ALEXA: a microarray design platform for alternative expression analysis

To the editor: Eukaryotic genomes are predicted to contain about 7,000–29,000 genes[1]. Each of these genes may be alternatively processed to produce multiple distinct mRNAs by alternative transcript initiation, splicing and polyadenylation (collectively referred to as alternative expression). Although analysis of available transcript resources indicates that up to ~75% of genes are alternatively processed, most microarray expression platforms cannot detect alternative transcripts[2].

Proof-of-principle experiments have described the use of oligonucleotide microarrays to profile transcript isoforms generated by alternative expression, but resources to create such arrays are lacking[3,4]. To address this limitation we created a microarray design platform for alternative expression analysis (ALEXA), which is capable of designing arrays that can detect all of the major categories of alternative expression.

The ALEXA platform facilitates selection and annotation of oligonucleotide probes representing alternative expression events for any species in the EnsEMBL database[1]. For each target gene, probes are selected within every exon, intron, exon junction and exon boundary. This approach allows for the detection of constitutive and alternative exons, canonical exon junctions, junctions of known or new exon-skipping events, alternative exon boundaries and retained introns (**Supplementary Fig. 1** online). We designed the platform to be flexible to the user's experimental interests and preferred array manufacturer. The user may limit probe selection to known alternative expression events or include all possible exon junctions and boundaries to drive the discovery of transcripts. Probes may be designed for an arbitrary subset of genes or for all genes. Most technical parameters of the design can be modified by the user, including: the amount and types of control probes; the use of varying or fixed probe length; and the thresholds for filtering of probe sequences. The probe design process begins with retrieval of genomic sequences from EnsEMBL, removal of pseudogenes, masking of repeat elements and extraction of probe sequences. Random probe sequences are generated to uniformly represent the melting temperature and length of all experimental probes. Extracted and randomly generated probes are scored according to their melting temperature, folding potential, complexity and specificity (**Supplementary Methods** online).

Although several publications have described using microarrays to study alternative expression in model organisms and specific tissues[2], to our knowledge ours is the first report of a resource that makes alternative expression microarray designs readily available. Using the ALEXA approach, we precomputed microarray designs representing ~100 million probe sequences for ten EnsEMBL genomes (**Supplementary Table 1** online).

We assessed the ALEXA approach by using a prototype human array to profile the expression of alternative mRNA isoforms in 5-fluorouracil (5-FU)-sensitive and resistant colorectal cancer cell lines[5] and compared the results to those from the Affymetrix 'GeneChip Human Exon 1.0 ST' array (see **Supplementary Results**, **Supplementary Fig. 2** and **Supplementary Table 2** online). Genes and exons differentially expressed between 5-FU–sensitive and resistant cells were identified by both platforms (with significant overlap), but ALEXA arrays provided additional information on the connectivity and boundaries of exons (**Table 1**). Furthermore, alternative expression events identified by ALEXA were significantly enriched for known alternative expression events represented in publicly available mRNA and expressed sequence tag (EST) databases (**Supplementary Results** and **Supplementary Data 1** online). Finally, we demonstrated the advantage of the ALEXA approach by identifying several differentially expressed known and predicted isoforms with potential relevance to 5-FU resistance (**Supplementary Fig. 3** and **Supplementary Tables 3** and **4** online).

The approach and resources described in this work have considerable potential to advance studies of gene regulation, transcript processing, human disease and evolutionary biology (**Supplementary Discussion** online). The source code, precomputed array designs and related materials to assist in the creation of custom alternative expression microarrays are available on the ALEXA website (http://www.alexaplatform.org).

Note: Supplementary information is available on the Nature Methods website.

Malachi Griffith, Michelle J Tang, Obi L Griffith, Ryan D Morin, Susanna Y Chan, Jennifer K Asano, Thomas Zeng, Stephane Flibotte, Adrian Ally, Agnes Baross, Martin Hirst, Steven J M Jones, Gregg B Morin, Isabella T Tai & Marco A Marra

Genome Sciences Centre, British Columbia Cancer Agency, 675 West 10th Avenue, Vancouver, British Columbia V5Z 1L3, Canada.
e-mail: mmarra@bcgsc.ca

1. Hubbard, T. *et al. Nuc. Acids Res.* **33**, D447–D453 (2005).
2. Griffith, M. & Marra, M.A. in *Genes, Genomes & Genomics* (eds., D. Thangadurai, W. Tang & T. Pullaiah.) 201–242 (Regency Publications, New Delhi, 2007).
3. Castle, J. *et al. Genome Biol.* **4**, R66 (2003).
4. Srinivasan, K. *et al. Methods* **37**, 345–359 (2005).
5. Tai, I.T., Dai, M., Owen, D.A. & Chen, L.B. *J. Clin. Invest.* **115**, 1492–502 (2005).

Table 1 | Differential expression events for genes profiled by Affymetrix and ALEXA platforms

	Differential expression event type	Total number of events profiled	Number of significant events	Number of significant events within ORF	Number of significant events affecting known protein features
Affymetrix	Gene-level	2,507	78	NA	NA
	Exon	49,681	1,117	978	589
	Intron	65,327	25	20	0
	Total	117,515	1,220	998	589
ALEXA	Gene-level	2,507	233	NA	NA
	Exon	32,164	2,703	2,537	1,544
	Canonical junction	27,046	2,310	2,260	1,277
	Exon skip	69,761	191	180	103
	Exon boundary	52,402	253	219	100
	Intron	472	0	0	0
	Total	184,354	5,690	5,196	3,024

Significant events had a multiple testing corrected P value < 0.05 and a fold change > 2. 'Canonical junction' refers to the connection of adjacent exons. NA, not applicable.