

Revealing Global Regulatory Features of Mammalian Alternative Splicing Using a Quantitative Microarray Platform

Qun Pan,^{1,5} Ofer Shai,^{2,5} Christine Misquitta,¹
Wen Zhang,^{1,3} Arneet L. Saltzman,^{1,3}
Naveed Mohammad,¹ Tomas Babak,^{1,3} Henry Siu,^{1,3}
Timothy R. Hughes,^{1,3,4} Quaid D. Morris,^{1,2}
Brendan J. Frey,² and Benjamin J. Blencowe^{1,3,4,*}

¹Banting and Best Department of Medical Research

²Department of Electrical and Computer Engineering

³Department of Molecular and Medical Genetics

⁴Program in Proteomics and Bioinformatics

C.H. Best Institute

University of Toronto

112 College Street

Toronto, Ontario M5G 1L6

Canada

Summary

We describe the application of a microarray platform, which combines information from exon body and splice-junction probes, to perform a quantitative analysis of tissue-specific alternative splicing (AS) for thousands of exons in mammalian cells. Through this system, we have analyzed global features of AS in major mouse tissues. The results provide numerous inferences for the functions of tissue-specific AS, insights into how the evolutionary history of exons can impact on their inclusion levels, and also information on how global regulatory properties of AS define tissue type. Like global transcription profiles, global AS profiles reflect tissue identity. Interestingly, we find that transcription and AS act independently on different sets of genes in order to define tissue-specific expression profiles. These results demonstrate the utility of our quantitative microarray platform and data for revealing important global regulatory features of AS.

Introduction

The splicing of precursor mRNA (pre-mRNA) to mRNA is a critical step in the expression of the majority of mammalian genes. Four small nuclear ribonucleoprotein particles (snRNPs) and numerous non-snRNP factors associate with pre-mRNA to form a spliceosome, which catalyzes the excision of intervening intron sequences and joining of the exon sequences (Burge et al., 1999; Graveley, 2000; Kramer, 1996; Will and Luhrmann, 1997). A typical human and mouse gene contains eight to ten exons, which can be joined in different arrangements by alternative splicing (AS) (Black, 2003; Smith and Valcarcel, 2000). Recent computational studies have estimated that one- to two-thirds of human and mouse genes contain at least one alternative exon (Brett et al., 2002; Croft et al., 2000; Kan et al., 2001; Lander et al., 2001; Mironov et al., 1999; Modrek and Lee, 2003; Okazaki et al., 2002). In contrast to budding yeast, which

only contains a few known examples of AS, the potential complexity of mRNA isoforms generated by AS in mammals and other metazoan organisms is vast. It is therefore widely assumed that AS is a key step in the generation of proteomic diversity in more complex organisms (Black, 2000; Graveley, 2001; Modrek and Lee, 2002).

Alternative splicing is known to play numerous critical roles in regulatory pathways in metazoans, including those controlling cell growth, cell death, differentiation and development, and its misregulation has been implicated in many life-threatening human diseases (Blencowe, 2000; Caceres and Kornblihtt, 2002; Cartegni et al., 2002; Faustino and Cooper, 2003; Smith and Valcarcel, 2000). A major goal of the postgenomic era is to identify physiologically- and disease-relevant AS events to determine where and when these occur, what their specific roles are, and how they are regulated. In contrast to the considerable advances made in recent years in the understanding of global regulatory properties of transcription, very little is understood as to how AS events are globally regulated and coordinated under different conditions. These fundamental questions require advances in the development of high-throughput technologies for monitoring alternative splicing. To date, most efforts in this direction have focused on computational analyses of AS using complementary DNA (cDNA) and expressed sequence tag (EST) data, and efforts to harness microarray technology for analyzing AS have begun.

Computational approaches with cDNA and EST data have allowed inferences to be made regarding the location, frequency, and conservation of AS events in different organisms (Brett et al., 2002; Modrek and Lee, 2003; Nurdinov et al., 2003; Pan et al., 2005; Sorek et al., 2004; Thanaraj et al., 2003; Xu et al., 2002). A limitation of these approaches is that AS can be detected only in transcripts for which there is sufficient sequence coverage. Moreover, the available sequences often are represented for only a limited number of cell and tissue types, which, in the case of human and mouse ESTs, are frequently derived from tumor or cell-line sources. Consequently, it is not possible to obtain substantial information on cell or tissue-type specific inclusion levels of alternative exons by using sequence-based computational approaches.

Recently, a number of groups have begun to explore the use of different experimental approaches for the high-throughput detection and monitoring of RNA processing events (Lee and Roy, 2004). Spotted-oligonucleotide microarrays employing probes designed to detect unprocessed and processed RNA have been used to monitor pre-mRNA splicing in yeast (Clark et al., 2002) and the processing of noncoding RNAs in yeast and mammals (Peng et al., 2003; Babak et al., 2004). A fiber optic-based array method (Yeakley et al., 2002), a polymerase colony assay (Zhu et al., 2003), and more conventional microarray-based approaches utilizing spotted cDNA fragments or oligonucleotides (Hu et al., 2001; Johnson et al., 2003; Wang et al., 2003) have been used for monitoring AS in mammalian cells. The most exten-

*Correspondence: b.blencowe@utoronto.ca

⁵These authors contributed equally to this work.

sive use of the latter approach was the application of “exon-junction” microarrays for the discovery of exon skipping events in human tissues and cell lines (Johnson et al., 2003). These authors used custom microarrays containing oligonucleotide probes complementary to mapped exon-exon junction sequences in RefSeq genes for the main purpose of discovering new AS events in human transcripts. Despite the progress described above, a system has not yet been described that permits the large-scale quantitative profiling of alternative splicing in mammalian cell and tissue sources. This is primarily due to limitations stemming from the design of existing microarrays and the lack of suitable algorithms for data analysis.

In this paper, we describe a microarray platform that permits the simultaneous quantification of the levels of thousands of alternative exons in mammalian cell and tissues sources. We have applied this system to the analysis of the regulation of 3126 sequence-verified AS events in diverse mouse tissues. The resulting data have generated hundreds of new inferences for functional roles of tissue-specific AS, insights into how the evolutionary origins of alternative exons relate to their inclusion levels in normal tissues, and information on global features of AS that underlie tissue-type specificity. This study therefore demonstrates the utility of a quantitative microarray platform for generating fundamental new insights into the global regulation of alternative splicing in mammals.

Results

A Custom Microarray for Quantitative Profiling of AS in Mammalian Cells

In order to perform large-scale quantitative analyses of functionally diverse AS events in mammalian tissues, we developed a custom microarray to represent sequence-validated AS events mined from mouse cDNA and EST sequence databases (refer to Experimental Procedures). To minimize representation of possible splicing errors or relatively low-abundance transcripts, we selected “cassette-type” AS events with the highest numbers of supporting cDNA and EST sequences from different cell and tissue sources. To enhance the sensitivity of detection and quantification of inclusion/exclusion levels of alternative exons, each AS event was measured by using six different oligonucleotide probes: one body probe for each exon sequence, designated as “C1, A and C2” probes (C, constitutive; A, alternative), and one junction probe for each of the three splice-junction sequences generated by AS, designated as “C1-A, A-C2 and C1-C2” probes (Figure 1A). In addition, a control probe specific to each intron sequence (located between C1 and A) was included to permit detection of unspliced pre-mRNA and/or contaminating genomic DNA in the hybridizations.

From an initial starting set of 4892 AS events in our database, 3126 AS events were selected for monitoring on a single ink-jet printed microarray, manufactured by Agilent Technologies (Figure 1B). The vast majority of the AS events correspond to cassette-type alternative exons, and additional events may correspond to mutually exclusive alternative exons. The 3126 AS events are

represented by 2647 distinct genes, with 413 of the genes containing two or more AS events. In addition, 54 of the AS events represented on the microarray are duplicates and were monitored by sets of probes that in some cases are complementary to different sequences within the same exons. These served as reproducibility controls (see below). The 2647 AS genes represented on the microarray are associated with 1118 distinct Gene Ontology Biological Process (GO-BP) categories among a total set of 2362 GO-BP categories assigned to 10,361 Mouse Gene Informatics (MGI) markers (refer to Experimental Procedures; see below). This indicates that the AS genes represented on the microarray are associated with a diverse range of biological functions in mammalian cells.

Quantitative Microarray Profiling of Alternative Splicing in Mouse Tissues

In order to assess the performance of our microarray system and to reveal global properties of alternative splicing in mammalian tissues, we hybridized random-primed Cy3/Cy5-labeled cDNA prepared from poly-A⁺ mRNA isolated from the following ten adult mouse tissues: brain, heart, intestine, kidney, liver, lung, muscle, salivary, spleen, and testis. The poly-A⁺ RNA was isolated from tissue samples pooled from several animals (CD1 strain) in order to collect sufficient quantities and to reduce detection of possible animal-to-animal variability. Fluor-reversals were performed for each hybridization to control for signal reproducibility and label incorporation biases. A multistep data-processing procedure was employed to normalize probe signals between the separately hybridized arrays and to remove spatial-bias artifacts (refer to Experimental Procedures).

