APPLICATION FOCUS:

Whole-transcript Expression Analysis

Gene Expression

Historically, microarrays have interrogated the few hundred bases proximal to the 3′ end of each gene, and used expression at the 3′ end to approximate expression of the entire gene. This approach is compatible with the 3′ oligo(dT)-based priming and labeling assays and provides valuable insight into global gene expression. However, this approach assumes that the 3′ end of each gene is clearly defined, that each transcript has an intact poly-A tail and that the entire length of the gene is expressed as a single unit.

These assumptions, however, do not apply to all genes or all samples. More than 60 percent of genes are known to be alternatively spliced[1,2,3,4], yielding hundreds of thousands of transcript variants with potentially distinct functions. As many as 50 percent of disease-related point mutations may result in splice pattern changes[5], and 20 percent of cancer-causing mutations can result in exon-skipping events.

Unfortunately, classical 3′ expression microarrays do not discriminate between alternatively spliced transcripts that have identical 3′ ends. Transcripts lacking a 3′ exon because of alternative splicing, non-polyadenylation, genomic deletions or other non-canonical genomic events are not detected in 3′ based expression experiments.

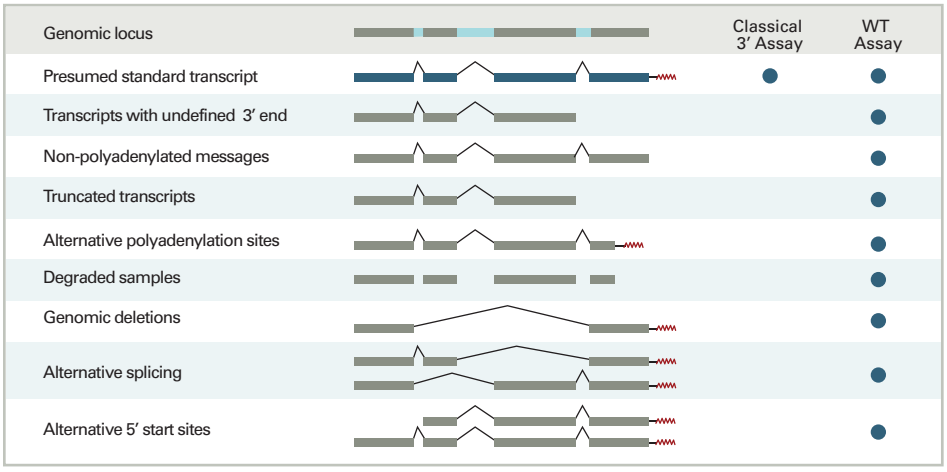

Figure 1: Types of transcripts captured by a whole-transcript assay. Most of these cannot be detected with the classical 3′ assay.

TABLE OF CONTENTS

APPLICATION FOCUS

This Application Note describes how researchers are using Affymetrix' new whole-transcript expression profiling tools to characterize disease etiology and molecular mechanisms with a new level of resolution and accuracy. By utilizing the GeneChip® Whole Transcript (WT) Assay and Affymetrix' high-density microarrays to explore diseases like colon and brain cancer, these experiments have already revealed new mechanistic pathways, suggested potential treatment targets and identified new splicing regulatory mechanisms.

Affymetrix' whole-transcript analysis approach enables researchers to detect not only the level of expression, but also precisely what is being expressed, including alternative isoforms or genomic deletions. This has opened the door to new insights at a resolution not possible with the classical 3′-based microarrays.

