# Reports

## Concordance among digital gene expression, microarrays, and qPCR when measuring differential expression of microRNAs

Sylvain Pradervand[1], Johann Weber[1], Frédéric Lemoine[2], Floriane Consales[1], Alexandra Paillusson[1], Mélanie Dupasquier[1], Jérôme Thomas[1], Hannes Richter[1], Henrik Kaessmann[2], Emmanuel Beaudoing[1], Otto Hagenbüchle[1] and Keith Harshman[1]

*[1]Genomic Technologies Facility, Center for Integrative Genomics, University of Lausanne, Genopode Building, Lausanne, Switzerland, and [2]Center for Integrative Genomics, University of Lausanne, Genopode Building, Lausanne, Switzerland*

*BioTechniques* 48:219-222 (March 2010) doi 10.2144/000113367

Keywords: microRNAs; digital gene expression; qPCR; microarrays; comparisons

Profiling microRNA (miRNA) expression is of widespread interest given the critical role of miRNAs in many cellular functions. Profiling can be achieved via hybridization-based (microarrays), sequencing-based, or amplification-based (quantitative reverse transcription-PCR, qPCR) technologies. Among these, microarrays face the significant challenge of accurately distinguishing between mature and immature miRNA forms, and different vendors have developed different methods to meet this challenge. Here we measure differential miRNA expression using the Affymetrix, Agilent, and Illumina microarray platforms, as well as qPCR (Applied Biosystems) and ultra high–throughput sequencing (Illumina). We show that the differential expression measurements are more divergent when the three types of microarrays are compared than when the Agilent microarray, qPCR, and sequencing technology measurements are compared, which exhibit a good overall concordance.

## Introduction

It has recently become standard practice to profile the expression levels of microRNAs (miRNAs). Researchers have many different technological options to comprehensively analyze miRNA expression, with each option having advantages and disadvantages. Digital gene expression (DGE) profiling—based on ultra high–throughput DNA sequencing—is increasingly popular since it allows for the discovery of new miRNAs along with quantitative expression analysis. In a recent communication, Linsen et al. showed that DGE profiling is strongly biased toward certain small RNAs, which makes DGE inappropriate for absolute quantification of miRNAs, but not for differential miRNA expression analysis (1). These biases were dependent on library preparation and were also observed in quantitative reverse transcription polymerase chain reaction (qPCR) amplifications. Therefore, when novel miRNA discovery is not a priority, alternative technologies to DGE can still be attractive.

Microarray-based techniques have the advantages of being relatively cost-effective, relatively quick from RNA labeling to data generation, and simple to use. Among the available commercial miRNA microarray platforms, the single-color array format is the most common. Agilent Technologies (Santa Clara, CA, USA) has developed a miRNA profiling platform that provides both sequence and size discrimination for mature miRNAs (2). This system generates results that are highly correlated with qPCR results and, therefore, is an excellent choice for miRNA profiling (3,4). Other major microarray manufacturers have produced single-color miRNA array platforms, including both Illumina and Affymetrix (5). Real-time quantitative PCR is another popular method for differential miRNA profiling. Compared with array platforms, it has superior sensitivity (6), and has recently been parallelized in an array-like format (Applied Biosystems, Foster City, CA, USA) allowing the profiling of 380 miRNAs in a single experiment (7). In this study, we compare miRNA expression in the same brain and heart RNA samples using three different array platforms, qPCR, and DGE.

## Materials and methods

### RNA samples

Human heart and brain total RNA were from Stratagene (MVP human normal adult tissue total RNA; La Jolla, CA, USA).

### Microarray

To assay technical reproducibility, four technical replicates from brain and heart RNA were hybridized on microarrays.

**Agilent.** One hundred nanograms of total RNA from each sample were labeled and hybridized on human Agilent miRNA v2 microarrays as described previously (3). Data were extracted and summarized using Agilent Feature Extraction Software. Then they were imported into

GeneSpring GX10 software (Agilent Technologies), quantile-normalized and $log_2$-transformed.

