



Data Sheet

GeneChip® Exon Array System for Human, Mouse, and Rat

Genome-Wide Gene Expression and Alternative Splicing Profiling on a Single Array

The GeneChip® Exon 1.0 ST (sense target) Arrays combine a novel array design strategy with the highest density array manufacturing capability. For the first time, both gene-level and exon-level expression profiling can be performed on the whole-genome scale using a single array. The Exon Array system solution presents new dimensions for researchers to explore the human, mouse, and rat genome so they can:

- Perform the most robust gene expression profiling with the most advanced design and higher sensitivity than other microarray platforms.
- Investigate global expression of individual exons to uncover and document novel alternative splicing events.
- Obtain more direct comparative genomic information with the syntenic sequences incorporated during the design of the Human, Mouse, and Rat Exon Arrays.

Features of this integrated solution include:

- Highest resolution expression analysis interrogating approximately one million exon clusters within the known and predicted transcribed regions of each genome, therefore providing the highest level of coverage at the gene level with more than 30 distinct probes representing most genes.
- Tailored GeneChip® Whole Transcript (WT) Assay and Reagents utilizing a random priming method for generating sense targets along the entire length of the transcripts.
- Flexible data analysis solutions and comprehensive annotation providing users the opportunity to explore and dissect the data in various, user-defined workflows.

Introduction

The GeneChip® Exon Arrays are the first experimental tools available to survey both gene expression and alternative splicing patterns on the whole-genome scale on a single array.

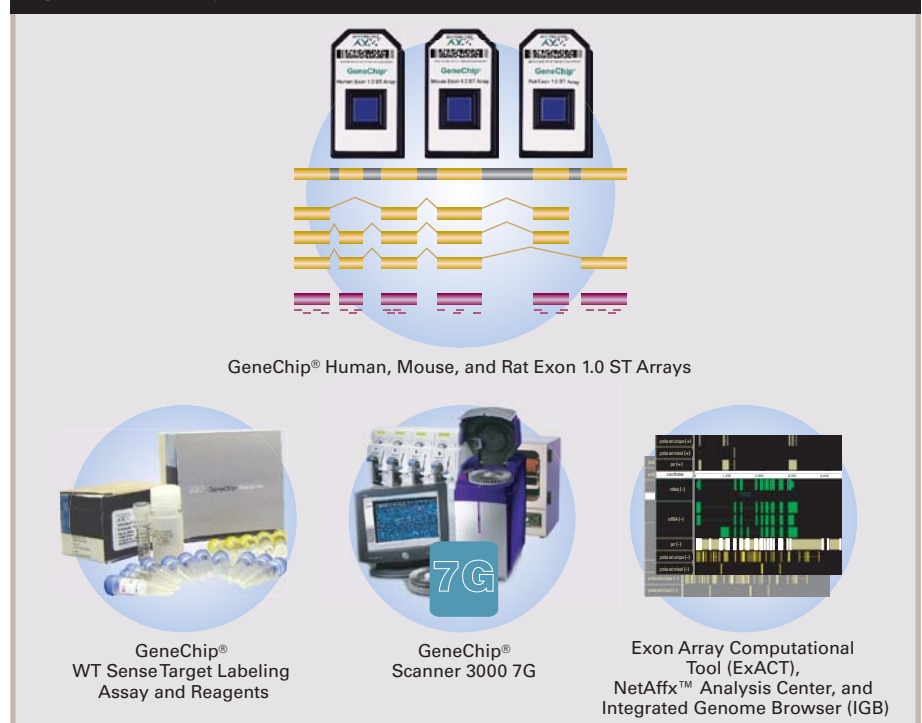
In addition to gene expression regulation, it is evident that alternative splicing plays a key role in biology:

- Tissue-specific and developmental stage-specific alternative splicing contributes to significant protein diversity.
- Disease-related deregulation of splicing may be critical in pathogenesis and contribute to disease diversity and complexity.

With over 5.5 million features on each array representing the main content, the Exon Arrays will deepen the understanding of biology for many discovery-focused applications including:

- Analysis of molecular mechanisms regulated by alternative splicing.
- Discovery of new splice variants of drug targets and mapping of their tissue-specific expression.
- Improved understanding of the downstream effects of genetic variations that may result in phenotypic changes in gene expression or alternative splicing patterns.

