Technical Note

Affymetrix®
Revolutionize life

GeneChip® miRNA 2.0 Array

Introduction

Since their discovery in 1993, small non-coding RNAs (microRNA, scaRNA, and snoRNA and others) have emerged as a major component in the regulatory circuitry that underlies the development and physiology of complex organisms. As a result, it is becoming increasingly important to complement messenger RNA (mRNA) gene expression studies with miRNA analysis to understand the biological context of differentially expressed genes. These key functional gene products are estimated to regulate approximately 30% of all protein-coding genes.

Small non-coding RNA molecules cover the broadest spectrum of developmental and physiological mechanisms in the cell, including:

- protein translation inhibition
- ribosomal RNA processing
- alternative splicing
- mRNA degradation

In response to the need for an expression microarray for detecting small non-coding RNA, Affymetrix introduced the first-generation miRNA array in 2009, which was based on miRBase version 11 and the available versions of snoBAse and Ensembl. In 2011, Affymetrix introduced a new version of the array that was based on miRBase version 15 and the updated information in snoBAse and Ensembl. Like the first-generation microarray, the new version of the array includes miRNA for all species on a single array as well as human, mouse, and rat snoRNA and scaRNA. A unique feature of the GeneChip® miRNA 2.0 Array is the inclusion of probe sets that provide the ability to identify human, mouse, and rat precursor miRNAs present in the sample.

The purpose of this technical note is to demonstrate GeneChip miRNA 2.0 Array performance, including:

- sensitivity
- specificity
- dynamic range
- reproducibility
- signal and fold change correlation to the GeneChip® miRNA Array
- fold change correlation to qPCR.

Lastly, this technical note describes the design strategy for the precursor miRNA probe sets to provide guidance for data interpretation of the precursor miRNA probe sets.

Experimental design

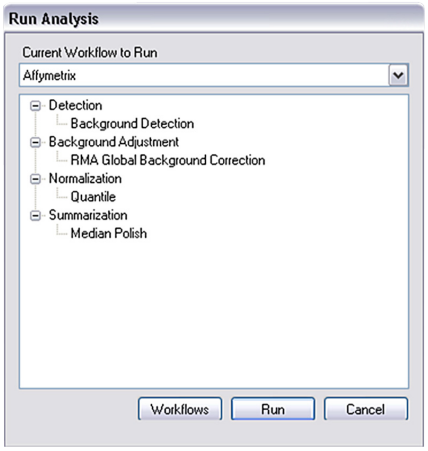
To focus the results from this experiment on the performance of the array, we used commercially available total RNA samples to minimize sample variability. All samples were prepared using the Genisphere® FlashTag® Biotin HSR RNA Labeling Kit for Affymetrix® GeneChip® miRNA Arrays using the protocol recommended by the manufacturer. The recommended protocol is available at the following URL: http://www.genisphere.com/pdf/FlashTag-Biotin-HSR-for-Affymetrix-Feb2011.pdf. Information about the commercially available total RNA samples and the amount used for each assay performed is listed in Table 1.

Table 1: Total RNA samples used

RNA	Supplier	Part no.	Lot no.	Amount per replicate
Human brain	Ambion®	AM7962	10060081	200 ng
Human lung	Ambion	AM7968	1007007	200 ng
Human liver	Ambion	AM7960	1006005	100 ng

