TECHNICAL BULLETIN

Gene Expression Profiling with Sentrix® Focused Arrays

1. INTRODUCTION

Microarrays have revolutionized the way researchers study the expression of genes. Many biologists have been able to utilize these technologies to identify targeted subsets of genes of interest, and would now benefit from simpler, more cost-effective methods to interrogate increasingly higher numbers of samples. By combining advances in the telecommunication and microfabrication industries with a novel method of building arrays, Illumina has developed a high-performance, state-of-the-art expression platform that allows researchers not only to study hundreds of genes in parallel, but also to screen hundreds or even thousands of different samples (*Figure 1*). This technical bulletin explains Illumina's core array technology, describes sample labeling and hybridization assays that have been implemented and highlights a number of validation studies.

2. WHY A FOCUSED ARRAY APPROACH?

Numerous advantages exist for applying a Focused Array approach to one's expression profiling studies.

- More manageable data from hundreds of genes, as opposed to tens of thousands, can mean less cumbersome data analysis.
- One can focus on the genes that are deemed important for a particular research question.
- A focused approach allows the capability to assay increased sample numbers at a lower cost.

Relevant gene groups, e.g., those that code for proteins in physiological pathways, can be used to comprise content to be deployed on Sentrix Array of Arrays™ formats (*Figure 2*). Illumina Focused Arrays can be used to query groups of up to 1,400 human genes. Researchers can design content for their own custom Focused Arrays or choose from a growing list of Illumina standard panels of Focused Arrays:

- Sentrix Human Sampler (orthologous to Mouse Sampler sequences)
- Sentrix Mouse Sampler (orthologous to Human Sampler sequences)
- Sentrix Arabidopsis Sampler
- Sentrix Human Toxicology

FIGURE 1: SENTRIX ARRAY MATRIX AND SENTRIX BEADCHIPS

The Sentrix Array Matrix (SAM) format (top) incorporates 96 fiber bundle arrays into a standard 96-well plate footprint. The Sentrix BeadChip format (bottom) can be used to process sixteen samples per BeadChip.

3. SENTRIX ARRAY FORMATS

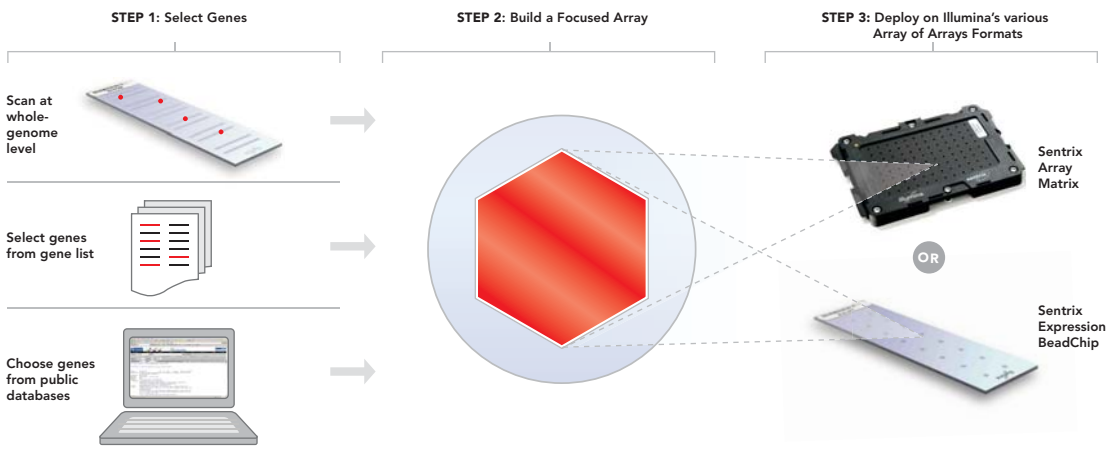
To meet the broadest possible range of researchers' needs, Illumina deploys pooled beads onto two different Array of Arrays formats: the Sentrix Array Matrix (SAM), and the Sentrix BeadChip.