**Illumina.** Five hundred nanograms of total RNA from each sample were labeled and hybridized on Human v2 MicroRNA Expression BeadChips (Cat. no. MI-102–1024; Illumina, San Diego, CA, USA), according to the manufacturers recommendations (Illumina MicroRNA Expression Profiling Assay Guide). BeadChips were scanned with the Illumina iScan Reader. Data were imported into GenomeStudio (Illumina), quantile-normalized and $log_2$-transformed in R (www.r-project.org).

**Affymetrix.** One microgram of total RNA from each sample was labeled with the FlashTag Biotin RNA Labeling Kit for Affymetrix GeneChip miRNA arrays (Genisphere, Hatfield, PA, USA). After labeling, the samples were hybridized on Affymetrix GeneChip miRNA arrays according to the manufacturer's recommendations (Affymetrix, Santa Clara, CA, USA). Hybridization, washing, and scanning of the slides were done according to Affymetrix's recommendations (Fluidics Protocol FS450_0003). Data were extracted from the images, quantile-normalized, summarized (median polish) and $log_2$-transformed with the miRNA QC tool software from Affymetrix.

## Quantitative real-time PCR

Seven hundred nanograms of total RNA were reverse-transcribed with the megaplex RT primers human pool A (Applied Biosystems). This pool contains specific stem-loop primers for 377 human miRNAs, 3 small RNAs, and 1 negative control, and are all based on miRBase v. 10.1. The resulting cDNA was transfered to a TaqMan Human MicroRNA A Array v2.0 (Applied Biosystems) and qPCR was performed on an Applied Biosystems 7900HT Sequence Detection system. Cycling conditions were 50°C for 2 min, 94.5°C for 10 min, and 40 cycles of 97°C for 30 s and 59.7°C for 1 min. Two technical replicates were performed per sample. Quantification cycle (Cq; standard name for Ct or Cp value) values were recorded with SDS version 2.3 software. Cq values ≥36 were considered beyond the limit of detection (a Cq value of 35 represents a single molecule template detection). miRNAs for which both brain duplicate or heart duplicate Cq values were ≥36 were removed. Cq values were imported into qbase$^{PLUS}$ version 1.3 software (Biogazelle, Ghent, Belgium), which is based on geNorm (8) and qBase (9). U6 and RNU48 were found to be the most stable reference genes and used to normalize the data. Mean RQs (relative quantities) were calculated for each tissue after removing the remaining undetected values.

## UHTS-based digital gene expression

Libraries of small RNAs were prepared using the DGE-Small RNA Sample Kit, Alternative v1.5 Protocol (Cat. no. FC-102–1009–1; Illumina) according to the protocol supplied by the manufacturer (Protocol Rev. A, published February 2009) and using 10 μg total heart or brain RNA. Two lanes of each library were sequenced on the Illumina Genome Analyzer IIx using Single-Read Cluster Generation Kit v2 (Cat. no. FC-103–2001; Illumina) and 36 Cycle Sequencing Kit v4 (Cat. no. FC-104–4002), generating 16.6 and 15.3 million quality tags for brain and heart libraries, respectively. Data were processed using the Illumina Pipeline Software v1.5.1. The miRNA data were further processed as follows. First, the data were cleaned by removing low complexity reads using the DUST algorithm as implemented in the NCBI C++ Toolkit (www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=toolkit). Second, adapter sequences were removed from the remaining reads using an alignment strategy that allows up to three mismatches if the adapter is >10 nucleotides, and up to 2 mismatches for shorter adapter sequences. Third, processed reads that were <15 and >26 nucleotides were discarded. The remaining reads were then mapped to the human genome (Ensembl 53) using Bowtie (10), without allowing mismatches and discarding reads that mapped to >10 positions in the human genome. Finally, we compared the position of the mapped reads to the genomic location of the mature miRNA sequences in miRBase v 12.0 (11). For this step, we allowed the reads to map at ±3 nucleotides of the 5′ end of the mature sequence and ±5 nucleotides relative to the 3′ end. Brain and heart libraries gave 1.13 and 2.27 million tags mapped to miRBase, respectively. For each mature miRNA sequence, we counted the number of reads that mapped to it. We distributed equally each read that mapped to more than one miRNA locus. Finally, we set the tag count to 1 for miRNAs with a tag count equal to 0. This threshold allowed $log_2$ transformation of tag counts before quantile-normalization of the data to remove systematic differences between the samples.