Figure 1. GeneChip® Human, Mouse, and Rat Exon 1.0 ST Array system solution.



Over 30 Probes for Most Genes for Most Robust Gene-Level Expression Profiling

The GeneChip Exon Arrays offer unique new insights for gene-level expression analysis:

- With probes selected across the entire length of the transcript, the Exon Arrays provide robust gene-level analysis that is not biased to the 3' end.
- With the highest density of coverage for most genes compared with all other microarrays, the Exon Arrays offer more sensitive measurement of gene expression quantitation, and more statistically robust results resistant to variability associated with probe performance, individual SNPs, alternative splicing events and array artifacts.

Exon-Level Profiling for New Dimensions in Expression Analysis

Using this array, expression profiles can also be studied at the exon level related to many of the alternative splicing events.

COMBINING EMPIRICAL SEQUENCES AND GENE PREDICTIONS FOR NOVEL DISCOVERIES

With the emergence of new information, it has been demonstrated that biologically significant transcription and alternative splicing events may not have been fully captured by the well-annotated sequences.

The Exon 1.0 ST Arrays support the most detailed survey of the human, mouse, and rat genomes and expand the discovery of novel transcript structures beyond the well characterized and annotated genes. Affymetrix took an all-inclusive design approach and the Exon 1.0 ST Arrays contain sequences from two primary sources:

- cDNA-based content including the more established human RefSeq mRNAs, GenBank® mRNAs, and ESTs from dbEST. Additional annotations were

created by mapping syntenic cDNAs to the human, mouse, and rat genomes using genome synteny maps from the UCSC Genome Bioinformatics group.

- Predicted gene structure sequences from:
 - GENSCAN
 - Ensembl
 - Vega
 - geneid and sgp
 - TWINSKAN
 - Exoniphy
 - microRNA Registry
 - MITOMAP
 - structural RNA predictions

ONE PROBE SET FOR EACH EXON FOR COMPLETE COVERAGE

As shown in Figure 2, most of the exons over 25 bp in length, along the whole transcript, are represented on the array with at least one probe set. The majority of the probe sets are comprised of four Perfect Match probes for each probe set to provide robust results.

ANCHORING THE DESIGN ON THE GENOME TO CORRELATE EXPRESSION DATA WITH GENETIC INFORMATION

Each probe on the array is designed from genomic sequence. Each probe is also anno-

tated with its genome coordinates and the references to its relationship to the associated exon, exon cluster, and transcript cluster. Projecting the design onto the genome facilitates the dynamic update of the design, reflecting the new knowledge of the genomic sequence and annotations.

In addition, this design and annotation approach enables more convenient correlation of the expression array data with DNA sequence information, and translated protein products. This makes it easier to dissect the mechanisms involved in transcription and alternative splicing regulation.

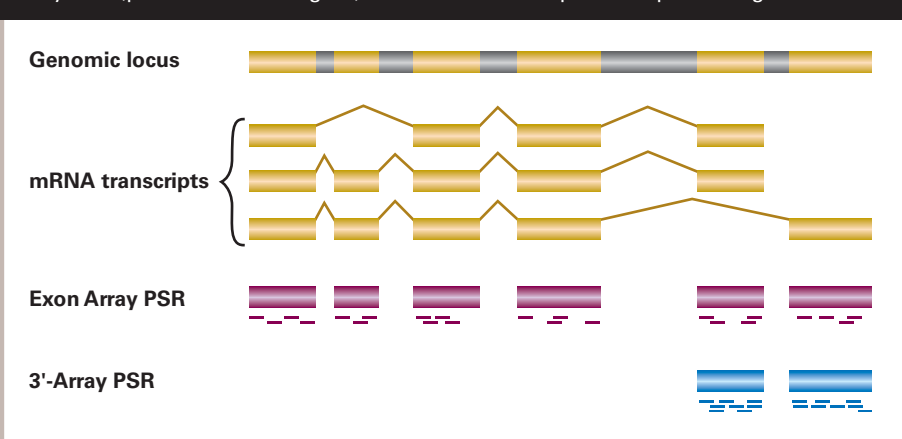
ONE MILLION EXONS ON A SINGLE ARRAY

Two key technological advancements have enabled the exon-level expression profiling providing cost and sample savings:

- Highest density array manufacturing at 5 µm feature size.
- More advanced and efficient background subtraction method utilizing surrogate mismatch probes*.