Because each microarray probe has unique hybridization characteristics and the probe signals can depend on various parameters including transcript abundance, tissue source, and noise, a direct analysis of the normalized probe signals (i.e., taking ratios) was not sufficient for producing accurate measurements of alternative splicing levels (data not shown). We therefore developed a new data analysis tool, referred to as the generative model for the alternative splicing array platform (GenASAP), to automatically generate the percent alternatively spliced exon exclusion (%ASex) values from the microarray data. This new tool, which employs machine learning and a Bayesian network, takes into account probe noise, outlier processes, abundance of transcript, and unanticipated correlations between probe values. In addition to outputting %ASex values for each AS event, GenASAP outputs a number indicating the confidence level of the %ASex value. These values were used to rank the 31,260 %ASex values (from profiling the ten mouse tissues) in order from highest (1) to lowest (31,260) confidence. In the presentation of the results below, GenASAP %ASex values across the ten tissues are assigned a tissue-specific rank (AS tissue rank; from 1 to 31,260) and also a single cumulative rank (AS rank; from 1 to 3126). The AS rank thus represents the rank for the same AS event across the ten tissues. Additional information on the GenASAP algorithm is given in the Experimental Procedures, and a detailed description of its derivation is described in a Technical Report (available at <http://www.psi.utoronto.ca/~ofer/AS-supp.pdf>).

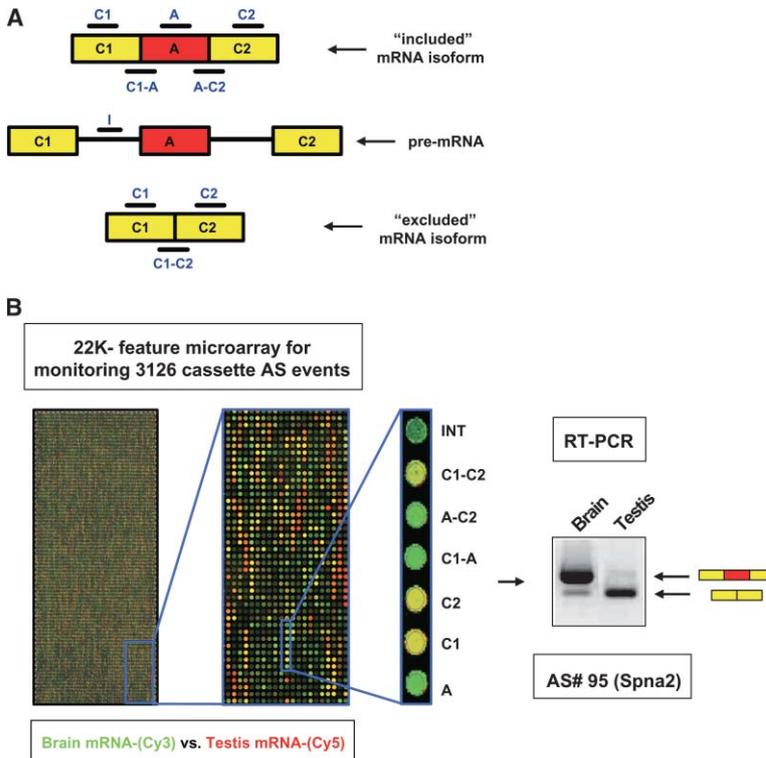


Figure 1. A Custom Microarray for Measuring AS Levels in Mouse

(A) Oligonucleotide probes used to monitor AS on custom microarrays. Each cassette-type AS event is measured with a combination of six probes: three specific for exon body sequences (C1, A, C2) and three specific for splice-junction sequences (C1-A, A-C2, and C1-C2). A single intron-specific probe (I) is also included for every AS event represented on the arrays.

(B) Scanned image of the entire hybridized region of an AS microarray, consisting of 3126 sets of seven probes as illustrated in (A). The microarray was hybridized with Cy3- and Cy5-labeled cDNA prepared from polyA⁺ mRNA isolated and pooled from multiple mouse adult brain and testis tissues, respectively. An enlarged view of the microarray is shown, where each set of seven probes for measuring AS are arranged in vertical strips. A set of probes for monitoring AS of exon 23 in the Spectrin α 2 (Spna2) gene is shown enlarged to the right. The probe fluorescence signals (without any normalization or algorithm analysis) suggest that Spna2 transcripts are expressed in both brain and testis (probes for C1 and C2 are yellow) but that exon 23 is only included in brain (probes for A, C1-A, and A-C2 are green). Skipping of exon 23 is apparent in both tissues (probe for C1-C2 is yellow). RT-PCR analysis using primers specific for Spna2 exons 22 and 24 confirm this expected tissue-specific AS pattern.

AS Microarray and GenASAP Validation

To assess the reproducibility of our array measurements for AS, the GenASAP values generated from duplicate AS events (54 in total) were compared against each other. The Pearson correlation coefficient (PCC) for the GenASAP %ASex values with AS tissue ranks from 81 to 31,234 is 0.9332 ($p < 10^{-24}$) (Supplemental Figure S1, available online at <http://www.molecule.org/cgi/content/full/16/6/929/DC1/>). This is in contrast to PCCs in the range of -0.3 to 0.3 obtained for GenASAP values generated from 100,000 trials of an equivalent number of randomly selected probe sets. The accuracy of the %ASex values and rankings generated by GenASAP were next assessed by performing over 200 RT-PCR reactions such that GenASAP values covering the spectrum of AS tissue ranks could be compared directly to corresponding %ASex measurements generated by an independent method. Most of the GenASAP values chosen for validation by RT-PCR assays were automatically selected by using a criterion that ensured a broad range of inclusion to exclusion for the same AS event across the ten tissues. In particular, most of the examples were selected such that at least one %ASex level was less than 33% and at least one was greater than 67%. A number of other interesting examples were selected manually. Figure 2A shows an example of the RT-PCR data representing a single AS event monitored across the ten mouse tissues. Data for additional reactions is shown in Supplemental Figure S2 and Figure 2E (see below).

There is a close correlation ($PCC > 0.8$) between the %ASex values generated by GenASAP and the RT-PCR assays for data that fall within the top 16,000 GenASAP

AS tissue ranks. Figure 2B shows a scatter plot comparing the GenASAP and RT-PCR %ASex values that have a AS tissue rank between 1 to 10,000, and Figure 2C shows the correlation levels when comparing GenASAP and corresponding RT-PCR %ASex values for AS events with AS tissue ranks from 23 to 29,053. Allowing for an error margin of $\pm 15\%$, the estimated cumulative false-positive rate (i.e., the rate at which the GenASAP %ASex value is more or less than $\pm 15\%$ from the value determined by RT-PCR) is less than 20% for GenASAP data with AS tissue ranks from 1 to 10,000. The cumulative false-positive rate rises to $\sim 30\%$ for GenASAP data with AS tissue ranks from 1 to 16,000, and to $\sim 40\%$ for data with AS tissue ranks near 30,000 (Figure 2D). These results demonstrate that the combined AS microarray and GenASAP system provide a reliable method for the quantitative profiling of thousands of alternative splicing events in mammalian tissues. For all subsequent analyses described below, we utilized GenASAP %ASex values with AS tissue ranks of 1 to 16,000 or, if analyzing AS events across all ten tissues, GenASAP %ASex values with AS ranks of 1 to 1600. The full dataset of GenASAP %AS values and their associated rankings can be found in Supplemental Table S1.

