

## Publications and References for Affymetrix® GeneChip® Exon and Gene 1.0 ST Arrays

GeneChip® Exon and Gene 1.0 ST Arrays are the first experimental tools that can profile both gene-level and exon-level expression on the whole-genome scale using a single array. Researchers worldwide have published data highlighting how whole-transcript expression arrays, featuring the most advanced design and highest sensitivity available, have enabled them to uncover novel gene expression and alternative splicing patterns.

### Analysis/methods development

■ Abdueva D., *et al.* Experimental Comparison and Evaluation of the Affymetrix Exon and U133 Plus 2.0 GeneChip Arrays. *PLoS ONE* 2(9):e913 (2007).

#### Key findings:

- Abdueva *et al.* compared the performance of GeneChip® Exon Arrays and Human Genome U133 Plus 2.0 Arrays through a series of spike-in hybridization experiments and demonstrated high comparability between the two arrays.
- “Despite several major technological changes, we observe a high concordance between these platforms; i.e., individual HuEx probes are capable to reliably detect concentration changes and thus provide unbiased expression measures at exon level.”

■ Benovoy D., *et al.* Effect of polymorphisms within probe-target sequences on oligonucleotide microarray experiments. *Nucleic Acids Research* 36(13):4417-4423 (2008).

#### Key findings:

- Benovoy *et al.* studied how probe-to-target hybridization affects exon and gene expression estimates derived from the GeneChip® Exon Array when polymorphisms are present in the probe-target sequences.
- The authors describe a simple masking procedure that effectively reduces the SNP-induced false positives with a very modest amount of lost coverage.

■ Bitton D. A., *et al.* Exon-level integration of proteomics and microarray data. *BMC Bioinformatics* 9:118 (2008).

#### Key findings:

- Bitton *et al.* compared quantitative protein mass spectrometry data with exon-level data from a non-tumorigenic human breast epithelial cell line to examine the correspondence between the two data types.
- The authors found much higher correspondence between the proteomic and exon array data than what had been seen previously between proteomic and 3' array data. They attributed this greater correlation to the ability to compare data at a finer, peptide exon level.

■ Hu Z., *et al.* Exon-Level Expression Profiling: A Comprehensive Transcriptome Analysis for Oral Fluids. *Clinical Chemistry* 54:824-832 (2008).

#### Key findings:

- The authors aimed to validate the use of saliva samples for high-resolution gene expression profiling by combining a universal mRNA-specific linear-amplification strategy with GeneChip Exon Arrays, and a multiplex pre-amplification method.
- Hu *et al.* demonstrate that “It is feasible to use samples containing fragmented RNAs to conduct high-resolution expression profiling with coverage of the entire transcriptome and to validate multiple targets from limited amounts of sample.”

■ Mao X., *et al.* Rapid high-resolution karyotyping with precise identification of chromosome breakpoints. *Genes, Chromosomes and Cancer* 46(7):675-83 (2007).

#### Key findings:

- Mao *et al.* used a combination of multiple color fluorescent *in situ* hybridization (M-FISH) and Affymetrix 500K Arrays for high-resolution karyotyping and identification of chromosome breakpoints in prostate cancer cell models.
- GeneChip Exon Arrays were used to verify translocation events that resulted in fusion genes or partial gene deletions.
- The authors demonstrated an approach that is capable of rapidly and precisely identifying most chromosomal rearrangements in individual tumors; this facilitates the identification of critical genes and genetic biomarkers in tumorigenesis.

■ Okoniewski M. J., *et al.* An annotation infrastructure for the analysis and interpretation of Affymetrix exon array data. *Genome Biology* 8(5):R79 (2007).

#### Key findings:

- Okoniewski *et al.* proposed a process to analyze GeneChip Exon Array data through a genome-level annotation database, called “X:MAP,” and a BioConductor/R package for exon array analysis, called “Exonmap.”
- X:MAP is used to efficiently handle fine-grained mapping of Affymetrix probe set sequences from GeneChip Exon Arrays to the genome and visualize data in a genome browser powered by a Google API interface.
- Exonmap is a Bioconductor/R package that is optimized for analysis of exon arrays, utilizing and combining data extracted from multiple tables in the database. This results in smaller data transfer overheads between client and server.

■ Okoniewski M. J. and Miller C. J. Comprehensive Analysis of Affymetrix Exon Arrays Using BioConductor. *PLoS Computational Biology* 4(2):e6 (2008).

#### Key findings:

- Microarrays, such as the GeneChip Exon Array, offer increasingly denser features to cover more of the transcribed genome. As a result, complex annotation is required to analyze the rich gene expression data.
- Okoniewski *et al.* describe an analysis strategy using BioConductor, an open-source software package, to apply annotation to the data after interesting probe sets have been selected, and in some cases to pre-filter data.