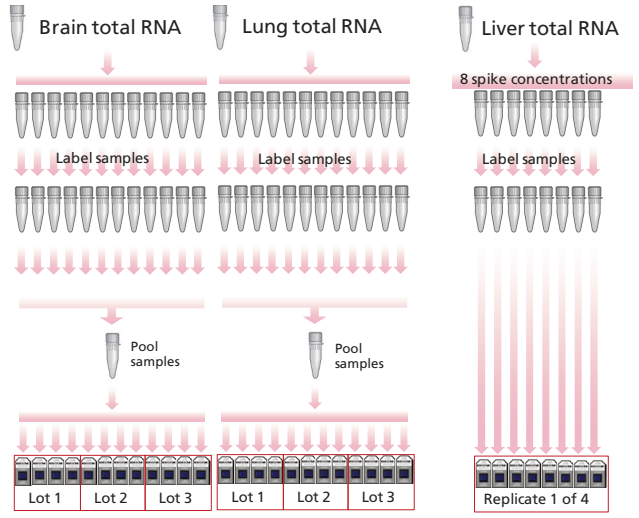
The experiments used to demonstrate array reproducibility and correlation to the first-generation GeneChip miRNA Array and qPCR were performed using 200 ng of human brain and lung RNA per labeling reaction. The sensitivity, specificity, and dynamic range experiment was performed using 16 synthetic miRNA transcripts spiked into 100 ng of human liver total RNA at different abundance levels (Table 2). The Affymetrix® miRNA QC Tool Software was used to perform data analysis using the analysis workflow described in the Genisphere FlashTag Labeling Kit protocol as shown in Figure 1.

Figure 1: Recommended analysis workflow



To demonstrate the reproducibility of the GeneChip® miRNA 2.0 Array, 12 samples of human brain and 12 samples of human lung total RNA were labeled using the Genisphere® FlashTag® Labeling Kit protocol. In order to minimize the sample variation, the 12 replicate labeling reactions of each tissue type were pooled together. The pooled samples were then hybridized to GeneChip miRNA 2.0 Arrays (Figure 2).

Figure 2: Sample labeling and array hybridization schematic



Sensitivity, specificity, and dynamic range

The GeneChip miRNA 2.0 Array demonstrated significant detection of transcripts present at levels as low as 0.1 attomoles spiked miRNA in 100 ng of human liver total RNA (Table 2). The low false positive and high true negative rate of the arrays that did not contain the 16 synthetic spikes demonstrates the high specificity of the array with low input amounts of total RNA. This experiment also demonstrates the broad dynamic range with a signal vs. abundance dose-response curve (Figure 3) showing linearity for more than 3 logs of signal (base 10).

Table 2: Sensitivity

No.	Amol	Percent dectected	No.	Amol	Percent dectected
1	0.0*	6% FP 94% TN	4	10	100%
2	0.1	38%	5	100	100%
3	0.5	81%	7	5,000	100%
4	1.0	94%	8	10,000	100%

Sixteen synthetic miRNA transcripts were spiked into 100 ng of human liver total RNA from 0.1–10,000 attomoles. Samples were labeled and hybridized in quadruplicate to GeneChip miRNA 2.0 Arrays. The average percentage of spikes detected at each abundance level is shown.

*For the sample without spikes, the percentage of false positive (FP) and true negative (TN) is shown.

Reproducibility of the GeneChip miRNA 2.0 Array

The GeneChip miRNA 2.0 Array provides the same high level of reproducibility provided by all Affymetrix expression arrays. The results from this experiment show the extremely high intra-lot and inter-lot signal correlation and extremely high intra-lot fold change correlation (Table 3). The signal and fold change plots in Figure 4 show the high intra-lot and inter-lot correlation within each class of probes (miRNA, snoRNA, and scaRNA) on the array.

Table 3: Signal and fold change correlation: GeneChip miRNA 2.0 Array

GeneChip miRNA 2.0	Median signal Brain (971)	Lung (869)	Fold change Brain vs. lung (971)
Lot 1	0.99	0.99	-
Lot 2	0.99	0.99	-
Lot 3	0.99	0.99	-
Lot 1 vs. lot 2	0.99	0.99	0.98
Lot 1 vs. lot 3	0.99	0.99	0.98
Lot 2 vs. lot 3	0.98	0.98	0.97

Signal and fold change correlations were calculated from detected human probe sets. Probe sets were defined as detected if the median *p*-value from 12 replicates across three lots was less than 0.06. The median Pearson product moment correlation coefficient was calculated from all pair-wise replicate comparisons within a manufacturing lot for intra-lot signal correlation. The Pearson product moment correlation coefficient was calculated from comparison of median signal between lots for inter-lot signal correlation. The Pearson product moment correlation coefficient was calculated from median fold change between lots.