*For more information on the surrogate mismatch probes, see the *GeneChip® Exon Array Design Technical Note*.

Figure 2. Schematic for coverage of probe sets across the entire length of the transcript. Golden regions are exons whereas the grey regions represent introns that are removed during splicing. The short dashes underneath the exon regions for the Exon Array and 3'-Array PSR (probe selection region) indicate individual probes representing that PSR.



GeneChip® Whole Transcript Sense Target Labeling Assay and Reagents for Optimal Sensitivity and Specificity

OPTIMIZED REAGENT SOLUTION FOR CONFIDENCE AND CONVENIENCE

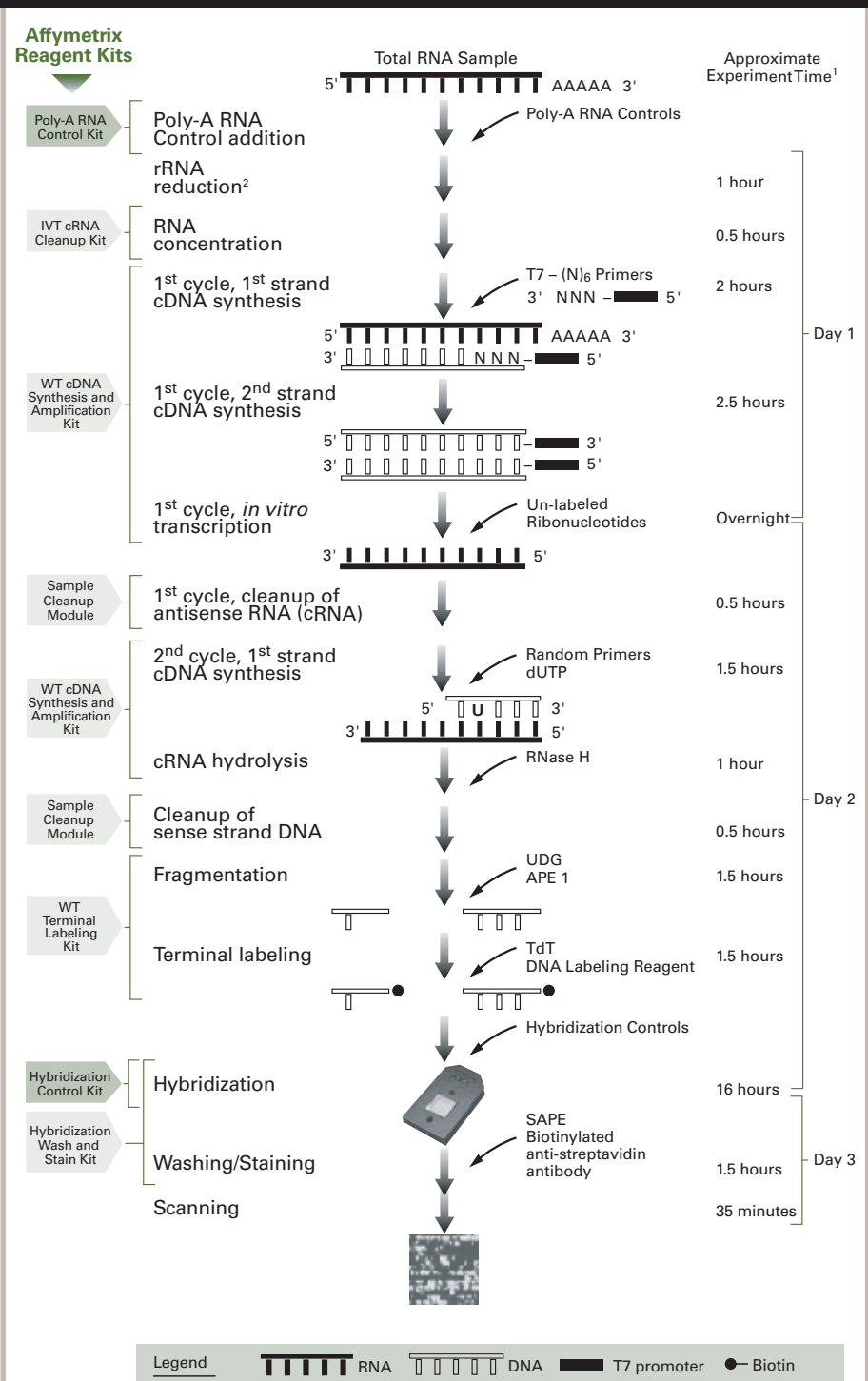
As a critical part of the integrated system solution, complete reagent kits have been developed, validated, and qualified specifically for the GeneChip Exon 1.0 ST Arrays. These reagent kits provide convenience as well as high-quality results and consistent data.