The SAM is configured with 96 fiber-optic bundles, each containing nearly 50,000 individual, light-conducting fiber strands that are chemically etched to create a 3 µm well at the end of each strand. Each fiber bundle comprises an individual array. Fiber bundles are configured to match the well spacing of standard 96-well plates. This unique format allows fast, simple and simultaneous analysis of samples on 96 arrays. The platform can be automated by implementing standard robotic equipment that reduces error, labor and resource requirements.

For users with more moderate throughput demands, Illumina has introduced the BeadChip format. With dimensions approximating a microscope slide (26 x 82.5 mm) and properties similar to the SAM, the BeadChip can be used to process up to sixteen samples per BeadChip.

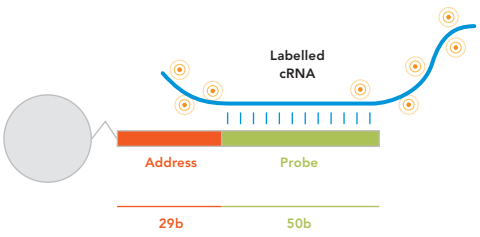
Independent of the array format, each bead in every array contains hundreds of thousands of copies of

FIGURE 2: ILLUMINA FOCUSED ARRAYS



covalently attached oligonucleotide probes. Greater than 1,500 unique bead types are represented, with an average 30-fold redundancy, in each array. Once introduced, the quantitatively pooled libraries of beads self-assemble into the chemically etched wells, resulting in the highest density array platform currently available. After bead assembly, a hybridization-based procedure is used to physically map (i.e., decode) the bead that resides in each well of the array. This final process also simultaneously validates the performance of each bead prior to shipment and provides a level of quality control unmatched in the microarray industry.

FIGURE 3: BEAD DESIGN



50-mer gene-specific probes are attached to beads, which are then assembled into the arrays. For simplicity, this diagram shows only one oligomer attached to the bead; actual beads have hundreds of thousands of copies of the same sequence attached.

4. BEAD DESIGN

Oligonucleotide probes attached to beads in Illumina's Focused Arrays are concatamers of a 29-base IllumiCode address and a 50-base gene-specific probe. The IllumiCode is used to decode the array; the gene-specific probe is used for analysis of expressed transcripts (*Figure 3*).

IllumiCode and probe sequence combinations are carefully selected bioinformatically, followed by a functional screen in the laboratory to eliminate cross-hybridization issues. Gene-specific probes are designed using a rigorous, multi-step algorithm that scores potential probes by screening the following parameters:

- similarity to other genes
- sequence complexity
- self-complementarity (e.g., hairpin structure formation)
- melting temperature
- distance from 3' end of transcript

Probes are also designed to query exon structure. When required, probes can be included for the quantitation of specific expressed splice isoforms or all known splice isoforms.

5. STANDARD CONTENT

Focused Arrays formats have been designed to contain genes of a broad range of interest. The Human Sampler contains probes for transcripts expressed from 516 genes that are orthologous to our companion Mouse Sampler, which containss probes for 503 genes. Genes were chosen from the following categories:

- housekeeping
- apoptosis
- drug metabolism
- GPCR
- cancer
- immunology/inflammation

The Arabidopsis Sampler contains probes for transcripts expressed by 502 genes, which represent gene classes of interest to plant biologists. The following gene categories are represented in the Arabidopsis Sampler:

- housekeeping
- metabolism
- pathogen response
- reproduction/development
- stress response
- transcription factors

Human Toxicology contains probes for transcripts expressed from 622 human genes related to toxicological screening. This panel includes genes involved in the following categories:

- genotoxicity
- oxidative stress
- inflammation
- phase 1 and 2 metabolism responses

5. CUSTOM CONTENT

For custom content, deployable onto SAM or BeadChip formats, customers may submit a list of desired genes, using Reference Sequence Identification (RefSeq ID) or Human Genome Organization (HUGO) gene names, to the Illumina Technical Support group. An Illumina scientist will collaborate with the customer and return a list of 'designable' genes (i.e., genes suitable for gene expression assessment on the Illumina platform). If quantitation of splice isoforms will be required, isoform-specific probes can be included in the custom array. After genes are selected from the list, Illumina Technical Support scientists design and return a fully customized array on the substrate of the customer's choosing (SAM or BeadChip).

