Splicing Express: A software suite for alternative splicing analysis using next-generation sequencing data

Jose E Kroll, Jihoon Kim, Lucila Ohno-Machado, Sandro J de Souza

Motivation: Alternative splicing events (ASEs) are prevalent in the transcriptome of eukaryotic species and are known to influence many biological phenomena. The identification and quantification of these events are crucial for a better understanding of biological processes. Next-generation DNA sequencing technologies have allowed deep characterization of transcriptomes and made it possible to address these issues. ASEs analysis, however, represents a challenging task especially when many different samples need to be compared. Some popular tools for the analysis of ASEs are known to report thousands of events without annotations and/or graphical representations. A new tool for the identification and visualization of ASEs is here described, which can be used by biologists without a solid bioinformatics background. Results: A software suite named Splicing Express was created to perform ASEs analysis from transcriptome sequencing data derived from next-generation DNA sequencing platforms. Its major goal is to serve the needs of biomedical researchers who do not have bioinformatics skills. Splicing Express performs automatic annotation of transcriptome data (GTF files) using gene coordinates available from the UCSC genome browser and allows the analysis of data from all available species. The identification of ASEs is done by a known algorithm previously implemented in another tool named Splooce. As a final result, Splicing Express creates a set of HTML files composed of graphics and tables designed to describe the expression profile of ASEs among all analyzed samples. By using RNA-Seq data from the Illumina Human Body Map and the Rat Body Map, we show that Splicing Express is able to perform all tasks in a straightforward way, identifying well-known specific events. Availability and implementation: Splicing Express is written in Perl and is suitable to run only in UNIX-like systems. More details can be found at: http://www.bioinformatics-brazil.org/splicingexpress.
Splicing Express: A Software Suite for Alternative Splicing Analysis Using Next-Generation Sequencing Data

José Eduardo Kroll ¹,², Jihoon Kim³, Lucila Ohno-Machado³ & Sandro José de Souza ²*

¹ Institute of Bioinformatics and Biotechnology, Natal, RN, Brazil.
² Brain Institute, UFRN, Natal, RN, Brazil.
³ Department of Biomedical Informatics, University of California San Diego, La Jolla, CA, USA.

* To whom correspondence should be addressed at Brain Institute – UFRN, av. Nascimento de Castro 2155, CEP 59056-450, Natal, RN, Brazil, sandro@neuro.ufrn.br.
| Abbreviation | Description |
|--------------|-------------|
| ASE          | Alternative Splicing Event |
| DAVID        | The Database for Annotation, Visualization and Integrated Discovery |
| DNA          | Deoxyribonucleic acid |
| FPKM         | Fragments Per Kilobase Of Exon Per Million Fragments Mapped |
| HTML         | HyperText Markup Language |
| NGS          | Next-Generation Sequencing |
| RefSeq       | Reference Sequence |
| RNA-seq      | Whole Transcriptome Shotgun Sequencing |
| UCSC         | University of California, Santa Cruz |
ABSTRACT

Motivation: Alternative splicing events (ASEs) are prevalent in the transcriptome of eukaryotic species and are known to influence many biological phenomena. The identification and quantification of these events are crucial for a better understanding of biological processes. Next-generation DNA sequencing technologies have allowed deep characterization of transcriptomes and made it possible to address these issues. ASEs analysis, however, represents a challenging task especially when many different samples need to be compared. Some popular tools for the analysis of ASEs are known to report thousands of events without annotations and/or graphical representations. A new tool for the identification and visualization of ASEs is here described, which can be used by biologists without a solid bioinformatics background.