■ Pradervand S., *et al.* Affymetrix Whole-Transcript Human Gene 1.0 ST Array is highly concordant with standard 3' expression arrays. *BioTechniques* 44:759-762 (2009).

#### Key findings:

- Pradervand *et al.* used MAQC (microarray quality control project) RNA samples to perform a series of comparative hybridizations between various expression arrays to examine reproducibility and differential gene expression call concordance.
  - According to the authors, "The high level of inter-platform correspondence observed in these analyses... provides strong evidence supporting the reliability of the Gene 1.0 ST Array as an excellent option for gene expression profiling."
- Purdom E., *et al.* FIRMA: a method for detection of alternative splicing from exon array data. *Bioinformatics* **24**(15):1707-1714 (2008).
- Key findings:**
- The authors present a method to detect alternatively spliced exons in individual samples, without replication, from GeneChip® Human Exon 1.0 ST Array data, and they evaluate the method using simulated data and two data sets of human tissue.
  - The FIRMA (finding isoforms using robust multichip analysis) method is shown to perform "well in detecting exon-specific changes in expression and therefore can contribute substantially to the detection of regulated alternative splicing."
- Robinson M. D. and T. P. Speed. A comparison of Affymetrix gene expression arrays. *BMC Bioinformatics* **8**:449 (2007).

#### Key findings:

- The authors demonstrated high concordance in generating gene-level signal estimates across the three Affymetrix expression platforms examined.
- Their data suggests that whole-transcript expression arrays may be more sensitive and may capture true biological variation that is missed with 3' arrays.
- "The HuEx array, an all-encompassing array, has the flexibility of measuring all known or predicted exonic content."

- Schutte M., *et al.* Exon Expression Arrays as a Tool to Identify New Cancer Genes. *PLoS ONE* **3**(8): e3007 (2008).

#### Key findings:

- The authors demonstrate that pattern-based correlation (PAC) is an effective high-throughput method for screening exon-skipping events to identify candidate cancer genes and novel exon-skipping events by applying the PAC strategy to 12 cancer cell lines.
- After expression profiles are generated with GeneChip Human Exon Arrays, the PAC algorithm is used to predict expression levels of each exon. Then outlier exons are identified.
- Under certain conditions, the PAC strategy is also shown to detect exon-skipping mutants in clinical cancer specimens and to identify recurrent exon-skipping mutations.

- Xing Y., *et al.* MADS: A new and improved method for analysis of differential alternative splicing by exon-tiling microarrays. *RNA* **14**:1470-1479 (2008).

#### Key findings:

- The authors describe microarray analysis of differential splicing (MADS), a new method for detecting alternative splicing events with improved sensitivity and specificity that exploits a series of low-level analysis algorithms to correct the major sources of noise in exon array probe intensities.
- In this study, MADS was used to analyze GeneChip Exon Array data on a mouse neuroblastoma cell line with improved results. From numerous novel PTB-dependent splicing events that were identified, 30 were tested by RT-PCR and 27 confirmed.

- According to the authors, the GeneChip Exon Array enables the discovery of new splicing patterns because of its high probe density, and the exon-tiling design does not depend on prior knowledge of splicing in target genes.

- Xing Y., *et al.* Probe Selection and Expression Index Computation of Affymetrix Exon Arrays. *PLoS ONE* **1**(1):e88 (2006).

#### Key findings:

- The authors proposed a new algorithm to identify constitutively expressed exons through empirical data sets, and generated a list of probes on GeneChip Exon Arrays corresponding to these constitutive exons.
- They used this empirically derived list, in combination with the well-established Li-Wong probe model (implemented in Bioconductor), to obtain accurate gene-level expression quantization from exon arrays that is more representative of the true transcriptional activity of each gene than classical microarrays.

- Yates T., *et al.* X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis. *Nucleic Acids Research* **36**:D780-D786 (2008).

#### Key findings:

- The authors introduced X:Map, a mapping project that maps GeneChip Exon Arrays and their corresponding genome data.
- X:Map provides detailed annotation of the intron-exon structure of each gene, their mappings to known transcripts and their location relative to exon array target sequences.

- Yoshida R., *et al.* A Statistical Framework for Genome-Wide Discovery of Biomarker Splice Variations with GeneChip® Human Exon 1.0 ST Arrays. *Genome Informatics* **17**(1):88-89 (2006).

#### Key findings:

- The authors proposed a new statistical method to identify differentially observed splicing variations from exon expression profiles.
- This work is an important first step toward the development of more advanced statistical algorithms for alternative splicing analysis.

## Gene expression analysis

- The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* doi:10.1038/nature07385 (2008).