AS Microarray versus cDNA/EST Sequence Data for Determining Tissue-Specific AS Levels

Because the AS events represented on the microarray were identified through EST and cDNA sequence data, it might be expected that sufficient numbers of EST and cDNA sequences from the same ten tissue sources used for microarray hybridization would permit %ASex levels

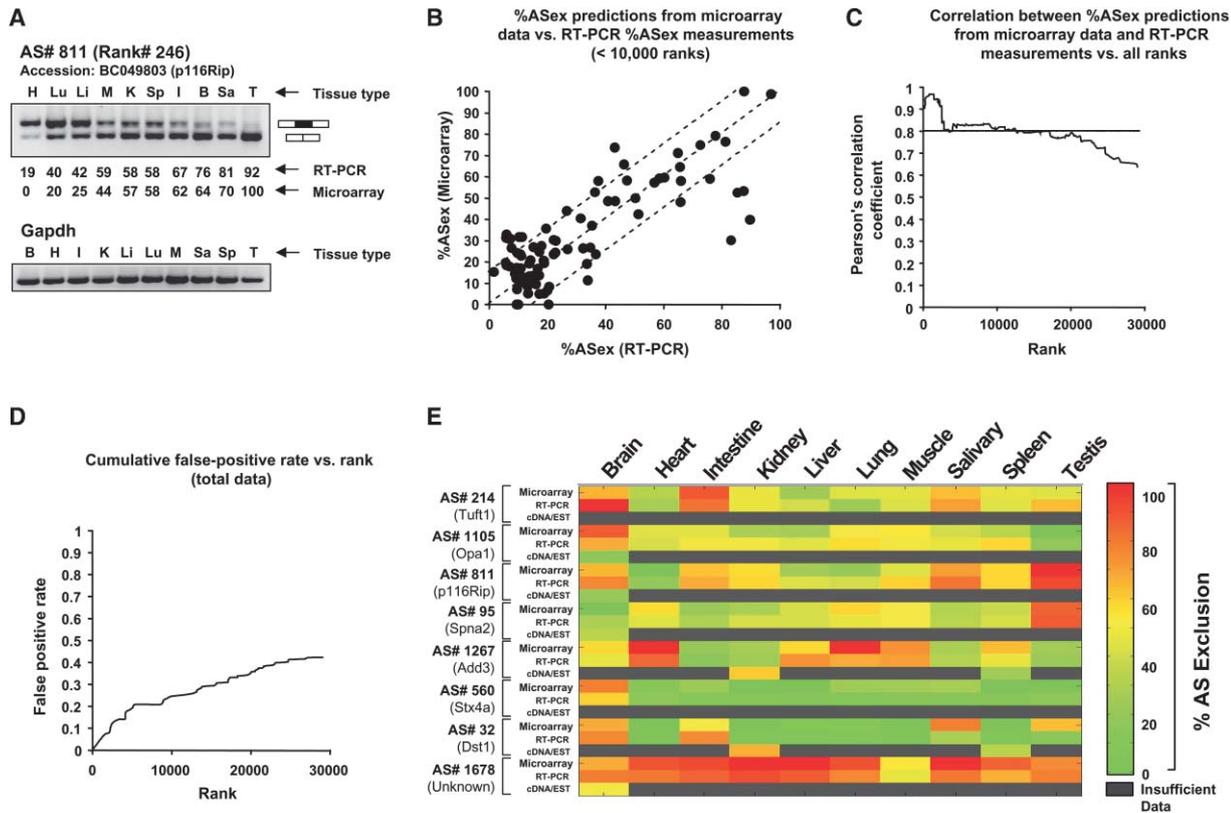


Figure 2. RT-PCR Validation of the GenASAP System for Generating the %ASex values from Microarray Data

(A) Over 200 RT-PCR reactions were performed to test GenASAP-derived %ASex values generated from the AS microarray data from ten major mouse tissues. A representative set of RT-PCR reactions are shown for a single cassette exon skipping event located in the Rho-interacting/F-actin binding cytoskeletal protein (P116Rip) gene (see Supplemental Figure S3 and legend for more information on this gene and AS event). The RT-PCR reactions were ordered from left to right according to the tissues showing lowest to highest %ASex levels values obtained from GenASAP. GenASAP and RT-PCR %ASex values for each of the ten tissues, and the “AS rank” are indicated. An additional 70 representative RT-PCR reactions are shown in Supplemental Figure S2. Positive control RT-PCR reactions, using mRNA from the same tissues, were performed with primers specific for Gapdh, which is expressed at comparable levels among the tissues profiled.

(B) Scatter plot showing %ASex values generated by the GenASAP algorithm versus %ASex measurements from RT-PCR assays (a subset of which are shown in [A] and in Supplemental Figure S2). %ASex measurements ranking in the top 10,000 (out of 31,260) portion of the data are plotted. Dotted lines indicate the $\pm 15\%$ boundaries.

(C) Correlation analysis of %ASex values generated by GenASAP and from >200 RT-PCR reactions, across the range of 31,126 GenASAP “tissue AS ranks”. GenASAP AS tissue ranks between 1-16,000 displayed a Pearson correlation coefficient (PCC) of >0.8, and were used for the majority of the subsequent analyses.

(D) Cumulative false-positive detection rate (allowing a $\pm 15\%$ margin of error) versus ranking for %ASex values generated by GenASAP.

(E) Comparison of 80 %ASex values from microarray data (generated by using GenASAP), from the RT-PCR assays shown in Supplemental Figure S2 and, where possible, from clusters of cDNA/EST sequences from the same mouse tissues. The %ASex values in each case were converted to a color scale (shown on the right). Percent ASex measurements from cDNA/EST clusters were only included if three or more sequences were available from the same adult mouse tissues. The gray bars indicate insufficient data (i.e., less than three cDNA/EST sequences were available).

to be measured independently. In order to assess the extent to which this might be possible, cDNAs and ESTs were sorted from the same ten tissues analyzed on the microarray (refer to the Experimental Procedures). The %ASex values were then determined from the ratio of sequences with or without inclusion of the alternative exon. A %ASex measurement was only made if there were at least three cDNA/EST sequences from the appropriate tissue. Using this criterion, we identified sufficient numbers of cDNA/EST sequences to generate measurements for only 2336 (7.5%) of the total 31,260 tissue-specific %ASex measurements from the microarray data, and 1818 (11.4%) of these fell within the top 16,000 ranked AS events (Supplemental Table S2). In

order to assess to what extent the measurements of %ASex based on cDNA/EST sequences potentially compare with %ASex values from GenASAP and RT-PCR assays, we compared available cDNA/EST-derived measurements for 80 of the tissue-specific AS events analyzed by RT-PCR assays (see Supplemental Figure S2).

Figure 2E shows the %ASex values (converted to color scales) for all three methods. Consistent with the limited tissue coverage of cDNAs/ESTs found in the total dataset, only eight (10%) of the 80 tissue-specific %ASex measurements were represented by three or more cDNAs/ESTs in the equivalent tissue. Four of these cases were in adult brain, consistent with the relatively

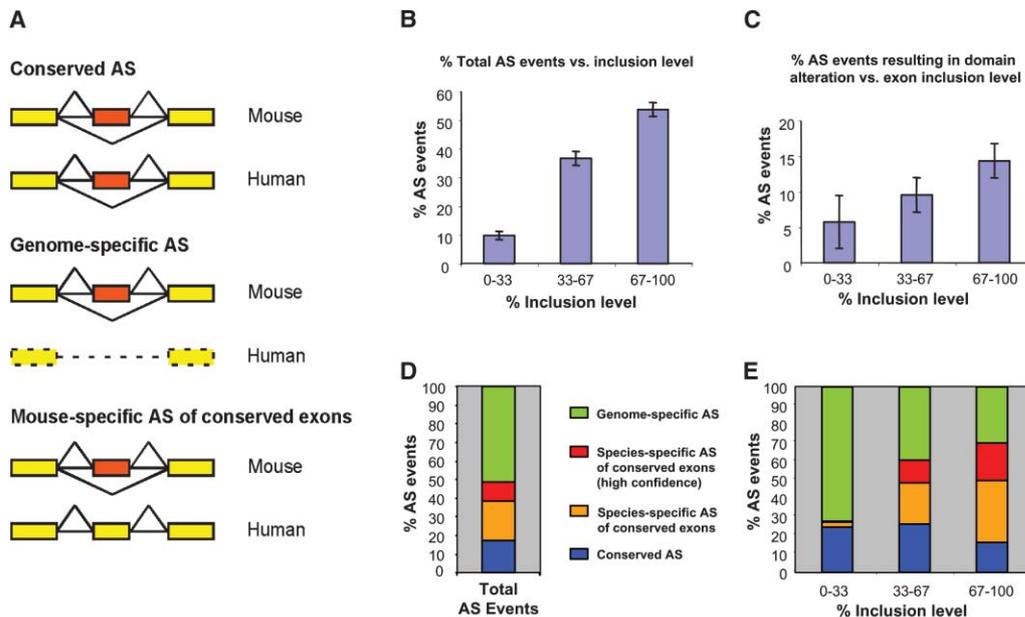


Figure 3. Evolutionary Features of Exons Impacting on Global AS Levels in Mouse Tissues and Implications for Protein Structure and Function (A) Alternative exons represented on the microarray were classified as: conserved AS (detected in mouse and human transcript sequences), genome-specific AS (detected at the transcript and genomic level in mouse but not human), or mouse-specific AS of conserved exons events (a mouse AS event is detected as a constitutive splicing event in human), based on a scoring scheme comparing relative cDNA/EST coverage of the ortholog exons between mouse and human (refer to main text and Experimental Procedures for more information). High-confidence cases have a $\sim 100\%$ probability of being correctly identified as species-specific AS of conserved exons events based on the scoring system, as validated by RT-PCR reactions analyzing these events in human and mouse (data not shown; Pan et al., 2005). (B) The percentage of the total AS events monitored on the microarray, which, based on the GenASAP %ASex data, display inclusion levels within the ranges indicated. (C) The percentage of alternative exons at different inclusion levels that have the potential to alter (i.e., upon skipping modify, disrupt, or create) conserved protein domains. (D) The percentage of the total AS events monitored on the microarray that correspond to each of the different classes of AS exon shown in (A). (E) The percentage of alternative exons at different inclusion levels which correspond to the types of alternative exons shown in (A).