Results: A software suite named Splicing Express was created to perform ASEs analysis from transcriptome sequencing data derived from next-generation DNA sequencing platforms. Its major goal is to serve the needs of biomedical researchers who do not have bioinformatics skills. Splicing Express performs automatic annotation of transcriptome data (GTF files) using gene coordinates available from the UCSC genome browser and allows the analysis of data from all available species. The identification of ASEs is done by a known algorithm previously implemented in another tool named Splooce. As a final result, Splicing Express creates a set of HTML files composed of graphics and tables designed to describe the expression profile of ASEs among all analyzed samples. By using RNA-Seq data from the Illumina Human Body Map and the Rat Body Map, we show that Splicing Express is able to perform all tasks in a straightforward way, identifying well-known specific events.

Availability and implementation: Splicing Express is written in Perl and is suitable to run only in UNIX-like systems. More details can be found at: http://www.bioinformatics-
INTRODUCTION

Alternative splicing is involved in many biological phenomena and has been better evidenced due to the development of next-generation sequencing (NGS) technologies. RNA-Seq technologies have facilitated the deep characterization of transcriptomes of several species. However, the huge amount of data generated in each experiment brings challenges for the analysis and interpretation of alternative splicing. This is particularly critical for researchers or research groups that lack bioinformatics expertise. Despite the fact that some bioinformatics tools have already been developed for the identification of ASEs in NGS-derived data, challenges remain. For example, most of the tools available rely on complex interfaces or command line inputs, not to mention their complex output. Overall, there is a lack of user-friendly tools that are suitable for genome-wide ASE analysis, especially when several samples need to be compared.

We present here a software suite named Splicing Express, which was designed to provide a simple and effective way for the identification, annotation and visualization of ASEs. It requires GTF files as input, as those created from Cufflinks. By using RNA-Seq data from the Illumina Human Body Map Project and the Rat Body Map, we show that Splicing Express is able to extract meaningful information from deep transcriptome data. Both analyses using Splicing Express are available at http://www.bioinformatics-brazil.org/ASE_HBM.

MATERIAL AND METHODS

Sequences. The human genome reference sequence (GRCh37/hg19) was downloaded from
UCSC Genome Bioinformatics portal (http://genome.ucsc.edu), as well as RefSeq sequences and their respective annotations (files refFlat.txt.gz and refSeqAli.txt.gz). RNA-Seq data from the Illumina Human Body Map Project Version 2 were downloaded from the EMBL-EBI portal (accession E-MTAB-513). All available tissues were used: adrenal, adipose, brain, breast, colon, heart, kidney, liver, lung, lymph, ovary, prostate, skeletal muscle, testes, thyroid and white blood cells. Analyses were restricted only to the 2x50 bp paired-end sequences, which represent approximately 80 million read pairs per tissue.

Transcriptome assembly. Reads from all 16 tissues from the Illumina Human Body Map Project were mapped to the genome using Tophat version 2. A conservative parameter was maintained (--splice-mismatches 0), and a reference GTF file, containing RefSeq alignments from the UCSC portal, was set for helping the mapping process. The resulting BAM files created by Tophat were, afterwards, used as input for Cufflinks version 2. Use of that software was carried out using default parameters and without any reference file as input. Additional processing, such as sequence clustering, was performed by Splicing Express.

Sequencing clustering. The expressed sequences were clustered by Splicing Express, using a strategy that has been used previously by us. Basically, the genome coordinates of all expressed sequences were compared to the genome coordinates of Reference Sequences (RefSeqs) entries. Sequences presenting multiple exons and sharing at least one exon-intron boundary with a RefSeq sequence were merged. In the specific case of Splicing Express, all remaining sequences were excluded from the alternative splicing analysis, but were taken into account for measuring the total gene expression. The intronless sequences were merged by
checking their overlap to RefSeqs, which should be higher or equal than 30 bp.

**Alternative splicing analysis.** The GTF files resulting from the transcriptome assembly of the Illumina Human Body Map Project were processed using Splicing Express. The expressed sequences were clustered to a set of RefSeq sequences, which comprised 26,318 genes and a total of 50,129 entries. The processing steps are better described in the “Implementation” section. In-house Perl scripts were developed for all large-scale analyses, and all events described in the “Analysis” section were visually identified by using Splicing Express. A basic analysis using Splicing Express was also performed for the Rat Body Map\(^5\), which is available at the Splicing Express website.