#### Key findings:

- Through integrative analysis of multi-dimensional genomic data from large sample cohorts, the Cancer Genome Atlas reports the benefits of a more systematic examination of the glioblastoma genome and how they were able to define core pathways and gain new insights into the roles of *ERBB2*, *NF1*, *TP53* and *PIK3R1* and a link between *MGMT* and clinical response and outcome.
- The authors were able to detect EGFR activation with high-resolution genomic and exon-specific transcriptomic profiling, using the Affymetrix® Genome-Wide Human SNP Array 6.0 and GeneChip Exon Array.

- Chahrour M., *et al.* MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Transcription. *Science* **320**:1224-1229 (2008).

#### Key findings:

- To explore the role of *MeCP2* in the neurodevelopmental disorder Rett syndrome, Chahrour *et al.* focused on gene expression patterns in a discrete brain region, the hypothalamus. From this study, the authors suggest that *MeCP2* can both activate and repress transcription.

- The GeneChip® Mouse Exon 1.0 ST Array was used to discover significant gene expression changes in hypothalmi of *MeCP2* mouse models for this study.
  
- Chiao E., *et al.* Isolation and Transcriptional Profiling of Purified Hepatic Cells Derived from Human Embryonic Stem Cells. *Stem Cells* 26:2032-2041 (2008).  
**Key findings:**
  - To better understand the molecular mechanisms governing liver development, the authors used the GeneChip Exon 1.0 ST Array, “which produces improved gene-level expression measurements when compared to standard 3’ arrays,” to generate a genome-wide transcriptional profile of hepatic cells differentiated from embryonic stem cells.
  - The authors aimed to determine whether the AFP promoter driving eGFP expression could be used to mark and isolate hepatic cells and examine the transcriptional profile of isolated cells to identify additional cell markers that could be used to further define the cell lineages arising during hepatic cell differentiation.
  - Calculations from the GeneChip Exon Array data “accurately represent the expression of genes with a wide range of expression levels, supporting previous work indicating that exon arrays provide highly sensitive and accurate quantitative measurements of the levels of gene expression.”
  
- Douglas D., *et al.* BMI-1 Promotes Ewing Sarcoma Tumorigenicity Independent of CDKN2A Repression. *Cancer Research* 68(16):6507-6515 (2008).  
**Key findings**
  - The authors performed gene expression profiling of Ewing sarcoma family of tumor (ESFT) cells to investigate if *BMI-1* functions as an oncogene.
  - *BMI-1* loss did not alter ESFT proliferation or death, but it promotes tumorigenicity, independent of *CDKN2A* repression, and regulates pathways involved in cell differentiation, development and cell adhesion.
  
- Duan S., *et al.* Genetic Architecture of Transcript-Level Variation in Humans. *American Journal of Human Genetics* 82:1101-1113 (2008).  
**Key findings:**
  - Duan *et al.* examined expression values of 12,747 transcript clusters with more than 2 million SNPs across two HapMap populations to determine associations between expression quantitative trait loci and expression quantitative nucleotides.
  - “By using the GeneChip Human Exon 1.0 ST Array, our study has the advantage of determining expression levels of probes across the whole gene... which is considered a more accurate measure of gene expression.”
  
- Fish J., *et al.* *miR-126* Regulates Angiogenic Signaling and Vascular Integrity. *Developmental Cell* 15:272-284 (2008).  
**Key findings:**
  - Extensive mRNA expression profiling was performed using the GeneChip® Gene 1.0 ST Array to identify and study endothelial-enriched microRNAs to better understand how specific microRNAs might modulate angiogenic signaling.
  - The authors demonstrate that angiogenic signaling and vascular integrity can be disrupted by modulating *miR-126* expression, and they demonstrate that *miR-126* expression is required to maintain vessel integrity during zebrafish vascular development.
  - Data also suggests that endothelial cell migration, organization of the cytoskeleton, capillary network stability and cell survival is regulated by *miR-126*, and by inhibiting *SPRED1* and *PIK3R2*, *miR-126* promotes vascular endothelial growth factor signaling.
  
- French D., *et al.* Concordant Gene Expression in Leukemia Cells and Normal Leukocytes Is Associated with Germline cis-SNPs. *PLoS ONE* 3(5):e2144 (2008).  
**Key findings:**
  - From 92 patients with acute lymphoblastic leukemia (ALL), the authors identified a set of genes with concordant expression in leukemia cells and normal leukocytes from the same individuals, and they observed that these 20 genes were more likely to exhibit expression variation that was influenced by germline *cis*-SNPs.
  - To validate the study, cell samples from 134 pediatric patients with ALL and HapMap lymphoblastoid cell line data of European and African descent were assessed using the GeneChip Human Exon 1.0 ST Array.
  
- Ge X., *et al.* Genome-wide analysis of antisense transcription with Affymetrix exon array. *BMC Genomics* 9:27 (2008).  
**Key findings:**
  - Ge *et al.* examined antisense transcription across exonic loci using GeneChip Exon Arrays with a modified protocol.
  - The authors were able to identify antisense transcription at 2,088 exonic loci across 1,516 UniGene clusters.
  
