AFFYMETRIX®

Technical Note

GeneChip® Expression Platform: Comparison, Evolution, and Performance

Advances in scanner technology, array manufacturing, analysis algorithms, sequence selection, probe design, and assay conditions have all contributed to improving the amount and quality of data obtained from a single GeneChip® brand array. Additionally, a reduction in feature size has increased the data density on GeneChip arrays. As a result, Affymetrix was the first company to enable researchers to conduct expression analysis of the known (human, mouse, rat) genomes on single arrays.

The latest generation of GeneChip expression arrays is represented by the new 2.0 Platform which includes the Human Genome U133 Plus 2.0 Array, Mouse Genome 430 2.0 Array, and the Rat Genome 230 2.0 Array. The new 2.0 Platform collectively includes arrays with smaller feature size (11 microns), new instrumentation (GeneChip® Scanner 3000 and Fluidics Station 450), new software (GREX), and new assays (GeneChip® IVT Labeling Kit). This Technical Note describes advances of the new 2.0 Platform, and compares its performance with that of the previous generation 18-micron arrays, instrumentation, and assays.

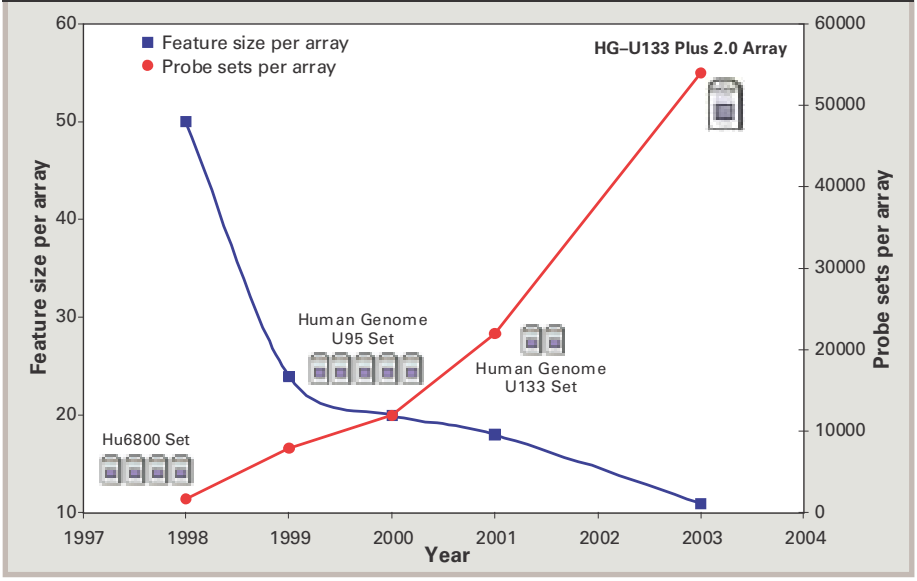
Affymetrix GeneChip® Array Evolution

Gene expression monitoring on the Affymetrix platform has evolved over the years by increasing the information content on each array. While feature size reduction has been the primary means of increasing data density per array, other advances in scanner technology, array manufacturing, analysis algorithms, probe and sequence selection, and assay conditions have all contributed to improving the amount and quality of data obtained from a single array. Throughout this evolution, the system has relied on the use of multiple probes of varying sequence to quantify an individual transcript. The "Power of the Probe Set" has provided the most rigorously tested and statistically qualified microarray expression data available. The latest array designs for expression analysis now enable the analysis of the known genome on a single array. This evolution of the platform can be traced through human arrays, with mouse and rat designs following similar patterns.

The first commercially available array set was the human GeneChip® Hu6800 Set, released in 1998 (Figure 1). This set was comprised of four arrays (Hu6800subA-Hu6800subD) representing approximately 8,000 full-length genes. The Hu6800 Set had 50-micron features and 20 probe pairs per probe set, with 1,700-1,800 probe sets per array. Decreasing the feature size to 24 microns enabled the Hu6800 Set to be condensed to a single array, the HuGeneFL.

In 2000, there were decreases in feature size (20 microns) and number of probe pairs (to 16) for the design of the Human

Figure 1. Evolution of feature size per array and number of probe sets per array.



Genome U95 Set, based on UniGene build 95. This was a five-array set representing predominantly full-length transcripts (~10,000) on the A array and ESTs on the B-D arrays. Each array had more than 12,000 probe sets.

Advances in sequence selection, sequence clustering, probe modeling, probe selection, analysis algorithms, and array manufacturing enabled a major step forward in the release of the Human Genome U133 Set in 2001. In addition, this design incorporated the first complete draft of the human genome. This was a two-array set with over 44,000 probe sets representing more than 39,000 transcripts. This set had 18-micron features and 11 probe pairs per probe set. The reduction in probe pair number was made possible by improvements in probe selection and expression analysis algorithms.

The latest iteration of the human expression arrays is represented by the Human Genome U133 Plus 2.0 Array. Enhancements in array manufacturing, new scanner technology, and improvements in data acquisition allowed the further reduction in feature size from 18 microns to 11 microns. The Human Genome U133 Plus 2.0 Array contains over 54,000 probe sets representing approximately 38,500 genes on a single array. This increase in feature density allows the expression of all known transcripts of an organism to be analyzed on a single array.

This Technical Note describes many of the advances, novel features, and new instrumentation that are currently employed for the analysis of the new 2.0 arrays. Collectively, the smaller feature size arrays, instrumentation, software, and assay comprise the 2.0 Platform. The major differences between the platforms are summarized in Table 1. The performance of the new 2.0 Platform will also be compared and contrasted to the previous generation 18-micron arrays, instrumentation, and assays.

Array Technology

GeneChip® brand probe arrays are manufactured through a unique combination of photolithography and combinatorial chemistry that results in probe densities unmatched in the microarray field. This robust and automated production process is based on technologies adapted from the semiconductor industry. Over the years, Affymetrix has been implementing innovations in its manufacturing process to produce GeneChip arrays with smaller feature sizes, increased information content, and consistent performance.

MANUFACTURING PROCESS

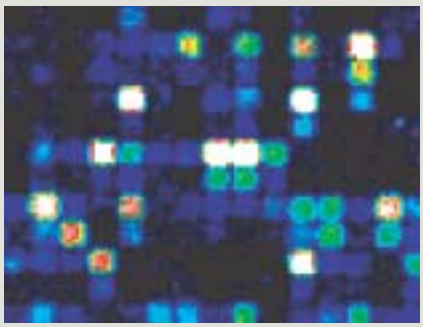
The latest generation of expression arrays has 11-micron feature spacing (also referred to as "11-micron features"), compared to the 18-micron features from the previous generation arrays. Manufacturing enhancements include improvements to the mask design and the use of wafers with a permanent anti-reflective coating (ARC) as the starting point for the manufacturing of the arrays. These improvements enable increased feature definition by reducing the synthesis of overlapping probes between adjacent features. The aforementioned improvements have resulted in higher resolution of 11-micron features when compared to 18-micron features (Figure 2). Also noted in this Technical Note is the enhanced performance in assays, as described in the

Table 1. Comparing platforms.

