridized to their target cDNA sequences simultaneously. An allele-specific primer extension step is then carried out: the assay oligos are preferentially extended if their 3'-bases are complementary to their cognate sequence in the cDNA templates. This additional enzymatic step helps enhance the discrimination among homologous miRNA and mRNA sequences, and provides the assay with a second level of specificity, as compared to other methods which only rely on DNA sequence-mediated hybridization specificity. This enzymatic discrimination step is particularly important for miRNA measurement as their small sizes (~21 nt) greatly limit the design of optimal assay probes. The same enzymatic biochemistry has been widely used in genotyping of single nucleotide polymorphisms (31,34).

The specifically extended products are amplified using a universal PCR which provides the assay with high sensitivity. Because the PCR primers are shared among all target sequences and amplicons are of uniform size, this approach allows unbiased amplification of the extended oligo population. The labelled amplification products are hybridized to a universal array bearing complementary address sequences (27,30). The signal intensity at each address location on the array corresponds to, and reflects the relative abundance of, the respective miRNAs targeted in the original sample.

miRNA assay reproducibility

To assess assay reproducibility, we ran six RNA samples, extracted from four cell lines (PC-3, MCF-7, 293 and HeLa) and two tissues (liver and brain), in four to six replicates over two operators and three 96-sample array matrices with 200 ng total RNA input (96 samples total).

We used a quantile normalization method and computed R^2 within the same array matrix and/or operators as well as between array matrices and/or operators. On average, across all samples, R^2 varies from >0.97 (different array matrix and operators) to >0.98 (same array matrix and same operator) (Figure 2 and Supplementary Table 2; the microarray intensity data are provided in Supplementary Tables 3 and 4). Highly reproducible expression profiles have been generated by other groups using the same platform (J.L. Schultze and J.M. Cunningham, personal communications). We believe the assay reproducibility can be further improved by increasing the number of probes designed for each miRNA.

To assess assay performance at different input levels, we ran PC3 total RNA in two to four replicates at seven different input amounts: 2, 5, 10, 25, 50, 100 and 200 ng. We used the 'Average' normalization method and computed R^2 to assess reproducibility for each input level.