GeneChip® miRNA 2.0 Array correlation to qRT-PCR

The differential expression between human brain and lung samples hybridized to GeneChip miRNA 2.0 Arrays was compared to fold change between brain and lung data generated using qRT-PCR data published by the supplier (Ambion TechNotes 14(2) – May 2007: miRNA Expression in FirstChoice® Human Brain Reference RNA).

The analysis of data from 196 common transcripts demonstrated a correlation coefficient of 0.85. It is important to note that the

Figure 3: Dynamic range. The median signal vs. abundance is plotted for 16 synthetic miRNAs spiked into 100 ng of human liver total RNA from 0–100 attomoles.

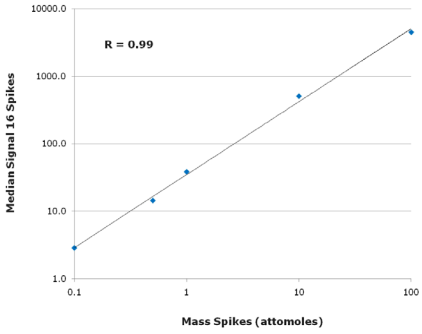
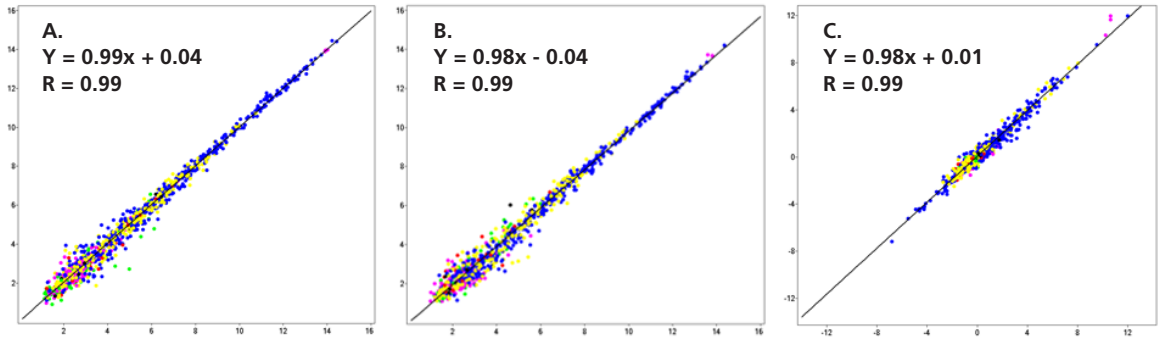


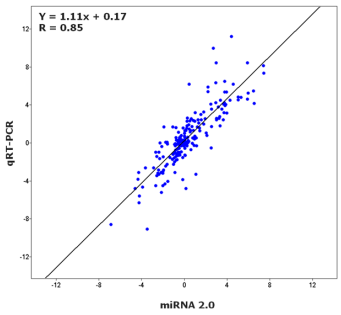
Figure 4: Sample labeling and array hybridization schematic



A. Intra-lot and **B.** Inter-lot signal correlation plots of detected human probe sets from brain. **C.** Inter-lot fold change comparison (brain vs. lung). **Blue:** miRNA; **Yellow:** CD box; **Red:** HAca box; Green: snoRNA; **Black:** scaRNA; **Magenta:** Stem-loop pre-miRNA.

same lots of human brain and lung total RNA were not used for both studies. We expect that the direction of change should be consistent for the miRNAs measured in human brain and lung, while magnitude of change may vary slightly between the studies due to slight lot-to-lot variations in the abundance of miRNA transcripts.The high concordance observed between array and qRT-PCR and the GeneChip® miRNA 2.0 Array results has raised the level of confidence in the biological accuracy of the results.