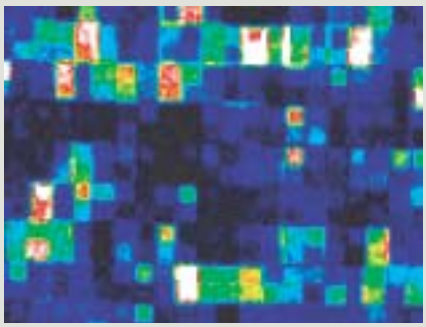
	Previous Generation	2.0 Platform
Array Technology	• 18-µm features • Edge minimization mask strategy	• 11-µm features • Chrome setback mask design strategy • ARC
Image Analysis	Global gridding	Feature extraction (in addition to global gridding)
Data Management	MAS / LIMS	GCOS Client / Server
Analysis	MAS Statistical Algorithm	GREX including PLIER algorithm (in addition to MAS Statistical Algorithm)
Scanning Technology	GeneArray® 2500 or GeneChip® Scanner 3000	GeneChip® Scanner 3000 (high resolution)
Fluidics	Fluidics Station 400/Fluidics Station 450	Fluidics Station 450
AutoLoader	Not available on GeneArray® 2500 (optional for GeneChip® Scanner 3000)	Optional for GeneChip® Scanner 3000
Reagents	• 3rd-party cDNA reagents • Enzo labeling kits	• GeneChip® One- and Two-Cycle cDNA Kits • GeneChip® IVT Labeling Kit

Figure 2. Improved feature definition. **A.** Image from a portion of a GeneChip® Human Genome U133 Plus 2.0 Array .DAT file, showing 11-µm features scanned with the GeneChip Scanner 3000. **B.** Image from a portion of a Human Genome U133A Array .DAT file, showing 18-µm features scanned with the GeneArray® 2500 scanner. Note the improvement in the sharpness, definition, and signal uniformity of the features in Panel A.

Panel A. HG-U133 Plus 2.0 Array



Panel B. HG-U133A Array



Platform Comparative Data section.

IMPORTANCE

Increased feature density enables the inclusion of all the probe sets from the previous generation GeneChip Human Genome U133A and U133B (HG-U133A and HG-U133B) Arrays, as well as 10,000 additional probe sets, on a single HG-U133 Plus 2.0 Array. The new array manufacturing process for the 11-micron feature expression arrays was validated for commercial arrays in 2003 (human, mouse, rat, Drosophila, and canine, and others to follow) and is also available for custom designs.

Scanning Technology

The GeneChip® Scanner 3000 design enables incorporation of future hardware and software advances, as GeneChip technology continues to evolve, to address new and emerging applications in genetic analysis and next-generation, whole-genome research. A comparison of the GeneChip Scanner 3000 and GeneArray® 2500 is provided in Table 2.

The Affymetrix GeneChip Scanner 3000 incorporates a "flying objective" confocal microscope that represents a radical departure from the previous GeneArray 2500 scanner. The "flying objective" employs a large NA (numerical aperture) (0.62) objective providing a four-fold improvement in collection efficiency over the GeneArray 2500. This improved collection efficiency eliminates the need for multiple pass scans.

PERFORMANCE AND BENEFITS

High-performance instrumentation must be able to accurately detect a range of signals (high and low). This performance metric is referred to as the dynamic range and is usually specified as a ratio of the largest data value that can be expressed divided by the smallest data value that can be resolved or digitized. The GeneChip Scanner 3000 is implemented with a full 16-bit resolution which can resolve >65,000 different levels of fluorescence. This allows for a 41 percent improvement in data precision over the 46,000 different levels delivered by the GeneArray 2500.

The GeneChip Scanner 3000 incorporates an auto-zero sub-system that is designed to measure and compensate for long-term offset effects in the scanner hardware. When adjusted, the lower boundary of the system noise envelope can be driven very close to zero. This performance is sufficient to satisfy the requirement for positive, non-zero data values while leaving the largest possible dynamic range available for fluorescence signal acquisition. The GeneChip Scanner 3000

Table 2. GeneArray® 2500 vs. GeneChip® Scanner 3000 Scanner Platform comparison.

Feature \ Platform	GeneArray® 2500	GeneChip® Scanner 3000
Optical Architecture	Pre-objective Line Scan System, small NA	Flying Objective™, large NA
Spot Size (1/e²)	6.8 µm	3.5 µm
One or Two Pass Validation	Two pass scanning	One pass scanning
Image Pixelation	3.0 µm, plus additional lower resolution modes	1.5625 µm, 2.5 µm
Photobleaching	25% at 3.0 µm, two pass scan	2.7% at 2.5 µm and 4% at 1.5625 µm
Scan Time (49 format, HG-U133A)	~10 minutes	~5½ minutes
Image Uniformity	15-20% across the field	Estimated at 1-3% across the field
Signal Resolution	15½ bits (46,000:1)	16 bits (65,000:1)
Measurement Offset	30 to 50 counts, typical	14 to 19 counts, typical
Auto-Zero	No, set at install and preventative maintenance	Yes, every scan

also employs an Auto-Set Laser Power feature. The laser power is accurately sampled prior to every scan and, if necessary, adjusted to default values preset at Affymetrix. This allows for greater consistency in scanning over time.

Advanced design improvements in the Affymetrix GeneChip Scanner 3000 result in significant improvements in scan speed, signal resolution, noise, and dynamic range, as well as reduction in lab benchtop space requirements.

AutoLoader

The GeneChip® AutoLoader is a sample carousel in a temperature-controlled environment, designed for use with the GeneChip Scanner 3000 for unattended scanning of up to 48 arrays.

PERFORMANCE BENEFITS

The AutoLoader is designed to maximize array throughput in the lab by enabling walk away automation. The AutoLoader has been validated internally at Affymetrix and shown not to compromise data quality. Benefits of the AutoLoader include:

- Temperature-Controlled Environment: extensive testing on temperature stability of the GeneChip arrays has shown that data concordance and assay integrity are maintained for arrays cooled at 15° C for up to 16 hours
- 48-Array Carousel: removable carousel enables unattended scanning of arrays for complete walk away automation
- Ease of Use: front-loading position is designed to maximize ease of operation with a simple user interface controlling the AutoLoader within Affymetrix GeneChip® Operating Software (GCOS)
- Additional features include integrated array tracking through a barcode reader, space saving design, multiple modes of operation (automatic, manual, disabled), and automatic e-mail notification for system failures

The GeneChip AutoLoader not only produces high-quality data for gene expression analysis with GeneChip microarrays, but also provides a critical link between high-quality array data and high-throughput automated array analysis.