## Probe mapping

Microarray probes were mapped using miRNAs miRBase entry name (11). Agilent miRNA v2 microarrays are based on miRBase v. 10.1, Illumina MicroRNA Expression v2 BeadChips on miRBase v. 12, and Affymetrix GeneChip miRNA arrays on miRBase v. 11 (11). Only two human miRNAs changed in name or sequence between miRBase version 10.1 and 12 and these were not contained in the set of common miRNAs to all three microarray platforms. The Illumina platform contains some probes that target more than one miRNA. These probes have a one-to-many relationship between Illumina and the other two platforms.

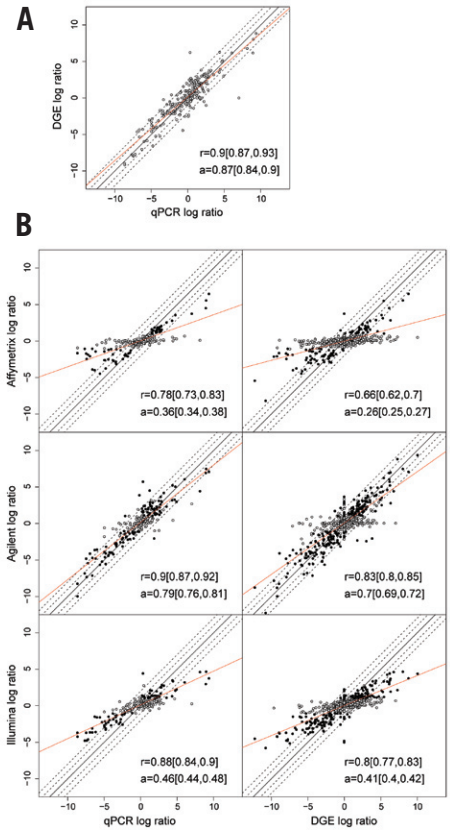


**Figure 1. Scatter plots of miRNAs $log_2$ fold differences (heart versus brain).** (A) Log ratios from DGE compared with qPCR. (B) Log ratios from microarray technologies compared with qPCR or DGE. Black dots are miRNAs detected in at least two of the eight brain and heart arrays with a 2-fold change and $P < 0.01$ (*t*-test, $n = 4$ per group) with the particular microarray technology. Solid lines represent the 45° lines of complete concordance. Dashed lines represent 2- and 4-fold change differences. Red lines represent the results of the regression analysis. Correlation coefficient (r), and regression slope (a) of regression lines are indicated. 95% confidence intervals are indicated in square brackets. Confidence intervals for correlation coefficients were calculated using Fisher's transformation.

# Results and discussion

DGE and qPCR are widely considered to be more accurate and quantitative methods for miRNA expression profiling