6. STANDARD PROTOCOLS & TOOLS

Illumina has designed Focused Arrays to be fully compatible with the standard reagents, techniques and tools that are used extensively throughout the gene expression community, from sample labeling methods to analysis software. Sample labeling, based on the principles of the Eberwine[1] technique, is carried out using the Illumina-modified protocol for the Message AmpII™ kit (Ambion®, Inc.) in a 96-well plate format. For sample labeling with reduced sample numbers, the Illumina RNA Amplification Kit (for 24 reactions; Ambion, Inc.) can be used. Both protocols involve a first and second strand reverse transcription step followed by a single in vitro transcription (IVT) amplification that incorporates biotin-labeled nucleotides. Subsequent steps include array hybridization, washing, blocking and streptavadin-Cy3 staining. Fluorescence emission by Cy3 is quantitatively detected with 0.85-micron resolution for downstream analysis by the Illumina BeadStation. Illumina's data extraction software provides results in standard flat file formats that can be readily processed with most commercial expression analysis software programs. Performance metrics reported for experiments in this technical bulletin obtained using SAMs have also been confirmed using BeadChips.

7. DETERMINING ASSAY PERFORMANCE

To characterize the assay's sensitivity (i.e., limit of detection) and dynamic range, dose-response experiments were performed. Six artificial, polyadenylated RNA transcripts were generated in vitro from cloned bacterial and viral genes and labeled using Illumina's recommended protocol. The labeled RNA was then spiked at twelve different concentrations (0, 0.1, 0.15, 0.3, 1.0, 1.5, 3.0, 10, 15, 30, 100 and 200 pM) across a panel of twelve aliquots of labeled human cRNA generated from HepG2 human hepatocellular carcinoma cell line mRNA. One microgram of each spiked, labeled sample was hybridized to eight different bundles of a SAM. The dose-response hybridization intensity values observed and their variation across replicates provided the basis for the calculation of performance metrics.

Normalized values for the six tested genes are shown (*Figure* 4A). Two-fold change discrimination across a range exceeding 2.8 logs of concentration was observed. This demonstrates that a wide dynamic range of transcripts can be detected. Across the range, the majority of probes show better than 1.3-fold discrimination of