- Guryev V., *et al.* Distribution and functional impact of DNA copy number variation in the rat. *Nature Genetics* 40(5):538 -545 (2008).  
**Key findings:**
  - Guryev *et al.* combined computational and experimental methods to catalog CNVs in commonly used rat strains and demonstrate the feasibility of using the rat as a model organism for studying disease phenotypes.
  - The authors identified 626 CNVs with the GeneChip® Rat Exon Array, and all de novo CNVs—one of which might represent a CNV hot spot—were experimentally confirmed.
  
- Huang R. S., *et al.* A genome-wide approach to identify genetic variants that contribute to etoposide-induced cytotoxicity. *PNAS* 104(23):9758-9763 (2007).  
**Key findings:**
  - Huang *et al.* aimed to identify potentially functional SNPs and/or haplotypes associated with chemotherapeutic agent-induced cytotoxicity using genotype, gene expression and cytotoxicity data.
  - Exon 1.0 ST Arrays were used to obtain whole-genome expression data to correlate using linear regression with SNPs found to be associated with etoposide cytotoxicity.
  - Analysis identified 63 genetic variants that contribute to etoposide-induced toxicity through their effect on gene expression.
  
- Huang R. S., *et al.* Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans. *Molecular Cancer Therapeutics* 7(9):3038-3046 (2008).  
**Key findings:**
  - To evaluate genetic variants contributing to carboplatin-induced cytotoxicity in individuals of African descent, Huang *et al.* took a whole-genome approach, integrating genotypes, gene expression and sensitivity to carboplatin.
  - The authors propose an integrative, high-resolution approach to assess the cumulative effects of genome-wide, tumor-specific changes in DNA methylation that may play a role in deregulation