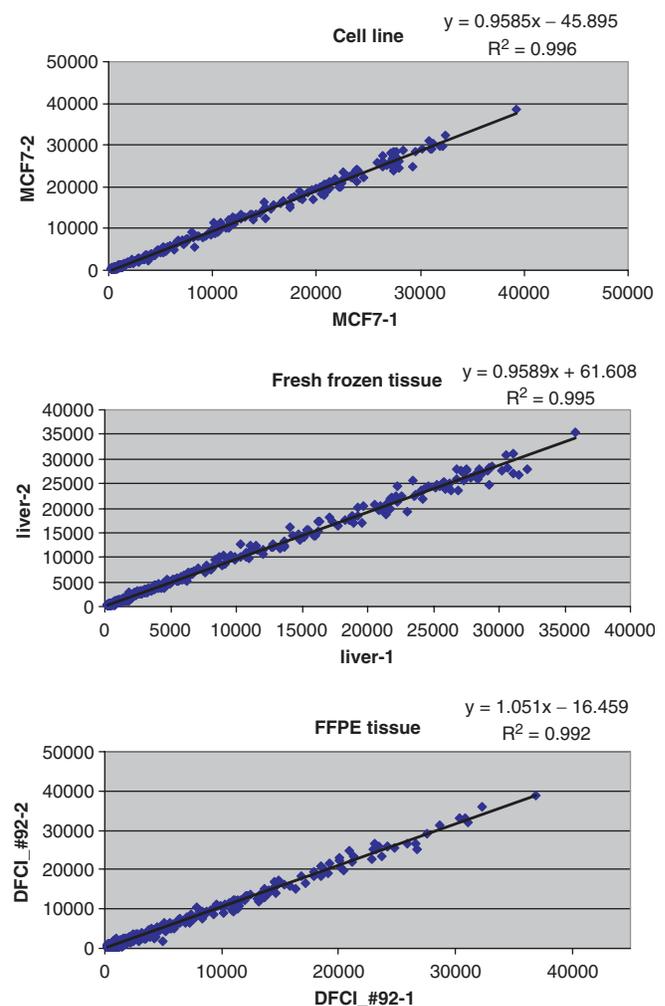


Figure 2. miRNA assay reproducibility. Assay intensity measured for the 735 miRNAs in one technical replicate (replicate #1; x-axis) is plotted against the assay intensity measured in another technical replicate (replicate #2; y-axis). Two hundred nanograms total RNAs extracted from a cell line (MCF7), fresh frozen tissue (liver) and formalin-fixed, paraffin-embedded (FFPE) tissue (ovarian tumor sample) were used for each array experiment.

Profiles generated with 200, 100 and 50 ng were highly correlated ($R^2 > 0.98$; Supplementary Table 5). Quite reproducible data was obtained in technically replicated experiments with as little as 2 ng input total RNA ($R^2 > 0.94$), although correlation between the profiles generated with 2 and 200 ng RNA input was only modest ($R^2 = 0.6$); this input level is 10- to 100-fold less than that required by some other array platforms, which makes it possible to assay small tissue samples, including archival tissue samples (see later). Furthermore, about 90% of the miRNAs that were detected ($P < 0.05$) with 200 ng total RNA input were also detected when 2 ng total RNA was used (data not shown). Similar results were obtained in independent studies (J.L. Schultze and J.M. Cunningham, personal communications).

miRNA profiling with partially degraded RNA samples

We artificially degraded the four cell-line RNAs by heating them at 90°C for 30 and 150 min. Profiles generated with these degraded samples and the corresponding intact samples (200 ng) were compared after average normalization; a correlation of $R^2 > 0.96$ was obtained between the intact and the 30-min degradation samples, and a correlation of $R^2 = \sim 0.9$ was obtained between the intact and the 150-min degradation samples. In addition, highly reproducible results ($R^2 > 0.99$) have been obtained with archived tissue samples (Figure 2). In a separate study, we generated miRNA expression profiles in 130 FFPE ovarian cancer samples (data not shown).

miRNA assay specificity

To estimate assay specificity, we designed perfect match and central mismatch probes for miRNAs hsa_let-7a, let-7c, let-7f, miR-152 and miR-182, and 3'-end mismatch probes for small nucleolar RNAs RNU24 and RNU66. We calculated the intensity ratio of the perfect match versus the mismatch probes across 24 samples. On average, we obtained a high signal/noise ratio ranging from 12 to 65 (Table 1).

In addition, we obtained high concordance ($R^2 > 0.97$) between profiles generated with total RNA and low molecular weight (LMW) RNA enriched with Invitrogen's PureLink miRNA Isolation Kit (Figure 3; the microarray

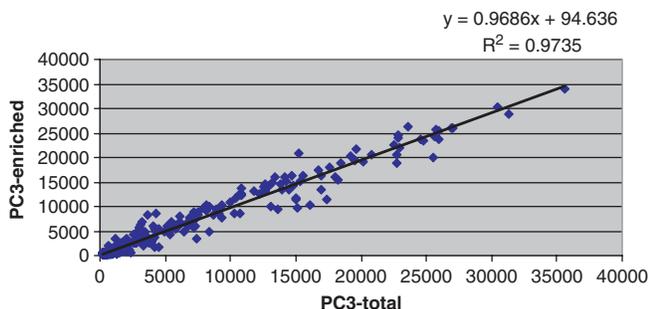


Figure 3. miRNA expression profiles generated with PC3 total RNA and enriched LMW RNA. Assay intensities obtained with 200 ng total RNA (x-axis) is plotted against the intensities obtained with the corresponding enriched small RNA (equivalent to 1 μ g of total RNA) (y-axis).

intensity data are provided in Supplementary Table 6). This result suggests that the assay is very specific, in which the presence of total RNA including mRNAs and ribosomal RNAs (rRNAs) background did not affect overall miRNA profiles.