high number of cDNAs/ESTs from this tissue source. However, with the exception of the cDNA/EST measurements for AS#32 and AS#95 in spleen and brain, respectively, which agree fairly well with microarray and RT-PCR data, the other six %ASex levels measured using cDNAs/ESTs did not correspond closely to either the GenASAP or the RT-PCR %ASex values, which do correlate well with each other. Based on these observations, it is apparent that the vast majority of our microarray measurements for tissue-specific %ASex levels cannot be derived independently from the analysis of cDNA and EST sequences. Moreover, the fact that the AS events represented on our microarray represent diverse cellular functions (as determined by GO-BP labels; refer to Experimental Procedures) indicates that our method for selection of AS events for microarray profiling generally is not biased toward the detection of AS events in specific tissues or in specific gene functional categories. These features, together with accuracy of the GenASAP data, permit the global regulation of AS to be investigated.

Evolutionary Features of Alternative Exons Influence Their Inclusion Levels in Adult Tissues

Using the microarray platform described above, we determined how the origin of cassette alternative exons impacts on their average inclusion levels in mouse tis-

sues. We also determined the extent to which AS of these exons in different tissues results in alterations to conserved protein domain coding sequences. In this analysis, we separated AS events into three classes: (1) genome-specific AS events (i.e., the alternative exon is detected in genome and cDNA/EST sequences in mouse but not in human), (2) conserved AS events (i.e., the same AS event is detected in cDNA/EST sequences in both human and mouse), and (3) mouse-specific AS of conserved exons events (i.e., cDNA/EST sequence data indicates that a conserved exon undergoes AS in mouse and constitutive splicing in human) (see Figure 3A). We identified high-confidence cases in the final category using a scoring system based on relative cDNA/EST coverage of human and mouse ortholog exons (Pan et al., 2005; refer to the Experimental Procedures). In this parallel study, 100% RT-PCR-verification was observed for the high-confidence cases identified using the scoring system.

The majority (53.6%) of the cassette AS exons monitored using the AS microarray are highly ($\geq 67\%$) included across the ten mouse tissues, whereas fewer (36.6%) are included at intermediate (33%–67%) levels and even fewer (9.8%) are included at low ($\leq 33\%$) levels (Figure 3B). Similar results were observed in each individual tissue (data not shown). Thus, at steady state, the majority of cassette AS exons do not undergo pro-

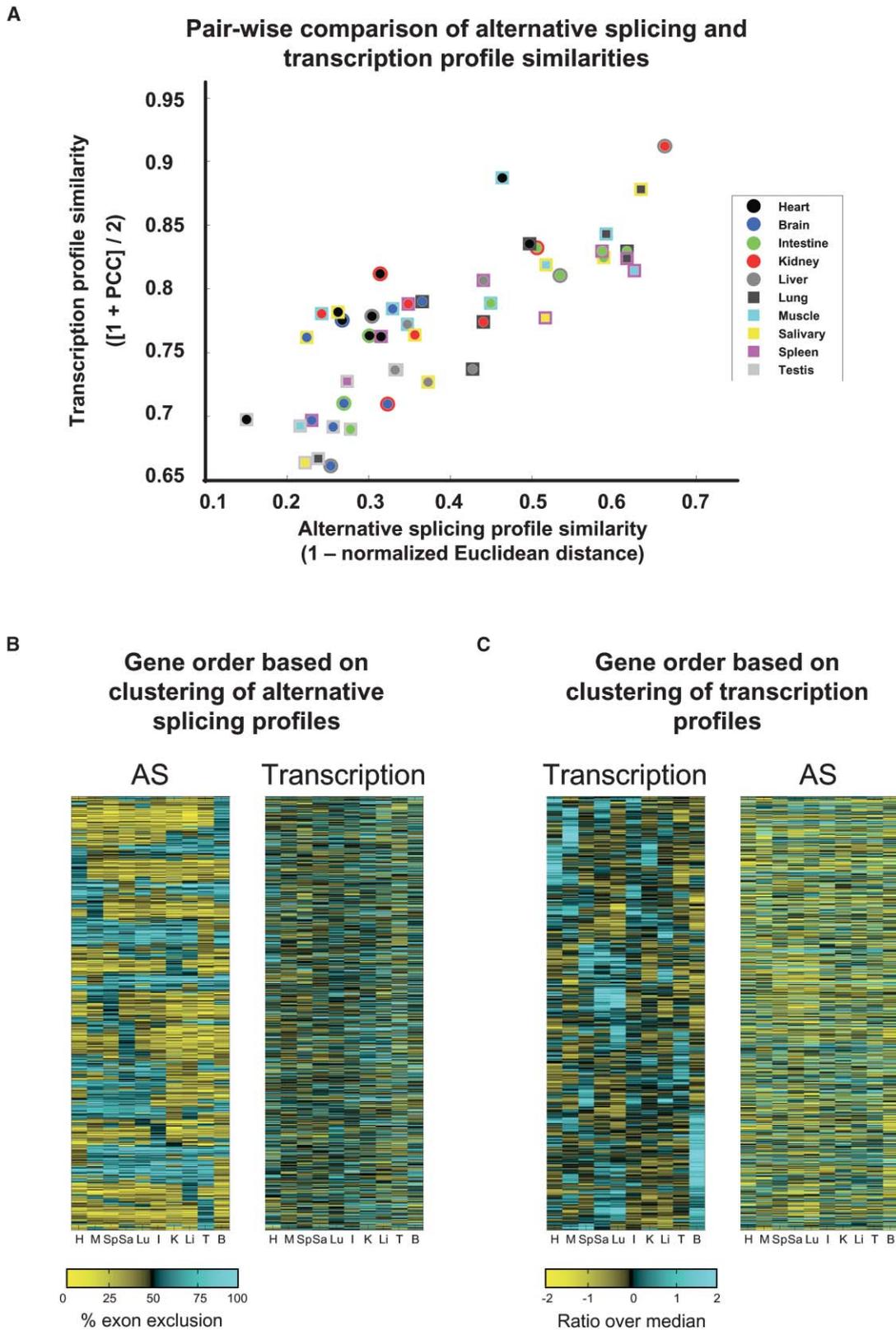


Figure 4. Independent Roles for AS and Transcription in the Generation of Tissue-Specific Gene Expression Profiles

(A) The level of similarity between AS and transcription profiles for every pairwise combination of the ten tissues analyzed in this study were compared. AS profiles were obtained from GenASAP values for each tissue (top 1600 AS ranks), and transcription profiles were measured using the average values from the C1 and C2 probes for the same set of genes (see Figure 1A). Each tissue is assigned a color and shape, as indicated, and each point with a different combination of colored shapes represents the two tissues being compared. The resulting scatter plot shows the level of similarity between the AS profiles on the x axis and the level of similarity between the transcription profiles on the y

nounced skipping in major mouse tissues. Surprisingly however, there is a parallel increase between the steady-state inclusion levels of alternative exons and the frequency at which they alter (i.e., modify, delete, or create) conserved structural and functional domains in proteins (Figure 3C; see discussion below).

We observe significant differences in the frequency at which the different classes of alternative exon described above are included at high, intermediate, and low levels (Figures 3D and 3E). The majority (>70%) of alternative exons included at low steady-state levels are represented by “genome-specific AS events,” whereas fewer (~24%) of these AS exons are represented by “conserved AS events,” and even fewer (3.2%) are represented by “mouse-specific AS of conserved exons events.” Alternative exons included at intermediate, steady-state levels across tissues are represented to similar extents by these three classes of AS events. In contrast, ~70% of highly included exons are represented by AS events that involve conserved exons, the majority of which correspond to mouse-specific AS of conserved exons events, of which at least 20% of the total are in the high-confidence category (Figure 3E).

These results confirm experimentally a conclusion based on the computational analyses of cDNA/EST data from mixed normal and tumor/cell line sources (Modrek and Lee, 2003), that genome-specific AS exons are generally weakly included, and further demonstrate that this property applies in normal adult tissues. Our results extend the observations in this previous study by demonstrating experimentally that there are significant differences in the inclusion levels of conserved exons depending on whether they correspond to conserved AS events or mouse-specific AS events. Moreover, we observe that cassette alternative exons that are highly included at steady-state levels, which primarily correspond to the species-specific AS of conserved exons class, tend to target conserved protein domains more frequently than alternative exons included at low, steady-state levels. Thus, in the majority of cases, AS events in adult tissues would not be expected to affect the major fraction of transcripts and may therefore play a predominant role in the fine-tuning of structural and functional activities associated with conserved protein domains. Nevertheless, because AS events that involve highly included exons appear more often to target conserved domains in proteins, they may be responsible for relatively frequent modulation of tissue-specific activities associated with these domains. In addition, because many of the highly included alternative exons belong to the mouse-specific AS of conserved exons

class, it is possible that this class of AS event plays a significant role in defining mouse-specific characteristics.