FIGURE 4A: DOSE-RESPONSE CURVES

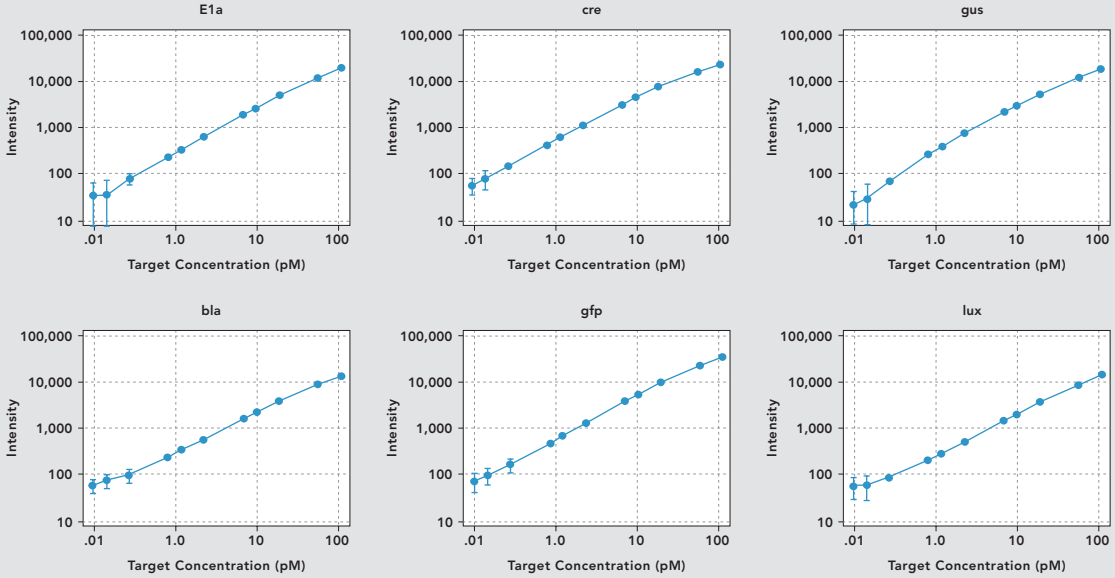
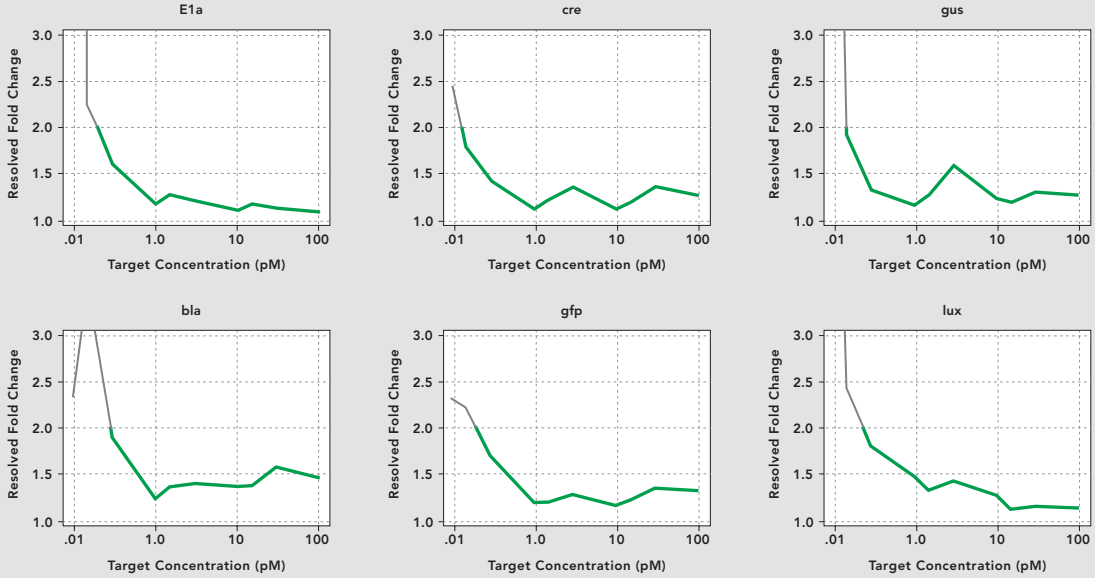


FIGURE 4B: PRECISION AND DYNAMIC RANGE



4A: Data points represent the mean of eight replicate arrays following normalization. A rank-invariant normalization algorithm was used to compute a linear least-squares fit. The fit was based on genes with a less than three percent relative rank change. Error bars represent 90% two-sided confidence intervals for a single reading based on the spread of the replicates. Based on this performance, dynamic range (the concentration range over which two-fold differences are distinguishable from a single replicate with 95% confidence) was measured as 2.8 logs.

4B: Each data point represents the ability to distinguish a specific concentration from its neighboring concentrations, with 95% confidence for a single reading. Values below two-fold are colored green, while values greater than two-fold are colored gray.

concentration changes for their cognate genes (*Figure 4B*). Taken together, these data show that *changes* within a wide dynamic range of expression levels (e.g., low or high abundance transcripts) can be detected with a high degree of confidence.