GENERATING SENSE DNA TARGETS ALONG THE ENTIRE LENGTH OF THE TRANSCRIPT

Using a random priming strategy, in combination with *in vitro* transcription-based linear amplification and a novel end-point fragmentation and labeling assay scheme, the whole transcript (WT) assay provides a robust method for target labeling.

As a standard recommendation, 1 µg of total RNA is used as starting material following the assay described in Figure 3. Alternative procedures bypassing the initial rRNA reduction step may be carried out by users with limited sample quantity starting with 100 ng of total RNA.

Figure 3. Schematic for WT Sense Target Labeling Assay for 1 µg of total RNA.



¹Assume that only one sample is carried through the assay. The time may be longer if multiple samples are processed simultaneously.

²RiboMinus Human/Mouse Transcriptome Isolation Kit is required for this step, and needs to be ordered directly from Invitrogen.

Flexible Data Analysis Solutions Supporting Customizable Workflows

ADDED DIMENSIONS OF ANALYSIS FOR NEW INSIGHTS

New algorithms have been developed specifically to support Exon Array data analysis for:

- Providing the exon-level expression metrics including signal and feature residuals.
- Providing similar metrics for gene-level expression by leveraging the multiple probe sets tiled for most genes.

In addition, new methods have been developed to integrate the exon-level and gene-level signal information for detecting gene loci that have a high probability of exhibiting differential alternative splicing.

A sample workflow for Exon Array data analysis is shown in Figure 4.