Alternative Splicing and Transcription Primarily Regulate Independent Sets of Genes to Generate Tissue Specificity

We find that AS profiles in mouse tissues, like transcription profiles, reflect tissue type. Previous global analyses of human and mouse mRNA abundance levels using microarrays have revealed that more closely related tissues (i.e., sharing common physiologies) in general have more closely correlated transcription profiles than less related tissues (Miki et al., 2001; Su et al., 2004; Zhang et al., 2004). This is also apparent in our data, where transcription profiles (measured using the C1 and C2 probes; refer to Experimental Procedures) for the top 1600 ranking GenASAP genes are correlated to a greater extent for pairs of tissues with more closely related physiologies (i.e., kidney + liver and heart + skeletal muscle) than less related tissues. To directly compare how overall AS and transcription profiles compare between tissues, we coplotted the similarity levels between tissue %ASex profiles (from GenASAP data) and the similarity levels between transcription profiles for every pairwise combination of the tissues analyzed with the AS microarray. Figure 4A shows the resulting pairwise comparison plot, where every tissue combination is color coded as indicated. Notably, it is apparent that pairs of tissues with more closely correlated transcription profiles also have more closely related AS profiles (seen as the overall grouping of points along the diagonal line in the plot in Figure 4A). Thus, overall, in a manner similar to transcription profiles, AS profiles reflect tissue identity.

However, an important question relating to the global regulation of gene expression is whether transcription and AS are coordinated to control specific sets of common genes to achieve tissue specificity, or whether these steps in gene expression primarily act independently on separate sets of genes. To address this question, we ordered the genes by clustering them according to their AS profiles and then compared both the AS profiles and the transcriptional profiles using the same ordering. As shown in Figure 4B, groupings of coregulated AS profiles are not preserved in transcription profiles. Because the genes used in this analysis represent %ASex values with correlation coefficients >0.8 between microarray and RT-PCR data, this result is not due to noise but due to differences between gene-specific AS and transcription profiles. We repeated this pro-

axis, for each pair of tissues. Transcription profile similarity is measured using the Pearson correlation coefficient (PCC). AS profile similarity is measured as $1 - \text{normalized Euclidean distance}$, reflecting consistency of overall AS levels. Pairs of tissues that have the most similar AS and transcription profiles between pairs of tissues are located at the upper-right portion of the plot (e.g., liver + kidney and heart + skeletal muscle).

(B and C) Hierarchical agglomerative clustering, followed by optimal leaf ordering (Bar-Joseph et al., 2001), was used to cluster genes (within the top 1600 GenASAP AS ranks). Only the 745 genes showing a change in %ASex of greater or less than 50% between any pair of tissues were included in the clustering analysis. In the left panel in (B), GenASAP %ASex values were clustered based on the PCC of binarized profiles (thresholded at 50%) of different AS events (rows) and tissues (columns) monitored on the microarray. Transcription levels are shown in the adjacent panel for the same genes, maintaining the same order in the vertical axis as shown in the left panel. In the left panel of (C), genes were clustered according to PCCs of the transformed transcription profiles. The right panel shows %ASex levels for the same genes, maintaining the same order in the vertical axis as shown in the left panel.

Table 1. Analysis of Overlap between Genes that Define Tissue-Specific AS and Transcription Profiles

Tissue Pair Compared	Number of Unique Genes (and GO-BP Categories) in AS Profile	Number of Unique Genes (and GO-BP Categories) in Transcription Profile	Number of Overlapping Genes (and GO-BP Categories)
Liver + Kidney	89 (206)	89 (227)	11 (50)
Heart + Muscle	82 (230)	82 (169)	18 (90)
Lung + Salivary	86 (225)	86 (223)	14 (64)
Testis + Spleen	87 (222)	87 (237)	13 (81)

Sets of 100 genes, from the four tissue pairs that display the most closely correlated AS and transcription profiles shown in Figure 4A, were selected and analyzed to determine the extent of their overlap. Available GO-BP annotations for these gene sets were also analyzed to determine the extent of their overlap. The numbers of unique and overlapping genes and GO-BPs are shown.

cedure, clustering genes according to their transcription profiles, and found that groupings of coregulated transcription profiles generally are not preserved in AS profiles (Figure 4C). In both pairs of clustergrams, tissues are arranged in the same order along the horizontal axis, according to their overall profile similarities. It is evident from comparing these clustergrams that there is very little, if any, correspondence between the clustering patterns obtained for AS and transcription levels. Thus, genes that cluster according to correlated AS levels do not display parallel clusters, indicating that the same genes are significantly coregulated at the transcription level, and the converse situation is also true.

We next asked to what extent genes that define tissue-specific profile similarities at the levels of AS and transcription overlap. To address this question, we selected sets of 100 genes in the data that most strongly define tissue-specific profiles at both the transcription and AS levels (refer to Experimental Procedures). In this case, the gene sets were selected from the tissue pairs that display the most closely correlated AS and transcription profiles (which include liver + kidney, heart + skeletal muscle, lung + salivary gland, and spleen + testis), as indicated by their clustering in the upper-right-most quadrant of the scatter-plot in Figure 4A. The gene sets for each tissue pair were next analyzed to determine the extent of their overlap. As shown in Table 1, there is a very similar extent of gene overlap for each pair of tissues. The overlap ranges from 11 genes between liver + kidney to 18 genes between heart + skeletal muscle, and the average overlap for the four tissue pairs is 14 genes. This corresponds closely to the mean overlap of 12.2 genes obtained from a hypergeometric distribution, which represents the overlap obtained from independent selections of genes from the same dataset. This result is consistent with the clustergram analysis in Figures 4B and 4C and suggests that AS and transcription primarily operate as independent processes on different sets of genes in order to establish tissue-specific expression profiles.

By using the available GO-BP annotations associated with the genes in our dataset, we also asked whether the sets of 100 genes that most define the tissue-specific AS and transcription profiles in Figure 4A are represented by functional categories that overlap more or less than would be expected based on random sampling. The numbers of different GO-BP labels associated with the genes are similar between the tissue pairs, and the overlap is also comparable (Table 1). The overlap of GO-BP annotations is not significantly different than

would be expected based on the overlap between randomly selected sets of genes (data not shown). This observation indicates that cassette exon AS and transcription also primarily act independently on different functional processes to generate tissue-specific expression patterns.

In conclusion, these results provide quantitative evidence that regulated transcription and AS of cassette alternative exons operate primarily as independent processes on genes to establish adult tissue-specific gene expression patterns.

Inferring the Functions of Tissue-Specific Alternative Splicing

Besides affording new insights into global aspects of AS regulation, our microarray data also provides a useful resource for generating experimentally testable hypotheses on the functional roles of tissue-specific AS in genes of interest. In particular, functional inferences can be readily generated if information is available on the structure and/or function of the protein (or RNA) sequence overlapping the AS event monitored on the microarray. To this end, we have provided a comprehensive database of conserved protein domains that are modified by tissue-specific AS, as detected in the ten mouse tissues surveyed in this study (see Supplemental Table S3). Representative examples of inferences generated from the data are shown in Supplemental Figure S3 (see legend for more information) and are discussed below. We have also clustered the microarray data to identify examples of coordinated AS events that are specific to individual tissues or pairs of related tissues. An example cluster containing brain-specific exon skipping events is shown in Figure 5, and selected examples from this cluster that allow interesting functional inferences are also discussed below. Thus, from profiling the inclusion levels of thousands of cassette-type alternative exons in major mouse tissues, we have been able to discern new features of global gene regulation, as well as to generate a database that can be mined by researchers who wish to explore the structural and functional consequences of tissue-specific changes in alternative splicing, in genes of interest.