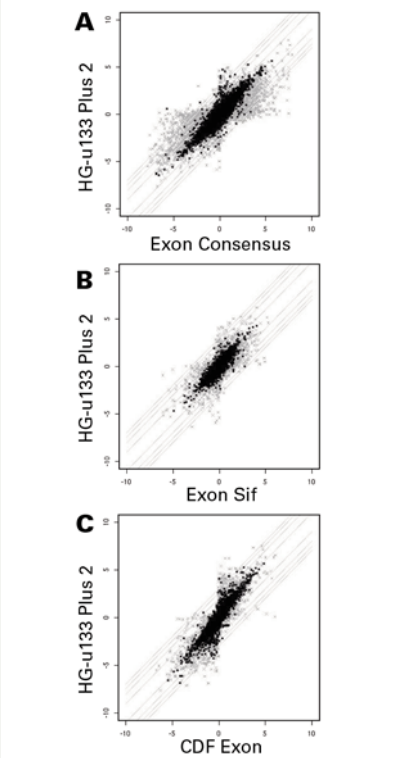

Figure 2: Scatterplots showing high correspondence between Affymetrix exon and standard expression arrays (Okoniewski M., *et al.*[6]).

"Since the classical microarrays have already been repeatedly validated experimentally, this provides strong evidence that exon arrays are also reliable, not only for probesets that can be successfully mapped to the existing arrays, but also for the many thousands of additional probesets that provide a more detailed coverage of the transcriptome."
— Okoniewski M., *et al.*[6]

The combination of Affymetrix' Whole Transcript (WT) Assay and high-density arrays, including GeneChip® Exon 1.0 ST Arrays and GeneChip® Gene 1.0 ST Arrays, provides a more complete and more accurate picture of overall gene expression, enabling researchers to detect transcript isoforms they didn't know existed, and to discriminate between transcripts that previously appeared to look the same (Figure 1).

Affymetrix' Complete Product Solution for High-definition Expression Profiling

Affymetrix provides a complete product solution, standardized procedures, and world-class technical support to help researchers easily and quickly obtain biologically significant results in their research. The complete high-definition expression profiling system includes:

- High-density microarrays
- Conveniently packaged target amplification and labeling, array processing and various process control reagents
- Automated array processing with the Fluidics Station
- High-resolution scanner
- Basic analysis software
- Freely accessible online probe sequence and annotation resource
- Genome context design and array results viewer
- Complementary third-party statistical and pathway analysis tools

Human, Mouse and Rat Exon 1.0 ST Arrays

With approximately four probes per exon and roughly 40 probes per gene, GeneChip® Exon 1.0 ST Arrays provide data for two complementary levels of expression analysis in a single experiment—"exon-level" and "gene-level" analysis.

Exon-level analysis of multiple probes per exon provides the highest resolution microarray analysis, with the ability to analyze alternative splicing and differential expression of each exon within a gene. On a whole-genome level, this enables researchers to detect not only the level of expression, but also precisely what is being expressed. These differences may play a critical role in disease susceptibility and etiology.

In gene-level expression analysis, multiple probes on different exons are summarized into a single expression-level data point that represents all transcripts derived from each gene.

Exon arrays provide the most comprehensive coverage of all microarray designs, including empirically supported and predicted transcripts. The high level of coverage maximizes researchers' ability to identify known and novel splicing events and make groundbreaking discoveries. Options are available to restrict analysis to subclasses of sequences, for a faster preview of the biology.

In addition, exon array probe design is based on genomic sequence, not UniGene clusters, so sequence annotations can be updated easily with each new genome build. With the annotations of each probe on the array anchored to the genome, the design makes it easy to integrate various types of genomic information—SNPs, mutations, chromosome deletions and amplifications—by referring to the same genomic structure. This enables rapid scientific advances from a systems biology perspective.

GeneChip® Human Gene 1.0 ST Array

The Human Gene 1.0 ST Array, supported by the same whole-transcript assay as the exon array, is the latest addition to the family of Affymetrix' expression arrays offering whole-transcript coverage. Each of the well-annotated genes is represented by approximately 26 probes on the array spread across the full length of the gene, providing a more complete and more accurate picture of gene expression than 3'-based expression array designs. The Human Gene 1.0 ST Array utilizes a subset of the probes selected from the Exon 1.0 ST Array and focuses on well-annotated content at the gene level.

