ALEXA – A computational platform for alternative expression analysis

Malachi Griffith, Michelle Tang, Susanna Chan, Jennifer Asano, Obi Griffith, Thomas Zeng, Stephane Flibotte, Agnes Baross, Martin Hirst, Isabella Tai and Marco Marra

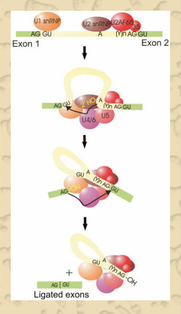
British Columbia Cancer Agency
Vancouver, British Columbia, Canada

Canada's Michael Smith Genome Sciences Centre
www.bcgsc.ca

1. Abstract

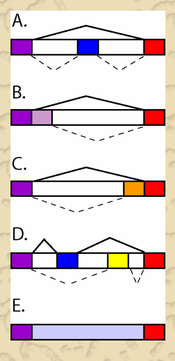
Introduction: The human genome contains approximately 25,000 genes. Recent estimates suggest that as many as 74% of human genes undergo alternative splicing (AS) and this process is thus an important mechanism for encoding a diversity of functions from a single genomic locus. We hypothesize that this diversity has functional implications for biology and for the development of chemotherapy resistance in cancer and other models of cancer progression.
Methods: A computational platform for Alternative Expression Analysis (ALEXA) was developed for the design and analysis of splicing microarrays for any species in EnsEMBL. To test this platform, we designed microarrays consisting of 2.5 million 26 to 46-mer (isothermal) oligonucleotide probes, selected to measure potential AS events for all multi-exon, protein-coding human genes. The design consists of probes for almost every annotated exon, intron and exon-intron boundary as well as candidate exon-exon junctions. Preliminary experiments to test the ALEXA platform were conducted by hybridizing reference mRNA samples to a prototype array synthesized by NimbleGen Systems Inc. Additional experiments are underway to compare the utility of ALEXA designs to the Affymetrix Human 1.0 ST Exon Array platform using mRNA samples isolated from 5-FU sensitive and resistant colorectal cancer cell lines.
Results: Hybridization data was analyzed by recording raw signal intensities in ALEXA and processing with Bioconductor and novel algorithms. Preliminary analysis has focused on: (1) evaluating experimental and control probe behavior, (2) optimizing hybridization conditions, (3) assessing background correction strategies, (4) identifying differentially expressed AS events, and (5) comparing the sensitivity and specificity of ALEXA arrays to Affymetrix exon arrays.
Conclusions: The approach described here represents an extension of current microarray design strategies with the ability to identify novel alternatively spliced isoforms representing potential therapeutic targets. We are characterizing the differential expression, transcript structure and function of isoforms associated with the transition from chemotherapy sensitivity to resistance in colorectal cancer. We have so far identified 20 differentially expressed genes and 150 differentially spliced exons in our drug resistance model.

Basic Splicing Machinery



RNA splicing is accomplished by the 'spliceosome' which consists of a complex of several nuclear components including a number of snRNPs. Through the action of RNA-RNA, RNA-protein and protein-protein interactions, this complex recognizes a pair of exon-intron boundaries and catalyzes sequential trans-esterification reactions to remove an intron and join two exons (Kalnina et al. 2005 Gen. Chrs. Cancer).

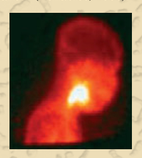
Types of Alternative Splicing



A.) Exon skipping/inclusion
B.) Alternate 3' splice site usage
C.) Alternate 5' splice site usage
D.) Mutually exclusive exons
E.) Intron retention

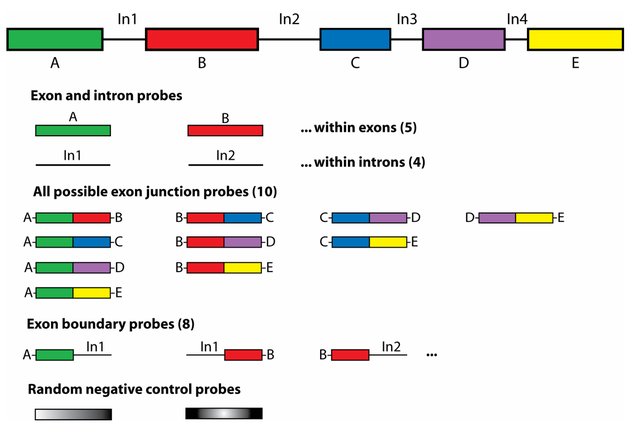
Alternative Transcription and Cancer

Alternative transcription (initiation, splicing and polyadenylation) generates multiple isoforms for most human genes. Specific isoforms may represent useful therapeutic targets or diagnostic markers. The image below shows a throat and neck cancer which has been targeted by a therapeutic antibody specific to a single alternate CD44 exon which is aberrantly expressed in the cancer (Venables, 2006).