Discussion

In this report, we describe a custom microarray and associated data analysis tool that permits the reliable quantification of thousands of alternative exon inclusion levels in mammalian cells and tissues. By using this

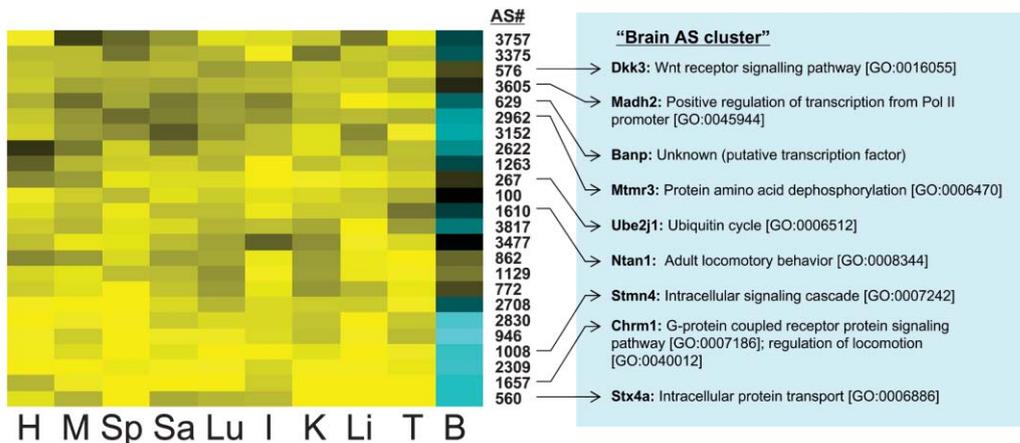


Figure 5. Identification of Brain-Specific AS Events from Microarray Data

A cluster of brain-specific exon skipping events is shown. The names and associated GO-BP annotations for representative genes within this cluster are indicated. Tissue abbreviations are the same as those in Figure 2. Selected AS events in the brain AS cluster were confirmed by RT-PCR assays to have increased exon skipping, specifically in brain tissue (see list in Figure; Supplemental Figure S2; data not shown). Functional implications of a subset of the RT-PCR-verified brain-specific skipping events discovered in the cluster are described in the Discussion section. %ASex levels are indicated using the same color scale as shown in Figure 4B.

microarray system, we have elucidated global regulatory features of alternative splicing (AS) and have identified thousands of tissue-specific AS events in functionally diverse genes. The results shed light on evolutionary features of alternative exons that determine their inclusion levels in major mammalian tissues, information on how AS operates in conjunction with transcription to generate overall adult tissue-specific gene expression profiles, as well as many new inferences for the functional roles of tissue-specific AS. These results thus demonstrate the application of a microarray system for addressing important questions in the postgenomic era, namely, the location, extent, and functional consequences of tissue-specific AS events.

A Microarray System for the Quantification of AS Levels in Mammals

Several microarray-based methods have been reported previously for the monitoring and discovery of RNA processing events (see also Introduction). Fu and colleagues described a fiber optic microarray system for monitoring several AS events in different human cell lines (Yeakley et al., 2002). Ares and colleagues used spotted oligonucleotide microarrays, employing exon and junction-specific probes, to monitor globally the constitutive splicing of yeast introns in different mutant backgrounds (Clark et al., 2002). Hughes and colleagues extended this approach to monitoring the processing of yeast noncoding RNAs (Peng et al., 2003). Changes in AS, primarily in the 3' UTR regions of rat transcripts have been retroactively inferred from data generated from an Affymetrix microarray designed for profiling transcription (Hu et al., 2001). Various researchers have demonstrated the use of “exon tiling” arrays to measure exon signals across human genes, although, to date, this approach has not yielded significant information on AS (Shoemaker et al., 2001; Kapranov et al., 2002; Rinn et al., 2003). Recently, researchers at Rosetta Inpharmatics used custom “ink-jet”-printed microarrays (Hughes et al., 2001) containing probes targeted to exon-exon junc-

tions of RefSeq genes in order to discover new AS events in human transcripts (Johnson et al., 2003). This system was capable of monitoring exon skipping as a “yes/no event”, with a false-positive detection rate of >50%, and comparisons of multiple adjacent splice junction signals between tissues were required to detect a skipping event.

The microarray system described in the present study, which also employs the ink-jet printing technology developed at Rosetta Inpharmatics, differs from the previous approaches in several important ways and offers several advantages. First, our microarray specifically monitors the inclusion levels of sequence-verified AS events, such that a higher density of AS events can be monitored simultaneously on a single array. Because the microarray-monitored AS events were selected from transcripts expressed in a wide range of cell and tissue types and are associated with functionally diverse genes, the resulting data is unlikely to be significantly biased toward the detection of AS events in specific tissues or in specific functional classes of genes (e.g., see Figure 2E). Second, the use of sets of T_m -matched probes to both exon body and splice-junction sequences, representing both skipped and included mRNA isoforms (see Figure 1), significantly improves the detection specificity and permits the quantification of AS levels. This approach also circumvents the requirement for detection of additional flanking exons/splice-junction signals to monitor exon skipping events. Third, the development of a method for data analysis, GenASAP, based on unsupervised learning and Bayesian inference, was successfully applied to modeling of the microarray data and permitted the prediction of AS levels with considerable accuracy. Based on validation experiments using semiquantitative RT-PCR assays, our mouse AS microarray and GenASAP system generated approximately 16,000 values for exon exclusion levels across ten major mouse tissues, that are expected to agree well with actual %ASex levels. The generation of quantitative information on AS using our new microarray

system permitted the discovery of new global properties of AS.

Global Features of AS Levels in Mammalian Tissues and Consequences for the Regulation of Protein Structure and Function

A surprising feature of the majority of cassette-type AS exons surveyed in major mouse tissues is that they are in general highly (>67%) included at steady state levels. However, by separating the cassette alternative exons represented on the microarray according to those that are conserved (in human) and those that are mouse specific, we observed significant differences in the behavior of each class. In this case, by using a recently devised scoring system based on comparisons of cDNA and EST coverage of ortholog exons between human and mouse (see Experimental Procedures; Pan et al., 2005), we have classified species-specific AS events into two types: (1) genome-specific AS, which correspond to alternative exons that are detected at the transcript and genomic level in mouse only, and (2), species-specific AS of conserved exons, which correspond to exons that undergo AS in mouse, but constitutive splicing in human (refer to Figure 3A).

Consistent with the results of Lee and colleagues (Modrek and Lee, 2003), based on computational analyses of collections of mRNA and EST sequences combined from normal and tumor/cell-line derived sources, we find that genome-specific AS events are generally weakly (<33%) included. In contrast, we find that the species-specific AS of conserved exons class (Figure 3A), which were not analyzed in the Modrek and Lee study, are generally highly included and that conserved AS events, which are those detected in both mouse and human transcript sequences, on average display intermediate (33%–67%) inclusion levels across mouse tissues. Thus, the microarray data generated in the present study confirm and extend previous findings, based on computational analyses, relating the evolutionary history of exons to their inclusion levels. It is important to emphasize in this case that the present study represents the use of an experimental system to address these questions, and that it provides global information on inclusion levels of different classes of cassette alternative exons in “normal” mammalian tissues.

Although the majority of cassette AS events do not undergo pronounced changes in inclusion levels between the adult tissues that we have surveyed, we observe that alternative exons that are on average highly included are more likely to target structurally and functionally conserved regions in proteins. This suggests that a predominant role for the regulation of tissue-specific AS in adult tissues is to fine-tune the activities of proteins associated with these domains. Moreover, since this type of tissue-specific regulation often involves conserved exons that undergo mouse-specific AS, it is possible that many species-specific characteristics could arise from relatively subtle but frequent changes in AS involving conserved exons, as well as genome-specific exons. These findings, which are supported by a parallel computational analysis of cDNA and EST sequences (Pan et al., 2005), represent experimental confirmation that regions in transcripts encoding

conserved domains in proteins are relatively frequently targeted by alternative cassette exons that do not undergo pronounced differential inclusion levels in adult mammalian tissues.

Independent Roles for AS and Transcription in the Generation of Tissue-Specific Gene Expression Profiles

Our data reveal that overall, AS and transcription profiles both reflect tissue identity (Figure 4). This finding is consistent with a conclusion from the study of Johnson et al. (2002), who found that exon skipping events inferred from junction probe microarray data were more similar in physiologically-related tissues, than in distantly related tissues. However, the global transcription and AS profiles from the same dataset were not directly compared in this previous study. Because we have focused our study on determining the inclusion levels of alternative exons in differentiated adult tissues, we asked to what extent steady-state AS profiles that most define tissue identity occur in genes that are also regulated to define tissue identity at the transcription level. We find that the overlap between the genes that are regulated at the levels of AS and transcription to define tissue type follows a hypergeometric distribution, which represents the overlap obtained from independent selections of sets of genes from the same dataset. Similar results were obtained when surveying GO-BP categories corresponding to the gene sets that define tissue-specific transcription and AS profiles, which also did not show a significant difference in overlap compared to what would be expected based on independent sampling of GO-BP-annotated genes from the larger gene set. Thus, despite the fact that AS and transcription act in parallel to define tissue-specific profiles (Figure 4A), regulation of AS of cassette exons and regulation of transcription appear to act on separate genes and processes to contribute tissue specificity in adult mammalian organs (Figures 4B and 4C, and Table 1). It is therefore interesting to consider that AS may provide an additional layer of control, primarily on independent sets of genes, to refine tissue-specific expression patterns established at the level of transcription.