The Human Gene 1.0 ST Array—the most advanced and cost-effective gene expression profiling option for new microarray users—allows researchers to integrate microarray-based gene expression profiling more routinely in their research.

Whole-transcript Array Results Validated by 3' Expression Microarrays

Recent studies demonstrate that results from high-resolution, whole-transcript Affymetrix arrays are consistent with those of Affymetrix' proven 3' gene expression microarrays. Okoniewski, *et al.*[6] recently compared the gene expression profiles of two established cell lines on the Exon 1.0 ST Array and classical 3'-based design, the GeneChip® Human Genome U133 Plus 2.0 Array. Using three mapping techniques, the two arrays showed a high degree of correspondence in terms of fold changes (Figure 2).

EXON-LEVEL ANALYSIS IN PRACTICE

In addition to more accurate gene expression analysis, exon arrays provide the unique opportunity to unveil changes that occur along the entire length of the gene. Researchers are already detecting specific alterations in exon utilization that may play a critical role in disease mechanism and etiology.

Uncover Aberrant Splicing Events Linked to Disease State

Using Exon 1.0 ST Arrays, researchers at Millennium Pharmaceuticals have confirmed that specific "cassette" exons in the CD44 gene are highly expressed in primary tumor colon cancer cell lines, but not in metastatic colon cancer or HeLa cell lines. The team was able to monitor different exons independently because the array design covers the entire CD44 transcript, including the central part of the gene where these variant cassette exons are located.

Their results suggest that CD44 splice variants might be used as a diagnostic or prognostic marker for colon cancer.

Figure 3 is a schematic representation of the Exon 1.0 ST Array's coverage of the CD44 locus, created with the Affymetrix Integrated Genome Browser (IGB). It shows different known RefSeq splice variants of CD44 in green. Exon 1.0 ST Array cov-

PREVALENCE OF ALTERNATIVE SPLICING EVENTS

"A number of studies by different groups all reported finding alternative splice forms in a surprisingly large fraction of human genes, ranging from 40 percent to 60 percent."
— Lee C. and Roy M.[1]

"Adding to previous studies, the results indicate that at least 74 percent of human multi-exon genes are alternatively spliced." — Johnson, *et al.*[2]

IMPORTANCE OF ALTERNATIVE SPLICING TO DISEASE RESEARCH

"Recent studies indicate that 70 to 88 percent of alternative splices change the protein product. The majority of these changes appear to be functionally interesting."
— Modrek B. and Lee C. J.[4]

erage of the CD44 gene is shown in blue and HG-U133 probe sets are shown in pink. The Exon 1.0 ST Array provides broad coverage of the entire CD44 transcript, including, but not restricted to the established RefSeq exons. This coverage enables researchers to monitor differential exon expression and uncover previously unidentified novel events.

Additionally, Gardina, *et al.*[7] demonstrated the ability of the Human Exon 1.0 ST Array to detect alternative splicing and differential gene expression in colon cancer samples compared with matched normal control tissues.

Their results suggest that aberrant splicing might be the mechanism of action for colon cancer, after finding several aberrant splicing events affecting the same functional pathways of cell architecture and the extracellular matrix. These results correlate extremely well with known cancer genes, pathways and many of the splicing events identified by the microarray data were subsequently verified by RT-PCR.

Discover Causative Genes of a Disease by Identifying Unique Exon-skipping Events

Drs. Pim French and Justine Peeters of Erasmus Medical Center in the Netherlands discussed their work using the Human Exon 1.0 ST Array to classify glial tumors in the Summer 2006 issue of the *Affymetrix Microarray Bulletin*[8].

French and Peeters used exon arrays to discover expression profiles with distinct splice variants that can be used to distinguish glioblastomas from oligodendrogliomas and help clinicians more accurately diagnose the multiple classes and variable prognoses of brain cancer.