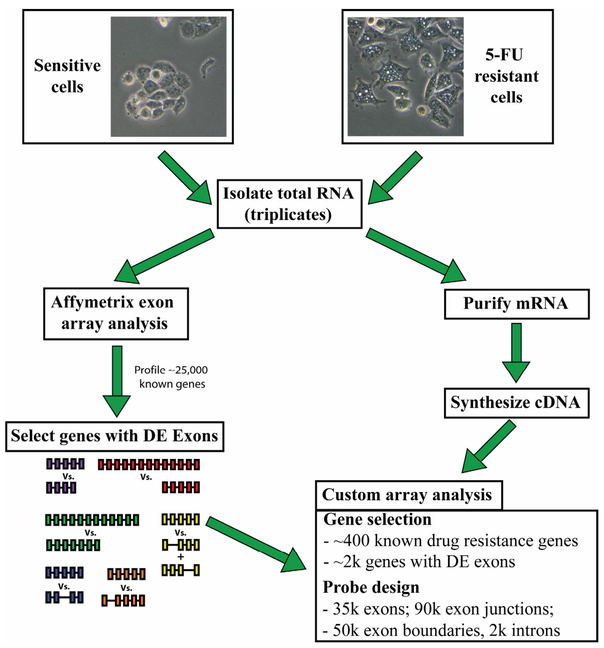


2. Methods

Array design strategy



Experimental overview



3. Results

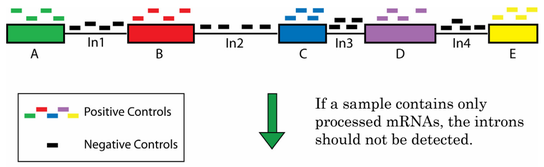
Platform description

ALEXA is a computational platform for the design of custom microarrays for Alternative Expression Analysis (alternative transcription initiation, splicing and polyadenylation). ALEXA consists of Perl and R modules interacting with a mySQL relational database and is capable of creating custom array designs for any EnsEMBL species. The design process consists of: (1) extracting probe sequences corresponding to exons, introns, exon junctions, exon boundaries and random sequences with minimal homology to any known sequence; (2) filtering probes according to their specificity and thermodynamic parameters; (3) selecting gene targets and generating an array design file for submission to a custom array manufacturer. The user may select fixed or variable length (isothermal) probes.

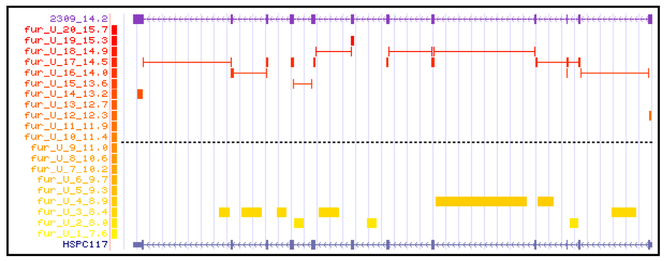
Pre-computed designs are currently being generated for a variety of metazoan species. To test the utility of ALEXA designs, a prototype human array was generated and hybridized with RNA extracted from chemotherapy sensitive and resistant cell lines. The same biological triplicates were analyzed on both the Affymetrix Exon and ALEXA array platforms. ALEXA arrays were manufactured by NimbleGen Systems Inc. For each platform the manufacturer's recommendations for sample processing were followed.

Assessing array performance with control genes

The performance of each hybridization was assessed by examining expression estimates from positive control probes (exonic) and negative control probes (intronic) for 100 housekeeping genes routinely used as controls in Affymetrix arrays. These genes were also targeted in the custom ALEXA design to allow direct comparison of array performance. The following figures show the arrangement of control probes and log2 expression values observed in an ALEXA experiment for the exon and intron probes of a single gene.



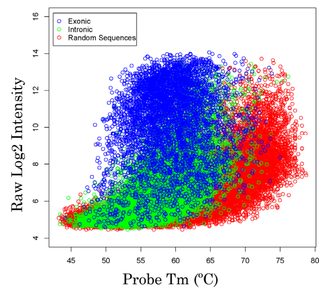
Mean expression values for the control gene, HSPC117 (Mean of 5FUR triplicates)



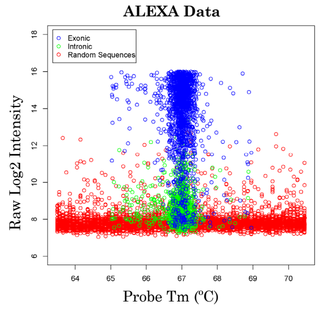
Comparing Affymetrix and ALEXA sensitivity/specificity

The mean expression value for all exon probes was considerably higher than that for intron probes (**21.2 +/- 3.3** fold higher in Affymetrix and **59.2 +/- 18.3** fold higher in ALEXA). The following figures show: (1) The distribution of exonic and intronic probe intensities for 100 control genes contrasted with the intensities observed for random sequence probes (plot is shown for a single hybridization on each platform). (2) A receiver-operator curve (ROC) illustrating the ability of each array platform to correctly identify exons as expressed and introns as not expressed (three replicates for each platform are shown). The area under the curve (AUC) is a measure of overall performance. (3) The correlation of Affymetrix and ALEXA differential expression estimates for all ~2,500 genes targeted by both platforms.