Fluidics Station

The GeneChip® Fluidics Station 450 is the next generation of fluidics stations for processing GeneChip probe arrays. The washing and staining operations performed on the GeneChip arrays are crucial to the delivery of consistent results. By improving the critical mechanical and fluidic interfaces between the Fluidics Station and the GeneChip array, the Fluidics Station 450 provides new capabilities and robust performance to users of the GeneChip system.

PERFORMANCE BENEFITS

Extensive internal and external testing of the GeneChip Fluidics Station 450 has demonstrated concordance with Fluidics Station 400, enabling comparison with existing data and seamless integration into laboratory workflow. Benefits of the Fluidics Station 450 include:

- Simultaneous loading and detection of the three solution vials required for staining provides greater convenience and maximizes array throughput in the lab
- Enhanced Fluid Detection ensures that fluidics protocols are completed according to the selected fluidic script, which eliminates errors caused by residual fluid
- Improved GeneChip Cartridge Loading ensures proper cartridge orientation and trouble-free loading
- Modular Design enables users to quickly and easily replace modules in the field

The design improvements offered by the new Fluidics Station 450 provide improved reliability with decreased user intervention, which allows for better integration of the Fluidics Station 450 into a high-throughput environment.

Image Analysis (Feature Extraction)

One of the challenges encountered when dealing with decreased feature size is identifying pixels associated with each cell. The decrease in feature size has necessitated new image analysis methods. Historically, a global grid method has been used for gridding Affymetrix® expression arrays. The global grid method works by placing grid anchors at the 4 corners of the array image, which is followed by a linear interpolation based on expected feature spacing to determine the pixel coordinates of the center of each feature. This method has worked well for arrays with 18-micron or greater feature spacing. With the 2.0 Platform, an additional step has been introduced to the gridding method to provide additional robustness on arrays with 11-micron feature spacing. This additional step – called Feature Extraction – fine-tunes the initial estimate of the location (in pixel coordinates) of each feature.

After the initial global grid is placed or addressed, the Feature Extraction algorithm is applied to each cell on the array. A fixed-size rectangle is moved through a pre-defined range of possible pixel positions centered on the initial estimate of the center of the feature. At each considered position, the Coefficient of Variation (CV) of pixel intensities is computed and the position of minimum CV is taken as the final estimate of the feature's center. The size of the rectangle and the positions through which the rectangle is moved are dependent on the type of array. For example, when scanning 11-micron features, the rectangle is 5x5 pixels in size and is moved up to two pixels in each direction on both the x and y axes, for a total of 25 considered positions.

PERFORMANCE

Table 3 presents data as a result of an evaluation of Feature Extraction's performance. Three different arrays were scanned in triplicate (r1, r2, r3 in Table 3); for two of

Table 3. Evaluation of Feature Extraction performance. Results are based on probe level intensities from an internal test array.

Distortion	Replicate	Correlation, Feature Extraction Disabled	Correlation, Feature Extraction Enabled
Type I	r1	0.781	0.871
Type II	r1	0.780	0.876
Type I	r2	0.882	0.932
Type II	r2	0.870	0.930
Type I	r3	0.732	0.754
Type II	r3	0.749	0.799

the three scans an intentional nonlinearity was introduced to produce a slightly distorted image (labeled Type I or Type II in Table 3). For each distorted image, feature intensities were estimated – with Feature Extraction enabled and disabled – and the resultant intensities were compared with those estimated from the original unperturbed scan (not listed in Table 3). In all cases, use of Feature Extraction led to better correlation with intensity estimates from the reference image.

Figure 3 presents data demonstrating the benefit of Feature Extraction at the level of probe set signal estimates. A set of images was deliberately perturbed to simulate the effect of an image distortion and a paired-sample t-test was performed comparing signals between the distorted and original images. Any differences found to be significant would constitute false positives in a real experiment comparing two treatments. The tests were conducted in two contexts – with Feature Extraction turned off and then with it turned on. As can be seen in Figure 3, Feature Extraction greatly reduces the number of false positives.

A specific example of Feature Extraction's utility can be seen in Figure 4, in which the initial estimate from the global grid is off by approximately two pixels, as demonstrated in the image on the left. Without Feature Extraction, part of the intensity for the cross-shaped pattern in this image is incorrectly assigned to features to the right, which should appear dim. With Feature Extraction enabled, the grid is adjusted to correct the initial placement and pixel intensities are attributed to the proper features.

Feature Extraction is automatically applied to the data when running a 2.0 array that has the high-resolution scanner and GCOS v1.1.1 or GCOS v1.0 with the High-Resolution Scanning Patch applied. It is important to note here that after Feature Extraction has been applied, the global grid does not update the scanned (.DAT) image to display the new gridlines or the area by which the algorithm quantifies the features.

At smaller feature sizes you may see grid misalignment in your arrays from time to time. However, Feature Extraction on the 2.0 Platform will correct for up to 2 pixels of misalignment.

Labeling and Hybridization

The new 2.0 Platform also includes a novel and proprietary labeling reagent based on a biotinylated pseudouridine compound. Specific conditions for the hybridization, stain, and wash steps were optimized for use with the new labeling kit in conjunction with other components of the 2.0 Platform. Although the new reaction follows a similar setup and purification process as the previous protocol, a number of studies on different 2.0 arrays have confirmed that the new labeling protocol results in improved detection of exogenous spikes.

Figure 3. Feature Extraction provides robustness at level of probe set signal estimate. A set of images was perturbed to simulate the effect of an image distortion. A paired-sample t-test was conducted on signal estimates to determine differences between the distorted and original images. Any differences found to be significant represent false positives. The test was conducted twice, once with Feature Extraction enabled and once with it disabled. Use of Feature Extraction greatly reduces the number of false changes found.