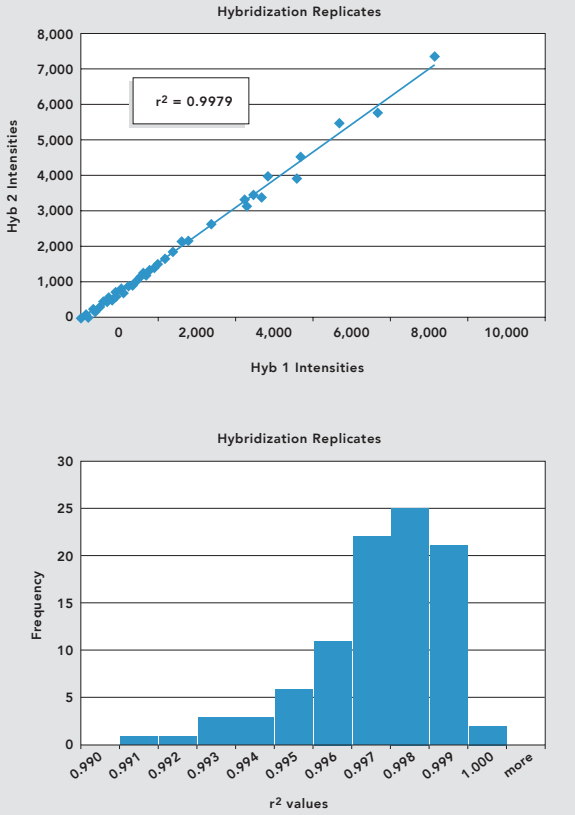
8. HYBRIDIZATION REPRODUCIBILITY

To draw sound conclusions from array results, a high degree of hybridization reproducibility across different arrays must be demonstrated. In a study with researchers at the National Institute of Environmental Health Sciences (NIEHS), total RNA from 96 different human lymphoblastoid cell samples (100 ng each) was labeled and amplified. For 95 samples, labeling and amplification procedures yielded sufficient material for hybridization to at least two arrays on a SAM. For each pair of replicate arrays, the correlation of signals for the 633 genes represented was calculated. The minimum correlation (r^2) was 0.990; the median was 0.997 (*Figure 5*). All replicates yielded extremely high correlations demonstrating a high degree of reproducibility across arrays.

9. LOW SAMPLE INPUT REQUIREMENTS

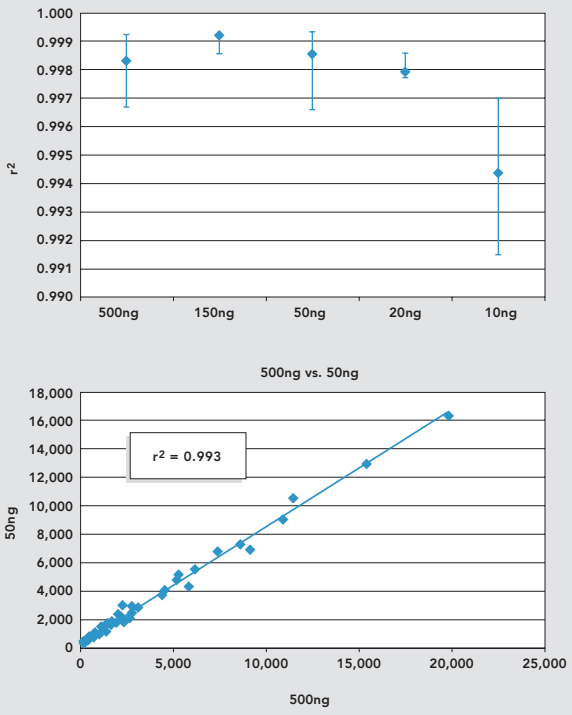
An additional advantage of using Illumina expression arrays is the reduced requirement for sample input. To evaluate optimum input starting material, Illumina's sample labeling protocol was used to label 10, 20, 50, 150 and 500 ng of total RNA from mouse spleen. Each reaction was run in quadruplicate and hybridized to separate arrays on a SAM. The mean and range for each correlation is shown (*Figure 6*, top panel); all r^2 values exceeded 0.99. *Figure 6*, bottom panel, shows a scatter plot of the signal correlation from labeling 50 ng and 500 ng total