**Statistics.** Splicing Express performs statistical analysis aiming to give significance values to the fold differences between the splicing variants. That analysis is only performed when the GTF files have data regarding the FPKM confidence intervals, which are standardly present in any GTF file created by Cufflinks\(^5\). Those values are gathered to calculate the FPKM Standard Error (SE), given by the equation:

\[
SE = \frac{C_{hi} - C_{lo}}{2 \times 1.96},
\]

where \(C_{hi}\) and \(C_{lo}\) are the FPKM values at the lower and higher limit of the 95% FPKM confidence interval, respectively. Moreover, it is known that more than one expressed sequence can share the same splicing event in a sample. Hence, the FPKM of those expressed sequences have to be summed. Its respective SE is then calculated by the equation:

\[
SE_x = \sqrt{SE_1^2 + \ldots + SE_n^2},
\]

where \(x\) refers to a splicing cluster and \(n\) refers to the expressed sequences from the respective
Finally, a Welch's t-test is done between the two clusters of splicing variants, defined here as \( a \) and \( b \). The t-test equation is given by:

\[
t = \frac{(X_a - X_b)}{\sqrt{SE_a^2 + SE_b^2}},
\]

where \( X \) is the FPKM of the sample. Probabilities are then calculated from t-values through the Perl Statistics::Distributions library using a degree of freedom equal one.

Availability. Splicing Express source code is available through a repository manager platform at http://www.bioinformatics-brazil.org/splicingexpress, where tutorials (written and video), sample files, and a bug track system are provided.

Figure 1. A) Table generated by Splicing Express showing expression level and number of variants in a gene-by-gene basis. B) Schematic representation of the computational pipeline used...
by Splicing Express. C) Graphic representation of the expression pattern of NRXN3. D) Graphic representation of the differential expression of one isoform (exon skipping) of NRXN3. The exon skipping isoform is represented in red, while the isoform including the respective exon is colored in green.
RESULTS AND DISCUSSION

Implementation

*Splncing Express* is a suite composed of Perl scripts that executes several processing steps, as schematized in Figure 1B. Prior to running *Splncing Express*, parameters about the reference genome, the output directory and the input GTF files must be firstly set. After that, GTF files are checked for consistency and converted to a workable file format. GTF files created by Cufflinks are recommended to be used as input due to some annotation standards referring to sequence names and FPKM expression. Similarly, chromosome names and genome positions from the GTF files must be compatible with the respective genome data provided by the UCSC genome portal.

Expressed sequences are then clustered through a known strategy (see Material & Methods). Clustering is needed for identifying the gene of origin of each expressed sequence contained in the GTF files, and is also very important to standardize the annotation process using gene symbols according to the HUGO (Gene Nomenclature Committee). *Splncing Express*, therefore, allows the use of GTF files showing any kind of gene annotation since those information will be posteriorly rewritten.

The ASEs identification is done through an algorithm based on pairwise comparison, a very common approach used for many other similar tools⁴,⁸,⁹. *Splncing Express*, moreover, uses an strategy implemented previously by us in *Sploce*⁸, which is based on regular expressions that are capable of identifying all well-known simple and combined ASEs. In that strategy, the expressed sequences are represented as binary sequences, where exons and introns are represented by 1 and 0, respectively. Those sequences are then pairwisely compared in an exhaustive manner, creating numerical patterns which represent their splicing differences.
Currently, Splicing Express is able to identify patterns of simple alternative splicing events, such as exon skipping, intron retention and alternative 5' and 3' splicing borders.

Afterwards, a graphic representation for the expression level (FPKM) is created for each gene and for each identified ASE. Graphics are created using an external SVG library, which is available at CPAN (Comprehensive Perl Archive Network). They can be opened in all recent releases of internet browsers, zoomed in or out without loss of quality. Last, those graphics are gathered together with additional information in HTML files. An example of some graphical displays are shown in Figure 1C and 1D. A main file (index.html) is created at the root of the results directory tree, providing advanced tools for exploring and filtering events (figure 1A). A basic statistical summary and a help section are also provided.