- of gene expression, genomic imbalance contributions and gene expression in two osteosarcoma cell lines relative to normal human osteoblasts. The study used a combination of Affymetrix Promoter Tiling Arrays for DNA methylation and Affymetrix Gene 1.0 ST Arrays for gene expression analysis.
- Sadikovic *et al.* provide genomic DNA methylation maps and identify genes with aberrant DNA methylation in two osteosarcoma cell lines.
- Huang R. S., *et al.* Genetic Variants Contributing to Daunorubicin-Induced Cytotoxicity. *Cancer Research* **68**(9):3161-3168 (2008).  
**Key findings:**
- Using International HapMap cell lines, the authors designed a genome-wide study that integrates genotype, gene expression and cytotoxicity data to identify potentially functional single nucleotide polymorphisms associated with chemotherapy-induced cytotoxicity.
  - The authors "successfully identified 53 genetic variants significantly associated with daunorubicin-induced cytotoxicity through the expression of 61 genes in CEU and/or YRI cell lines."
- Huang R. S., *et al.* Identification of genetic variants and gene expression relationships associated with pharmacogenes in humans. *Pharmacogenetics and Genomics* **18**(6):545-549 (2008).  
**Key findings:**
- To better understand genetic effects on drug therapy, this genome-wide association study used GeneChip Human Exon 1.0 ST Arrays and 2 million HapMap SNPs to identify SNPs that potentially affect the expression or gene-gene interactions of very important pharmacogenes.
  - The study identified very important pharmacogenes selected by the Pharmacogenetic Research Network which are significantly associated with SNPs in European and African populations. This public information can be used to identify potential genetic determinants of gene expression and to explore the genetic contributions to drug therapy.
- Huang R. S., *et al.* Identification of Genetic Variants Contributing to Cisplatin-Induced Cytotoxicity by Use of a Genome-wide Approach. *American Journal of Human Genetics* **81**:427-437 (2007).  
**Key Findings:**
- Huang *et al.* examined expression profiles and SNP patterns across 176 HapMap cell lines (87 CEU and 89 YRI) to understand the genetic basis of cisplatin-induced cytotoxicity.
  - The authors identified 17 SNPs significantly associated with the differential expression of 26 genes and cisplatin-induced cytotoxicity.
- Kapur K., *et al.* Exon arrays provide accurate assessments of gene expression. *Genome Biology* **8**(5):R82 (2007).  
**Key Findings:**
- Kapur *et al.* developed a strategy for estimating gene expression on GeneChip Exon Arrays, a first step toward creating a baseline to judge the expression of individual exons.
  - This method includes probe-specific background correction and a probe selection strategy. It is based on the MAT algorithm developed for GeneChip Tiling Arrays.
  - The authors propose that using GeneChip Exon Arrays with their model (called GeneBASE) offers more accurate measurements of gene expression than using traditional 3' arrays.
- Kosiol C., *et al.* Patterns of Positive Selection in Six Mammalian Genomes. *PLoS Genetics* **4**(8):e1000144 (2008).  
**Key findings:**
- This genome-wide study of ~16,500 genes in six mammalian genomes applied phylogenetic and population genetic methods to detect which genes positive selection has affected, key differences between species, how patterns of selection have changed and the effects of gene expression patterns on positive selection.
  - The authors identified 544 genes that showed evidence of positive selection and found that positively selected genes (PSGs) are expressed at lower levels and in a more tissue-specific manner than non-PSGs.
  - Kosiol *et al.* used the GeneChip Human Exon 1.0 ST Array, "which contains probes for nearly all of our genes and permits accurate estimation of expression levels."
- Sadikovic B., *et al.* In Vitro Analysis of Integrated Global High-Resolution DNA Methylation Profiling with Genomic Imbalance and Gene Expression in Osteosarcoma. *PLoS ONE* **3**(7):e2834 (2008).\*  
**Key findings:**
- The authors propose an integrative, high-resolution approach to assess the cumulative effects of genome-wide, tumor-specific changes in DNA methylation that may play a role in deregulation of gene expression, genomic imbalance contributions and gene expression in two osteosarcoma cell lines relative to normal human osteoblasts. The study used a combination of Affymetrix Promoter Tiling Arrays for DNA methylation and Affymetrix Gene 1.0 ST Arrays for gene expression analysis.
  - Sadikovic *et al.* provide genomic DNA methylation maps and identify genes with aberrant DNA methylation in two osteosarcoma cell lines.
- Sandberg R., *et al.* Proliferating Cells Express mRNAs with Shortened 3' Untranslated Regions and Fewer MicroRNA Target Sites. *Science* **320**(5883):1643-1647 (2008).  
**Key findings:**
- Sandberg *et al.* performed a global analysis of alternative 3' untranslated regions (UTR) isoforms during T cell activation, and describe a method for probe-level alternative transcript analysis using the GeneChip Mouse Exon 1.0 ST Array.
  - The authors found increased expression of mRNAs with short 3' UTRs and suggest that UTR-based mRNA plays a role in the regulatory network of cell proliferation.
- Sequeira A., *et al.* Coding SNPs included in exon arrays for the study of psychiatric disorders. *Molecular Psychiatry* **13**:363-365 (2008).  
**Key findings:**
- Using cell lines from 10 human samples and GeneChip Human Exon 1.0 ST Array, the authors study the genetic variation of coding exons in psychiatric disorders.
  - The "experiment illustrates the power and specificity of exon arrays to detect... allele-specific expression within an exon."
- Sindhi R., *et al.* Genetic Variants in Major Histocompatibility Complex-Linked Genes Associate With Pediatric Liver Transplant Rejection. *Gastroenterology* **135**:830-839 (2008).  
**Key findings:**
- The authors propose a novel, multi-step approach to perform a genome-wide association study with a limited sample of children with primary liver transplantation and their biologic parents to identify causal variants of transplant rejection.

- The GeneChip Human Exon 1.0 ST Array was used to measure differential splicing patterns in archived RNA isolated from 29 of the children studied.
- Plausible candidates contributing to pediatric liver transplant rejection are the minor allele of the SNP rs9296068, a dysfunctional HLA-DOA gene product and the B-lymphocyte in which it is exclusively expressed.

■ Vokes S. A., *et al.* A genome-scale analysis of the *cis*-regulatory circuitry underlying sonic hedgehog-mediated patterning of the mammalian limb. *Genes & Development* 22:2651-2663 (2008).

**Key findings:**

- The authors combine whole-genome chromatin immunoprecipitation (ChIP-on-chip) and transcriptional profiling to study the roles of Shh signaling in the developing mouse limb.
- Expression data from the GeneChip Mouse Exon 1.0 ST Array was used to identify gene networks associated with Gli3-binding.
- Vokes *et al.* provide a framework of the regulatory network in the developing limb and identify Gli activator signaling and associations between Shh and other signaling pathways that drive limb growth.

■ Wang X., *et al.* The Expression of MicroRNA *miR*-107 Decreases Early in Alzheimer's Disease and May Accelerate Disease Progression through Regulation of  $\beta$ -Site Amyloid Precursor Protein-Cleaving Enzyme 1. *Journal of Neuroscience* 28(5):1213-1223 (2008).

**Key findings:**

- Wang *et al.* examined expression profiles on RNA extracted from brain tissue from individuals with various states of Alzheimer's disease.
- The authors found BACE1 mRNA levels to be negatively correlated with *miR*-107 levels and positively associated with the progression of Alzheimer's.

■ Xing Y., *et al.* Assessing the Conservation of Mammalian Gene Expression Using High-Density Exon Arrays. *Molecular Biology and Evolution* 24:1283-1285 (2007).