Fold-difference detection

To estimate a fold change, we used mixtures of HeLa/293 RNAs in the following percentages: 90/10, 75/25, 50/50 and 25/75% (combined input = 200 ng total RNA). We chose seven miRNAs that were expressed in HeLa cells but not in 293 cells, and then calculated the average signal intensities for these seven targets in each mixture and computed the fold difference in the mixtures compared to the 100% HeLa RNA sample (Supplementary Figure 1). We detected a 1.14-fold change (i.e. between 90% and 100% mixtures) for the best performing miRNAs with a P -value of 3.59×10^{-5} (Supplementary Figure 1). On average, 1.2- to 1.3-fold differences were detected with 95% confidence.

Characterization of system LOD and dynamic range

To measure the assay dynamic range and limit of detection (LOD), we spiked eight different amounts (ranging from 10^3 to 10^{10} molecules) of four synthetic RNAs into a 200 ng total RNA background and assayed the samples in duplicate. Using an algorithm described previously (33), we show that our assay was able to clearly detect 10^5 molecules and started to saturate after 10^9 molecules (Supplementary Figure 2). This corresponds to a ~ 4 log (10^5 – 10^9 molecules) dynamic range.

Quantitative RT-PCR

For comparison with another assay platform, expression levels of 12 miRNAs were measured in the four cancer cell lines by a stem-loop-based RT-PCR method (12); high concordance ($R^2 = 0.90$) was obtained between the array results and the RT-PCR results, when 'fold-difference' was compared (Figure 4; the original microarray and RT-PCR data are provided in Supplementary Table 7).

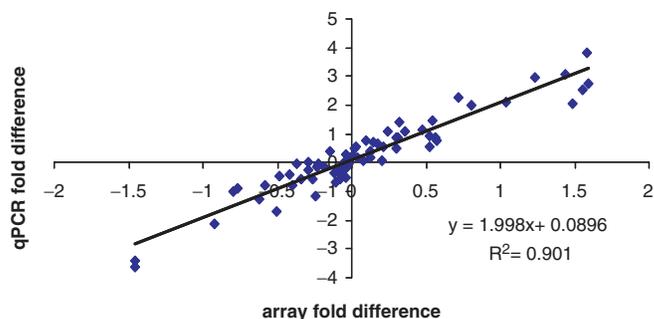


Figure 4. Fold-difference detected by array and RT-PCR. High concordance ($R^2 = 0.90$) was obtained between the miRNA array results and RT-PCR results. The logarithmic fold difference in abundance in pair-wise comparisons between four cancer cell lines (PC3, 293, MCF7 and HeLa) was estimated for 12 miRNAs in both the Illumina miRNA assay (fold difference in array intensity, x-axis) and RT-PCR (fold difference in abundance derived from crossover threshold, y-axis).