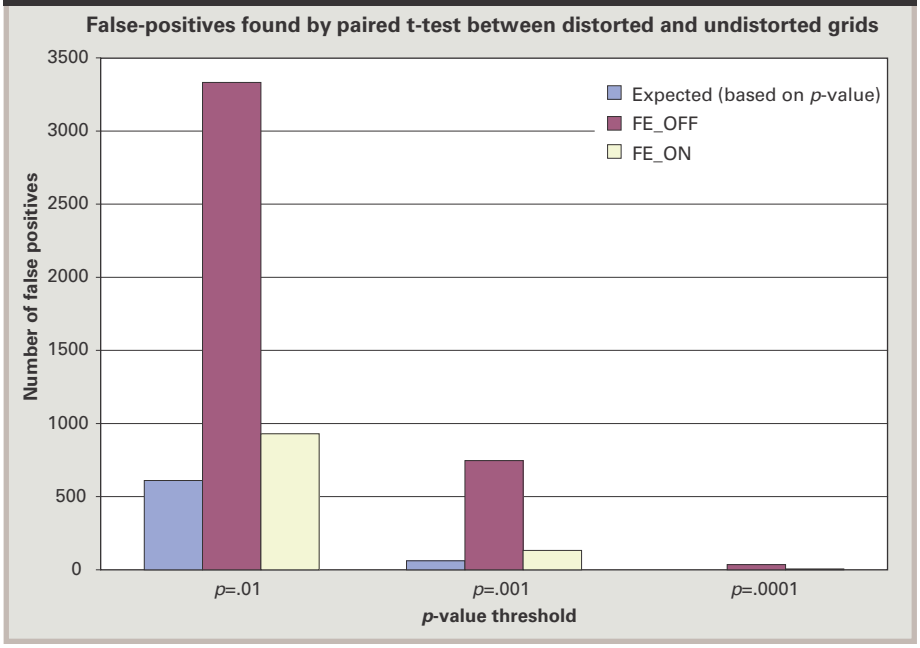
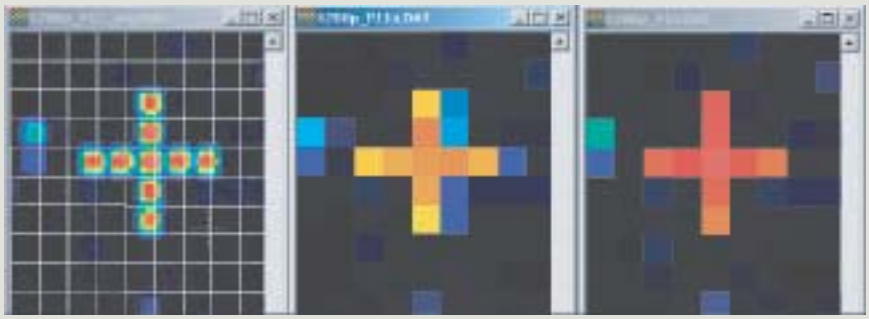


Figure 4. Feature Extraction at work: note that the .DAT file view can be altered within GeneChip® Operating Software with the hot keys A and M to view how Feature Extraction has adjusted the grid.



Measured Image (M) | Average Image, Feature Extraction turned OFF (A) | Average Image, Feature Extraction turned ON (A)

PERFORMANCE

The new GeneChip® 3'-Amplification Reagents for IVT Labeling Kit is based on a T7 RNA polymerase-mediated reaction like the previous system, only now the users have the convenience of an overnight reaction. The new IVT (*in vitro* transcription) Labeling Kit is also optimized to enable customers to start with as little as 1 µg of total RNA. Overall cRNA yields are comparable to the previous labeling system. The overall biotin incorporation rates with the single-labeled nucleotide were slightly reduced from the previous system (Figure 5). However, more stringent hybridization and wash conditions were introduced in the 2.0 Platform to create a system with improved sensitivity in the detection of low concentration transcripts. Ten percent DMSO was also added to the hybridization buffer and a new fluidic script was developed for 49- and 64-format arrays. The new conditions improved the overall discrimination ((Perfect Match - Mismatch) / (Perfect Match + Mismatch)) signal ratios for the probe sets as shown in Figure 6. An increase in the discrimination ratio results in an improved overall sensitivity of the assay for detecting low-expressed genes.

IMPORTANCE

The combination of the new IVT labeling kit and its associated hybridization and wash conditions has helped produce better performance than the previous platform in terms of low-end sensitivity. While most users will find the improved sensitivity to be an asset in their current studies, users should be aware that differences exist between the labeling systems. As in any experimental design, it is essential to minimize experimental variation to allow focus on the biological differences. Samples run on the same array under the same hybridization conditions minimize the system variation.

Data Analysis (PLIER and GREX)

Another key component of the 2.0 Platform is the new probe level analysis method, Probe Logarithmic Intensity ERror estimation (PLIER). This algorithm incorporates many of the concepts that have been recently published within the field of GeneChip data analysis, which includes model-based expression analysis and non-linear normalization techniques.

The PLIER algorithm utilizes experimental data generated across multiple arrays in order to identify and account for observed patterns in probe intensities. Like other model-based approaches, PLIER

Figure 5. Biotin label incorporation rates from the new GeneChip® IVT Labeling Kit and the previous Enzo® Kit across various tissue samples. Labeled cRNA samples were digested into mononucleosides, resolved on an analytical HPLC, and the specific activities of the targets were plotted as a function of the number of biotins incorporated out of every 100 bases.

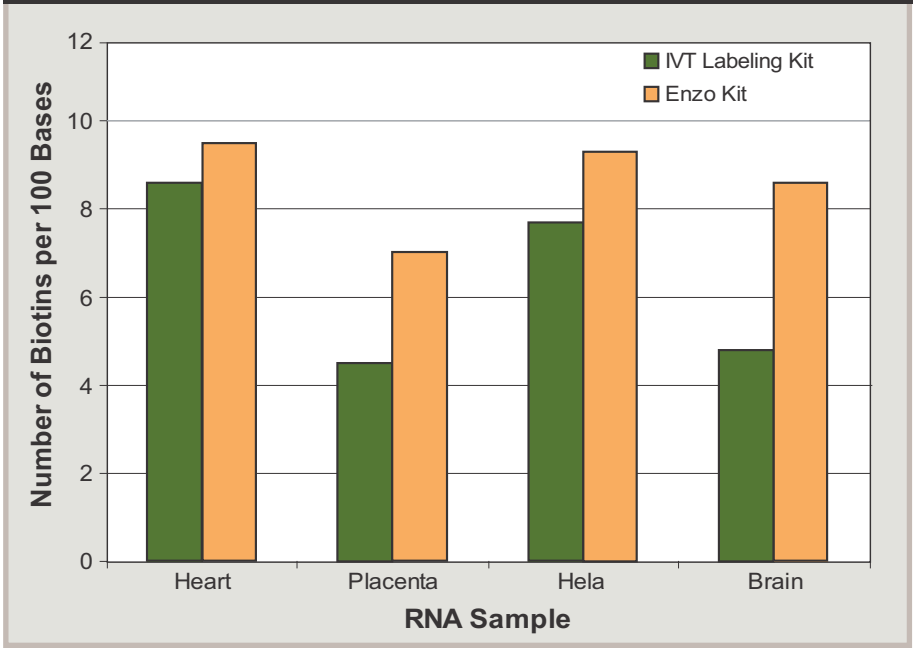
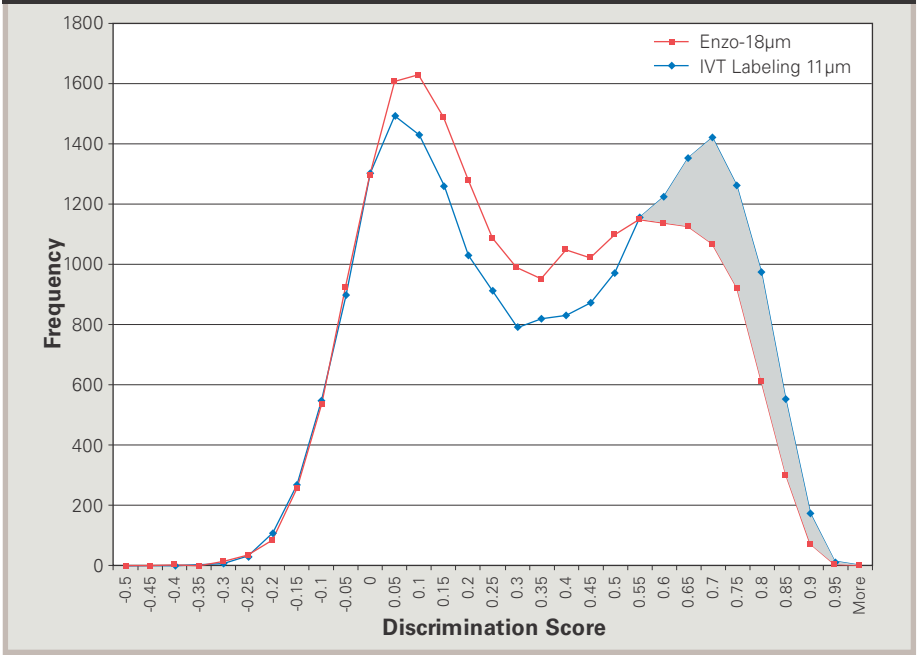