Figure 5: GeneChip miRNA 2.0 Array correlation to qRT-PCR



GeneChip miRNA 2.0 Array: To determine fold change, lung average Log_2 signal was subtracted from brain average Log_2 signal.

qRT-PCR: Ct values were downloaded from *Ambion TechNotes 14(2) – May 2007: miRNA Expression in FirstChoice® Human Brain Reference RNA*. To determine fold change, brain Ct values were subtracted from lung Ct values. Brain/lung fold changes were plotted for 196 miRNAs in common.

The first miRNA microarray to include precursor miRNA probe sets

A new feature on the GeneChip miRNA 2.0 Array is the inclusion of probe sets interrogating human, mouse, and rat precursor miRNAs. To facilitate the analysis and interpretation of your data, we have added a prefix and a suffix to the standard miRNA ID numbers found in miRBase. Table 4 lists the probe sets that interrogate human precursor miRNA hsa-mir-124-1 and the two mature miRNAs arising from this precursor: hsa-miR-124 and hsa-miR-124*.

Table 4: GeneChip miRNA 2.0 Array probe set nomenclature

Probe set	miRBase ID	Prefix	Suffix	Affymetrix probe ID
Mature	hsa-mir-124	-	_st	hsa-miR-124_st
Mature	hsa-mir-124*	-	_s_st	hsa-miR-124_s_st
Precursor	hsa-mir-124-1	hp_	_st	hp-hsa-mir-124-1_st
Precursor	hsa-mir-124-1	hp_	_s_st	hp-hsa-mir-124-1_s_st
Precursor	hsa-mir-124-1	hp_	_x_st	hp-hsa-mir-124-1_x_st
Precursor	hsa-mir-124-2	hp_	_st	hp-hsa-mir-124-2_st
Precursor	hsa-mir-124-2	hp_	_s_st	hp-hsa-mir-124-2_s_st
Precursor	hsa-mir-124-2	hp_	_x_st	hp-hsa-mir-124-2_x_st
Precursor	hsa-mir-124-3	hp_	_s_st	hp-hsa-mir-124-3_s_st
Precursor	hsa-mir-124-3	hp_	_x_st	hp-hsa-mir-124-3_x_st

_st signifies "sense target"

The probe set names have an "hp" at the beginning to denote that they represent the precursor hairpin (e.g., hp-hsa-mir-124-2_st). Typically, there are multiple probe sets on the array for each precursor miRNA as well as probe sets specific to the mature miRNA(s) derived from the precursor miRNA. Moreover, there can be multiple precursor miRNAs from different genomic loci that give rise to mature miRNAs with the same sequence. For instance, there are three different precursor miRNA sequences present in miRBase v15 that produce hsa-mir-124 and hsa-mir-124*. These three precursor miRNAs include hsa-mir-124-1, hsa-mir-124-2, and hsa-mir-124-3. The same probe set suffix nomenclature used for expression arrays is used for the GeneChip miRNA 2.0 Array and can be used for interpreting results (Table 5).

Table 5: Probe set suffix nomenclature

Suffix	Description
_st	All probes in the probe set are complementary to one known transcript.
_s_st	All probes in the probe set are complementary to transcripts from different genes.
_x_st	Some probes in the probe set are complementary or similar to transcripts from different genes.

The reason for such high probe coverage and multiple probe sets per precursor miRNA is due to the high degree of sequence homology between different miRNAs. Due to the sequence homology between miRNAs, it is not possible to design probes that are complementary to only one transcript.

To overcome this difficulty with interrogating sequences that are complementary or similar to transcripts from different genes, we have grouped these probes into a probe set that provides a single expression value for every probe in the set. On the GeneChip® miRNA 2.0 Array, there are 1,121 probe sets designed to interrogate 911 precursor miRNA sequences.