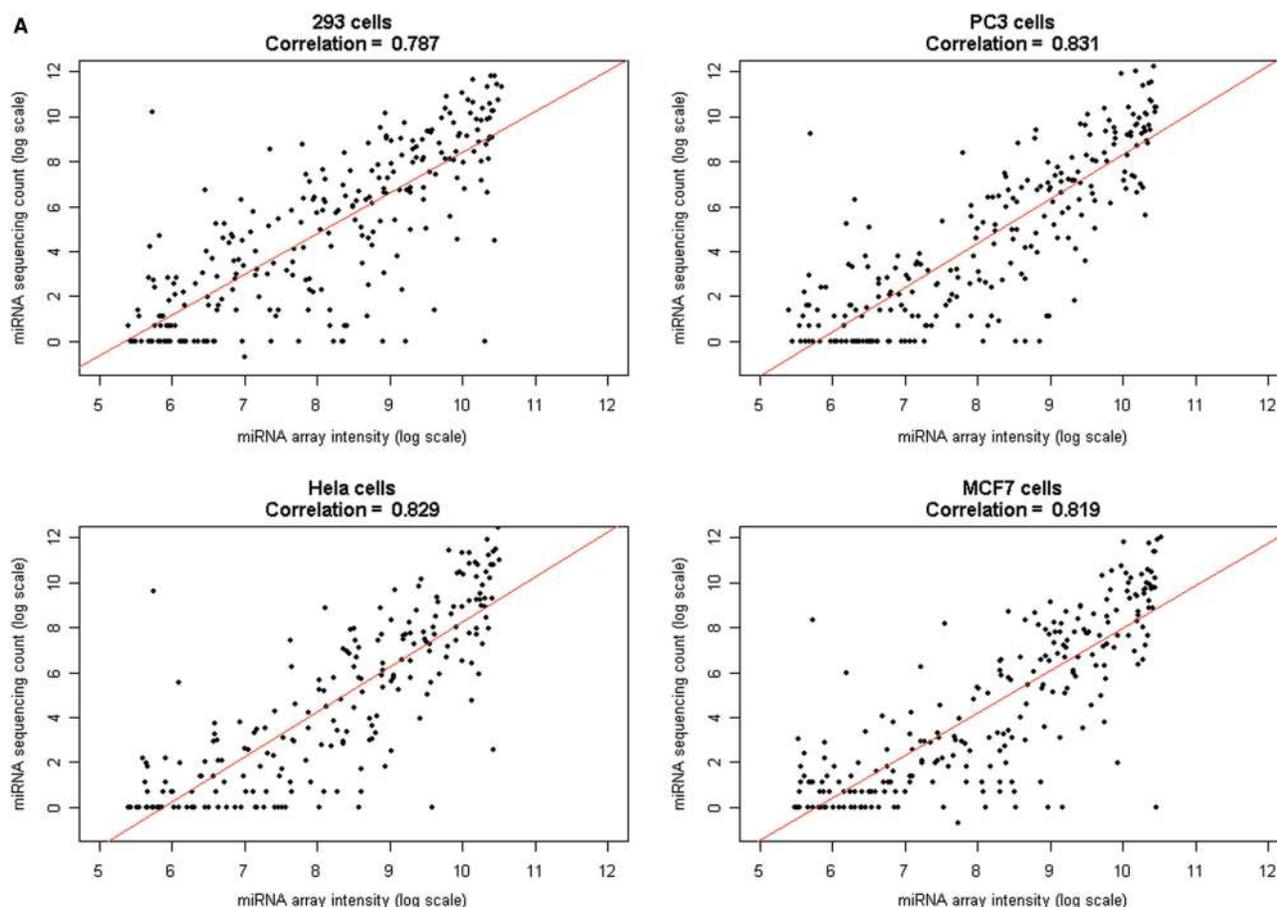


Figure 5. Cross-platform comparisons. (A) Correlation between the array intensity (x -axis) and sequencing count (y -axis). A natural log conversion was used for the plot. (B) Correlation between the array intensity ratio (x -axis) and sequencing count ratio (y -axis) for each possible comparison between four cell lines. A natural log conversion was used for the plot.

It is worthwhile to point out that the fold-difference in miRNA abundance as determined by RT-PCR was larger than that determined by the microarray analysis. This type of underestimating bias has been reported previously for both oligonucleotide arrays and cDNA arrays (27).

Validation with sequencing-based digital gene expression profiling

While the array versus RT-PCR comparison was done in only 12 miRNAs, a more comprehensive validation of our array method could result from a sequencing-based measurement. To this end, we sequenced 8.5–10.5 million small RNAs for each of the four cell lines described earlier, using the Illumina Genome Analyzer system. Of which, 2 849 000, 3 207 245, 4 059 495 and 5 104 527 sequence tags from 293, HeLa, PC3 and MCF7, respectively aligned to the 735 mature miRNA sequences using the Eland sequence-matching algorithm (see ‘Methods’ section and Supplementary Table 8). There was a good correlation ($R = 0.78$ – 0.83) between the absolute array intensity and sequencing count (Figure 5A). In all these cell lines, there were a small number of miRNAs that were detected only by the array but not by sequencing (i.e. the

dots along the x -axis of Figure 5A), presumably the result of limited cross-hybridization from the microarray measurement.

Differential expression between each pair of the samples was calculated (e.g. miRNA sequence tag counts in sample A/miRNA sequence tag counts in sample B) and compared with array results (e.g. miRNA intensity in sample A/miRNA intensity in sample B), and an overall correlation $R = 0.87$ – 0.89 was obtained (Figure 5B). This correlation is slightly lower than our previous results for mRNA versus sequencing comparison ($R = 0.89$ – 0.93) (S. Luo and G. Schroth, unpublished data), which we believe is due to the limited choice to design optimal probes for miRNAs. However, this correlation range is quite similar to the cross-microarray platform comparisons done in the MAQC study (35).