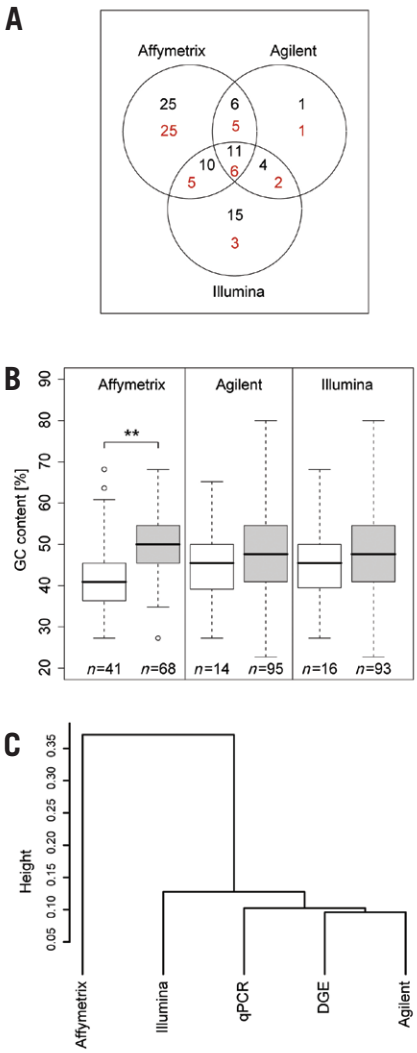


**Figure 2. Comparison of microarray performance characteristics.** (A) Venn diagram of miRNAs among the 218 miRNAs detected by qPCR in both brain and heart tissues that showed a 2-fold change or higher with both DGE and qPCR and that were missed by at least one microarray platform. Black numbers are missed miRNAs based on three criteria: $P < 0.01$, absolute fold change >2, and detection by the platform in at least two of the eight brain and heart arrays. Red numbers are missed miRNA based on two criteria: $P < 0.01$ and detection by the platform in at least two of the eight brain and heart arrays. (B) Boxplot of the percentage of GC content in mature miRNAs sequences. White boxes are miRNAs that were missed by the microarray platform with the criteria indicated above (no fold change criteria). Gray boxes are miRNAs that showed a 2-fold change or higher with both DGE and qPCR and that were identified as differentially expressed by the microarray platform with the criteria indicated above (no fold change criteria). **, $P < 0.001$, two-sample *t*-test with unequal variance. (C). Hierarchical clustering of the heart versus brain fold change comparisons for the five different measurement methods. $\log_2$ fold changes of the 218 miRNAs detected by qPCR in both brain and heart tissues were used. Pearson correlation distance metric and Ward clustering were used.

than DNA microarrays. Therefore, we first determined the correlation of miRNA expression differences measured by DGE (using the Illumina Genome Analyzer IIx) and qPCR (the Applied Biosystems qPCR-array). The $\log_2$ fold differences of miRNA expression between human brain and heart samples, focusing on the 218 miRNAs that were detected by qPCR assay in both tissues, gave a correlation of $0.9 \pm 0.03$, showing that these two methods are highly consistent (Figure 1A). This correlation is very similar to data published by Linsen et al. (1) who analyzed miRNA expression in rat spleen and liver by qPCR (Applied Biosystems) and DGE (SOLiD, Applied Biosystems) and observed a correlation of 0.87 between the two technologies.

Next, we looked at the correlation of $\log_2$ fold difference between qPCR or DGE and the Affymetrix, Agilent, and Illumina microarray platforms (Figure 1B). For the comparison with DGE, we took the 718 miRNAs that were common across the three microarray platforms and that were found in Release 12 of miRBase (11). We found that for each specific microarray platform, the microarray-qPCR correlation was consistent with the microarray-DGE correlation. However, each microarray platform gave a unique correlation pattern. The best qPCR-microarray correlation was seen with the Agilent platform. This platform also had a regression line whose slope was the closest to 1, meaning that the fold changes measured on this platform have the least compression. The Illumina platform showed good correlations with DGE and qPCR, but the regression slopes showed a strong fold change compression, which may result from the PCR amplification step that is performed during the target preparation (5). With one common miRNA between them, the Agilent and Illumina platforms respectively called 14 and 15 miRNAs as differentially expressed that were called unchanged by DGE (due to tag counts of zero). Ten of these miRNAs were in the set of miRNAs analyzed by qPCR. Of these 10 miRNA, eight had a Cq value of ≥34. Therefore, these miRNAs were either poorly expressed and not detected by DGE/qPCR, or their microarray signal generated cross-hybridization with other miRNAs sequences of the microarrays.