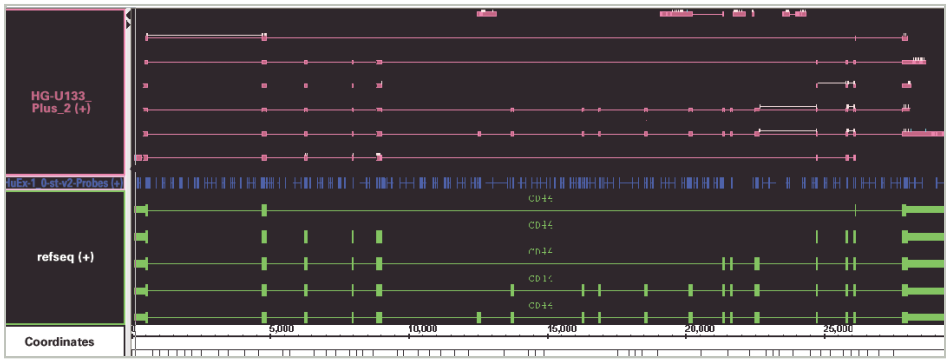

Figure 3: The Exon 1.0 ST Array provides coverage of the entire CD44 gene. Known RefSeq variants of CD44 are shown in green. Exon 1.0 ST Array coverage of the CD44 gene is shown in blue and HG-U133 probe sets are shown in pink.

The Erasmus team also used exon arrays to systematically survey the genome and compare different tumor samples to identify novel exon-skipping events and associated genes in individual patients. French and Peeters estimated that approximately 20 percent of mutations in glial tumors cause exon skipping; aberrant transcripts affected by exon skipping are potentially cancer-causing. They confirmed the findings from the exon arrays by sequencing regions surrounding the genes in which exon skipping occurred. The team identified mutations in those regions that could contribute to the exon-skipping phenotype, leading to discovery of novel potential targets for glial cancer therapies.

Uncover Splicing Regulatory Mechanisms

A recent proof-of-concept study by Bruno, *et al.*[9] at The University of Texas M. D. Anderson Cancer Center used exon arrays to examine exon skipping in the FGFR gene in glioma cell lines.

The team hypothesized that binding of the splicing inhibitory factor, PTB, to intronic splicing regulatory sequences within the FGFR gene, caused skipping of exon 3 in glioma cells. This hypothe-

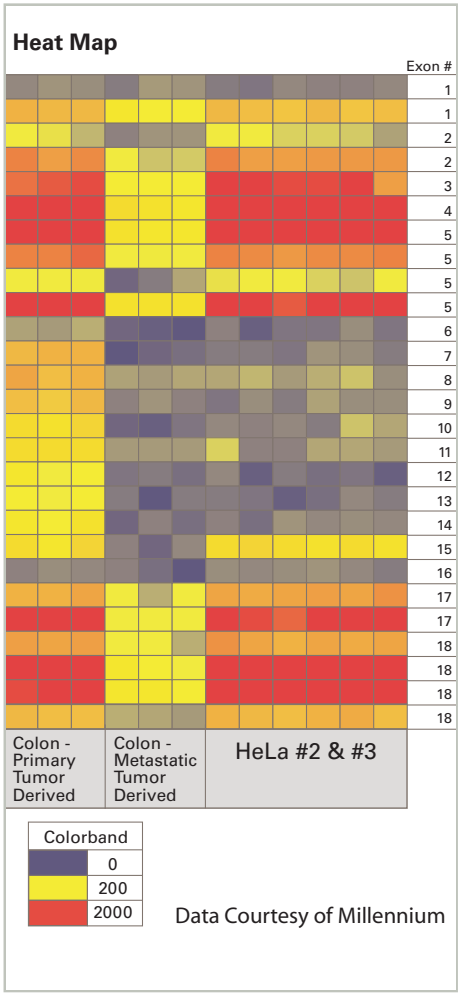

Figure 4: This heatmap image from Millennium scientists illustrating differential expression of CD44 exons in metastatic tumor versus primary tumor versus control HeLa cells.

sis was supported by evidence that PTB was over-expressed in glioma cells.

