Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

SplicerEX: A tool for the automated detection and classification of mRNA changes from conventional and splice-sensitive microarray expression data

TIMOTHY J. ROBINSON,[1] ELEONORA FORTE,[2,6] RAUL E. SALINAS,[2] SHAAN PURI,[2] MATTHEW MARENGO,[2,3,7] MARIANO A. GARCIA-BLANCO,[2,3,4,5] and MICAH A. LUFTIG[2,3,4,5,8]

[1]Molecular Cancer Biology Program, [2]Department of Molecular Genetics and Microbiology, [3]Center for RNA Biology, [4]Department of Medicine, and [5]Center for Virology, Duke University Medical Center, Durham, North Carolina 27710, USA

ABSTRACT

The key postulate that one gene encodes one protein has been overhauled with the discovery that one gene can generate multiple RNA transcripts through alternative mRNA processing. In this study, we describe SplicerEX, a novel and uniquely motivated algorithm designed for experimental biologists that (1) detects widespread changes in mRNA isoforms from both conventional and splice sensitive microarray data, (2) automatically categorizes mechanistic changes in mRNA processing, and (3) mitigates known technological artifacts of exon array-based detection of alternative splicing resulting from 5′ and 3′ signal attenuation, background detection limits, and saturation of probe set signal intensity. In this study, we used SplicerEX to compare conventional and exon-based Affymetrix microarray data in a model of EBV transformation of primary human B cells. We demonstrated superior detection of 3′-located changes in mRNA processing by the Affymetrix U133 GeneChip relative to the Human Exon Array. SplicerEX-identified exon-level changes in the EBV infection model were confirmed by RT-PCR and revealed a novel set of EBV-regulated mRNA isoform changes in caspases 6, 7, and 8. Finally, SplicerEX as compared with MiDAS analysis of publicly available microarray data provided more efficiently categorized mRNA isoform changes with a significantly higher proportion of hits supported by previously annotated alternative processing events. Therefore, SplicerEX provides an important tool for the biologist interested in studying changes in mRNA isoform usage from conventional or splice-sensitive microarray platforms, especially considering the expansive amount of archival microarray data generated over the past decade. SplicerEX is freely available upon request.

Keywords: SplicerEX; alternative splicing; exon arrays; gene expression; software; alternative mRNA processing

INTRODUCTION

The key postulate that one gene encodes one polypeptide chain (one enzyme) has been overhauled with the discovery that one gene can generate multiple RNA transcripts (and indirectly many different polypeptide chains) through a process referred to as ''alternative mRNA processing'' (Blencowe 2006). Alternative processing defines a range of events, including alternative splicing (AS) and alternative polyadenylation (APA), that result in distinct mRNA species. Recent deep sequencing studies indicate that 94% of all protein-coding genes generate multiple mRNA transcripts (Wang et al. 2008). A substantial proportion of human genetic diseases are likely caused by aberrant mRNA splicing (Krawczak et al. 1992; Lopez-Bigas et al. 2006). Functional consequences of alternative processing have been shown across a wide variety of biological processes including drug metabolism, stem cell self-renewal, neurological disease, autoimmune disease, and cancer (for review, see Garcia-Blanco et al. 2004; Venables 2006; Cooper et al. 2009). Our understanding of the global regulation of alternative processing remains incomplete (Takeda et al. 2006). Despite the known importance of alternative processing in cancer, our ability to fully harness alternative processing as a tool in prognosis, diagnosis, and treatment is limited.

The pace of mRNA processing research has increased following the introduction of new genome-wide approaches to the study of alternative mRNA processing (Wang et al. 2008; Venables et al. 2009; Pickrell et al. 2010), and has been particularly invigorated following the commercial availability of splice-sensitive microarrays such as the Affymetrix Human

[6]**Present address:** Department of Microbiology-Immunology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA
[7]**Present address:** GrassRoots Biotechnology, Durham, NC 27701, USA
[8]**Corresponding author**
E-mail micah.luftig@duke.edu
Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.033621.112.

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

Exon 1.0 ST array (HuEx) (Gardina et al. 2006; Cheung et al. 2008; Thorsen et al. 2008; Xi et al. 2008). The data generated by these genome-wide transcript assays have been accompanied by a number of analytical approaches and software (Affymetrix 2005; Cline et al. 2005; Sandberg et al. 2008; Xing et al. 2008; Laajala et al. 2009). These algorithms all share a similar statistical approach to evaluate changes in alternative mRNA processing: They evaluate the null hypothesis that all features within a gene are expressed under the same, whole-transcript-based (aka nonalternatively spliced) distribution. Any genes that are not regulated at the whole-transcript level reject this null hypothesis and are considered to be alternatively processed. We have previously developed a fundamentally different method of analyzing alternative mRNA processing, SplicerAV (Pearson et al. 2008; Robinson et al. 2010), in which we explicitly characterize alternative splicing as an alternative hypothesis with specific expectations as to probe set behavior.