Using miRNA hsa-mir-124-2 as an example, you can see the hairpin structure of the precursor in Figure 6. The sections of sequence in yellow indicate the positions of the two mature miRNAs derived from this precursor miRNA: hsa-miR-124* and hsa-miR-124, respectively.

Figure 6: Structure of precursor miRNA hsa-124-2

Stem-loop sequence MI0000443

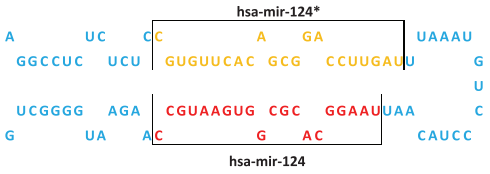
Accession: MI0000443

ID: hsa-mir-124-2

Symbol: HGNC:MIR124_1

Description: Homo sapiens miR-124-1 stem-loop

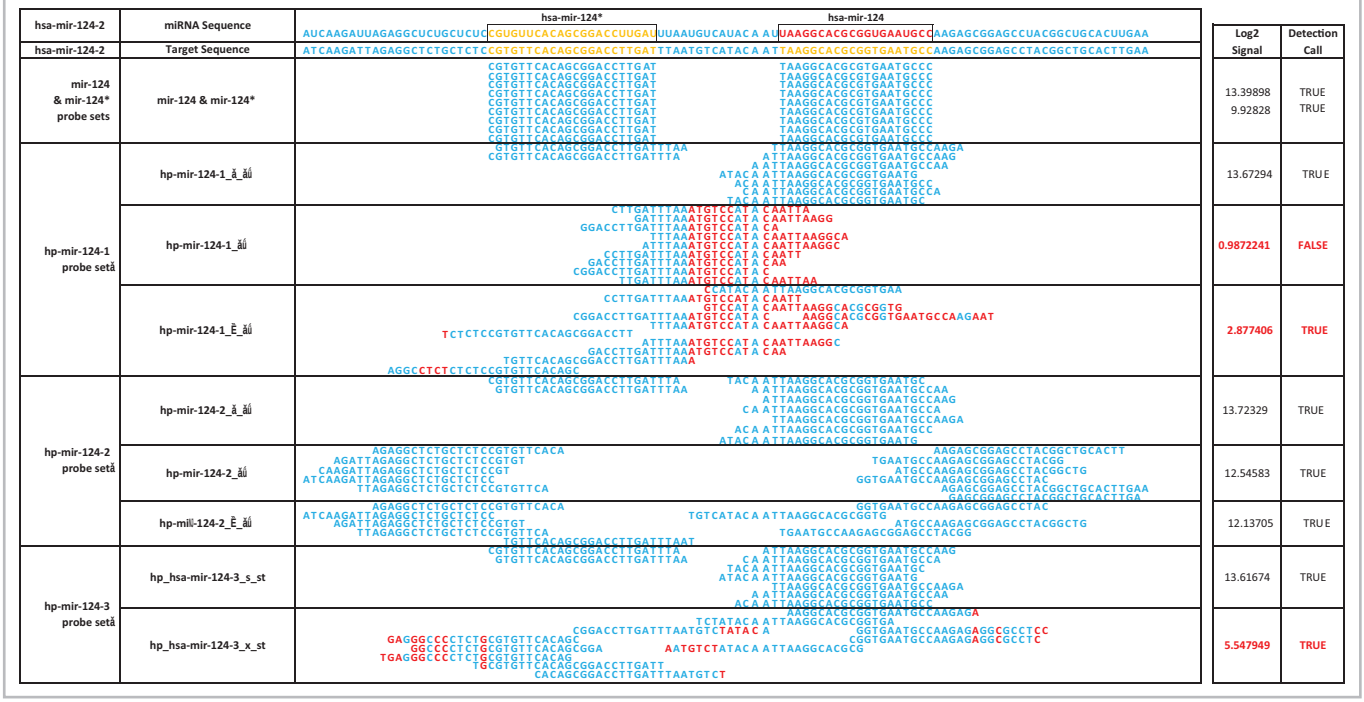
Stem-loop structure



Stem-loop sequence

Information on has-mir-124-2 taken from miRBase (www.miRBase.org).