EXON ARRAY COMPUTATIONAL TOOL (ExACT) FOR FLEXIBLE ANALYSIS NEEDS

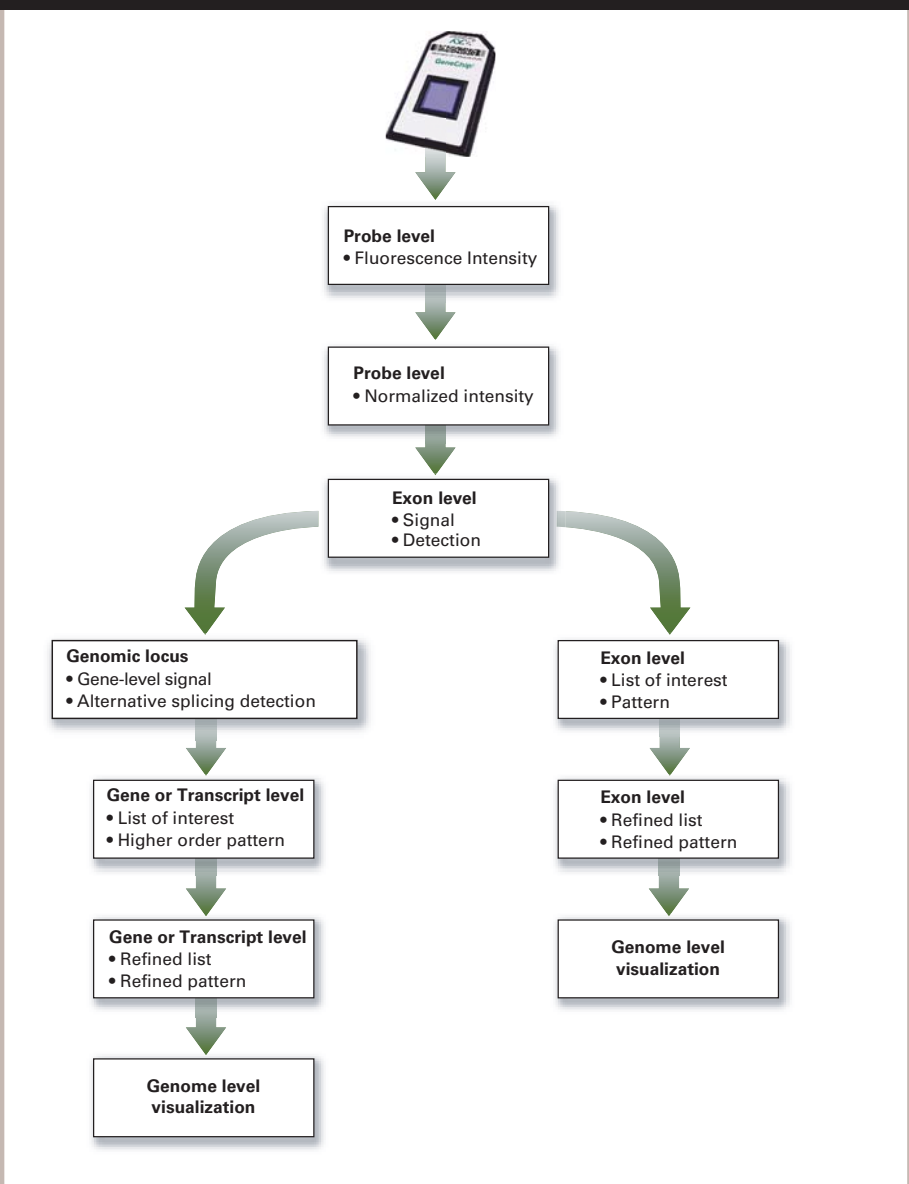
With approximately one million data points from a single array, the key challenge scientists face is how to apply various filters to quickly narrow down their search, and rapidly focus on the loci or exons of interest.

Incorporating the new algorithms for exon-level and gene-level signal analysis, ExACT provides an architecture that maximizes flexibility. Therefore, users can customize their analysis by filtering the array data for further analysis based on annotations, such as supporting evidence type, supporting sequence depth, chromosomal location, etc.

Furthermore, the packages support custom input. This helps further refine the data mining process, incorporating prior knowledge into the analysis, and rapidly turning array results into biologically significant interpretations.

The output files from ExACT can be imported to other commercially available advanced statistical analysis tools for secondary analysis. The basic workflow in ExACT is shown in Figure 5.

Figure 4. Exon-level array data analysis workflow.



ExACT is freely downloadable from www.affymetrix.com. Affymetrix is planning to maintain and update ExACT to incorporate feedback and input from users.

VISUALIZATION AND ANNOTATION THROUGH IGB AND THE NETAFFX™ ANALYSIS CENTER
To seamlessly incorporate various biological information and annotations as part of the data analysis workflow, the unique

web-based Affymetrix® NetAffx™ Analysis Center offers a summary of the array design information, including probe sequence and supporting evidence, as well as functional annotations at the gene level. All the information can be queried and searched. Selected subsets of the genes/probe sets can be obtained in an output file suitable for input into ExACT for further array data analysis (Figure 6).

Visualization of the Exon Array design and experimental results in the context of the genome is available through IGB with dynamic links between IGB and the NetAffx Analysis Center. These tools in combination facilitate faster understanding of the complex biological systems and enable users to obtain a complete picture of genomic information.

ACCESS TO ExACT SOURCE CODE SUPPORTS CUSTOMIZED SOLUTION

The Affymetrix® Developers' Network (ADN) is a forum for Affymetrix customers and third-party vendors whose goals include building, customizing, and integrating software applications and systems with the Affymetrix software platform. Affymetrix will continue to support this effort and provide Software Developer Kits with source code accessible to customers with unique interest in developing their own tools.