DISCUSSION

We have developed a highly sensitive and specific method for miRNA expression profiling, which has a 3.5–4 log dynamic range, over which an average 1.2- to 1.3-fold differences in miRNA abundance can be detected with 95% confidence. The method has three specific features: (i) an enzymatic 3'-end discrimination (governed by an

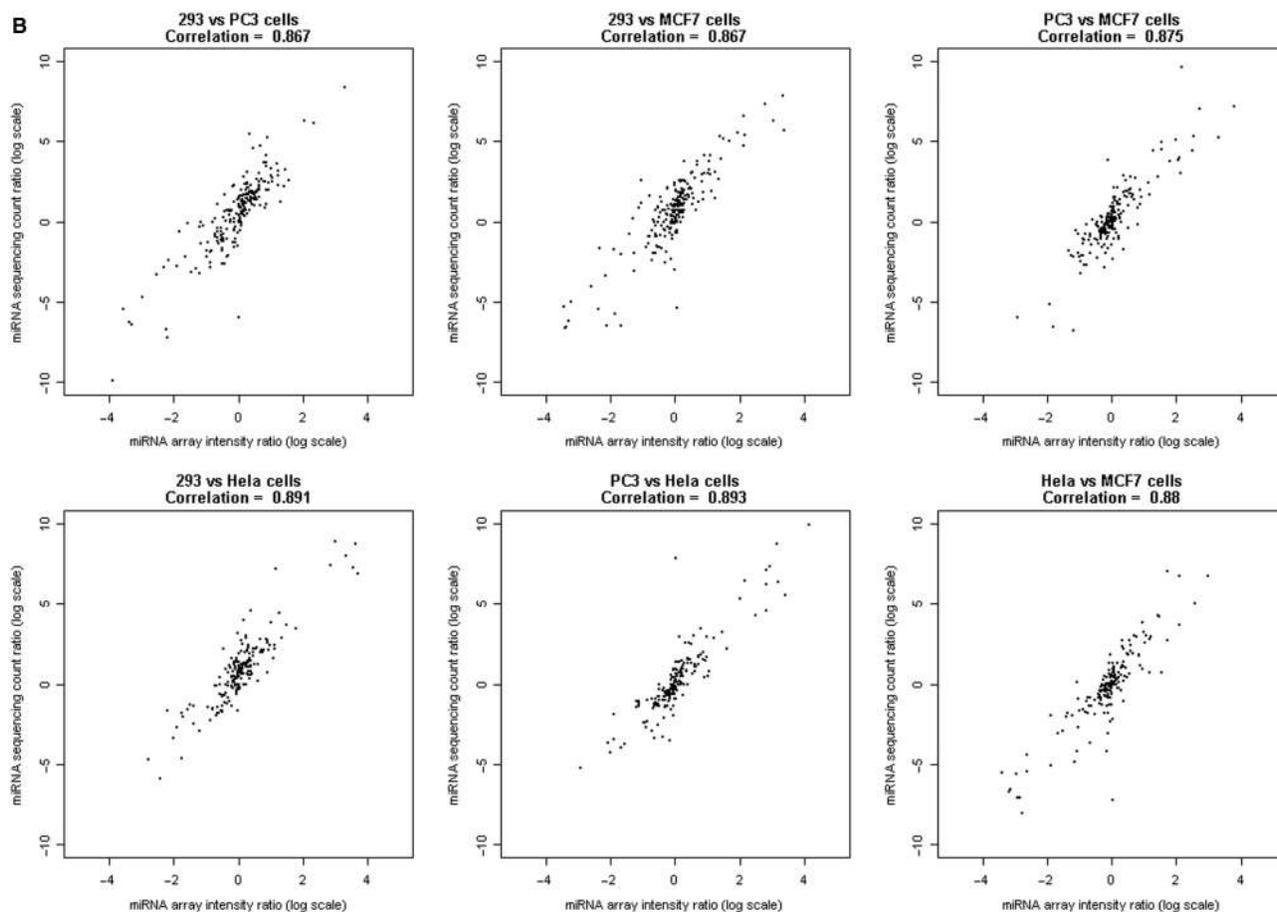


Figure 5. Continued.

allele-specific extension step), in addition to DNA sequence-mediated hybridization specificity; (ii) a solid-phase cDNA selection with proven multiplexing capacity for expression profiling (27,36) and (iii) a universal PCR amplification which renders the method highly sensitive. Similar amplicon size with universal PCR primers has proven to be a faithful signal amplification method (27).

