TECHNICAL BULLETIN

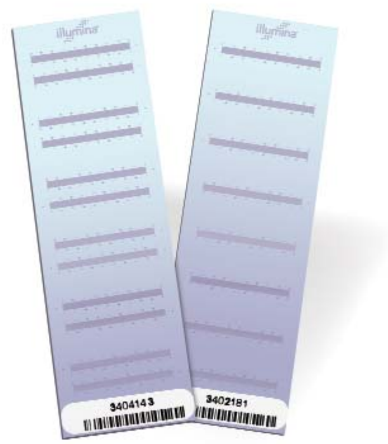
Whole-Genome Expression Analysis Using the Sentrix® Human-6 and HumanRef-8 Expression BeadChips

1. INTRODUCTION

Microarrays have grown in popularity over the past decade as the only tool that provides a truly comprehensive view of gene activity in biological samples. With the introduction of high-density platforms incorporating tens of thousands of sequences, microarray technology has recently reached the point where such measurements can be carried out on every identified gene in the genome. Despite progress, challenges are still associated with the technology that have prevented it from being broadly utilized. The expense of microarray technology is one of these challenges, both for the arrays and for the associated labeling procedures. Current commercially available technologies are also limited by the number of samples that can be interrogated in parallel. Often this leads to compromises in experimental design, such as inadequate replicates or the omission of control samples. Further, high sample input requirements or the need for multiple rounds of amplification have limited the kinds of samples scientists routinely apply to arrays.

In response to these issues, Illumina has designed two new genome-scale array products, the Sentrix Human-6 and HumanRef-8 Expression BeadChips (*Figure 1*). The Human-6 Expression BeadChip contains six arrays on a single BeadChip, each with >46,000 probes derived from human genes in the National Center for Biotechnology Information (NCBI) Reference Sequence (RefSeq) and UniGene databases. The HumanRef-8 Expression BeadChip contains eight arrays of >24,000 probes derived from the RefSeq database. These new Expression BeadChips provide high-quality performance at significantly lower cost per sample. In addition, a minimal amount of total RNA (50-100 ng) is required for the single-round in vitro transcription (IVT) reaction. This technical bulletin explains this technology and presents the results of quantitative experiments that demonstrate the utility of this product.

FIGURE 1: SENTRIX HUMAN-6 AND HUMANREF-8 EXPRESSION BEADCHIPS



2. BEADARRAY™ TECHNOLOGY

At the heart of Illumina's expression product portfolio is a fundamentally different way of building arrays: the random self-assembly of microspheres into ordered microwells. Illumina has leveraged technological advances in the semiconductor manufacturing industry to build substrates containing millions of wells in highly ordered, pre-defined patterns. Oligonucleotide probes are immobilized onto beads to generate the array elements. These beads are quantitatively pooled and introduced to the etched microwell substrates. Once introduced, the beads spontaneously assemble into the wells. The result is the highest density array platform commercially available.

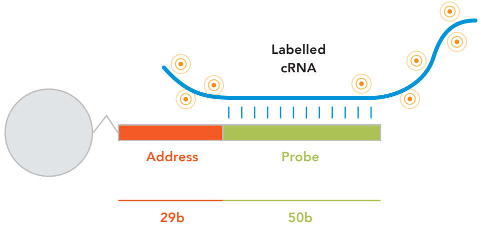
More than 48,000 different bead types, each with a unique sequence, are represented on the Human-6 Expression BeadChip, >24,000 on the HumanRef-8 Expression BeadChip. Each bead contains hundreds of thousands of copies of covalently attached oligonucleotide probes. Beads are assembled into >1.6 million pits, each measuring 3 µm in diameter, generating an average 30-fold redundancy for each sequence represented on the array. This means that each reading is taken multiple times across the array, increasing the accuracy of the measurement. The high quality inherent in this redundancy is bolstered by our manufacturing process. After bead assembly, a hybridization-based procedure is used to map the array, determining which bead type resides in each well. This final process validates the hybridization performance of every bead on every array and provides 100% array QC.