THIRD PARTY ALGORITHMS AND SOFTWARE SOLUTIONS TO COMPLEMENT THE AFFYMETRIX TOOLS

Affymetrix is working proactively with the academic research community to facilitate and catalyze efforts in further developing new algorithms and novel methods to effectively analyze new types of array data. With the active participation of the community, more refined data analysis solutions and practices will more quickly emerge, resulting in more streamlined and mature ways to turn array results into biological meaning.

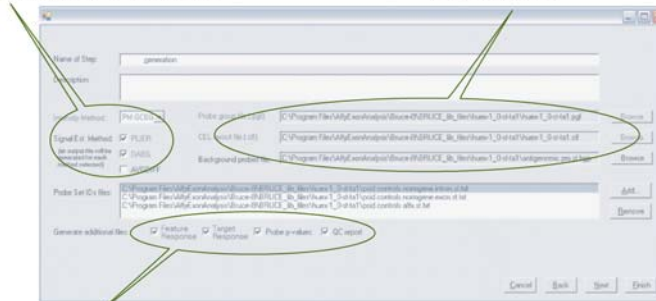
Refer to the Affymetrix website (www.affymetrix.com) for a complete listing of GeneChip-compatible™ Software Providers that offer commercially available tools specifically for Exon Array analysis.

Figure 5. Basic analysis workflow in ExACT.

A. Input of basic analysis parameters

Various probe set-level analysis algorithms may be selected

Input files defining the probe sets to be analyzed may be browsed and selected



A variety of statistical analysis output files may be generated including the QC report

B. Analysis log

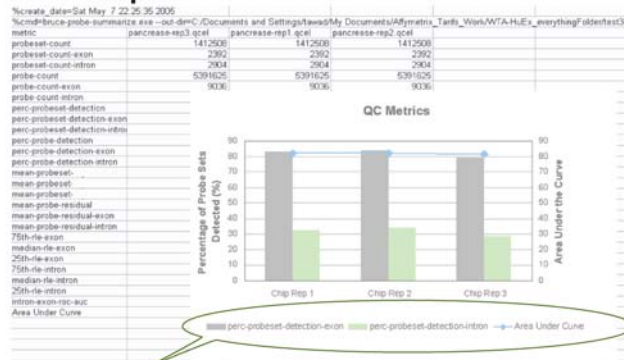
Analysis logs, input, and output files are captured



