

Publications and References for Affymetrix® GeneChip® Exon Arrays

The GeneChip® Exon Array, the most powerful and comprehensive expression array on the market, is the first experimental tool available to profile both gene-level and exon-level expression on the whole-genome scale using a single array. Researchers worldwide have published data highlighting how GeneChip® Exon Arrays, featuring the most advanced design and highest sensitivity available, have enabled them to uncover novel gene expression and alternative splicing patterns.

Peer-Reviewed Publications

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.”

■ 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.

■ 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.

■ 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.”

■ 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 ON EXON ARRAYS

■ 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.

■ 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.* 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.

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

Key Findings:

- The authors compared the gene expression profiles between two established cell lines and compared results obtained on GeneChip Exon and classical U133 Plus 2.0 Arrays.
- With three different mapping techniques, the two arrays showed a high degree of correspondence in terms of fold changes.
- The authors concluded that "since the classical microarrays have already been repeatedly validated experimentally, this provides strong evidence that exon arrays are also reliable, not only for probe sets that can be successfully mapped to the existing arrays, but also for the many thousands of additional probe sets that provide more detailed coverage of the transcriptome."

■ 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 Quotes from the Manuscript:

- "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."

■ Wang X., *et al.* The Expression of MicroRNA miR-107 Decreases Early in Alzheimer's Disease and May Accelerate Disease Progression through Regulation of β -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.

■ 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.

ALTERNATIVE SPLICING ANALYSIS ON EXON ARRAYS

■ 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 knock-down 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* **2007** **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.

■ 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.