Figure 7: Alignment of probe sets interrogating precursor miRNA, hsa-mir-124-1, hsa-mir-124-2, and hsa-mir-124-3



This figure graphically represents the alignment of all complementary probes in each of the five probe sets.

To identify the full length precursor miRNA (pre-miRNA), we have designed additional probe sets that cover the space between the two mature miRNA, probe sets that flank the 3′ and 5′ ends of the mature miRNA, and a probe set that spans the full length of the precursor miRNA. Figure 7 shows the five probe sets representing human mature and precursor miRNA hsa-mir-124-2 and the alignment of the probes for every probe set relative to the precursor sequence. The probes with mismatched bases relative to the precursor (hsa-mir-124-2) sequence are in red.

The nine probes in hsa-miR-124_st and probes for hsa-miR-124-star_st probe sets are 100% complementary to the mature and all three premature miRNA sequences. The hp_hsa-mir-124-1_s_st probe set contains probes that flank mature miRNA sequences and is 100% complementary to both hsa-mir-124-1 and hsa-mir-124-2. If hsa-mir-124-2 were present in the sample, you would expect a positive detection call with strong signal.

The precursor probe set hp_hsa-mir-124-1_st centers around the hairpin structure of the stem-loop sequence and contains a large number of mismatches with the hsa-mir-124-2 sequence. If hsa-mir-124-2 were present in the sample (in the absence of hsa-mir-124-1), you would not expect a positive detection call with low signal because the probes are complementary to one transcript.

The precursor probe set **hp_hsa-mir-124-1_x_st** is complementary or similar to other known transcripts from other genes, so low signal would be expected. The signal for this probe set would be high if hsa-mir-124-1 were present.

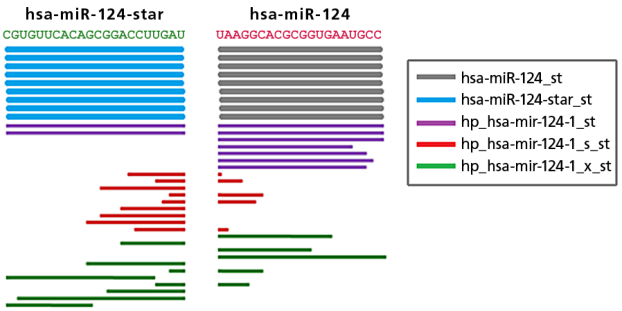
The probe set hp_hsa-mir-124-3_s_st is 100% complementary to both hsa-mir-124-2 and hsa-mir-124-3. If either of the precursors is present, you would expect to see a positive detection call with high signal. However, hp_hsa-mir-124-3_x_st contains a large number of mismatches, but it is also similar to transcripts from other genes, so you might expect low signal and possible positive detection call if hsa-mir-124-2 were present in the sample.

To demonstrate the utility of the precursor miRNA probe sets for detecting precursor miRNA in a sample, Genisphere conducted a number of synthetic precursor miRNA spike-in experiments. In one of these experiments, 2 ng of hsa-mir-124-2 was spiked into 25 ng of liver total RNA. Figure 6 shows the $\log_2$ signal and detection calls for the hsa-miR-124, hsa-miR-124*, hsa-mir-124-1, hsa-mir-1242, and hsa-mir-124-3 probe sets.

In the hsa-mir-124-2 example, if precursor miRNA is present, you would expect to observe signal above background and significant *p*-values (< 0.06) for the two mature miRNA probe sets and the three hsa-miR-124 precursor probe sets.