Analysis steps are organized hierarchically and saved into a project file

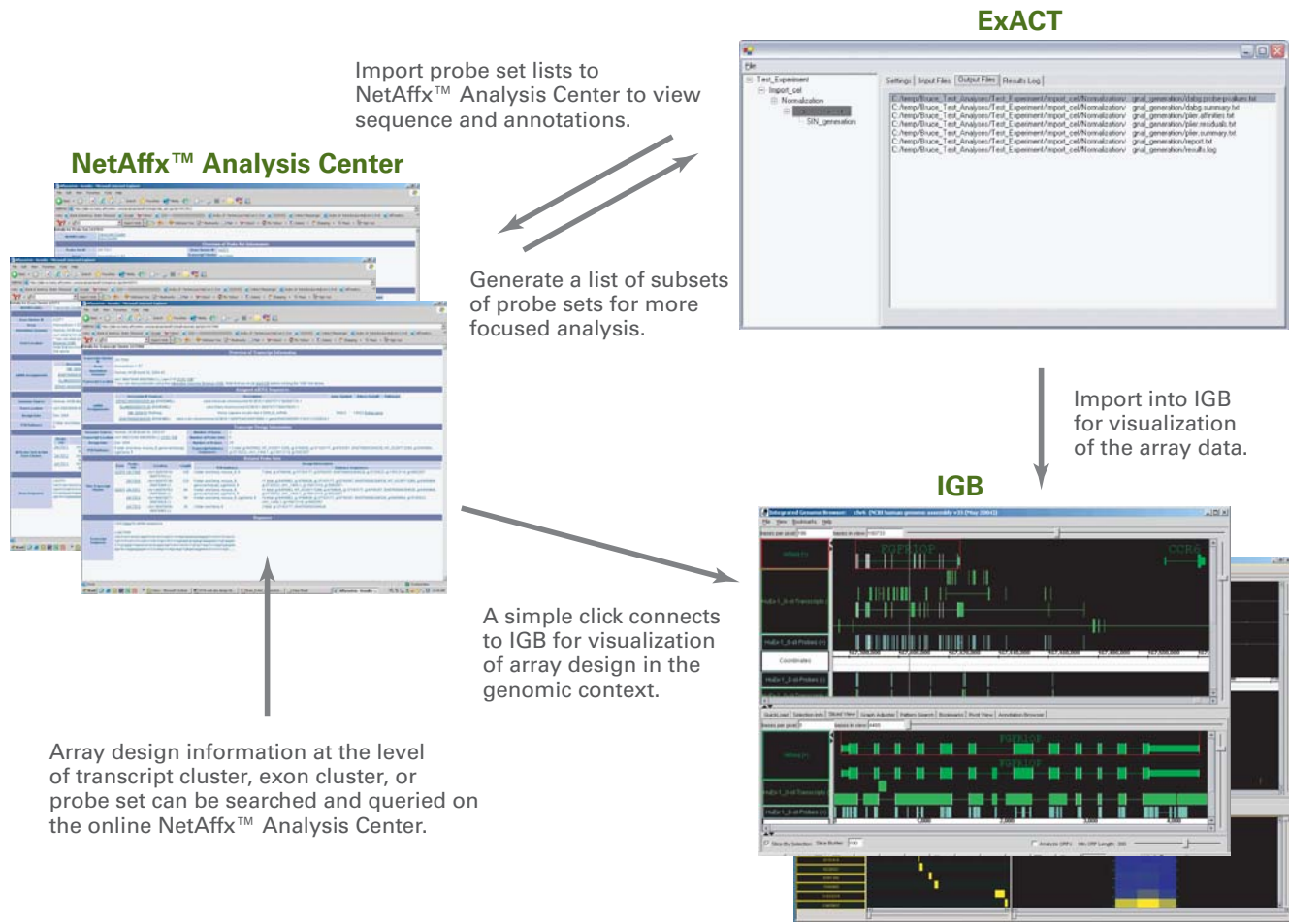
Output files can be further analyzed with other statistical or visualization tools

C. Standard QC reports from ExACT viewed in Excel



New metrics are implemented to examine the expression of introns and exons of normalization control genes as an indication of the sample and array quality

Figure 6. Integration of array data analysis, annotation, and visualization.



Exon Array System

Array	GeneChip® Human, Mouse, and Rat Exon 1.0 ST Arrays
Target Labeling Assay	GeneChip® Whole Transcript (WT) Sense Target Labeling Assay and Reagents
Fluidics Station	GeneChip® Fluidics Station 400 or Fluidics Station 450
Scanner	GeneChip® Scanner 3000 7G or higher
Operating Software	GeneChip® Operating Software v1.3 or higher
Data Analysis Software	Exon Array Computational Tool (ExACT)
Visualization	I ntegrated G enome B rowser (IGB)
Sequence and Annotations	NetAffx™ Analysis Center

GeneChip® Human Exon 1.0 ST Array Design Statistics Summary

	Human	Mouse	Rat
Probe sets	1.4 million	1.2 million	1 million
Exon clusters	>1 million	~1 million	850,000
Supported by putative full-length mRNA	289,961 probe sets	266,200 probe sets	92,038 probe sets
Supported by Ensembl transcripts	306,583 probe sets	266,791 probe sets	195,943 probe sets
Supported by EST	665,175 probe sets	554,003 probe sets	211,451 probe sets
Supported by syntenic mRNA	220,262 probe sets ³	214,763 probe sets ⁴	272,061 probe sets ⁵
Supported by gene prediction	883,105 probe sets	835,897 probe sets	875,666 probe sets
Probe selection region	Along the entire length of the transcripts		
Probes/Probe selection region	4 ¹		
Background subtraction strategy	Median fluorescence intensity of up to 1,000 background probes with the same GC content		
Total features per array	> 5,500,000		
Interrogated strand	Sense ²		

1. About 10 percent of the exon probe sets have fewer than four probes due to the probe selection region length and sequence constraints.
2. The probes tiled on the array are designed in the anti-sense orientation, requiring sense-strand labeled targets to be hybridized to the array. By convention, the array is called ST array representing the necessity of using sense targets (the labeled sample to be hybridized to the array).
3. Supported by mouse or rat mRNA.
4. Supported by human or rat mRNA.
5. Supported by human or mouse mRNA.