The Human-6 and HumanRef-8 Expression BeadChips are arranged in an Array of Arrays™ format, providing multiple arrays on each slide, six in the case of the Human-6 product and eight for HumanRef-8. The arrays are separated from one another by a seal, so that each array can be hybridized to a different sample *(Figure 2)*. Six or eight samples can than be interrogated simultaneously. This reduces experimental costs and decreases handling, as all steps downstream of hybridization are performed in parallel on each BeadChip. Human-6 and HumanRef-8 Expression BeadChips are scanned on the Illumina BeadArray™ reader, a sub-micron resolution scanner that can scan the 3 µm high-density features in Human-6 and HumanRef-8 Expression BeadChips. A special adapter tray accommodates three BeadChips (18 or 24 arrays) to be scanned in a single loading.

FIGURE 2: BEADCHIP HYBRIDIZATION CARTRIDGE

The BeadChip Hybridization Cartridge allows simultaneous hybridization of each sample within the Human-6 or HumanRef-8 Expression BeadChip.

FIGURE 3: BEAD DESIGN



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.

3. BEAD CONTENT DESIGN

The oligos covalently attached to beads in Human-6 and HumanRef-8 Expression BeadChips contain a 29-base address concatenated to a 50-base gene-specific probe. The address is used to map and decode the array, while the probe is used to quantify expression levels of transcripts.

3.1 Algorithms

Each address and probe sequence combination has been carefully selected bioinformatically and functionally screened in the laboratory to ensure the absence of cross-hybridization. Gene-specific probes were designed using a multi-step algorithm scoring the following parameters:

- Similarity to other genes
- Absence of highly repeated sequence in the genome
- Sequence complexity
- EST coverage (Genome Annotation-RefSeq genes)
- Self-complementarity for hairpin structure prediction
- Melting temperature for hybridization uniformity
- Distance from 3’ end of the transcript

The design also took into account exon structure. Probe design incorporated splice isoforms that have been identified and documented in the RefSeq database.

3.2 Content Sources

The Human-6 and HumanRef-8 Expression BeadChips contain content from a variety of public data sources *(Table 1)*. Because no proprietary data sources are used, access to the source sequences used in the design will remain public. The data sources used include:

TABLE 1: CONTENT SOURCES

Category	Human-6	Ref-8
Curated RefSeq (Release 4 and Build 34)	19,730	19,730
Genome Annotation RefSeq (Release 4 and Build 34)	6,368	4,627
Gnomon (Build 34)	9,576	-
Unigene-163	11,622	-
	47,296	**24,357**

The core data source for the HumanRef-8 Expression BeadChip was the Curated RefSeq Database from the NCBI. This database has many benefits over other sources, including high-quality sequence, annotation tracking, and curation by field experts. RefSeq content was supplemented by sequences from other sources. These other sources primarily contain genes that are less well characterized and less widely expressed. To ensure the selection of quality sequences within this supplementary content, the content was extensively characterized on the basis of the following:

- Match to human genome sequence
- Multiple hits to expressed sequence tag (EST) databases
- Hits to cDNA library databases
- Open reading frame (ORF) prediction to ensure correct strand choice
- Cross-referencing among multiple data sources
- Avoidance of duplicating redundant records
- Avoidance of pseudogenes

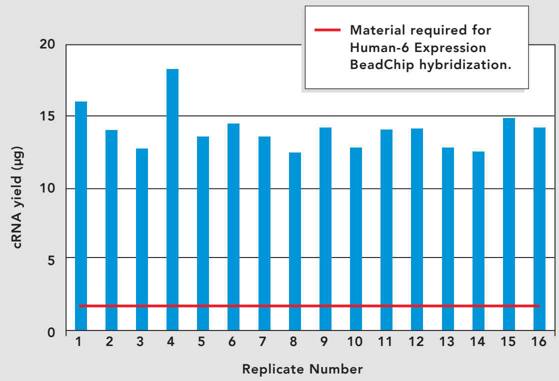
3.3 Probe Screening

Following bioinformatic probe design, probes were selected on the basis of an empirical screening process. Multiple probes were made for each gene (three for RefSeq content, two for genes from other sources). These probes were then hybridized to a panel of RNA samples representing 26 human tissue types. The data were analyzed by a proprietary algorithm (patent pending) to select the best probe for each gene on the basis of its responsiveness and specificity. More than 98% of probe content for the Human-6 and HumanRef-8 Expression BeadChips was selected using this rigorous screening process.