**Key findings:**

- "Since 3' expression microarrays use a small number of probes for each gene's 3' end, it is misleading to directly compare absolute expression estimates between human and mouse 3' arrays, which have completely independent probe designs for orthologous genes."
- "Unlike 3' expression arrays, exon arrays show highly correlated expression levels for orthologous genes in corresponding human and mouse tissues, suggesting a strong stabilizing selective pressure on transcript abundance."
- "Our analysis also demonstrates the power of high-density Exon Array technology, in particular for evolutionary studies of gene expression."

■ Zhang W., *et al.* Evaluation of Genetic Variation Contributing to Differences in Gene Expression between Populations. *The American Journal of Human Genetics* (In press, 2008).

**Key findings:**

- Zhang *et al.* examined expression profiles between two HapMap populations, one of African and one of European descent, to identify differentially expressed genes and to understand what biological pathways and processes were associated with the two populations.

- The authors identified 383 differentially expressed transcript clusters between the two populations. Among the differentially expressed genes, ribosome biogenesis and antimicrobial humoral response pathways were found to be enriched. The authors also identified a list of proximal and distal SNPs associated with differentially expressed genes.
- Overall, the authors' work shows that population-level differences in gene expression may be associated with genetic susceptibility to disease and drug toxicity.

■ Zhang X., *et al.* Comparison of smoking-induced gene expression on Affymetrix Exon and 3'-based expression arrays. *Genome Informatics* 18:247-257 (2007).

**Key findings:**

- Zhang *et al.* hybridized bronchial airway epithelial RNA from smokers and nonsmokers to GeneChip Exon Arrays and GeneChip® Human Genome U133 Plus 2.0 Arrays for performance comparison.
- While both platforms showed correlation for detecting smoking-related changes, the exon array provided more robust genome-wide measurements, measured partially degraded mRNA more reliably and enabled the detection of alternative splicing events.

## Alternative splicing analysis

■ Cheung C., *et al.* Global analysis of aberrant pre-mRNA splicing in glioblastoma using exon expression arrays. *BMC Genomics* 9:216 (2008).

**Key findings:**

- The authors used a GeneChip Human Exon 1.0 ST Array to measure glioma-specific splicing in 24 GBM and 12 non-tumor brain samples. They assessed the specificity, sensitivity and feasibility of this array-based, genome-wide approach to identify alternative splicing events.
- According to the authors, "The U251 experiments confirmed that targeted changes in alternative splicing of cassette exons would be reflected in high AS scores and *p*-values < 0.05 on the GeneChip Exon Array. This data indicated the general feasibility of the exon array and our analytical approach to identify cassette exon changes on a genome-wide level."
- Cheung *et al.* confirmed 14 genes with glioma-specific splicing, seven of which were novel events identified by the GeneChip Exon Array. The authors concluded, "The number of genes showing exclusive tumor-specific isoforms was fewer than suggested previously by *in silico* mining."

■ Clark T. A., *et al.* Discovery of tissue-specific exons using comprehensive human exon microarrays. *Genome Biology* 8(4):R64 (2007).

**Key findings:**

- This paper is the original Affymetrix publication examining alternative splicing using the prototype GeneChip Exon Array.
- The authors showed results for tissue-specific alternative splicing events as well as significant expression outside of known exons and well-annotated genes. This data is only available because of the comprehensive design of GeneChip Exon Arrays.
- Additionally, a splicing index algorithm is offered to identify alternative splicing events, whose efficacy was confirmed with RT-PCR validation on brain-enriched exons.

■ Das D., *et al.* A correlation with exon expression approach to identify cis-regulatory elements for tissue-specific alternative splicing. *Nucleic Acids Research* **35**(14):4845-4857 (2007).

**Key findings:**

- Das *et al.* performed a correlation with expression approach to identify cis-regulatory motifs for alternative splicing.
- Examining 56 cassette exons that exhibited higher transcript-normalized expression in muscle than in other tissues, the authors found multiple candidate regulatory motifs for muscle-specific splicing.
- “We anticipate that correlation with exon expression will provide valuable insights into the cis-regulation of alternative splicing as additional datasets of tissue-specific exons become available for analysis.”

■ French P. J., *et al.* Identification of differentially regulated splice variants and novel exons in glial brain tumors using exon expression arrays. *Cancer Research* **67**:5635-5642 (2007).

**Key findings:**

- French *et al.* set out to identify splice variants that are differentially expressed between histological subgroups of glial brain tumors.
- The results showed that using Human Exon 1.0 ST Arrays could help molecular classification of subgroups of gliomas based on their histological appearance.
- Exon-level profiling also identified more than 700 novel exons and a significant number of exons that are differentially spliced between glioblastomas and oligodendrogliomas, many of which were validated using RT-PCR.

■ 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).