The miRNA assay method is highly sensitive; highly reproducible miRNA expression profiles were generated with 100–200 ng total RNA input. Due to the high assay specificity, this method generated very similar expression profiles with total RNA and enriched low molecular weight RNA; therefore, the purification of small RNA species prior to sample labelling is unnecessary. This feature will enable high-throughput miRNA analysis in a clinical setting where only limited amounts of biopsy material may be available.

With the current assay design, the method may offer advantages for detecting mature miRNAs more effectively than pre-miRNAs. First, the cDNA synthesis may be more complete with short mature miRNA templates as compared to pre-miRNAs. The extension step also favours mature miRNAs because longer sequences will not achieve complete extension to the same degree as mature miRNAs under the experimental condition we use (data not shown). More importantly, the well-known stem-loop structure of pre-miRNAs could interfere with

cDNA synthesis and assay oligo annealing, which should also enhance the relative detection of mature miRNAs.

Expression profiles generated by this method are highly comparable to those obtained with RT-PCR and direct sequence counting of small RNA molecules. Indeed, digital miRNA expression profiling (DGE) provides the most comprehensive and rigorous cross-platform comparison: (i) it measures all miRNAs, i.e. it is a complete and un-biased approach and (ii) since it is sequencing-based, it avoids any cross-hybridization issues which may exist between array versus array and array versus RT-PCR comparisons. Judging by this comparison, our method appears to provide a specific and quantitative measurement of miRNA abundance (Figure 5). We believe this kind of microarray versus DGE comparison should provide an objective assessment for all different miRNA microarray platforms. To help facilitate this kind of cross-platform comparison, we are currently generating both array and sequencing data on some of the MAQC samples as well as some standard human cell lines from ATCC and Coriell Institute.

As new miRNA sequences are constantly discovered and deposited in public databases, a flexible array platform that can swiftly handle array content updates is highly desirable. Since our assay uses universal array read-out, new miRNA assay probes can be easily added to existing assay oligo pools without new array development

and manufacture. This feature simplifies array content upgrades. Using the same design scheme, we have also developed an assay panel targeting 380 well-annotated mouse miRNA sequences derived from Sanger miRBase version 9.1 (February 2007 Release) (data not shown).

We have used the method described in this paper to profile a diverse set of human embryonic stem cells, somatic stem cells and differentiated cells (37). In addition, the technology has been used to study a variety of human tissue samples including archived ovarian cancer (J. Fan, unpublished data), colon cancer (S. Thibodeau, personal communication), post-mortem brain tissues (R. Thompson, personal communication) and peripheral blood of patients with diseases such as chronic lymphatic leukaemia (CLL), scleroderma, bacteremia or lung cancer, and healthy individuals (J.L. Schultze, personal communication). We believe, when coupled with the 96-sample array matrix platform, this method should prove useful for high-throughput expression profiling of miRNAs in large numbers of tissue samples and help to unveil fundamental disease mechanisms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Shawn Baker, Tanya Boyaniwsky, Kirt Haden, Mark Staebell, Christopher Streck, Scott Taylor, Joanne Yeakley and John Stuelpnagel at Illumina, Louise Laurent and Jeanne Loring at The Scripps Research Institute, Renee Rubio, Kristina Holton and John Quackenbush at Dana-Farber Cancer Institute, Hua Gu at Columbia University and Guoping Fan at UCLA for helpful discussions. Funding to pay the Open Access publication charges for this article was provided by Illumina.

Conflict of interest statement. The authors are employees and shareholders of Illumina, where this study was conducted.