If the precursor is not present in the sample, but both mature miRNAs were present, you would expect to observe signal above background and significant *p*-values (< 0.06) for both mature miRNA probe sets and precursor probe sets hp_hsa-mir-1_s_st and hp_hsa-mir-124-1_x_st but **not** probe set hp_hsa-mir-124-1_st. This is because the hp_hsa-mir-124-1_st probe set (red) has very little overlap with either of the two mature miRNA sequences, as shown in Figure 7. If a precursor miRNA is present in your sample, all three precursor miRNA probe sets should be detected in the sample.

Figure 8: Probe set sequence overlap with mature miRNA



This example was applied to data generated with the brain total RNA samples (Figure 8). The data generated is consistent with the presence of both mature miRNA but not the precursor.

The signal for probe set hp_hsa-mir-124-1_s_st (purple) most closely matches the signal for probe set hsa-miR-124_st (grey), which is not surprising, since seven out of nine probes from hp_hsa-mir-124-1_s_st overlap the hsa-miR-124 sequence. The signal from probe set hp_hsa-mir-124-1_x_st (green) is relatively low. Only five out of the 11 probes significantly overlap the mature miRNA sequences. The two mature miRNA probe sets and the two precursor miRNA probe sets that overlap the mature miRNAs gave detection calls of TRUE. Probe set hp_hsa-mir-124-1_st (red), with minimal overlap to the mature miRNAs, gave a detection call of FALSE.

In this example, if precursor miRNA hsa-mir-124-2 were present, you would expect to observe signal above background and significant *p*-values (< 0.06) for **all** three precursor probe sets **and** the two mature miRNA probe sets. The low observed signal for the probe set that overlaps the complete precursor pre-miRNA sequence (hp_hsa-mir-124-1_x_st) and the negative detection call on the probe set that covers the hairpin (hp_hsa-mir-124-1_st) are extremely strong evidence that the precursor miRNA is not present in the sample.

Comparison of results from the GeneChip® miRNA 2.0 Array and GeneChip® miRNA Array

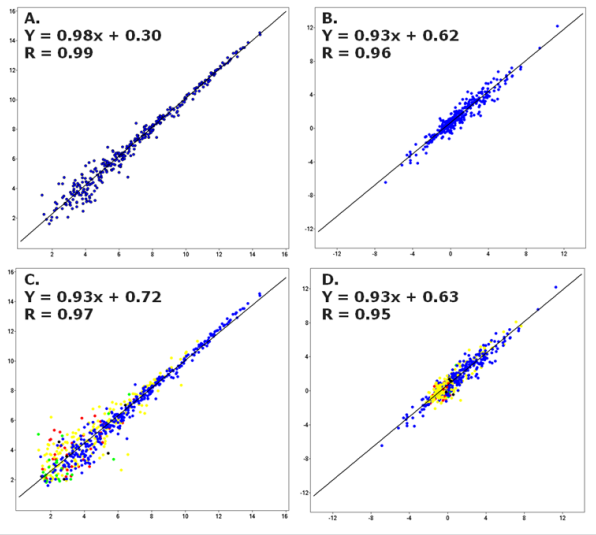
Human brain and lung total RNA samples were used in the

Table 6: Signal and fold change comparison: GeneChip® miRNA 2.0 Array and miRNA Array

Median signal comparison (N=4)	miRNA 2.0 Array mfg. lot	No. probe sets used (human miRNA)	R (human miRNA)	No. probe sets used (all human)	R (all human)
Brain	1	360	0.99	631	0.97
Brain	2	360	0.99	631	0.97
Brain	3	360	0.99	631	0.97
Lung	1	306	0.99	555	0.97
Lung	2	306	0.99	555	0.97
Lung	3	306	0.99	555	0.97
Median fold change comparison					
Brain vs. lung	1	360	0.97	631	0.95
Brain vs. lung	2	360	0.96	631	0.95
Brain vs. lung	3	360	0.97	631	0.95

Target preparation replicates were pooled and hybridized to four arrays per sample. The same pooled target was hybridized to both the miRNA 2.0 Array and the miRNA Array. The Pearson product moment correlation coefficient was calculated from median signal or fold change for detected matched probe sets on the miRNA 2.0 Array compared to the miRNA Array. Probe sets were defined as detected if the median *p*-value from four replicates was less than 0.06. Two subsets of matched probe sets were used: human miRNA (human miRNA probe sets where the probe sequence is identical on both the miRNA Array and miRNA 2.0 Array) and all human miRNA (miRNA probe sets and sno/sca RNA probe sets, which represent the same transcript but do not necessarily share identical probe sequences). One manufacturing lot of miRNA Arrays and three manufacturing lots of miRNA 2.0 Arrays were used in the study.