The scatter plots of Affymetrix platform data show a considerable number of miRNAs with fold changes significantly different from 0 but lower than the fold changes measured by qPCR or DGE (black circles), and other miRNAs not significantly different from 0 but with strong differential expression as determined by qPCR or DGE (some gray circles). The latter miRNAs may represent false-negative calls by the Affymetrix platform. The Agilent and Illumina platforms also exhibited a smaller number of miRNAs with no fold change that were clearly differentially expressed in qPCR or DGE measurements. We examined more closely the miRNAs with DGE and qPCR fold change values of 2-fold or higher but that were missed by microarrays. We applied two different sets of criteria to the microarray data—a *P* value cutoff of 0.01 plus fold change cutoff of 2 or only the *P* value cutoff—and classified those miRNAs that did not meet these criteria as "false negatives" (Figure 2A). In the case of the Illumina data, the majority of the miRNAs were removed from the false-negative class when the 2-fold change criteria were not applied. This indicates that the Illumina platform primarily suffered from a fold change compression. In contrast, miRNAs uniquely classified as false negatives by Affymetrix were not affected by the elimination of the fold change criteria, indicating that the Affymetrix platform lacked sensitivity for these miRNAs.

The signal intensities of these false negative miRNAs were near or equalled background level (data not shown). To further investigate why these miRNAs were not detected by the Affymetrix platform, we compared their GC content

with the GC content of miRNAs showing a 2-fold change or higher difference with both DGE and qPCR and that were identified as differentially expressed by the microarray ("true positives"). We observed that for Affymetrix, but not for the other two platforms, the GC content of these false negatives (mean GC content: 42.4%) was significantly lower than the GC content of the true positives (mean GC content: 50.5%, Figure 2B). The Agilent probes are designed to equalize the melting temperature ($T_m$). This is done by adding a guanine nucleotide to the 5′ end of each probes so that the $T_m$ for nearly all miRNAs on the array is >55.5°C, Additionally, the hybridization length is shortened for any probe with a $T_m$ >57.5°C (2). The $T_m$ of the probes on the Illumina platform are homogenized by shortening the probe lengths (5). The method for the design of the miRNA gene probes on the Affymetrix platform has not been published. However, using the sequence information in the Affymetrix probe set data file available on the Affymetrix web site to calculate the $T_m$, we obtained a mean $T_m$ of 51.9°C for the "false negative" probes and 55.1°C for the "true positive" probes ($P < 0.001$, two-sample *t*-test with unequal variance). Although not included in this study, a microarray platform based on locked nucleic acid (LNA)–modified probes designed to achieve a balanced $T_m$ also performs well when compared with DGE (correlation of fold changes of 0.93) (12).

In order to assess the overall similarity between the results obtained from the five platforms, we performed a hierarchical clustering analysis using $\log_2$ fold change values of the 218 miRNAs detected by qPCR (Figure 2C). We observed more dissimilarity among the three microarray platforms than among DGE, qPCR and the Agilent microarray, indicating that the technology type is not the major variability factor. It is likely that other factors, such as microarray probe design, target preparation, or hybridization stringency are more important. Indeed, miRNA microarrays face challenges that are more severe than those of microarrays designed for messenger RNA analysis. These challenges include differentiating between mature and precursor miRNAs and differentiating among different miRNAs that differ by only one or a few nucleotides, as well as the relatively short target sequence and the associated limited latitude to adjust probe design and hybridization conditions to account for the unequal $T_m$ of mature miRNAs (13). Furthermore, the target preparation in itself is challenging due to the fact that miRNAs are relatively short. PCR-based labeling methods such as that used by Illumina are sensitive, but may introduce amplification bias (14). Direct labeling methods, as used by Agilent, may be less prone to biases. Moreover, in addition to balancing the $T_m$ of probe-target hybrids, the Agilent platform, which gave the closest results to qPCR and DGE, allows for both sequence and size specificity in its probe design (2).

In conclusion, our results show that microarray technology is a good alternative to sequencing-based methods for miRNAs differential expression measurement, not only for ease of use, but because it has been shown to be a highly reproducible technology (3,15). In this study, the average correlation coefficients among technical replicates were 0.978, 0.989, and 0.978 for the Affymetrix, Agilent, and Illumina microarrays, respectively; 0.981 for QPCR; and 0.999 for DGE (analyzing the same library on different flow cells). Within the group of three microarray technologies evaluated, there was significant variation in both the number of miRNAs identified as differentially expressed, as well as the fold change level determined. The issue of low inter-platform concordance for miRNA microarrays was also raised by another study (15). In our analysis, we show that the Agilent platform outperforms the Illumina and Affymetrix platforms, due to its greater accuracy in fold change measurement and its accurate profiling of miRNAs that differ in GC content.