To directly test the hypothesis, the M. D. Anderson team treated human glioblastoma cells with antisense oligonucleotides targeting the intronic splicing regulatory sequences within the FGFR gene, to compete for binding with the PTB factor. If PTB contributes to the skipping event, then the antisense treatment should reverse the phenotype and trigger increased inclusion of exon 3. They evaluated whether exon arrays could be used to detect a change in exon 3 inclusion. Statistical analysis of the exon array data correctly identified the differential exon skipping of the FGFR gene with enhanced exon 3 inclusion as a result of antisense oligo treatment (Figure 5).

Associate Splicing Patterns with Inherited SNPs

Dr. Jacek Majewski of McGill University and Genome Québec Innovation Centre discussed the detection of known and novel cases of alternative splicing using Affymetrix exon arrays during an October 24, 2006 *Affymetrix Microarray Bulletin* Symposia conference call[10].

His group ran the CEPH samples from the International HapMap Project on the Affymetrix Exon 1.0 ST Array and demonstrated several differences in splicing patterns supported by linkage and/or association analyses, suggesting that they have underlying genetic causes.

Majewski's team also investigated how exon expression profiles segregated according to HapMap families. In several cases, they found SNP changes that resulted in exon skipping and alternative splicing.

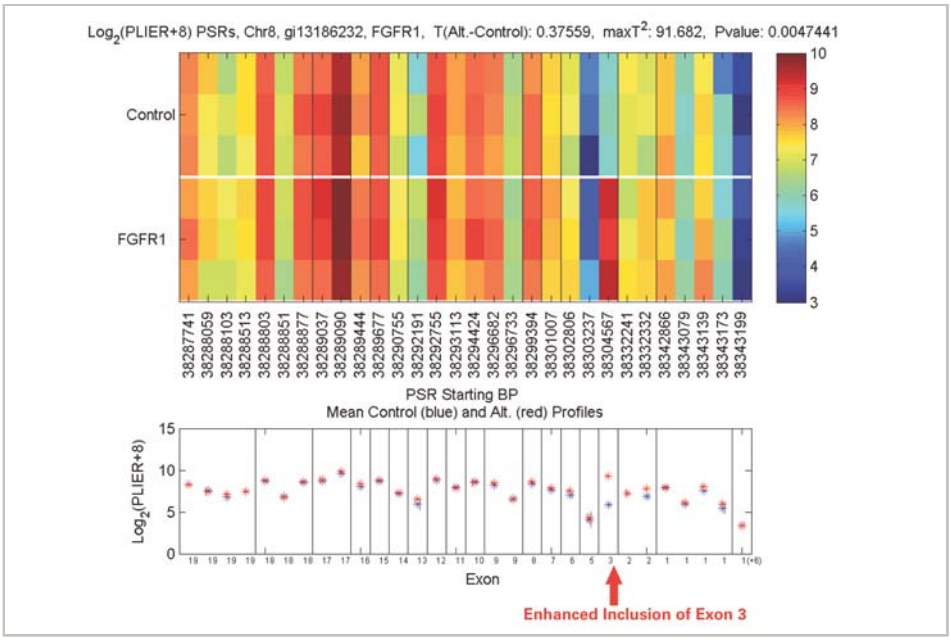

Figure 5: Heatmap image showing enhanced inclusion of exon 3 in FGFR alleles8.

AFFYMETRIX HIGH-DEFINITION EXPRESSION PROFILING TOOLBOX

Assay and Reagent Solutions

Affymetrix provides a single source for highly reproducible, robust and consistent reagents. The complete system solution for Exon 1.0 ST and Gene 1.0 ST Arrays includes WT assay reagent kits for sample amplification and labeling. Random-priming and linear amplification combined with a novel, robust fragmentation and labeling strategy enables researchers to generate targets along the entire length of the transcript.