Figure 6. Discrimination score histograms for the GeneChip® Human Genome U133A Array probe sets show a shift of a significant number of signals to higher value score bins for the new platform. HeLa sample data are shown with the gray-shaded area representing 1,755 probe sets with higher discrimination scores for the new platform.



accounts for the difference between probes by means of a parameter termed "probe affinity." Probe affinity is a measure of how likely a probe is to bind to a complementary sequence across a range of concentrations, as all probes have different thermodynamic properties and binding efficiencies. Probe affinities determine the signal intensity produced at a specific target concentration for a given probe, and are calculated using experimental data across multiple arrays. By accounting for these observed differences, all of the probes within a set can be easily compared. For example, if one probe is consistently twice as bright as others in the set, PLIER appropriately scales probe intensities. In the case of a probe set, this enables all set members to be compared and combined accurately.

PLIER also utilizes an error model that assumes error is proportional to the probe intensity rather than the target concentration. At high concentrations, error is approximately proportional to target concentration, since most of the intensity is due to target hybridization signal. At the low end, however, error is approximately proportional to background hybridization intensity, which is the largest component of the observed intensity. Due to the latter effect, it is inaccurate to treat error as a proportion of target concentration in all circumstances. The PLIER error model smoothly transitions between the low end, where error is dependent upon background, and the high end, where error is dependent on signal.

PERFORMANCE BENEFITS

While the Affymetrix® Microarray Suite 5.0 statistical algorithm treats all experiments independently, PLIER supports a multi-array approach that enables replicate analysis. PLIER utilizes consistent probe behavior across experiments to improve the quality of results in any one experiment and discount outliers. Specific benefits of this approach include:

- Higher reproducibility of signal (lower coefficient of variation) while retaining accuracy
- Higher differential sensitivity for low expressors

The PLIER algorithm will be incorporated into a new software application called GeneChip RNA EXpression Analysis (GREX), which represents a significant upgrade from the Data Mining Tool product. GREX offers the following new features in addition to Microarray Suite 5.0 statistical algorithm and PLIER:

- Quantile Normalization
- ANOVA for replicate analysis
- Integrated retrieval of NetAffx™ Analysis Center annotations for faster identification of biological relevance
- Enhanced Gene Ontology Mining Tool for derivation of biological significance
- Hierarchical clustering and Venn diagrams for pattern finding
- Performance enhancements for quicker data access
- Wizards and built-in help for a more intuitive user interface

With these new features, the application addresses a full workflow from probe level analysis to biological interpretation.

Platform Comparative Data

Data derived from three recent 11-µm products (HG-U133 Plus 2.0, Mouse 430 2.0, and Rat 230 2.0 Arrays) were used to assess differences between them and their 18-µm counterparts. These differences were analyzed using the Affymetrix® Microarray Suite 5.0 statistical algorithm, unless otherwise stated. Tissue panels consisting of approximately ten samples were used to evaluate general trends, such as Present and Absent calls between the platforms. Tissue panel data were also used for the Two-Way ANOVA that examined platform similarities and differences. Exogenous spikes were used to address system sensitivity and specificity through generating Receiver-Operator Characteristic (ROC) curves.

Table 4. Common probe sets: GeneChip® IVT Labeling Kit vs. Enzo® Kit.

Average % Present: Common probe sets
Rat 230 2.0 Array (IVT Labeling) Rat 230A/B Arrays (Enzo)

Tissue	230 2.0	230A	230 2.0	230B
Brain	76%	64%	60%	45%
Embryo	76%	64%	56%	41%
Heart	67%	55%	43%	33%
Kidney	70%	59%	49%	37%
Liver	64%	52%	41%	30%
Lung	74%	66%	56%	43%
Muscle	65%	53%	41%	32%
Ovary	75%	65%	59%	45%
Spleen	70%	58%	53%	41%
Testicle	65%	52%	45%	32%
Thymus	71%	61%	53%	41%

PRESENT/ABSENT CALLS

The new 2.0 Platform generally produces a higher percentage of Present calls as observed in the overall performance using tissue samples prepared and analyzed with either platform (Table 4). The new 2.0 Platform shares a high proportion of the Present calls with the previous platform, while producing additional Present calls (Table 5). These new Present calls come from relatively low-signal probe sets. In the analysis of rat tissue samples, the Rat 230B Array probe sets on the Rat 230 2.0 Array show an increased percentage of Present calls relative to the Rat 230A Array probe sets when comparing the two platforms (Table 5, %PA column).

DETECTION CALL ROC CURVES

Previous observations suggest that the new 2.0 Platform is more sensitive than the previous generation 18-micron arrays and assays with low-signal probe sets. Spike experiments were used to further explore the notion of improved sensitivity with the 2.0 Platform. Forty-six different mouse cRNA spikes were analyzed in a Latin square experiment at 0, 0.19, 0.38, 0.75, 1.5, and 3.0 pM concentrations. The hybridizations also included background cRNA from a cell line shown to be devoid of these transcripts by quantitative RT-PCR.

Data from Latin Square experiments were analyzed as a ROC curve using the Detection Call Algorithm (Figure 7). In this graph, the Detection Call sensitivity

Table 5. Detection call concordance: GeneChip® IVT Labeling Kit vs. Enzo® Kit.