By using the system we have described in this report, it will be of interest to determine to what extent transcription and AS might be coordinated in other types of regulatory situations, for example, involving transient changes in cell physiology in response to specific signaling events, and involving other types of AS events, including alternative 5' and 3' splice site selection. The microarray system we have described in the present study provides the opportunity to address these and other important questions concerning the global regulation of AS.

Functional Inferences

Our microarray data has also provided a wealth of new information on the potential roles of tissue-specific AS in mammals, which can be used as a resource by investigators who wish to investigate experimentally the regulation of genes of interest. To this end, we have created a database that can be used to retrieve information on the tissue-specific levels of AS of exons of query genes,

and the conserved protein domain sequences that overlap or are proximal to each AS event represented on the microarray. By using this information, we have already noted numerous cases of AS events that target different proteins associated with the cytoskeleton (e.g., see Supplemental Figure S3 and legend), and can therefore infer from these observations that regulated AS likely performs important roles in establishing differences in cytoskeletal architecture that contribute to defining tissue characteristics.

By clustering tissue-specific AS events, we have identified many examples of widely expressed genes that contain alternative exons uniquely skipped in specific adult tissues. For example, in the brain AS cluster shown in Figure 5 are examples of widely expressed genes that contain alternative exons displaying pronounced skipping only in brain. One of these AS events corresponds to exon 3 in *Madh2* (also known as *Smad2*), which is a transcription factor that transduces signaling pathways by transforming growth factor (TGF)- β , and thus regulates multiple important cellular processes, including cell proliferation, apoptosis, and differentiation (ten Dijke and Hill, 2004). Skipping of exon 3 results in the creation of an intact MH1 (DNA binding) domain and is a key event in promoting activity and altered specificity of *Madh2* for gene targets (Yagi et al., 1999). To our knowledge, an increased level of skipping of exon 3 in brain tissue has not been reported, allowing us to speculate that the activity and target specificity of *Madh2* may be very different in brain compared to other tissues. Among the other genes in the brain AS cluster are *Dickkopf-3* (*Dkk3*), which has been implicated in brain development (Diep et al., 2004), and the N-terminal-asparagine-amidase (*Ntan1*), which functions in regulated protein turnover by the N-terminal rule pathway. Interestingly, mice deficient for *Ntan1* have altered activity, social behavior, and spatial memory ability (Balogh et al., 2000, 2001; Grigoryev et al., 1996; Kwon et al., 2000). Thus, it is possible to infer that these newly identified brain-specific exon skipping events contribute to controlling the neural-specific activities of these widely expressed genes. Such information generated by our microarray data will facilitate new experimentation directed at defining how important tissue-specific activities of proteins are established and controlled in mammals.

Experimental Procedures

Computational Identification and Analysis of AS Events from cDNA and EST Databases

AS Identification and Filtering

AS events were identified by clustering cDNA and EST sequences to contigs of exons extracted from alignments of transcript and genomic sequences using Sim4 (Florea et al., 1998). Details of databases and filtering criteria used are provided in the Supplemental Data.

Sorting of ESTs According to Tissue Sources

Cell and tissue source information, where available, was retrieved from the GenBank flat file for cDNA sequences and from the EST Library Browser for EST sequences by using expression searches with manually compiled key words for 54 different cell/tissue source categories (Pan et al., 2005).

Identification of Conserved and Mouse-Specific

Alternative Exons

To determine whether the mouse AS events were conserved in human, we Blasted "probe" sequences corresponding to included

and skipped isoform sequences against the human mRNA/EST sequences (<ftp://ftp.ncbi.nih.gov/repository/UniGene/>), by using an e value cut-off of $1e-3$ (refer to Supplemental for additional details).

AS Gene Annotation and Detection of Protein Domains Altered by AS

Retrieval and analysis of GO-BP categories and MGI markers associated with the AS genes profiled on the microarray is described in the Supplemental Data. GO annotations are provided in Supplemental Table S4. Skipped and included mRNA isoform sequences were translated into all possible reading frames by using the EMBOSS getorf application (Rice et al., 2000) and searched against the Conserved Domain Database (Marchler-Bauer et al., 2003) by using an e value cut-off of 0.001. Domain "alteration" was scored when a domain was detected in one splice isoform but not the other. A full list of annotated domains altered by the 3126 AS events monitored on the microarray is shown in Table S3.

Custom Microarray Design

Oligonucleotide Probe Design

Exon and flanking intron sequences were repeat masked through RepeatMasker (<http://ftp.genome.washington.edu/RM/RepeatMasker.html>). A program was developed to automatically select optimal oligonucleotide probe sequences with the following criteria: (1) exon body and intron probes have a matching T_m of $66^\circ\text{C} \pm 2^\circ\text{C}$, and junction probes are positioned over a splice junction such that the T_m on either side of the junction is equal, and the overall T_m is $66^\circ\text{C} \pm 2^\circ\text{C}$; (2) each probe sequence has a similar AT/GC base composition and simple repeats, especially GGG and CCC are avoided; (3) sequences with self-annealing potential are avoided; (4) sequences with four or more consecutive internal base pairings are avoided; and (5) probe sequences have the maximal differential between the target sequence and second-best match from BLAST searches of mouse sequence databases are prioritized. Probe sets represented on the microarray were prioritized according to a ranking scheme based on the design parameters listed above. No more than 12 T's were added to the 3'-end of each probe sequence; the total length of the probe sequence cannot be more than 60 nt. Microarrays were manufactured by Agilent Technologies.

Microarray Hybridization, Scanning, Data Extraction, and Normalization

Labeling and Hybridization

cDNA labeling and microarray hybridization was performed essentially as described by Zhang et al. (2004). All hybridizations were performed in duplicate with fluor reversal, and microarrays were scanned with a 4000A microarray scanner (Axon Instruments, Union City, California).

Image Processing and Normalization

TIFF images were quantitated with GenePix 3.0 (Axon Instruments). Individual channels were spatially denoised (i.e., overall correlations between spot intensity and position on the slide removed) by using spatial trend removal (STR) with 10% outliers (Shai, Morris, and Frey, in preparation). We applied variance stabilizing normalization (VSN) by using 25% of the genes to normalize all single channels to each other. Finally, we manually identified and removed data from residual artifacts apparent on microarray images and measurements that showed high inconsistency between dye-swaps.

RT-PCR Assays

Primer pairs were designed to have a matching T_m (59°C) and were targeted to constant exon sequences flanking each alternative exon. RT-PCR assays were performed with the OneStep RT-PCR Kit (Qiagen). Reactions were performed in 25 μl volumes containing 1.0 ng polyA⁺ mRNA, 7.5 units porcine RNAGuard (Amersham), and 600 pM each of the forward and reverse primers. After 30 rounds of amplification, the reaction products were separated on 2% agarose gels stained with ethidium bromide. Several representative AS genes were assayed over a dilution range of input mRNA in order to assess the linearity of the amplifications. The ratios of exclusion to inclusion were generally maintained over the titration ranges. Inverted black and white images of the gels were recorded with a Syngene gel documentation system and GeneSnap software.

AS Microarray Data Analysis

Algorithm Development

Details of the GenASAP algorithm used to analyze the AS microarray data are provided in a technical report (www.psi.utoronto.ca/~ofer/AS-supp.pdf) and in the Supplemental Data.

Comparison of Gene Sets Defining Tissue-Specific

AS and Transcription Profiles

The top 1600 ranked AS events were filtered to remove events for which no GO-BP annotation was available, and only one AS event per gene was analyzed. 820 AS events remained, representing 769 GO-BP annotations, after up-propagation. Using a "greedy" algorithm, the top 100 AS events were selected whose profiles most closely reflect tissue similarity between (1) kidney and liver, (2) heart and muscle, (3) lung and salivary, and (4) testis and spleen, while at the same time distinguishing these tissue-pair profiles from the rest of the tissues. (Further details of the greedy approach are given in Supplemental Data.)

Acknowledgments

Correspondence relating to GenASAP can be addressed to B.J.F. and Q.D.M. We thank Richard Aronovitz, John Calarco, Richard Collins, Jim Friesen, Christos Ouzounis, and Susan McCracken for helpful discussions and comments on the manuscript. We also thank Ursula Skalska and Sanie Mnaimneh for their excellent assistance. This work was funded by the Canadian Institutes of Health Research Proof-of-Principle and Operating grants (to B.J.B. and T.R.H.), Ontario Premier's Research Excellence Awards (to B.J.F. and B.J.B.), a National Cancer Institute of Canada operating grant (to B.J.B.), and Canadian Foundation for Innovation infrastructure grants (to T.R.H., B.J.F. and B.J.B.). We also thank Jim Friesen for his generous support during the course of this work.