**Performance**

Splicing Express analysis does not aim speed or efficiency since the major computational processing is done previously by tools like Tophat and Cufflinks. Overall, Splicing Express processing time is fast, and it is able to process 15 samples in about half-hour using an average desktop personal computer. Here, we performed a simple performance analysis. No accuracy and sensitivity parameters were measured because the ASE identification algorithm here proposed is based on an exhaustive approach. As a testing platform, an i7 Core processor with 16GB RAM machine was chosen, although Splicing Express uses only a single core and needs no more than 1 Gb of RAM memory.
Sixteen GTF files representing each tissue from the Illumina Human Body Map Project were analyzed. The performance was measured for each processing step of Splicing Express as a function of the number of analyzed samples (Figure 2). The processing was done in triplicate and the results shown represent their average. Total processing time increased almost linearly as more samples were added to the analysis. The linear trend is particularly shaped by the processing time of “Data preparation” and “Annotation” steps. The “ASE identification” step, on the other hand, is based on a pairwise strategy, which may consume exponentially more processing time as more samples are analyzed. The “ASE identification” processing, however, was very fast and the exponential trend could be barely observed. In turn, the “HTML & Graphics” processing time theoretically decreases as the identification of new alternative splicing
events becomes saturated (less graphics are created), meaning that it tends to stabilize as more samples are analyzed. However, it could not be observed for the current data. “HTML & Graphics” step also depends on a moderate use of disk I/O due to a large amount of files being written. Hence, slow downs may be observed when outputing Splicing Express results directly to pen drives or other kind of slow devices.

Analysis

| Event Type            | Known | Partially known | Unknown | Total |
|-----------------------|-------|-----------------|---------|-------|
| Exon skipping         | 9175  | 8147            | 1012    | 18334 |
| Alt. 5' border        | 2368  | 2766            | 1101    | 6235  |
| Alt. 3' border        | 2755  | 3845            | 1019    | 7619  |
| Intron retention      | 891   | 8168            | 2060    | 11119 |
| Total                 | 15189 | 22926           | 5192    | 43307 |

Table 1. Frequency of alternative splicing events by type and by degree of knowledge.

Data from the Illumina Human Body Map Project was used for validating the use of Splicing Express in finding specific events among different samples. In total, 16 distinct tissues, showing high sequence coverage, were processed. As result, 43,307 events were identified, from which 36% comprehended well-known events (Table 1). Meanwhile, partially known and unknown events accounted for 64% of the data, as expected. Known events were defined as any event showing both variants supported by a RefSeq. Partially known events, in turn, were defined as events showing only one of both variants supported by a RefSeq. In order of
frequency, taking into account only known events, the most frequent event type was exon skipping, followed by alternative splicing borders and intron retention, as observed before. A more detailed analysis, considering tissue-specific alternative splicing events (tsASEs), showed that brain, adrenal gland, lymphnode and testis, among all analyzed tissues, had the highest amount of specific variants (t(df=14)=9.644117, p < 0.00001) (Table 2). TsASEs were defined as any splicing variant showing a relative expression equal or higher than 75% of the total expression in all analyzed tissues. Only TsASEs showing a FPKM higher or equal than 1 were selected.