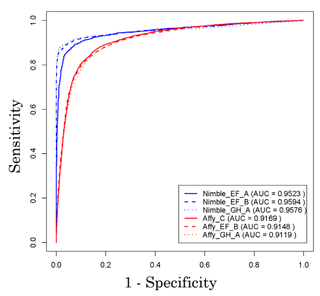
Affymetrix Data



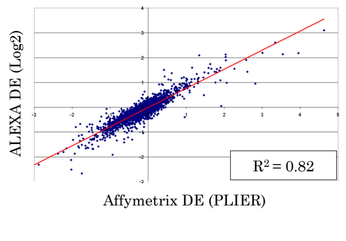
ALEXA Data



ROC Curves – Affymetrix/ALEXA



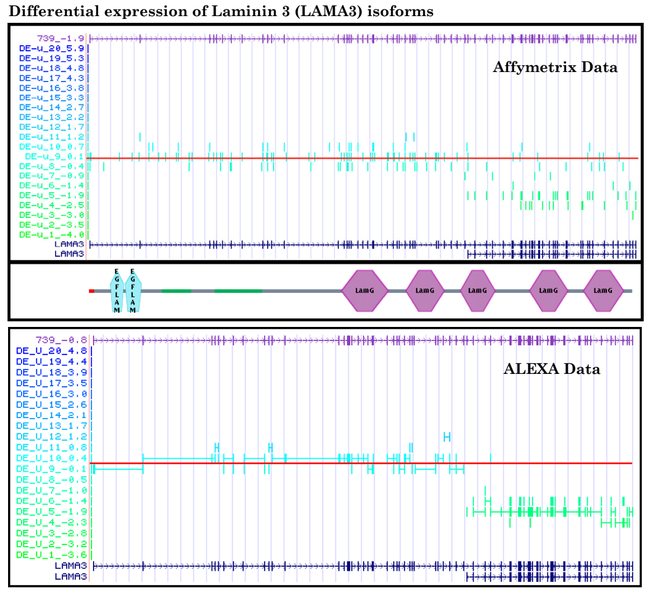
Affymetrix versus ALEXA estimates of Differential Gene Expression



4. Candidate genes

Differential expression of exons and genes between chemotherapy sensitive and resistant cell lines was determined by direct comparison of exon- and gene-level expression estimates as well as by use of a Wilcoxen Rank statistical test followed by multiple testing correction. A total of ~3,100 exons and ~3,000 exon junctions were found to be significantly changed between sensitive and resistant states after correcting for multiple testing (Benjamini and Hochberg, one-step FDR correction). These differentially expressed exons and junctions correspond to ~850 genes. The following figures illustrate examples of candidate genes identified by this approach.

Differential expression of Laminin 3 (LAMA3) isoforms



Gene Name	Gene Description	Description of Event
GNG2 (*)	G-protein gamma. Membrane component. Involved in signal transduction and cell proliferation.	Differential expression of two known isoforms, one with two additional 5' exons and an extra internal exon relative to the shorter isoform.
H19 (*)	Imprinted maternally expressed untranslated mRNA. Largely unknown function.	Differential expression (15.98 +/- 1.36 fold change) is observed for all 6 exons of this gene. There is also evidence for retention of the first intron. Expression of this gene is high in sensitive cells and almost completely lost in resistant cells
HHIP (*)	'Hedgehog interacting protein'. Regulatory component of the hedgehog pathway.	Differential expression of two known isoforms, a long isoform of 13 exons and a short isoform of only the first 4 exons. Expression of the long isoform is lost in the resistant state. The short isoform lacks two EGF motifs and one transmembrane domain compared to the long isoform.
LAMA3 (*)	Member of the laminin family. Component of the ECM. Involved in cell adhesion, signal transduction and differentiation.	Differential expression of two known isoforms, a long isoform of 76 exons and a short isoform of only the last 38 exons. The short isoform lacks several of the protein motifs found in the long isoform (including an EGF-like domain).

(*) Events observed in both the Affymetrix and ALEXA datasets

5. Conclusions and future work

The approach described in this work identified a list of candidate differentially expressed isoforms associated with the transition from 5-FU sensitivity to resistance. qRT-PCR will be conducted on a subset of candidates to confirm the accuracy of exon- and junction- level expression profiling; cloning and sequencing will be used to determine the precise structure of predicted DE isoforms; and functional assays will be conducted to study the potential role of selected genes in chemotherapy resistance. Further analysis is required to determine the benefits of additional probe types used in ALEXA designs compared to Affymetrix Exon Arrays.

6. Acknowledgments

Thanks | In addition to the authors listed above I would like to thank the GSC MGC Closure Group, Anca Petrescu, Steve Jones, Joseph Connors, and Gregg Morin

Funding | I am grateful for support provided by the Natural Sciences and Engineering Research Council, the Michael Smith Foundation for Health Research and Genome British Columbia

NSERC CRSNG · Michael Smith Foundation for Health Research · GenomeBritishColumbia