4. DIRECT HYBRIDIZATION ASSAY

Whole-genome gene expression was assessed using a direct hybridization approach. To amplify starting material, a single-round IVT amplification was carried out using the Illumina RNA Amplification Kit, manufactured by Ambion®, Inc. The Illumina RNA Amplification Kit was developed based on the principles of the popular Eberwine technique[1], using reverse transcription followed by a T7 RNA polymerase-based linear amplification step. The efficiency of this amplification, coupled with the small volume requirements of BeadArray technology, results in a very low sample input requirement. *Figure 4* shows the results from a sample labeling experiment. Sixteen 100-ng aliquots of Universal Human Reference RNA (Strategene®) were added into separate labeling reactions using the Illumina RNA Amplification Kit, following manufacturer's instructions. The output of labeled cRNA for each reaction was measured by fluorometry using the Molecular Probes' RiboGreen® kit. One whole-genome hybridization required only 1.5 µg of material, as indicated by the horizontal red line *(Figure 4)*. This amount is approximately 12% of the lowest yielding reaction. Based on these and similar results from other experiments not shown, a total of 50-100 ng of total RNA is recommended. Thirty-six samples can be processed in three days with only three hours of hands-on time *(Figure 5)*.

FIGURE 4: PRODUCT YIELDS GENERATED WITH THE ILLUMINA RNA AMPLIFICATION KIT



Mean output for 16 replicates was 14.0 +/- 1.5 µg using one round of IVT amplification. The lowest yield was 12.4 µg. The red line indicates 1.5 µg, the amount required for one whole-genome hybridization. Input was 100 ng total RNA for each replicate.

FIGURE 5: ASSAY WORKFLOW

Times listed are those required for processing 36 samples (six Human-6 Expression BeadChips). Values shown are start-to-finish times; hands-on times are listed in parenthesis.

5. ASSAY PERFORMANCE DATA

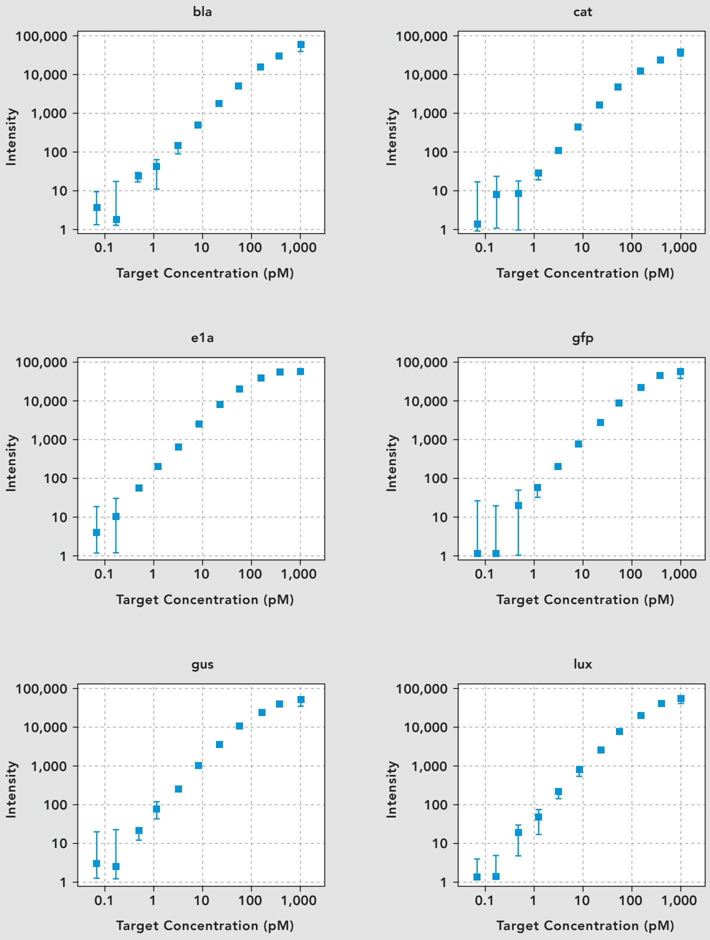
5.1 Dose-Response and Dynamic Range

To characterize the assay's performance, a dose-response study was completed (*Figure 6*). Known quantities of reporter gene RNAs that lack sequences present in human RNA were spiked into human RNA samples. The range over which differences can be measured was assessed by doping a series of samples with increasing concentrations of spiked genes. This dose-response curve was generated for each gene upon hybridization to the array. Sample labeling was performed using the single round IVT amplification method previously described (*Section 4*).