FIGURE 5: ARRAY BATCH REPRODUCIBILITY



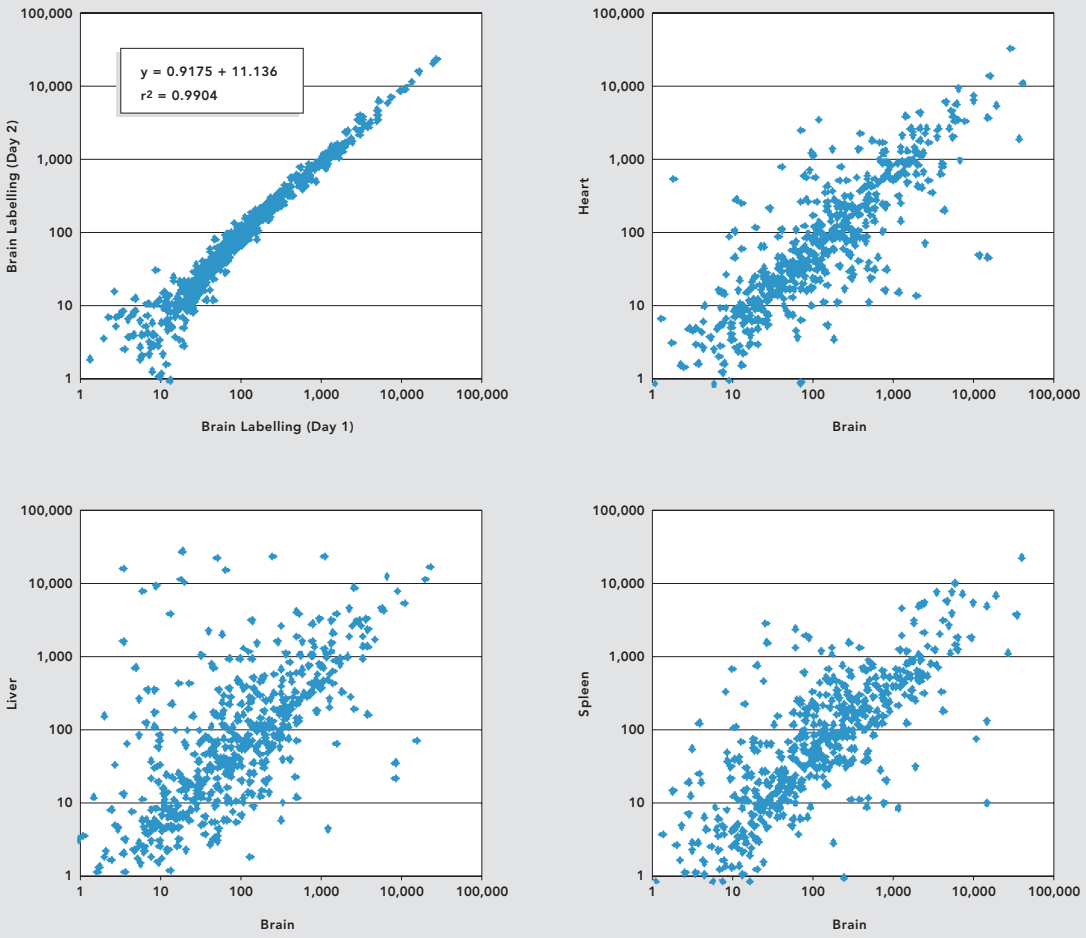
Top: Scatter plot of the median pairwise correlation between replicate hybridizations. Bottom: Histogram of r^2 values for all pairwise correlations.