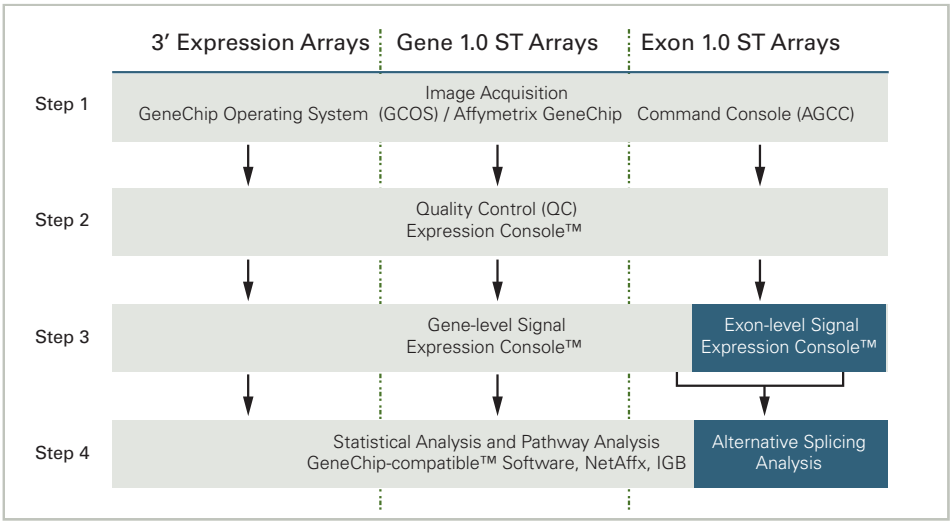

Figure 6: The data analysis workflow for Gene 1.0 ST and Exon 1.0 ST Arrays is similar to that of 3′ expression arrays, utilizing GCOS or Affymetrix GeneChip Command Console (AGCC), Expression Console, GeneChip-compatible software, the NetAffx Analysis Center and IGB.

The 1 µg total RNA labeling protocol, which includes an initial rRNA removal procedure for optimal sensitivity, is required on exon arrays to obtain gene expression and alternative splicing information. The 100 ng total RNA labeling protocol—which omits the rRNA reduction step, offering the same high-level performance on the gene array with reduced hands-on time—is recommended for use with gene arrays.

Affymetrix WT Reagents include:

- PolyA Control Kit
- WT cDNA Synthesis and Amplification Kit
- WT Terminal Labeling Kit
- Sample Cleanup Module
- IVT cRNA Cleanup Kit
- Hybridization Controls
- Hybridization, Wash and Stain Kit

DATA ANALYSIS SOLUTIONS

Gene-level Analysis

Identifying and prioritizing gene-level expression changes using Exon 1.0 ST and Gene 1.0 ST Arrays is as simple as 3' expression analysis, and the workflows are similar. The GeneChip® Operating System (GCOS) or Affymetrix GeneChip® Command Console (AGCC) is used for initial image acquisition. Affymetrix Expression Console™ software then provides an easy-to-use analysis workflow for quality control (QC) and probe set summarization to attain gene-level signal data.

Affymetrix partners with a number of GeneChip®-compatible™ software providers, who provide statistical analysis solutions for generating lists of differentially regulated genes. Additionally, multiple software providers offer solutions for pathway analysis—building and visualizing potential gene interactions by leveraging databases of published literature.

The NetAffx™ Analysis Center, freely available from Affymetrix, enables researchers to correlate these results with array design and annotation information. The Integrated Genome Browser (IGB) provides visualization tools to further explore genomes and corresponding annotations from multiple data sources.