For this experiment, eight artificial polyadenylated RNAs were generated in vitro from cloned bacterial and viral genes. Each of these RNAs was labeled and spiked at twelve different concentrations (0, 0.01, 0.03, 0.1, 0.3, 1, 3, 10, 30, 100, 300, 1000 pM) across a panel of twelve aliquots of labeled human cRNA. This background RNA was generated from Universal Human Reference RNA using the Illumina RNA Amplification Kit. Each spiked, labeled sample (1.5 µg per sample) was hybridized in quadruplicate across 48 arrays on eight different Human-6 Expression BeadChips. Each sample replicate was hybridized on a different BeadChip to avoid the possibility of BeadChip bias. The dose-responses seen and their variation across replicates provided the basis of calculating assay performance metrics.

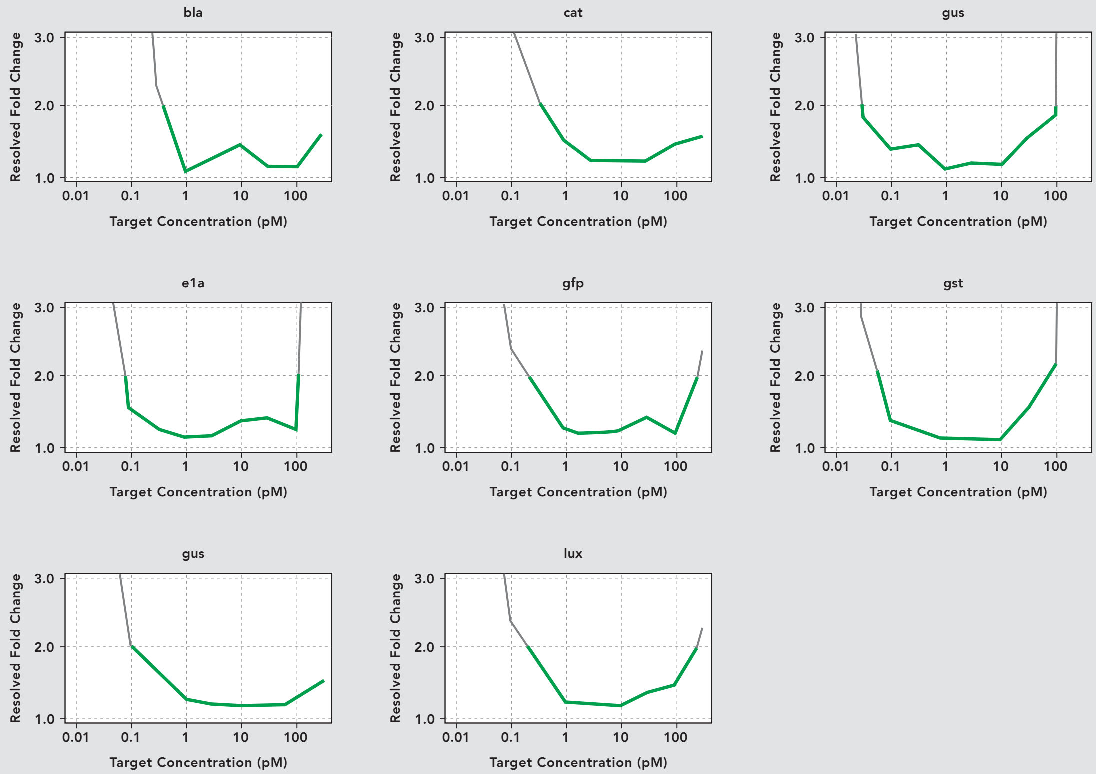
The dose-response curves for normalized hybridization intensities for all eight genes tested are shown (*Figure 6*). Based on these curves, and the spread around each data point, the calculated median limit of detection for this experiment was 0.2 pM (99% confidence). The median dynamic range, defined as the concentration range over which 2-fold changes can be detected with 95% confidence, was measured as 3.05 logs, and the median concentration change detectable with 95% confidence was measured as 1.3-fold (*Figure 7*). This experiment was repeated on HumanRef-8 Expression BeadChips and performance is reflected in the metrics reported in *Table 2*.

FIGURE 6: DOSE-RESPONSE CURVES



Data points represent the mean hybridization intensity of eight different Expression BeadChips. Error bars represent two-sided symmetric intervals (with 90% confidence) for a single reading, calculated based on the spread of four separate readings. All points contain error bars, but some are too small to be resolved at the plotted scale.