To further explore the use of Splicing Express, we focused in the analysis of brain, since it is known to show one of the highest levels of skipped exons, alternative 5' and 3' splice sites. Moreover, a high frequency of tsASEs was found for brain (Table 2). A gene ontology (GO) enrichment analysis of biological processes using DAVID showed that genes showing tsASEs for brain are more enriched with terms related to brain functions, such as “axonogenesis”, “cell morphogenesis involved in neuron differentiation” and “neuron projection morphogenesis” (p-value < 6 x 10^-3) (Figure 3). MACF1, for example, is a microtubule actin crosslinking factor known to determine neuronal positioning by regulating microtubule dynamics and mediating GSK-3 signaling during brain development, and it was observed to be more expressed in 6 other tissues than brain. However, three ASEs were specific to brain (two exon skipping and one 3' alternative splicing site). In genomic scale analysis, examples like that could only be related to a specific tissue through meticulous alternative splicing analysis, as that provided by Splicing Express.
| Tissue      | Exon skipping | Alt 3' Splice Site | Alt 5' Splice Site | Total | Unique |
|------------|---------------|--------------------|--------------------|-------|--------|
| Adipose    | 8             | 13                 | 7                  | 27    | 15     |
| Adrenal gland | 21          | 31                 | 32                 | 80    | 39     |
| Brain      | 27            | 24                 | 34                 | 78    | 46     |
| Breast     | 11            | 12                 | 13                 | 32    | 13     |
| Colon      | 8             | 7                  | 14                 | 22    | 11     |
| Heart      | 13            | 14                 | 13                 | 40    | 17     |
| Kidney     | 11            | 16                 | 21                 | 44    | 24     |
| Leukocyte  | 11            | 18                 | 18                 | 41    | 16     |
| Liver      | 8             | 12                 | 10                 | 29    | 17     |
| Lung       | 10            | 22                 | 14                 | 46    | 26     |
| Lymphnode  | 22            | 39                 | 26                 | 87    | 49     |
| Ovary      | 13            | 15                 | 17                 | 40    | 22     |
| Prostate   | 16            | 14                 | 13                 | 40    | 21     |
| Skeletal muscle | 16        | 10                 | 6                  | 28    | 15     |
| Testis     | 34            | 29                 | 20                 | 81    | 47     |
| Thyroid    | 20            | 21                 | 17                 | 51    | 33     |

Table 2. Number of genes showing tissue-specific alternative splicing events (tsASEs).

Among other events, an interesting one was observed for NRXN3, a neurexin related to
presynaptic cell adhesion. In brain, the skipping/inclusion of exon SS#4 is known to control the
postsynaptic AMPA receptor trafficking, being therefore an important factor for synaptic
plasticity\textsuperscript{13}. The specificity of NRXN3 and of its event to brain could be easily observed (Figure
1C and D). On the other hand, many genes do not show tsASEs, such as the CELF family of
proteins, which are involved in cell-specific and developmentally regulated alternative splicing
in brain\textsuperscript{14}. \textit{Splicing Express}, however, was able to confirm the specificity of CELF3 and CELF5
to brain through their respective gene expression. For more informations about those and other
genes, a full analysis of the Human Body Map and Rat Body map\textsuperscript{15} using \textit{Splicing Express} were
made available at http://www.bioinformatics-brazil.org/ASE_HBM.

\textbf{CONCLUSION}

\textit{Splicing Express} provides expression graphics in an intuitive fashion. The final HTML
results can be open locally (offline), hence they can easily be hosted on web servers, shared and
distributed. The complete automation in \textit{Splicing Express} makes it a useful tool for deep analysis
of ASE in any transcriptome. Now that deep sequencing is readily available and there is a
shortage of human resources with bioinformatics skills, \textit{Splicing Express} can make a significant
contribution to the community.

\textbf{ACKNOWLEDGMENTS}

We would like to thank Marbella Maria da Fonsêca for helping us on making the video
tutorial, available at the \textit{Splicing Express} website.
Figure 3. Gene ontology (GO) enrichment analysis of biological processes for the 78 genes showing tsASEs in brain.

REFERENCES
1. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq.
   Bioinformatics 2009; 25(9):1105-1111.

2. Liu Q, Chen C, Shen E, Zhao F, Sun Z, Wu J. Detection, annotation and visualization of
   alternative splicing from RNA-Seq data with SplicingViewer. Genomics 2012; 99(3):178-182.