**Key findings:**

- The authors analyzed the expression profiles of 10 colon cancer and 10 normal tissue samples with exon arrays.
- They found a correlation in gene-level signals between GeneChip Exon and U133 Plus 2.0 Arrays for genes that were significantly differentially expressed between tissue types.
- When reviewing differentially expressed genes between the cancer and control samples, they were able to identify 160 genes differentially expressed, and found that almost one-third of the up-regulated genes in cancer form a part of a tightly interconnected network involved in mitosis, cell cycle control, cell proliferation, invasion, matrix remodeling and Wnt signaling.
- They also identified a number of genes that are differentially spliced between cancer and normal groups. Eleven of these events were confirmed by RT-PCR. Interestingly, out of these 11 genes, 10 are involved in the organization of the cytoskeleton, or interaction with the matrix of other cells, forming a network that is regulated by splicing. These results could contribute to better understanding of cancer etiology and may provide therapeutic targets and diagnostic markers.

■ Hung L. H., *et al.* Diverse roles of hnRNP L in mammalian mRNA processing: A combined microarray and RNAi analysis. *RNA* **14**:284-296 (2008).

**Key findings:**

- Hung *et al.* used a combination of RNAi and GeneChip Exon Arrays to identify alternative splice changes based on the knockdown of a known mRNA splicing regulator.
- The authors predicted several novel splice variants for which no previous alternative splicing evidence had been available.

■ Based on GeneChip Exon Array data, the authors surmised that alternative poly-A site selection might act as a new regulatory mechanism where hnRNP L is involved.

■ Jhavar S., *et al.* Detection of TMPRSS2-ERG translocations in human prostate cancer by expression profiling using GeneChip Human Exon 1.0 ST Arrays. *Journal of Molecular Diagnostics* **10**:50-57 (2008).

**Key findings:**

- Jhavar *et al.* examined the ability of GeneChip Exon Arrays to detect a hybrid TMPRSS2-ERG transcript produced through a translocation event by monitoring the expression of individual *ERG* exons from 27 prostate cancer samples.
- The authors detected altered expression of the *ERG* gene in 15 out of 27 cancer samples, with increased expression of exons 4 to 11 in all 15 cases relative to exons 2 and 3. They hypothesized that this was indicative of a translocation event involving the fusion of *ERG* exon 4, and confirmed this via RT-PCR for all 15 cases.
- “Our results demonstrate the effectiveness of expression analysis using Exon 1.0 ST Arrays for detecting *ERG* translocations and provide novel insights into the mechanism of development of human prostate cancer.”

■ Kwan T., *et al.* Genome-wide analysis of transcript isoform variation in humans. *Nature Genetics* **40**(2):225:231 (2008).

**Key findings:**

- Kwan *et al.* examined the CEU HapMap population to perform a genome-wide analysis of common genetic variation controlling differential expression of transcript isoforms.
- The authors examined genes that displayed a correlation between flanking SNPs and transcript levels and determined that 39 percent reflected changes in whole-gene expression and 55 percent reflected transcript isoform changes such as splicing variants, differential 5' and 3' UTR use.
- The authors identified striking differences from previous reports that had examined expression profiles of the CEU population using 3' biased arrays. Their results show that previous studies incorrectly identified transcripts with shortened 3' UTRs and missed alternatively spliced transcripts altogether.
- “We show that tools such as the exon array, targeting probes to many regions of the gene, give a more complete picture of the true complexity of variation in gene expression than previously believed.”

■ Kwan T., *et al.* Heritability of Alternative Splicing in the Human Genome. *Genome Research* **17**:1210-1218 (2007).

**Key findings:**

- Kwan *et al.* investigated human variation in alternative splicing patterns from a HapMap population-derived cell line and identified several transcripts containing sequence-verified exon skipping, intron retention and cryptic splice site usage that are specific between individuals.
- The authors also identified several splicing events with no previous annotations, thus demonstrating that GeneChip Exon Arrays can identify both known and novel alternative splicing events.

■ McKee A. E., *et al.* Exon expression profiling reveals stimulus-mediated exon use in neural cells. *Genome Biology* **8**:R159 (2007).

**Key findings:**

- McKee *et al.* used GeneChip Exon Arrays to examine calcium-induced exon-level and transcript-level expression and made a

connection between extracellular stimuli and the transcription of specific exons.

- “Strikingly, many more transcripts demonstrated changes in expression levels of only a subset of exons, as compared with those showing changes throughout the entire transcript, suggesting that a pronounced alteration in exon usage occurs in response to KCl and TPG.”
- The authors found that stimulus-induced changes in alternative splicing act as a major contributor to gene regulation.

- Moore M. J. and Silver P. A. Global analysis of mRNA splicing. *RNA* 14:197-203 (2008).