REFERENCES

- Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, **116**, 281–297.
- Farh,K.K., Grimson,A., Jan,C., Lewis,B.P., Johnston,W.K., Lim,L.P., Burge,C.B. and Bartel,D.P. (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science*, **310**, 1817–1821.
- Lee,R.C., Feinbaum,R.L. and Ambros,V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, **75**, 843–854.
- Berezikov,E., Cuppen,E. and Plasterk,R.H. (2006) Approaches to microRNA discovery. *Nat. Genet.*, **38** (Suppl), S2–S7.
- Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
- Bentwich,I., Avniel,A., Karov,Y., Aharonov,R., Gilad,S., Barad,O., Barzilai,A., Einat,P., Einav,U., Meiri,E. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.*, **37**, 766–770.
- Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
- Lu,J., Getz,G., Miska,E.A., Alvarez-Saavedra,E., Lamb,J., Peck,D., Sweet-Cordero,A., Ebert,B.L., Mak,R.H., Ferrando,A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
- Calin,G.A. and Croce,C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.
- Yang,B., Lin,H., Xiao,J., Lu,Y., Luo,X., Li,B., Zhang,Y., Xu,C., Bai,Y., Wang,H. *et al.* (2007) The muscle-specific microRNA miR-1 regulates cardiac arrhythmogenic potential by targeting GJA1 and KCNJ2. *Nat. Med.*, **13**, 486–491.
- Kim,J., Inoue,K., Ishii,J., Vanti,W.B., Voronov,S.V., Murchison,E., Hannon,G. and Abeliovich,A. (2007) A MicroRNA feedback circuit in midbrain dopamine neurons. *Science*, **317**, 1220–1224.
- Chen,C., Ridzon,D.A., Broomer,A.J., Zhou,Z., Lee,D.H., Nguyen,J.T., Barbisin,M., Xu,N.L., Mahuvakar,V.R., Andersen,M.R. *et al.* (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.*, **33**, e179.
- Lu,C., Tej,S.S., Luo,S., Haudenschild,C.D., Meyers,B.C. and Green,P.J. (2005) Elucidation of the small RNA component of the transcriptome. *Science*, **309**, 1567–1569.
- Landgraf,P., Rusu,M., Sheridan,R., Sewer,A., Iovino,N., Aravin,A., Pfeffer,S., Rice,A., Kamphorst,A.O., Landthaler,M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
- Neely,L.A., Patel,S., Garver,J., Gallo,M., Hackett,M., McLaughlin,S., Nadel,M., Harris,J., Gullans,S. and Rooke,J. (2006) A single-molecule method for the quantitation of microRNA gene expression. *Nat. Methods*, **3**, 41–46.
- Lagos-Quintana,M., Rauhut,R., Yalcin,A., Meyer,J., Lendeckel,W. and Tuschl,T. (2002) Identification of tissue-specific microRNAs from mouse. *Curr. Biol.*, **12**, 735–739.
- Lau,N.C., Lim,L.P., Weinstein,E.G. and Bartel,D.P. (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, **294**, 858–862.
- Sempere,L.F., Freemantle,S., Pitha-Rowe,I., Moss,E., Dmitrovsky,E. and Ambros,V. (2004) Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biol.*, **5**, R13.
- Cummins,J.M., He,Y., Leary,R.J., Pagliarini,R., Diaz,L.A. Jr, Sjoblom,T., Barad,O., Bentwich,Z., Szafrańska,A.E., Labourier,E. *et al.* (2006) The colorectal microRNAome. *Proc. Natl Acad. Sci. USA*, **103**, 3687–3692.
- Castoldi,M., Schmidt,S., Benes,V., Noerholm,M., Kulozik,A.E., Hentze,M.W. and Muckenthaler,M.U. (2006) A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). *RNA*, **12**, 913–920.
- Liu,C.G., Calin,G.A., Meloon,B., Gamlie,N., Sevignani,C., Ferracin,M., Dumitru,C.D., Shimizu,M., Zupo,S., Dono,M. *et al.* (2004) An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc. Natl Acad. Sci. USA*, **101**, 9740–9744.
- Thomson,J.M., Parker,J., Perou,C.M. and Hammond,S.M. (2004) A custom microarray platform for analysis of microRNA gene expression. *Nat. Methods*, **1**, 47–53.
- Wang,H., Ach,R.A. and Curry,B. (2007) Direct and sensitive miRNA profiling from low-input total RNA. *RNA*, **13**, 151–159.
- Beuvink,I., Kolb,F.A., Budach,W., Garnier,A., Lange,J., Natt,F., Dengler,U., Hall,J., Filipowicz,W. and Weiler,J. (2007) A novel microarray approach reveals new tissue-specific signatures of known and predicted mammalian microRNAs. *Nucleic Acids Res.*, **35**, e52.
- Jiang,J., Lee,E.J., Gusev,Y. and Schmittgen,T.D. (2005) Real-time expression profiling of microRNA precursors in human cancer cell lines. *Nucleic Acids Res.*, **33**, 5394–5403.
- Nelson,P.T., Baldwin,D.A., Scearce,L.M., Oberholtzer,J.C., Tobias,J.W. and Mourelatos,Z. (2004) Microarray-based, high-throughput gene expression profiling of microRNAs. *Nat. Methods*, **1**, 155–161.
- Fan,J.B., Yeakley,J.M., Bibikova,M., Chudin,E., Wickham,E., Chen,J., Doucet,D., Rigault,P., Zhang,B., Shen,R. *et al.* (2004) A versatile assay for high-throughput gene expression profiling on universal array matrices. *Genome Res.*, **14**, 878–885.