FIGURE 6: COMPARISON OF INPUT MASS



Top: Mean (diamond) and range (error bars) values are plotted for each correlation. Bottom: Correlation between sample preparations from different starting amounts of total RNA. The linear correlations for the hybridization signals of expressed RNA represented on the array were calculated for all pairwise comparisons at each starting material point.

FIGURE 7: REPRODUCIBILITY AND TISSUE-SPECIFIC EXPRESSION



Plots demonstrate robust sample preparation reproducibility (BRAIN vs. BRAIN plot in upper left) and tissue-specific expression profiling (all other shown plots). Axes show absolute, non-normalized hybridization intensities.

RNA. The strong correlation demonstrates the robustness of the assay even with input material amounts differing by ten-fold. Recommended starting material input is 50-100 ng. Note the optimum input amount may vary depending on RNA integrity, as well as the tissue type from which RNA was isolated.

10. LABELING REPRODUCIBILITY ACROSS RNA ISOLATED FROM VARIOUS TISSUES

The reproducibility of Illumina's arrays across different sample labeling events was evaluated using total RNA purchased from Ambion, Inc. Total RNA from human brain, heart, liver and spleen tissue (100 ng each) was amplified and labeled on the same day. On a separate day, total RNA from brain (100 ng) was independently amplified and labeled. The samples were hybridized to different arrays on a SAM and scatter plots for 633 genes on the array were generated (*Figure 7*). The strong linear correlation of the samples derived from brain, labeled on different days, demonstrates the reproducibility of Illumina's labeling and hybridization procedures. Samples from the remaining tissues not derived from brain exhibited many differences in gene expression. These results also demonstrate the system's capability to

discriminate gene expression variations between disparate tissue types.

11. CONTROLS

Illumina incorporates a number of internal control beads into each array of every SAM (96 arrays) or BeadChip (16 arrays). The large number of internal controls, unique in the industry, makes array-to-array and sample-to-sample comparison not only possible, but a standard feature. The controls are designed to allow a researcher to routinely monitor the following:

- sample quality
- labeling efficiency
- hybridization stringency
- signal generation
- experimental noise

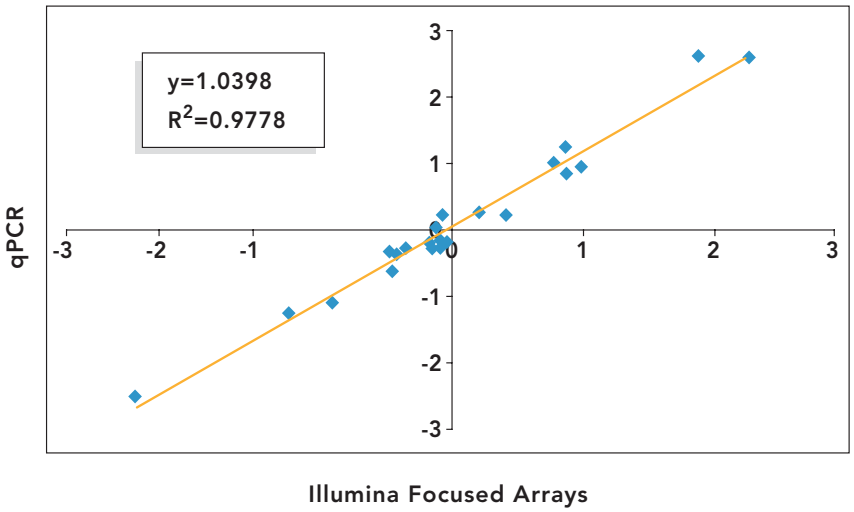
12. SAM ARRAY HYBRIDIZATION VS. QPCR

As an independent measure, relative gene expression patterns derived from SAMs were compared to results obtained from Applied Biosystems® Assays-on-Demand™ quantitative, real-time PCR gene expression system. Tissue-specific RNA expression levels were compared for twenty genes (human brain vs. human liver). Genes were selected to span changes from 1:100 to 100:1 (brain:liver). A scatter plot of log-transformed expression ratios from SAM hybridization and quantitative PCR showed a strong correlation (r^2 =0.97, *Figure 8*).