Technical Support Literature

Available online at www.affymetrix.com

Manuals

- GeneChip® Whole Transcript (WT) Sense Target Labeling Assay Manual
- ExACT User Manual

Technical Notes

- GeneChip® Exon Array Design
- GeneChip® Human Exon 1.0 ST Array and GeneChip® WT Sense Target Labeling Assay for Genome-Wide, Exon-Level Expression Analysis

White Papers

- Alternative transcript analysis methods for Exon Arrays
- Exon Array background correction
- Exon probe set annotations and transcript cluster groupings
- How-to for gene level summarization on Exon Arrays
- Gene signal estimates from Exon Arrays
- Quality assessment of Exon Arrays

Ordering Information

GeneChip® Exon 1.0 ST Array and WT Reagents

GeneChip® Exon 1.0 ST Array Starter Pack

(Highly recommended for users new to the Exon Array system solution)

- 900654 GeneChip® Human Exon 1.0 ST Array Starter Pack
 - 900831 GeneChip® Mouse Exon 1.0 ST Array Starter Pack
 - 900848 GeneChip® Rat Exon 1.0 ST Array Starter Pack
- GeneChip® Exon 1.0 ST Array, 30 arrays
GeneChip® WT Sense Target Labeling and Control Reagents, 30 rxn
On-site training including the following arrays and reagents for the training:
- GeneChip® Exon 1.0 ST Array, 10 arrays
 - GeneChip® WT cDNA Synthesis and Amplification Kit, 10 rxn
 - GeneChip® WT Terminal Labeling Kit, 10 rxn
 - GeneChip® Sample Cleanup Module, 30 rxn
 - GeneChip® Eukaryotic Poly-A RNA Control Kit, ~100 rxn
 - GeneChip® Hybridization Control Kit, 30 rxn

GeneChip® Exon 1.0 ST Array

- 900651 GeneChip® Human Exon 1.0 ST Array — Contains 30 Arrays
- 900819 GeneChip® Mouse Exon 1.0 ST Array — Contains 30 Arrays
- 900822 GeneChip® Rat Exon 1.0 ST Array — Contains 30 Arrays

GeneChip® WT Sense Target Labeling and Control Reagents, 30 rxn

- 900652 GeneChip® WT cDNA Synthesis and Amplification Kit, 30 rxn
- GeneChip® WT Terminal Labeling Kit, 30 rxn
- GeneChip® Sample Cleanup Module, 30 rxn
- GeneChip® IVT cRNA Cleanup Kit, 30 rxn
- GeneChip® Eukaryotic Poly-A RNA Control Kit, ~100 rxn
- GeneChip® Hybridization Control Kit, 30 rxn

ExACT, NetAffx™ Analysis Center, and Integrated Genome Browser

Available online at www.affymetrix.com

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
UK and Others Tel: +44 (0) 1628 552550
France Tel: 0800919505
Germany Tel: 01803001334
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. 702005 Rev. 2

©2005-2006 Affymetrix, Inc. All rights reserved. Affymetrix®,  GeneChip®, HuSNP®, GenFlex®, Flying Objective™, CustomExpress®, CustomSeq®, NetAffx™, Tools to take you as far as your vision®, The Way Ahead™, Powered by Affymetrix™, GeneChip-compatible™, and Command Console™ are trademarks of Affymetrix, Inc. Array products may be covered by one or more of the following patents and/or sold under license from Oxford Gene Technology: U.S. Patent Nos. 5,445,934; 5,700,637; 5,744,305; 5,945,334; 6,054,270; 6,140,044; 6,261,776; 6,291,183; 6,346,413; 6,399,365; 6,420,169; 6,551,817; 6,610,482; 6,733,977; and EP 619 321; 373 203 and other U.S. or foreign patents.