FIGURE 7: PRECISION AND DYNAMIC RANGE



The resolvable fold-change is plotted vs. target concentration. Each data point represents the ability to distinguish a specific concentration from its neighboring concentrations, with 95% confidence for a single reading. Values below 2-fold are colored green, while values greater than 2-fold are colored grey. The assay shows two-fold change discrimination across a range exceeding 3 logs of concentration. The median fold-change resolved across the dynamic range for all genes in the experiment was 1.3.

TABLE 2: ASSAY PERFORMANCE

Values obtained from dose-response spiking experiments

Limit of Detection (LOD) 99% confidence	0.25 pM
Precision (fold-change) 95% confidence	≤ 1.3 fold
Dynamic Range 95% confidence	≥ 3 logs

These performance levels have been independently obtained on both the Human-6 and HumanRef-8 Expression BeadChips.

5.2 Comparison to qPCR

To obtain an independent measure of relative gene expression patterns determined by the Human-6 and HumanRef-8 Expression BeadChips, we performed a study comparing expression results to those obtained using the Applied Biosystems® Assays-on-Demand™ quantitative real-time PCR gene expression products (*Figure 8*). A total of 20 genes was chosen from a model system comparing human brain vs. liver expression. Genes were selected to span expression ratios from 1:1000 to 1000:1 brain:liver. A scatter plot of log-transformed hybridization intensity ratios from BeadChip hybridization and quantitative PCR shows a strong correlation ($R^2 = 0.9328$), indicating that accurate measurements of changes in gene abundance over a wide range of concentrations were observed.

FIGURE 8: COMPARISON TO qPCR

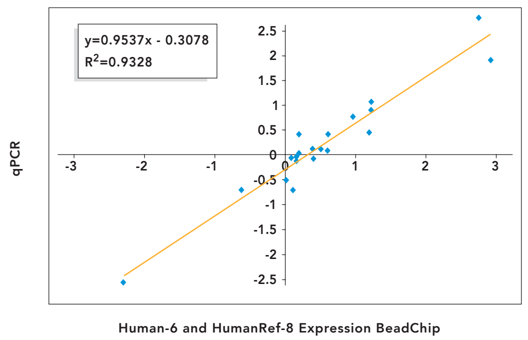
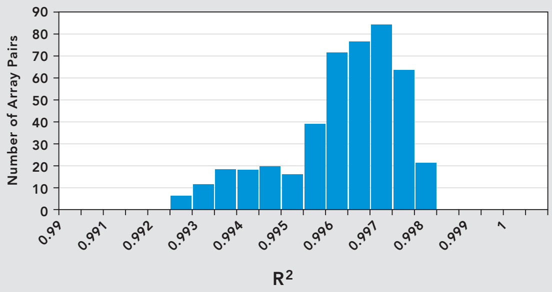


FIGURE 9: ARRAY BATCH REPRODUCIBILITY



Equal mass Universal Human Reference RNA hybridized to five Human-6 Expression BeadChips. Consistent batch-to-batch values were observed.

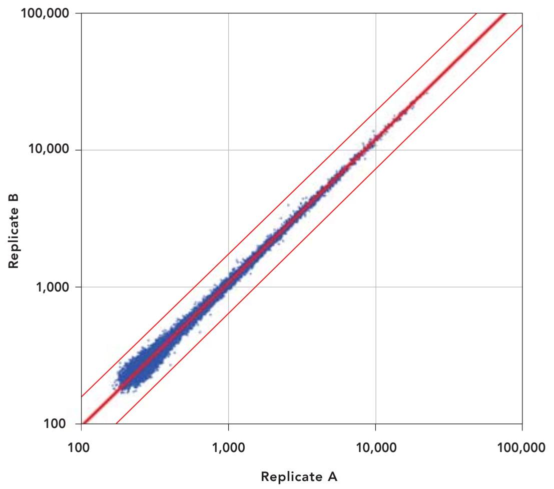
6. HYBRIDIZATION REPRODUCIBILITY

We investigated the reproducibility of data from BeadChips that originate from different manufacturing lots to establish BeadChip-to-BeadChip variability. Five Human-6 Expression BeadChips (30 whole-genome arrays) were randomly selected from three different lots spanning twelve days of production. Labeled cRNA (1.5 µg) generated from Universal Human Reference RNA was hybridized to each array. Linear gene signal correlations (R^2) were calculated for all possible 421 pairwise comparisons of these arrays. A histogram of these values is shown in Figure 9. Minimum, mean, and median values for the comparisons were 0.992, 0.996 and 0.996, respectively. These results demonstrate reliable batch-to-batch consistency.