Received: August 9, 2004

Revised: November 18, 2004

Accepted: December 8, 2004

Published: December 21, 2004

References

- Babak, T., Zhang, W., Morris, Q., Blencowe, B.J., and Hughes, T.R. (2004). Probing miRNAs with microarrays: tissue specificity and functional inference. *RNA* 10, 1813–1819.
- Balogh, S.A., McDowell, C.S., Tae Kwon, Y., and Denenberg, V.H. (2001). Facilitated stimulus-response associative learning and long-term memory in mice lacking the NTAN1 amidase of the N-end rule pathway. *Brain Res.* 892, 336–343.
- Balogh, S.A., Kwon, Y.T., and Denenberg, V.H. (2000). Varying intertrial interval reveals temporally defined memory deficits and enhancements in NTAN1-deficient mice. *Learn. Mem.* 7, 279–286.
- Bar-Joseph, Z., Gifford, D.K., and Jaakkola, T.S. (2001). Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics Suppl.* 17, S22–S29.
- Black, D.L. (2000). Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* 103, 367–370.
- Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* 72, 291–336.
- Blencowe, B.J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.* 25, 106–110.
- Brett, D., Pospisil, H., Valcarcel, J., Reich, J., and Bork, P. (2002). Alternative splicing and genome complexity. *Nat. Genet.* 30, 29–30.
- Burge, C., Tuschl, T., and Sharp, P.A. (1999). Splicing of Precursors to mRNAs by the Spliceosomes, Second Edition (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).
- Caceres, J.F., and Kornblihtt, A.R. (2002). Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet.* 18, 186–193.
- Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.* 3, 285–298.
- Clark, T.A., Sugnet, C.W., and Ares, M., Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* 296, 907–910.
- Croft, L., Schandorff, S., Clark, F., Burrage, K., Arctander, P., and Mattick, J.S. (2000). ISIS, the intron information system, reveals the high frequency of alternative splicing in the human genome. *Nat. Genet.* 24, 340–341.
- Diep, D.B., Hoen, N., Backman, M., Machon, O., and Krauss, S. (2004). Characterization of the Wnt antagonists and their response to conditionally activated Wnt signalling in the developing mouse forebrain. *Brain Res. Dev. Brain Res.* 153, 261–270.
- Faustino, N.A., and Cooper, T.A. (2003). Pre-mRNA splicing and human disease. *Genes Dev.* 17, 419–437.
- Florea, L., Hartzell, G., Zhang, Z., Rubin, G.M., and Miller, W. (1998). A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* 8, 967–974.
- Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. *RNA* 6, 1197–1211.
- Graveley, B.R. (2001). Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17, 100–107.
- Grigoryev, S., Stewart, A.E., Kwon, Y.T., Arfin, S.M., Bradshaw, R.A., Jenkins, N.A., Copeland, N.G., and Varshavsky, A. (1996). A mouse amidase specific for N-terminal asparagine. The gene, the enzyme, and their function in the N-end rule pathway. *J. Biol. Chem.* 271, 28521–28532.
- Hu, G.K., Madore, S.J., Moldover, B., Jatke, T., Balaban, D., Thomas, J., and Wang, Y. (2001). Predicting splice variant from DNA chip expression data. *Genome Res.* 11, 1237–1245.
- Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* 19, 342–347.
- Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. (2003). Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302, 2141–2144.
- Kan, Z., Rouchka, E.C., Gish, W.R., and States, D.J. (2001). Gene structure prediction and alternative splicing analysis using. *Genome Res.* 11, 889–900.
- Kapranov, P., Cawley, S.E., Drenkow, J., Bekiryanov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296, 916–919.
- Kramer, A. (1996). The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.* 65, 367–409.
- Kwon, Y.T., Balogh, S.A., Davydov, I.V., Kashina, A.S., Yoon, J.K., Xie, Y., Gaur, A., Hyde, L., Denenberg, V.H., and Varshavsky, A. (2000). Altered activity, social behavior, and spatial memory in mice lacking the NTAN1p amidase and the asparagine branch of the N-end rule pathway. *Mol. Cell. Biol.* 20, 4135–4148.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lee, C., and Roy, M. (2004). Analysis of alternative splicing with microarrays: successes and challenges. *Genome Biol.* 5, 231.
- Marchler-Bauer, A., Anderson, J.B., DeWeese-Scott, C., Fedorova, N.D., Geer, L.Y., He, S., Hurwitz, D.I., Jackson, J.D., Jacobs, A.R., Lanczycki, C.J., et al. (2003). CDD: a curated Entrez database of conserved domain alignments. *Nucleic Acids Res.* 31, 383–387.
- Miki, R., Kadota, K., Bono, H., Mizuno, Y., Tomaru, Y., Carninci, P., Itoh, M., Shibata, K., Kawai, J., Konno, H., et al. (2001). Delineating developmental and metabolic pathways in vivo by expression profiling using the RIKEN set of 18,816 full-length enriched mouse cDNA arrays. *Proc. Natl. Acad. Sci. USA* 98, 2199–2204.

- Mironov, A.A., Fickett, J.W., and Gelfand, M.S. (1999). Frequent alternative splicing of human genes. *Genome Res.* 9, 1288–1293.
- Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. *Nat. Genet.* 30, 13–19.
- Modrek, B., and Lee, C.J. (2003). Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat. Genet.* 34, 177–180.
- Nurtdinov, R.N., Artamonova, I.I., Mironov, A.A., and Gelfand, M.S. (2003). Low conservation of alternative splicing patterns in the human and mouse genomes. *Hum. Mol. Genet.* 12, 1313–1320.
- Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., et al. (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420, 563–573.
- Pan, Q., Bakowski, M.A., Morris, Q., Zhang, W., Frey, B.J., Hughes, T.R., and Blencowe, B.J. (2005). Alternative splicing of conserved exons is frequently species-specific in human and mouse. *Trends Genet.*, in press.
- Peng, W.T., Robinson, M.D., Mnaimneh, S., Krogan, N.J., Cagney, G., Morris, Q., Davierwala, A.P., Grigull, J., Yang, X., Zhang, W., et al. (2003). A panoramic view of yeast noncoding RNA processing. *Cell* 113, 919–933.
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277.
- Rinn, J.L., Euskirchen, G., Bertone, P., Martone, R., Luscombe, N.M., Hartman, S., Harrison, P.M., Nelson, F.K., Miller, P., Gerstein, M., et al. (2003). The transcriptional activity of human Chromosome 22. *Genes Dev.* 17, 529–540.
- Shoemaker, D.D., Schadt, E.E., Armour, C.D., He, Y.D., Garrett-Engle, P., McDonagh, P.D., Loerch, P.M., Leonardson, A., Lum, P.Y., Cavet, G., et al. (2001). Experimental annotation of the human genome using microarray technology. *Nature* 409, 922–927.
- Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* 25, 381–388.
- Sorek, R., Shamir, R., and Ast, G. (2004). How prevalent is functional alternative splicing in the human genome? *Trends Genet.* 20, 68–71.
- Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* 101, 6062–6067.
- ten Dijke, P., and Hill, C.S. (2004). New insights into TGF-beta-Smad signalling. *Trends Biochem. Sci.* 29, 265–273.
- Thanaraj, T.A., Clark, F., and Muilu, J. (2003). Conservation of human alternative splice events in mouse. *Nucleic Acids Res.* 31, 2544–2552.
- Wang, H., Hubbell, E., Hu, J.S., Mei, G., Cline, M., Lu, G., Clark, T., Siani-Rose, M.A., Ares, M., Kulp, D.C., and Haussler, D. (2003). Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics Suppl.* 19, i315–i322.
- Will, C.L., and Luhmann, R. (1997). Protein functions in pre-mRNA splicing. *Curr. Opin. Cell Biol.* 9, 320–328.
- Xu, Q., Modrek, B., and Lee, C. (2002). Genome-wide detection of tissue-specific alternative splicing in the human. *Nucleic Acids Res.* 30, 3754–3766.
- Yagi, K., Goto, D., Hamamoto, T., Takenoshita, S., Kato, M., and Miyazono, K. (1999). Alternatively spliced variant of Smad2 lacking exon 3. Comparison with wild-type Smad2 and Smad3. *J. Biol. Chem.* 274, 703–709.
- Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S., and Fu, X.D. (2002). Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.* 20, 353–358.
- Zhang, W., Morris, Q., Chang, R., Shai, O., Bakowski, M., Mitsakakis, N., Mohammad, N., Robinson, M., Eftekharpour, E., Grigull, J., et al. (2004). The functional landscape of mouse gene expression. *J. Biol.* 3, 21.
- Zhu, J., Shendure, J., Mitra, R.D., and Church, G.M. (2003). Single molecule profiling of alternative pre-mRNA splicing. *Science* 301, 836–838.

Accession Numbers

All microarray data have been deposited at GEO under accession number GSE2054.