3. Seok J, Xu W, Gao H, Davis RW, Xiao W. JETTA: junction and exon toolkits for
   transcriptome analysis. Bioinformatics 2012; 28(9):1274-1275.

4. Florea L, Song L, Salzberg SL. Thousands of exon skipping events differentiate among
   splicing patterns in sixteen human tissues. F1000Research 2012; 2.

5. Roberts A, Pimentel H, Trapnell C, Pachter L. Identification of novel transcripts in annotated
   genomes using RNA-Seq. Bioinformatics 2011; 27(17):2325-2329.

6. Galante PAF, Sakabe NJ, Kirschbaum-Slager N, de Souza SJ. Detection and evaluation of
   intron retention events in the human transcriptome. RNA 2004; 10(5):757-765.

7. Galante PAF, Vidal DO, de Souza JE, Camargo AA, de Souza SJ. Sense-antisense pairs in
   mammals: functional and evolutionary considerations. Genome biology 2007; 8(3):R40.

8. Kroll JE, Galante PA, Ohara DT, Navarro FC, Ohno-Machado L, de Souza SJ. SPLOOCE: a
279 new portal for the analysis of human splicing variants. RNA biology 2012; 9(11):1339-1343.

280

281 9. Sammeth M, Foissac S, Guigó R. A general definition and nomenclature for alternative
splicing events. PLoS Comput Biol 2008; 4:e1000147.

283

284 10. Yeo G, Holste D, Kreiman G, Burge CB. Variation in alternative splicing across human
tissues. Genome biology 2004; 5(10):R74.

286

287 11. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene
lists using DAVID Bioinformatics Resources. Nature Protoc 2009; 4(1):44-57.

289

290 12. Ka M, Jung EM, Mueller U, Kim WY. MACF1 regulates the migration of pyramidal neurons
via microtubule dynamics and GSK-3 signaling. Developmental biology 2014; 395(1):4-18.

292

293 13. Aoto J, Martinelli DC, Malenka RC, Tabuchi K, Südhof TC. Presynaptic neurexin-3
alternative splicing trans-synaptically controls postsynaptic AMPA receptor trafficking. Cell
2013; 154:75–88.

296

297 14. Ladd AN, Charlet-B N, Cooper TA. The CELF family of RNA binding proteins is implicated
in cell-specific and developmentally regulated alternative splicing. Molecular and cellular
2001; 21(4):1285-1296.

300

301 15. Yu, Ying, et al. A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4
302 development stages. Nature communications 2014; 5.
FIGURE LEGENDS

Figure 1. A) Table generated by Splicing Express showing expression level and number of variants in a gene-by-gene basis. B) Schematic representation of the computational pipeline used by Splicing Express. C) Graphic representation of the expression pattern of NRXN3. D) Graphic representation of the differential expression of one isoform (exon skipping) of NRXN3. The exon skipping isoform is represented in red, while the isoform including the respective exon is colored in green.

Figure 2. Performance analysis for each processing step of Splicing Express as a function of the number of analyzed samples.

Figure 3. Gene ontology (GO) enrichment analysis of biological processes for the 78 genes showing tsASEs in brain.
Table 1. Frequency of alternative splicing events by type and by degree of knowledge.

|                  | Known | Partially known | Unknown | Total  |
|------------------|-------|-----------------|---------|--------|
| Exon skipping    | 9175  | 8147            | 1012    | 18334  |
| Alt. 5' border   | 2368  | 2766            | 1101    | 6235   |
| Alt. 3' border   | 2755  | 3845            | 1019    | 7619   |
| Intron retention | 891   | 8168            | 2060    | 11119  |
| Total            | 15189 | 22926           | 5192    | 43307  |
Table 2. Number of genes showing tissue-specific alternative splicing events (tsASEs).