7. LOW FALSE RATE OF DIFFERENTIAL EXPRESSION

The Human-6 Expression BeadChip shows exceptionally low rates of false differential expression. To test this aspect of the system, replicate hybridizations of labeled Universal Human Reference RNA were performed under standard conditions for eighteen whole-genome arrays on three Human-6 Expression BeadChips. Data were analyzed for all genes for all pairwise comparisons. The false rate of differential expression was determined by counting the number of times any gene showed a change in raw intensity ≥ 2-fold between any two samples. Based on this analysis the mean false detection value was 0.017% for all comparisons. An example of one such comparison between two replicates is shown in *Figure 10*.

FIGURE 10: LOW FALSE RATE OF DIFFERENTIAL EXPRESSION



One comparison between two replicates is shown. The intensity fold-change was determined for all genes for all possible pair-wise comparisons of each of the 18 arrays. False differential expression values are defined as genes showing a ≥ 2-fold difference in raw hybridization intensity (up or down) between two samples. All possible pairwise comparisons among the 18 arrays were performed. The mean false differential expression value was 0.017% +/- .0062%. Red lines mark the thresholds for two-fold expression changes.

8. CONTROLS

Every array on each Human-6 Expression BeadChip includes >1,000 bead types as controls for every experiment, >775 for the HumanRef-8 Expression BeadChip. The controls allow all steps in the process to be monitored carefully, using the following parameters:

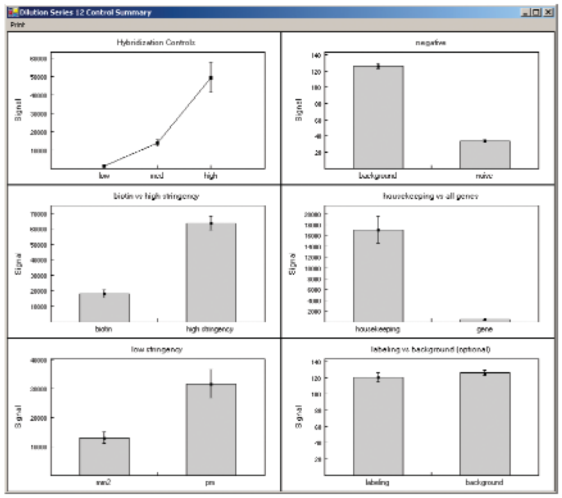
- Sample quality
- Labeling reaction success
- Hybridization stringency
- Signal generation

Illumina's BeadStudio Software has been designed to automatically report the performance of controls after processing each experiment, providing a uniform view of performance every time an experiment is run. An example view of a Control Summary is shown in *Figure 11*.

9. SUMMARY

Illumina's new Whole-Genome Expression Solution using the Human-6 and HumanRef-8 Expression BeadChips allows researchers to streamline array handling and reduce experimental cost by processing multiple samples in parallel on a single BeadChip. Excellent data quality and high reproducibility across arrays and across BeadChips have been demonstrated. Controls included with every array, 100% array QC and high feature redundancy all contribute to the reliability of the data. This reliability is further demonstrated by correlations of R^2>0.93 with qPCR. By utilizing a robust single-round amplification together with Illumina's high-density arrays, sample input of 50-100 ng total RNA is all that is required. With Illumina's high data quality, low sample input, reduced handling and cost, the Illumina Whole-Genome Expression Solution provides a novel, high-performance solution for gene expression research.

FIGURE 11: EXPERIMENTAL CONTROLS DASHBOARD



The Control Summary in BeadStudio Software enables experimental monitoring. A uniform view of the performance of controls is displayed with each experiment.

REFERENCES

(1) Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., Eberwine, J.H. Amplified RNA synthesized from limited quantities of heterogenous cDNA. *Proc Natl Acad Sci USA* 87: 1663-7, 1990).

ADDITIONAL INFORMATION

To learn more about the Sentrix Human-6 or HumanRef-8 Expression BeadChips or to download the gene list, visit our website or contact us at the address below:

Illumina, Inc.
Customer Solutions
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. 470-2005-004 28Jun05

illumina®
making sense out of life