Detection Call Concordance: Rat 230 2.0 Array (IVT labeling) vs. Rat 230A/B Arrays (Enzo)

Tissue	Array Comparison	% Concordance	% Discordance	% PP	% PA	% AP	% AA
Brain	Rat 230 2.0 vs. Rat 230A	86%	14%	62%	13%	1%	23%
	Rat 230 2.0 vs. Rat 230B	80%	20%	43%	17%	2%	37%
Embryo	Rat 230 2.0 vs. Rat 230A	85%	15%	62%	14%	1%	23%
	Rat 230 2.0 vs. Rat 230B	79%	21%	38%	18%	3%	41%
Heart	Rat 230 2.0 vs. Rat 230A	84%	16%	53%	14%	2%	32%
	Rat 230 2.0 vs. Rat 230B	83%	17%	29%	14%	3%	54%
Kidney	Rat 230 2.0 vs. Rat 230A	85%	15%	57%	13%	2%	28%
	Rat 230 2.0 vs. Rat 230B	83%	17%	34%	14%	3%	48%
Liver	Rat 230 2.0 vs. Rat 230A	83%	17%	50%	15%	2%	34%
	Rat 230 2.0 vs. Rat 230B	82%	18%	27%	15%	3%	55%
Lung	Rat 230 2.0 vs. Rat 230A	88%	12%	64%	10%	2%	24%
	Rat 230 2.0 vs. Rat 230B	82%	18%	40%	15%	3%	41%
Muscle	Rat 230 2.0 vs. Rat 230A	84%	16%	51%	14%	2%	33%
	Rat 230 2.0 vs. Rat 230B	84%	16%	29%	13%	4%	55%
Ovary	Rat 230 2.0 vs. Rat 230A	88%	12%	64%	11%	1%	24%
	Rat 230 2.0 vs. Rat 230B	82%	18%	42%	16%	2%	39%
Spleen	Rat 230 2.0 vs. Rat 230A	85%	15%	57%	13%	2%	29%
	Rat 230 2.0 vs. Rat 230B	82%	18%	38%	15%	3%	44%
Testicle	Rat 230 2.0 vs. Rat 230A	83%	17%	50%	15%	2%	33%
	Rat 230 2.0 vs. Rat 230B	81%	19%	29%	16%	3%	52%
Thymus	Rat 230 2.0 vs. Rat 230A	87%	13%	59%	11%	2%	28%
	Rat 230 2.0 vs. Rat 230B	82%	18%	38%	15%	3%	44%

Figure 7. ROC curve analyses of a spike-in experiment which assessed low-end platform sensitivity. In a cyclic Latin square design for each platform (previous generation: Mouse 430A, 2.0 Platform: Mouse 430 2.0), forty-six exogenous mouse transcripts were spiked into a cell line sample (D2N lymphoblast) which was empirically checked by quantitative RT-PCR to be absent of these transcripts. For each spike pool, triplicate arrays were hybridized across three independent lots of the respective array designs and detection calls were made by the Affymetrix® Microarray Suite 5.0 statistical algorithm across a range of alpha 1 cutoff values. Default alpha 1 value (0.05) is shown for each trace by the enlarged data point.

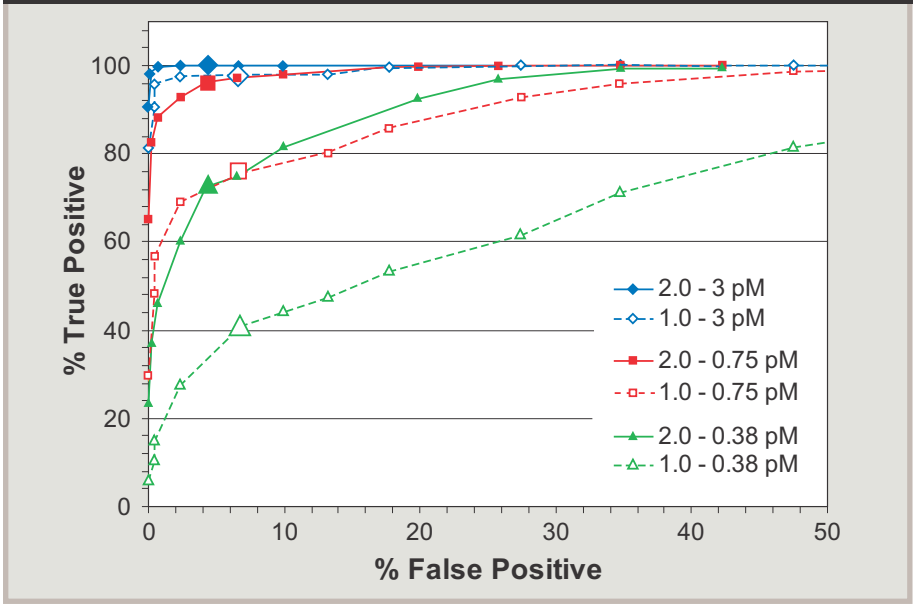
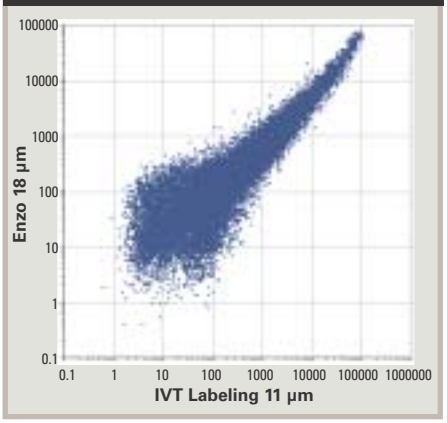


Figure 8. Scatter plot of probe set signal intensities shows the correlation between the two platforms. Data from GeneChip® Human Genome U133 Plus 2.0 Arrays (2.0 Platform) were filtered for GeneChip® Human Genome U133A Array probe sets and median normalized to the intensities of the previous generation platform (HG-U133A) to adjust for scaling factor differences. Signal intensities for both platforms were generated using MAS 5.0 statistical algorithm.



is plotted on the y axis while the False Positives (1 – Specificity) are shown on the x axis. The greater the sensitivity and specificity of an assay, the closer the curve will approach the upper left-hand corner of the graph.

At spike concentrations of 3.0 pM, both platforms have sensitivity values close to 100 percent, when using the default alpha 1 cutoff value as point of reference. At 0.75 pM the 2.0 Platform maintains a sensitivity of 95 percent while the previous generation platform has dropped to approximately 75 percent. At 0.38 pM the difference between the platforms continues to grow, as the 2.0 Platform has a sensitivity of 73 percent while the previous generation platform is at 41 percent. Moreover, the increase in detection sensitivity is bolstered by better specificity (i.e., lower false positives) for the new platform at similar alpha values, as shown in Figure 7. These results confirm that the new 2.0 Platform has improved low-end sensitivity.