Figure 9: Signal and fold change scatter plots: GeneChip miRNA 2.0 Array and miRNA Array



The Pearson product moment correlation coefficient was calculated from median signal or fold change for detected matched probe sets on the miRNA 2.0 Array compared to the miRNA Array. Probe sets were defined as detected if the median *p*-value from four replicates was less than 0.06. **A. miRNA brain signal correlation:** correlation of median RMA signal for detected human miRNA probe sets. **B. miRNA brain vs. lung fold change correlation:** correlation of median fold change for detected human miRNA probe sets. **C. Signal correlation, human RNA probe sets, brain:** detected miRNA and sno/sca RNA probe sets. **D. Brain vs. lung fold change correlation, human RNA probe sets:** detected miRNA and sno/sca RNA probe sets. **X-axis:** miRNA 2.0 Array. **Y-axis:** miRNA Array. Blue: miRNA; **Yellow:** CD box; **Red:** HAca box; **Green:** snoRNA; **Black:** scaRNA.

comparison of the miRNA and miRNA 2.0 Arrays. First we evaluated signal and fold change correlation for probe sets representing human miRNAs, where the probe sequences are identical on both array designs. Both array designs have 842 human miRNA probe sets in common. Next, we evaluated probe sets interrogating all human transcripts represented on both array designs, with 1,669 non-control probe sets in common. For the snoRNA and scaRNA, the probe sequences are not necessarily identical on both designs, even though they interrogate the same transcripts. For the correlation calculations, only detected probe sets were included.

Table 6 and Figure 9 demonstrate the high concordance observed between the two arrays. The differential expression, as measured by the fold change correlation, was also highly correlated between the two arrays ($R \geq 0.95$), indicating that users can expect similar biological results from the miRNA Array and the miRNA 2.0 Array.

Conclusions

The GeneChip® miRNA 2.0 Array was designed from miRBase build 15. This array enables users to interrogate miRNA from all 131 organisms present in this build. In addition to mature miRNAs, the array also contains probe sets representing precursor miRNA for human, mouse, and rat, as well as human snoRNA and scaRNAs. Users may transition easily from the predecessor miRNA Array, as very high correlation was observed for probe sets in common to both array designs. The miRNA 2.0 Array results demonstrated excellent reproducibility, sensitivity,

and specificity.

Publications

Kozomara A., Griffiths-Jones S. MiRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research* **39**(Database issue):D152-7 (2011).

Griffiths-Jones S., *et al*. miRBase: tools for microRNA genomics. *Nucleic Acids Research* **36**(Database issue):D154-8 (2008).

Griffiths-Jones S., *et al.* miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Research* **34**(Database issue):D140-4 (2006).

Griffiths-Jones S. The microRNA registry. *Nucleic Acids Research* **32**(Database issue):D109-11 (2004).

Affymetrix, Inc. Tel: +1-888-362-2447 ■ Affymetrix UK Ltd. Tel: +44-(0)-1628-552550 ■ Affymetrix Japan K.K. Tel: +81-(0)3-6430-4020
Panomics Solutions Tel: +1-877-726-6642 panomics.affymetrix.com ■ USB Products Tel: +1-800-321-9322 usb.affymetrix.com

www.affymetrix.com Please visit our website for international distributor contact information.
"For Research Use Only. Not for use in diagnostic procedures."

P/N EXP0158 Rev. 2