**Key findings:**

- Moore and Silver review experimental methods that have been used to investigate alternative splicing and provide a brief review of a few of the GeneChip Exon Arrays studies.

- Nembaware V., et al. Genome-wide survey of allele-specific splicing in humans. *BMC Genomics* 9:265 (2008).

**Key findings:**

- The authors performed a genome-wide scan for SNPs likely to influence splicing efficiency using publicly available expressed sequence tags and exon array data for which genome-wide genotype data are available.
- Nembaware et al. used the exon array data to: 1) analyze associations between mRNA isoforms and the genotype of putative *cis*-acting splicing polymorphisms and 2) investigate the heritability of splicing and compare it to the heritability of transcript-level expression.
- The study reveals a set of genes, both novel and previously described in former experiments, that show evidence of allele-specific splicing.

- Oberdoerffer S., et al. Regulation of CD45 Alternative Splicing by Heterogeneous Ribonucleoprotein, hnRNPL. *Science* 321(5889):686-691 (2008).

**Key findings:**

- The authors identified heterogeneous nuclear ribonucleoprotein (hnRNPLs) “as a critical inducible regulator of CD45 alternative splicing.”
- “Exon array analysis suggested that hnRNPL acts as a global regulator of alternative splicing in activated T cells.”

- Soreq L., et al. Identifying Alternative Hyper-Splicing Signatures in MG-Thymoma by Exon Arrays. *PLoS One* 3(6):e2392 (2008).

**Key findings:**

- The authors combined exon arrays with ad-hoc and post-hoc statistics, analysis of key transcripts and RT-PCR, FISH and MS validation tests to identify hyper-alternative splicing events characteristic of myasthenia gravis (MG)-thymomas and to discriminate them from colon cancer changes.

- The results suggest that alternative hyper-splicing contributes to the biological functions of specialized tumors, opening new potential approaches for research, diagnosis and treatment. The study “highlights potential benefits from using high-resolution arrays in the study of cancer.”
- The GeneChip Human Exon Array’s “massive increase in probes along with exon length-dependent probe numbers, ensures more robust detection of gene-level transcription changes and allows the discovery of potentially new transcripts and novel, predicted exons.”

- Thorsen K., et al. Alternative splicing in colon, bladder, and prostate cancer identified by exon-array analysis. *Molecular & Cellular Proteomics* 7:1214-1224 (2008).

**Key findings:**

- To identify tissue- and tumor-specific alternative splicing, the GeneChip Human Exon 1.0 ST Array was used to measure whole-genome exon expression in 102 normal and cancer tissue samples from three different organs: colon, urinary bladder and prostate.
- The authors identified and validated 10 tissue-related alternative splicing events in addition to seven genes with tumor-specific splice variants from colon, bladder and prostate, which may be potential biomarkers and drug targets.

- Yeo G. W., et al. Alternative Splicing Events Identified in Human Embryonic Stem Cells and Neural Progenitors. *PLoS Computational Biology* 3(10):e196 (2007).

**Key findings:**

- Yeo et al. introduced an outlier detection approach to identify 1,737 internal exons that are predicted to undergo alternative splicing in neural progenitor cells compared to human embryonic stem cells.
- The authors discovered that candidate-alternative exons are enriched in genes encoding serine/threonine kinases and helicase activities.
- By comparing genomic sequences across multiple mammals, the authors were able to identify dozens of conserved candidate-binding sites that were enriched proximal to exons predicted to be alternatively spliced.

- Zhang Z., et al. SMN Deficiency Causes Tissue-Specific Perturbations in the Repertoire of snRNAs and Widespread Defects in Splicing. *Cell* 133:585-600 (2008).

**Key findings:**

- The authors examined splicing patterns in spinal cord, brain and kidney tissues of mice with spinal muscular atrophy (SMA) and their healthy littermates to study how the survival of motor neurons (SMN) affects the biogenesis of small nuclear RNA (snRNA)-ribonucleoproteins.
- Total RNA from mice tissue were analyzed using the GeneChip Mouse Exon 1.0 ST Array and real-time RT-PCR on 31 genes confirmed the exon array data with a validation rate of 97 percent.



---

**Affymetrix, Inc.**  
3420 Central Expressway  
Santa Clara, CA 95051 USA  
Tel: 1-888-DNA-CHIP (1-888-362-2447)  
Fax: 1-408-731-5380  
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](http://www.affymetrix.com)** Please visit our website for international distributor contact information.

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

P/N 702477 Rev. 5

©2009 Affymetrix, Inc. All rights reserved. Affymetrix®, GeneChip®, NetAffx®, Command Console®, Powered by Affymetrix™, GeneChip-compatible™, Genotyping Console™, DMET™, and GeneTitan™ are trademarks or registered trademarks of Affymetrix Inc. All other trademarks are the property of their respective owners.