## Acknowledgments

We thank Patrick Descombes and the Genomics Platform of the National Center of Competence in Research (NCCR) Frontiers in Genetics Program for performing the hybridization and scanning steps for the Illumina arrays. We also thank Ioannis Xenarios, Roberto Fabbretti, and Li Long of the Vital-IT Project of the Swiss Institute of Bioinformatics for their assistance in developing and supporting the data analysis and storage infrastructure used for ultra high–throughput sequencing.

## Competing interests

The authors declare no competing interests.

## References

1. **Linsen, S.E., E. de Wit, G. Janssens, S. Heater, L. Chapman, R.K. Parkin, B. Fritz, S.K. Wyman, et al.** 2009. Limitations and possibilities of small RNA digital gene expression profiling. Nat. Methods *6*:474-476.
2. **Wang, H., R.A. Ach, and B. Curry.** 2007. Direct and sensitive miRNA profiling from low-input total RNA. RNA *13*:151-159.
3. **Pradervand, S., J. Weber, J. Thomas, M. Bueno, P. Wirapati, K. Lefort, G.P. Dotto, and K. Harshman.** 2009. Impact of normalization on miRNA microarray expression profiling. RNA *15*:493-501.
4. **Ach, R.A., H. Wang, and B. Curry.** 2008. Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. BMC Biotechnol. *8*:69.
5. **Chen, J., J. Lozach, E.W. Garcia, B. Barnes, S. Luo, I. Mikoulitch, L. Zhou, G. Schroth, and J.B. Fan.** 2008. Highly sensitive and specific microRNA expression profiling using BeadArray technology. Nucleic Acids Res. *36*:e87.
6. **Chen, C., D.A. Ridzon, A.J. Broomer, Z. Zhou, D.H. Lee, J.T. Nguyen, M. Barbisin, N.L. Xu, et al.** 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res. *33*:e179.
7. **Mestdagh, P., T. Feys, N. Bernard, S. Guenther, C. Chen, F. Speleman, and J. Vandesompele.** 2008. High-throughput stem-loop RT-qPCR miRNA expression profiling using minute amounts of input RNA. Nucleic Acids Res. *36*:e143.
8. **Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe, and F. Speleman.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. *3*:RESEARCH0034.
9. **Hellemans, J., G. Mortier, A. De Paepe, F. Speleman, and J. Vandesompele.** 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. *8*:R19.
10. **Langmead, B., C. Trapnell, M. Pop, and S.L. Salzberg.** 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*:R25.
11. **Griffiths-Jones, S., H.K. Saini, S. van Dongen, and A.J. Enright.** 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. *36*:D154-D158.
12. **Willenbrock, H., J. Salomon, R. Sokilde, K.B. Barken, T.N. Hansen, F.C. Nielsen, S. Moller, and T. Litman.** 2009. Quantitative miRNA expression analysis: comparing microarrays with next-generation sequencing. RNA *15*:2028-2034.
13. **Yin, J.Q., R.C. Zhao, and K.V. Morris.** 2008. Profiling microRNA expression with microarrays. Trends Biotechnol. *26*:70-76.
14. **Davison, T.S., C.D. Johnson, and B.F. Andruss.** 2006. Analyzing micro-RNA expression using microarrays. Methods Enzymol. *411*:14-34.
15. **Sato, F., S. Tsuchiya, K. Terasawa, and G. Tsujimoto.** 2009. Intra-platform repeatability and inter-platform comparability of microRNA microarray technology. PLoS One *4*:e5540.

Received 28 November 2009; accepted 26 January 2010.

Address correspondence to Sylvain Pradervand, Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland, e-mail: sylvain.pradervand@unil.ch.