| Tissue      | Exon skipping | Alt 3' Splice Site | Alt 5' Splice Site | Total | Unique |
|-------------|---------------|--------------------|--------------------|-------|--------|
| Adipose     | 8             | 13                 | 7                  | 27    | 15     |
| Adrenal gland | 21            | 31                 | 32                 | 80    | 39     |
| Brain       | 27            | 24                 | 34                 | 78    | 46     |
| Breast      | 11            | 12                 | 13                 | 32    | 13     |
| Colon       | 8             | 7                  | 14                 | 22    | 11     |
| Heart       | 13            | 14                 | 13                 | 40    | 17     |
| Kidney      | 11            | 16                 | 21                 | 44    | 24     |
| Leukocyte   | 11            | 18                 | 18                 | 41    | 16     |
| Liver       | 8             | 12                 | 10                 | 29    | 17     |
| Lung        | 10            | 22                 | 14                 | 46    | 26     |
| Lymphnode   | 22            | 39                 | 26                 | 87    | 49     |
| Ovary       | 13            | 15                 | 17                 | 40    | 22     |
| Prostate    | 16            | 14                 | 13                 | 40    | 21     |
| Skeletal muscle | 16    | 10                 | 6                  | 28    | 15     |
| Testis      | 34            | 29                 | 20                 | 81    | 47     |
| Thyroid     | 20            | 21                 | 17                 | 51    | 33     |
Figure 1: Overview of Splicing Express features

A) Table generated by Splicing Express showing expression level and number of variants in a gene-by-gene basis. B) Schematic representation of the computational pipeline used by Splicing Express. C) Graphic representation of the expression pattern of NRXN3. D) Graphic representation of the differential expression of one isoform (exon skipping) of NRXN3. The exon skipping isoform is represented in red, while the isoform including the respective exon is colored in green.
Figure 2: Performance analysis for each processing step of *Splicing Express* as a function of the number of analyzed samples.
Figure 3: Gene ontology (GO) enrichment analysis of biological processes for the 78 genes showing tsASEs in brain.
Table 1 (on next page)

Frequency of alternative splicing events by type and by degree of knowledge.
|               | Known | Partially known | Unknown | Total   |
|---------------|-------|-----------------|---------|---------|
| Exon skipping | 9175  | 8147            | 1012    | 18334   |
| Alt. 5' border| 2368  | 2766            | 1101    | 6235    |
| Alt. 3' border| 2755  | 3845            | 1019    | 7619    |
| Intron retention | 891    | 8168          | 2060    | 11119   |
| Total         | 15189 | 22926          | 5192    | 43307   |
**Table 2** (on next page)

Number of genes showing tissue-specific alternative splicing events (tsASEs).
| Tissue         | Exon skipping | Alt 3' Splice Site | Alt 5' Splice Site | Total | Unique |
|---------------|---------------|--------------------|--------------------|-------|--------|
| Adipose       | 8             | 13                 | 7                  | 27    | 15     |
| Adrenal gland | 21            | 31                 | 32                 | 80    | 39     |
| Brain         | 27            | 24                 | 34                 | 78    | 46     |
| Breast        | 11            | 12                 | 13                 | 32    | 13     |
| Colon         | 8             | 7                  | 14                 | 22    | 11     |
| Heart         | 13            | 14                 | 13                 | 40    | 17     |
| Kidney        | 11            | 16                 | 21                 | 44    | 24     |
| Leukocyte     | 11            | 18                 | 18                 | 41    | 16     |
| Liver         | 8             | 12                 | 10                 | 29    | 17     |
| Lung          | 10            | 22                 | 14                 | 46    | 26     |
| Lymphnode     | 22            | 39                 | 26                 | 87    | 49     |
| Ovary         | 13            | 15                 | 17                 | 40    | 22     |
| Prostate      | 16            | 14                 | 13                 | 40    | 21     |
| Skeletal muscle | 16            | 10                 | 6                  | 28    | 15     |
| Testis        | 34            | 29                 | 20                 | 81    | 47     |
| Thyroid       | 20            | 21                 | 17                 | 51    | 33     |