Exon-level Analysis

Expression Console also enables researchers to compute exon-level signal estimates for exon array data. To predict alternative transcript forms, GeneChip-compatible software that supports the analysis of Exon 1.0 ST Array data (Table 1) can be used to conduct statistical analysis of gene-level and exon-level sig-

	Biotique's Xray	Genomatix' ChipInspector	JMP Genomics	Partek® Genomics Suite™
Supports analysis of GeneChip Human Gene 1.0 ST Array	Yes	Yes	Yes	Yes
Supports analysis of GeneChip Exon Arrays	Yes	Yes	Yes	Yes
Supports analysis for 3' expression arrays	No	Yes	Yes	Yes
Other GeneChip-compatible offerings	—	Regulation (ChIP-on-chip)	SNP Analysis	• Tiling (Regulation, ChIP-on-chip) • SNP Analysis • Chromosomal Copy Number & LOH Analysis
Application Highlights	• Fast MS Excel add-in offers familiar Excel interface • Rigorous normalization and analysis for unlimited numbers of CEL files on computers with 500 MB ram • Dynamically generated methods, results and array quality document suitable for publications	• Analysis based on single probes • Single probe curation based on latest genome annotation • Currently 37 arrays (12 genomes) supported, including promoter tiling arrays • Tight integration with other downstream Genomatix tools to derive greater biological insight	• Powerful SAS analytics enhanced by interactive JMP graphics • Sophisticated, automated experimental design tools • Quality control (QC) tools including batch effect removal • Extensive predictive modeling capabilities	• Fast and memory-efficient • Comprehensive statistics and interactive visualization • Removes batch effects • Easy-to-use workflows for exon, copy number and tiling arrays • Integration of results from multiple GeneChip technologies such as expression, regulation and copy number • Build and validate diagnostic and prognostic classifiers
Website	www.orderXRAY.com	www.genomatix.de	www.jmp.com/genomics	www.partek.com

Table 1: GeneChip-compatible products supporting both gene expression and alternative splicing analysis.

nal data. Pathway analysis, annotations from the NetAffx Analysis Center and IGB analysis of exon array data can all be conducted using workflows similar to 3' expression array data. For an introduction on managing data provided by exon arrays, see the Technical Note, *Identifying and Validating Alternative Splicing Events*, located at www.affymetrix.com/support.

Partners Providing GeneChip-compatible Products

Several Affymetrix partners provide GeneChip-compatible products for gene expression analysis, pathway analysis and alternative splicing analysis at the exon level.

Additional Solutions and Resources

In addition to arrays, reagents and assays, Affymetrix offers instruments for streamlined array processing, including the GeneChip® Scanner 3000 7G for array scanning and GeneChip Fluidics Station 450.

Affymetrix also provides a variety of supporting materials, which can be accessed via the Exon 1.0 ST and Gene 1.0 ST Array product pages at www.affymetrix.com. This toolbox includes:

- Publications and references lists
- Webinars focused on alternative splicing and expression analysis using exon arrays
- GeneChip-compatible software demonstrations
- NetAffx Analysis Center, Expression Console, and IGB analysis demonstrations
- Technical Notes
- *Affymetrix Microarray Bulletin* interviews with researchers using exon arrays

Gene-level Analysis Solutions	Pathway Analysis Solutions
Applied Maths' GeneMaths XT	Ariadne Genomics' Pathway Studio®
BioDiscovery's GeneSight	GeneGo's MetaCore
Genedata's Expressionist®	Genomatix' BiblioSphere Pathway Edition
Genomatix' ChipInspector	Ingenuity Systems' Ingenuity Pathways Analysis (IPA)
IMC's TeraGenomics	Stratagene's PathwayArchitect®
SAS' JMP® Microarray	
Ocimum Biosolutions' Genowiz	
Partek® Genomics Suite	
Rosetta Biosoftware's Rosetta Resolver® System	
Spotfire's DecisionSite® for Functional Genomics	
Spotfire's DecisionSite® for Microarray Analysis	
Stratagene's ArrayAssist®	
VizX Labs' GeneSifter®	

Table 2: GeneChip-compatible products supporting pathway analysis and gene-level statistical analysis.