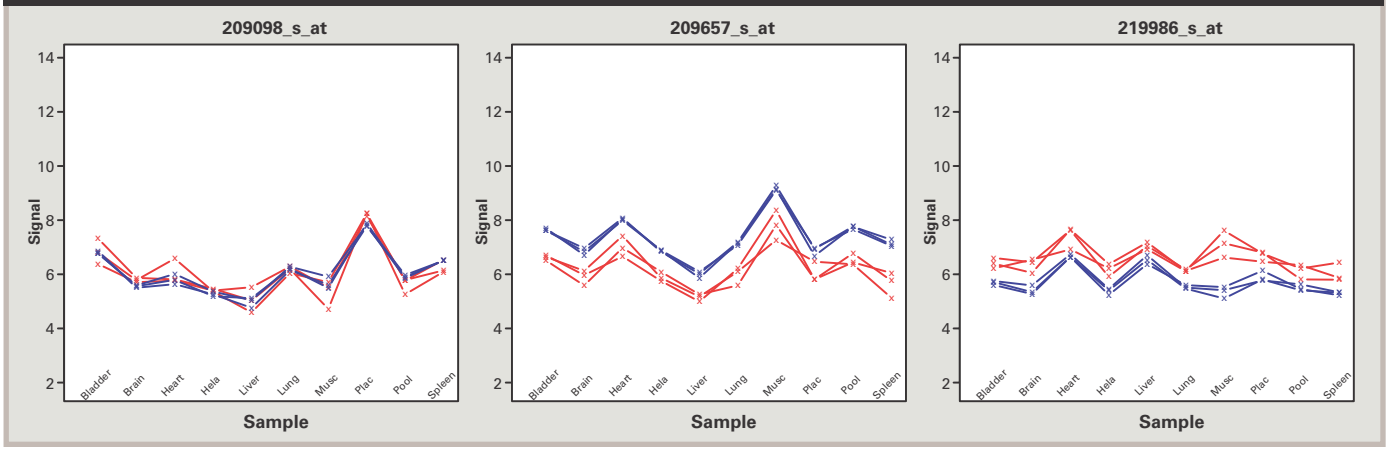
SIGNAL COMPARISON

While many customers rely on the output of the Affymetrix Detection Call Algorithm, still others are inclined to use the Signal value outputs, which are then analyzed by other software packages. An example of the degree of correlation that can be expected in comparing signals from the two platforms is shown in Figure 8. In this scatter plot, an R^2 value of 0.93 was observed. While this is indicative of a high degree of correlation, replicates within the same platform routinely produce R^2 values >0.98.

TWO-WAY ANOVA

Affymetrix evaluates signal values from a tissue panel to quantify similarities and differences between array designs and platforms. Two-Way ANOVA was used to quantify the observed signal differences. The two factors considered were platform (previous generation vs. 2.0) and tissue (a panel of 10 human tissues). Quantile normalization of signal estimates within each tissue type was found to be effective in reducing the variability between the two platforms, and was utilized for this analysis. The probe sets in common between the GeneChip Human Genome U133A and GeneChip Human Genome U133 Plus 2.0 arrays (>22,000) were compared. This analysis showed 10 percent of the variation was due to differences between platforms and 83 percent was due to tissue differences. The remaining 7 percent of the variation was attributed equally between tissue-platform interactions and residual unexplained variance. While there is a certain amount of systematic difference between the two platforms, differences between expression levels across tissues is, as expected, the most important source of variation.

Figure 9. Comparison of signal estimates between platforms. Data were analyzed using a quantile-normalized, PM-MM method. Ten tissues were analyzed in triplicate on each platform. The figure plots $\log_2$(signal) against tissue type for three representative probe sets. Estimates for the previous generation platform are in red (HG-U133A, Enzo, MAS 5.0 statistical algorithm) and those for the 2.0 Platform are in blue (HG-U133 Plus 2.0, IVT Labeling Kit, PLIER). Note the lower variance on the 2.0 Platform estimates. The first probe set, 209098_s_at, is an example of a probe set which behaves in essentially the same manner on both platforms. While the majority of the variation is due to tissue as opposed to platform differences, some probe sets exhibit an overall shift between platforms. Note how the shift can be in different directions for different probe sets as exemplified by probe sets 209657_s_at and 219986_s_at.



ANOVA is a sensitive approach to breaking down the observed variation into its source components. Behavior of some representative probe sets can be seen in Figure 9. Note how differences between platforms are not necessarily in the same direction from one probe set to another.

ADDITIONAL PLATFORM COMPARISONS

Comparison analysis between two unrelated tissues is another method of assessing signal changes between two platforms. Figure 10 shows the signal log ratio for the common probe sets plotted as a scatter plot, with the 2.0 Platform on the x axis and the previous generation platform on the y axis. Panel A shows the result of all probe sets included, while Panel B shows only those probe sets that are called Present in both tissues. Low correlation is observed in Panel A due to the inclusion of probe sets giving Absent calls that have inherently lower signals and greater variability with respect to signal log ratios. By filtering for probe sets with Present calls, the correlation is stronger between the two platforms (R^2 = 0.93). Among the probe sets with higher intensities (Present calls), signal changes are conserved with respect to direction (increase or decrease) and value (fold change).

SIGNAL ROC CURVES EMPLOYING THE PLIER ALGORITHM

The PLIER algorithm utilizes probe affinity and error modeling in combination with multiple array analyses to provide a more sensitive means of detection for low-abundance transcripts. An example of the improved performance made possible by applying PLIER to data is shown in Figure 11. In the left panel, signal data were generated from 18-µm arrays using the Affymetrix® Microarray Suite 5.0 statistical algorithm. On the right panel, signals were produced from 11-µm arrays using PLIER.

Figure 10. Correlation of differential fold change calculations (log base 2) across the two platforms for a brain versus skeletal muscle comparison (previous generation platform: HG-U133A, 2.0 Platform: HG-U133 Plus 2.0) **A.** When all 21,723 non-control probe sets are examined, results show a moderately low agreement of the magnitudes of fold changes between the platforms. **B.** However, when data are filtered for present calls only in both tissues on both platforms (7,015 probe sets), a much higher correlation is observed (R^2= 0.93).