13. PRODUCT RELIABILITY TESTING

To test the manufacturing quality and reliability of the SAM, labeled human RNA samples were hybridized to twelve SAMs from four independent production lots (*Table 1*). To set a baseline for assay performance, all SAM substrates from Lot 1 were hybridized on Day 1. Hybridization of the remaining SAM substrates was distributed over multiple days to separate manufacturing variation from experimental handling variation. The median array-to-array Coefficient of Variation (CV) across all detectable bead types was below ten percent; after data normalization, CV averaged eight percent. The average, normalized CV for twelve SAMs from four production lots across all detectable beadtypes was twelve percent. The lack of discernable correlation between variability and hybridization day is demonstrated in Figure 9 and further confirms the reliability of Illumina arrays.

FIGURE 8: COMPARISON OF ILLUMINA FOCUSED ARRAYS TO QPCR

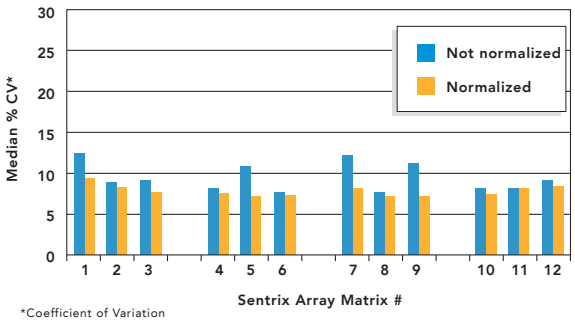


Log-log plot showing comparison of Illumina Focused Arrays expression ratios to qPCR.

TABLE 1: SENTRIX ARRAY MATRIX RELIABILITY TESTING

SAM	Production Lot	Hyb Day 1	Hyb Day 2	Hyb Day 3	Hyb Day 4
SAM1	Lot1	X			
SAM2	Lot1	X			
SAM3	Lot1	X			
SAM4	Lot2		X		
SAM5	Lot2			X	
SAM6	Lot2				X
SAM7	Lot3		X		
SAM8	Lot3			X	
SAM9	Lot3				X
SAM10	Lot4		X		
SAM11	Lot4			X	
SAM12	Lot4				X

FIGURE 9: ARRAY BATCH REPRODUCIBILITY



*Coefficient of Variation

Median values (non-normalized and normalized) are shown. All values were calculated after subtraction of a hybridization-based background value. The background value was calculated from twenty negative control probes that do not contain cognate sample targets. Normalization with respect to background facilitated accurate determination of sample-specific signal.

14. SUMMARY

Sentrix Array platforms combine high-quality results with unprecedented sample throughput, reduced sample input requirements and quality control of every feature on every array. Expression data obtained using both the Sentrix Array Matrix and the Sentrix BeadChip formats show strong correlation with each other and with qPCR, while also exhibiting outstanding array-to-array and batch-to-batch reproducibility. The scalability and interchangeability of the SAM and BeadChip formats allow scientists to select the format most suited to their needs. This provides a level of performance, flexibility and scalability to meet the demands of individual research labs to large-scale production environments.

REFERENCES

(1) Van Gelder, R.N., von Xastrow, M.E., Yool, A., Dement, D.C., Barchas, J.D., Eberwine, J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc Natl Acad Sci USA* **87**, 1663-1667.

ADDITIONAL INFORMATION

To learn more about Illumina's Focused Array products or to build a Custom Focused Array, contact us.

Illumina, Inc.
Technical Support
9885 Towne Centre Drive
San Diego, CA 92121-1975
1.800.809.4566 (toll free)
1.858.202.4566 (outside the U.S.)
techsupport@illumina.com
www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 470-2005-005 27Sep05

illumina®
making sense out of life