REFERENCES:

1. Lee C. and Roy M. Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biology* **5**:231-234 (2004).

2. Johnson J. M., *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **296**:916-919 (2003).

3. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**:860-921 (2001).

4. Modrek B. and Lee C. J. A genomic view of alternative splicing. *Nature Genetics* **30**:13-19 (2002).

5. Lopez-Bigas N., *et al.* Are splicing mutations the most frequent cause of hereditary disease? *FEBS Letters* **579**:1900-1903 (2005).

6. Okoniewski M., *et al.* High Correspondence Between Affymetrix Exon and Standard Expression Arrays. *Biotechniques* **42**(2):181-185 (2007).

7. Gardina P. J., *et al.* Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7**(1):325. (2006).

8. Bruno I. G., Jin W., Cote G. J. Correction of aberrant FGFR1 alternative RNA splicing through targeting of intronic regulatory elements. Hum Mol Genet 13(20):2409-20 (2004). Epub 2004 Aug 27. Erratum in: *Hum Mol Genet* **13**(21):2725 (2004).

9. Discovery of Novel Splice Variations Improves Glial Tumor Classification. *Affymetrix Microarray Bulletin* (Summer 2006). Available online at http://www.microarraybulletin.com/community/article.php?p=226

10. Majewski, J. Using the Affymetrix Exon Array to Investigate Variation in Alternative Splicing in a Human Population. *Affymetrix Microarray Bulletin* Symposia (October 24, 2006). Available online at http://www.microarraybulletin.com/community/article.php?p=257

ORDERING INFORMATION

Affymetrix WT Assay and Expression Arrays

Products	Species	Configuration	Content	Part Number
GeneChip® Exon1.0 ST Arrays – high-definition single array for expression and alternative splicing	Human	Starter Pack	30 Arrays 30 reactions of reagents On-site training	900654
		30-pack	30 arrays	900651
		6-pack	6 arrays	900650
		2-pack	2 arrays	900649
	Mouse	Starter Pack	30 Arrays 30 reactions of reagents On-site training	900831
		30-pack	30 arrays	900819
		6-pack	6 arrays	900818
		2-pack	2 arrays	900817
	Rat	Starter Pack	30 Arrays 30 reactions of reagents On-site training	900848
		30-pack	30 arrays	900822
		6-pack	6 arrays	900821
		2-pack	2 arrays	900820
GeneChip® Human Gene 1.0 ST Arrays for gene expression analysis	Human	30-pack	30 arrays	901087
		6-pack	6 arrays	901086
		2-pack	2 arrays	901085
		30-pack	30 arrays 30 reactions of reagents	901146
		10-pack	10 arrays 10 reactions of reagents	901147
Reagents	WT Sense Target Labeling and Control Reagents	30 reactions		900652
	Hybridition, Wash and Stain Kit	30 reactions		90072

	Kit		
Affymetrix Analysis Solutions	Expression Console™	Generate primary array analysis and QC	Freely available at www.affymetrix.com
	NetAffx™ Analysis Center	Query functional annotations of array content and probe sequence	
	Integrated Genome Browser	Align array design and array data in the genomic context to integrate various genomic information	

AFFYMETRIX Inc. USA
3420 Central Expressway
Santa Clara, CA 95051 USA
Tel: 408-731-5000
(1-888-DNA-CHIP)
Fax: 408-731-5380
www.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.
Affymetrix Japan K.K.
Mita NN Bldg 16 Floor
4-1-23 Shiba
Minato-ku, Tokyo 108-0014
Tel: *81-(0)3-5730-8100
Fax: *81-(0)3-5730-8101
salesjapan@affymetrix.com
supportjapan@affymetrix.com
www.affymetrix.com

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

For research use only. Not for use in diagnostic procedures. P/N 702503-2