28. Berezikov, E., Thuemmler, F., van Laake, L.W., Kondova, I., Bontrop, R., Cuppen, E. and Plasterk, R.H. (2006) Diversity of microRNAs in human and chimpanzee brain. *Nat. Genet.*, **38**, 1375–1377.
29. Berezikov, E., van Tetering, G., Verheul, M., van de Belt, J., van Laake, L., Vos, J., Verloop, R., van de Wetering, M., Guryev, V., Takada, S. *et al.* (2006) Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res.*, **16**, 1289–1298.
30. Fan, J.B., Gunderson, K.L., Bibikova, M., Yeakley, J.M., Chen, J., Wickham, E., Lebruska, L.L., Laurent, M., Shen, R. and Barker, D. (2006) Illumina universal bead arrays. *Methods Enzymol.*, **410**, 57–73.
31. Fan, J.B., Chee, M.S. and Gunderson, K.L. (2006) Highly parallel genomic assays. *Nat. Rev. Genet.*, **7**, 632–644.
32. Galinsky, V.L. (2003) Automatic registration of microarray images. II. Hexagonal grid. *Bioinformatics*, **19**, 1832–1836.
33. Chudin, E., Kruglyak, S., Baker, S.C., Oeser, S., Barker, D. and McDaniel, T.K. (2006) A model of technical variation of microarray signals. *J. Comput. Biol.*, **13**, 996–1003.
34. Fan, J.B., Oliphant, A., Shen, R., Kermani, B.G., Garcia, F., Gunderson, K.L., Hansen, M., Steemers, F., Butler, S.L., Deloukas, P. *et al.* (2003) Highly parallel SNP genotyping. *Cold Spring Harb. Symp. Quant. Biol.*, **68**, 69–78.
35. Shi, L., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C., Collins, P.J., de Longueville, F., Kawasaki, E.S., Lee, K.Y. *et al.* (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nat. Biotechnol.*, **24**, 1151–1161.
36. Bibikova, M., Talantov, D., Chudin, E., Yeakley, J.M., Chen, J., Doucet, D., Wickham, E., Atkins, D., Barker, D., Chee, M. *et al.* (2004) Quantitative gene expression profiling in formalin-fixed, paraffin-embedded tissues using universal bead arrays. *Am. J. Pathol.*, **165**, 1799–1807.
37. Laurent, L.C., Chen, J., Ulitsky, I., Mueller, F.J., Lu, C., Shamir, R., Fan, J.-B. and Loring, J.F. (2008) Comprehensive microRNA profiling reveals a unique human embryonic stem cell signature dominated by a single seed sequence. *Stem Cells*, Apr 10; [Epub ahead of print].