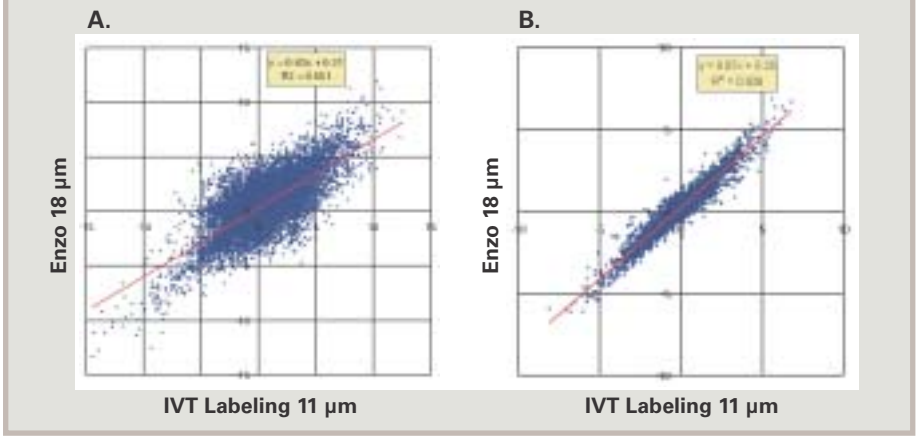
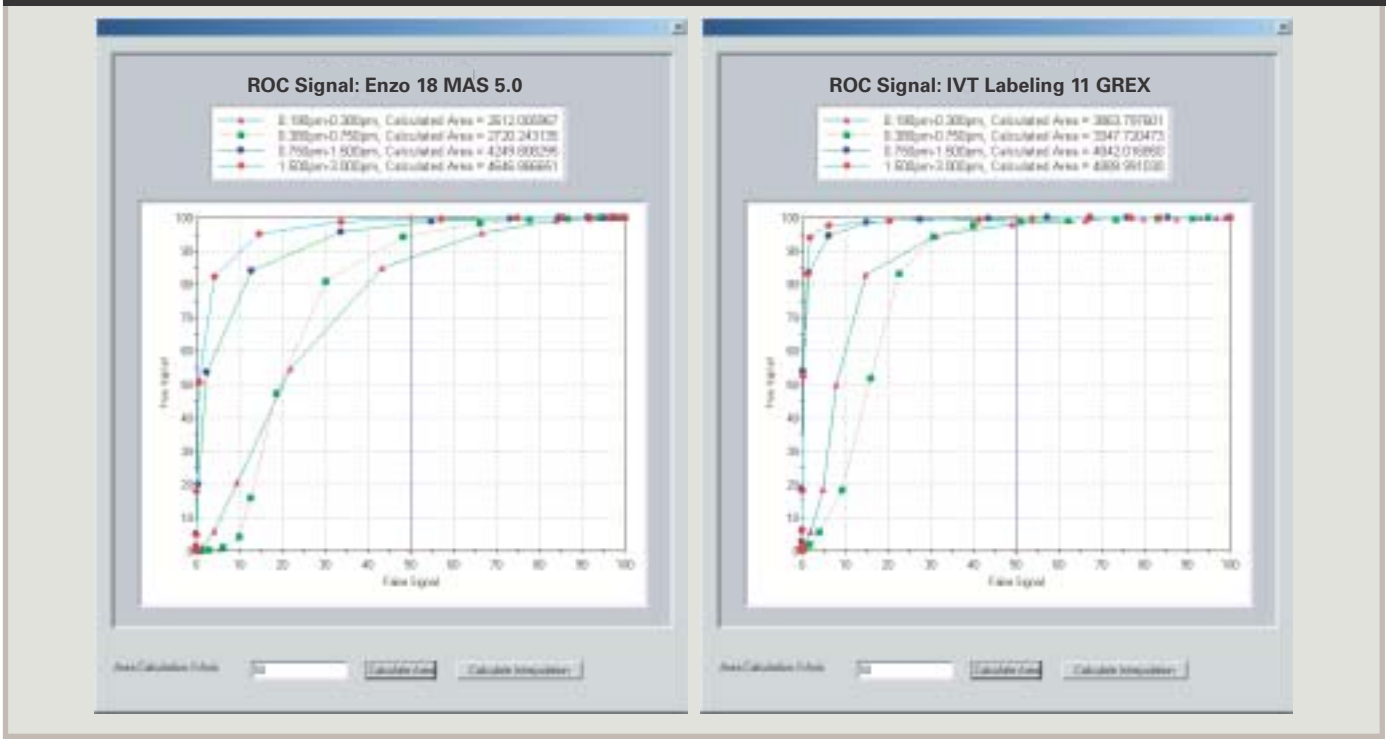


Figure 11. Signal ROC curves generated with Microarray Suite 5.0 statistical algorithm or PLIER Signal Intensities. Signal ROC curves were generated based on data from the Latin Square spike-in experiments. (Left Panel) Previous generation platform (Mouse 430A) .CEL data analyzed with Microarray Suite 5.0 statistical algorithm. (Right Panel) 2.0 Platform (Mouse 430 2.0) .CEL file data were analyzed using the PM-MM, no-quantile normalization PLIER algorithm. Signal ROC curves were plotted as a function of comparing signal differences at spike concentrations that differed by two-fold. Calculated Area = area under the curve, where the higher value represents the assay with the greater sensitivity and specificity.



Both systems were tested by evaluating signal differences between two different spike concentrations in a ROC curve analysis.

As noted in Figure 7, which shows Detection Call ROC data, the greater the sensitivity and specificity of the assay, the closer the curve approaches the upper left-hand corner of the graph. For all the concentrations evaluated (1.5 vs. 3.0 pM, 0.75 vs. 1.5 pM, 0.38 vs. 0.75 pM, and 0.19 vs. 0.38 pM), the 2.0 Platform data demonstrated greater sensitivity and specificity. When considered with the Figure 7 data, the new platform demonstrates improved detection of low-end expressors for both call or signal data.

Conclusion

This Technical Note highlights the changes and improvements that form the new 2.0 Platform. A reduction in feature size allows for complete representation of the known transcriptome on a single array for human, mouse, and rat expression studies. In addition to enabling genome-wide surveys of expression with a single hybridization, the new 2.0 Platform includes improved low-end sensitivity.

For a significant portion of probe sets, signals are not identical between the platforms when using statistical tests such as ANOVA and t-tests. Users with data sets generated on the previous platform need to be aware of these differences and, conservatively, data from the two platforms should not be compared. For customers initiating studies, the improved assay sensitivity and specificity, along with the increased number of data points per hybridization, provide a number of advantages to using the new platform.

For more details on the information provided in this Technical Note, please refer to:

- Technical Note: *Design and Performance of the GeneChip® Human Genome U133 Plus 2.0 and Human Genome U133A 2.0 Arrays*
- Technical Note: *The Affymetrix GeneChip® Scanner 3000: New Advances in Scanner Design for Superior Performance, Reliability, and Dynamic Range*
- Data Sheet: *AutoLoader for the GeneChip® Scanner 3000*
- Technical Note: *Performance Evaluation of the GeneChip® Fluidics Station 450*
- Technical Note: *Array Design for the GeneChip® Human Genome U133 Set*

AFFYMETRIX, INC.

3380 Central Expressway
Santa Clara, CA 95051 USA
Tel: 1-888-DNA-CHIP (1-888-362-2447)
Fax: 1-408-731-5441
sales@affymetrix.com
support@affymetrix.com

AFFYMETRIX UK Ltd

Voyager, Mercury Park,
Wycombe Lane, Wooburn Green,
High Wycombe HP10 0HH
United Kingdom
Tel: +44 (0) 1628 552550
Fax: +44 (0) 1628 552585
saleseurope@affymetrix.com
supporteurope@affymetrix.com

AFFYMETRIX JAPAN K.K.

Mita NN Bldg., 16 F
4-1-23 Shiba, Minato-ku,
Tokyo 108-0014 Japan
Tel: +81-(0)3-5730-8200
Fax: +81-(0)3-5730-8201
salesjapan@affymetrix.com
supportjapan@affymetrix.com

www.affymetrix.com Please visit our web site for international distributor contact information.

For research use only.
Not for use in diagnostic procedures.

Part No. 701632 Rev. 2