In this study, we describe the application and expansion of our previous methodology to exon arrays in the form of SplicerEX, a novel bioinformatics program that explicitly models alternative processing as a separate hypothesis and enables biologists to fine-tune the definition of alternative processing to reduce false positives and assist in identifying experimental models of alternative mRNA processing regulation. SplicerEX is capable of (1) detecting mRNA isoform changes in both conventional gene expression and splice-sensitive microarrays, (2) automatically categorizing changes in mRNA processing into distinct mechanistic and directional groups, and (3) mitigating the false-positive results found among HuEx array probe sets that interrogate the 5′ and 3′ ends of transcripts by preferentially selecting events whose probe set feature expression levels change in opposite directions. We demonstrate these capabilities in the analysis of the transition of resting primary human B cells to EBV-transformed lymphoblastoid cell lines.

RESULTS

The SplicerEX algorithm and defining alternative mRNA isoform usage

Previously, we created a program, SplicerAV, which uses a two Gaussian model to detect changes in mRNA processing from conventional Affymetrix 3′-IVT expression microarrays (Pearson et al. 2008; Robinson et al. 2010). SplicerAV compares changes in gene expression among any and all genes targeted by multiple probe sets and determines whether or not these probe sets exhibit congruous or disparate changes in expression between control and treatment samples. However, for this two Gaussian model to perform feature recognition and automatic categorization of mRNA isoform changes from Human Exon Arrays (HuEx), we had to first devise a method that would efficiently identify the two most distinguishing features of a gene's expression profile across a wide range of feature inputs. Therefore, we used a data reduction method to collapse HuEx probe sets with correlated expression within a given sample into meta-probe sets (Fig. 1A). A preliminary investigation exploring the combination of SplicerAV and collapse of correlated probe sets has been published previously (Miller et al. 2011). However, this previous implementation lacked the ability to generate automated annotation tracks and categorize mRNA isoform changes, which we have now incorporated into SplicerEX.

We tested SplicerEX using the model system of Epstein-Barr virus (EBV) transformation of primary human B cells into indefinitely proliferating lymphoblastoid cell lines (LCLs). Total mRNA from four independent human donors' primary B cells and resultant LCLs was queried using the HuEX and the U133 2.0 Plus Array (U133) and identified a collective 8006 or 8409 expressed genes detected in either state by the HuEx or U133 Arrays, respectively. These genes were detected by a total of 56,598 (HuEX) or 11,985 (U133) expressed probe sets. Collapsing HuEx probe sets by expression correlation reduced the total number of features per gene while preserving the ability to discern isoform-specific expression changes, which for any single alternative processing event can be described using two features

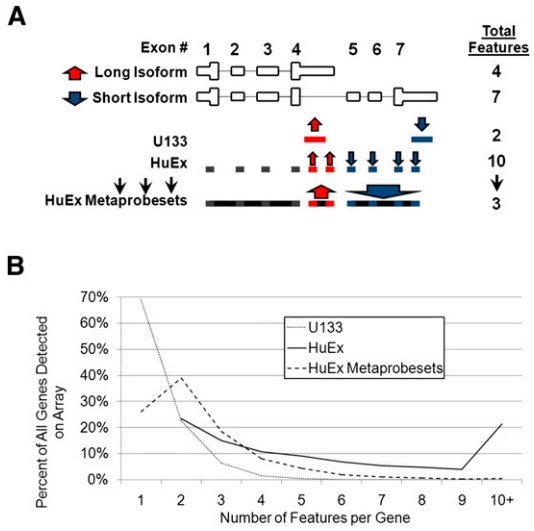

FIGURE 1. SplicerEX application of meta-probe sets to human exon (HuEx) arrays. (*A*) Diagram comparing interrogation by 3′ IVT (i.e., U133) versus HuEx microarrays of a hypothetical gene undergoing alternative 3′-terminal exon choice. U133 arrays possess fewer probe sets per gene than the HuEx array and preferentially target 3′ UTRs of known mRNA transcripts. Collapsing exon array probe sets into meta-probe sets on the basis of correlated expression patterns reduces the number of features per gene to a level similar to 3′-IVT arrays. SplicerEX then identifies the two most divergently expressed features within a gene (red and blue) to locate and characterize structural changes in mRNA isoforms. (*B*) Number of features per gene for the conventional U133, HuEx (Exon Array), and HuEx meta-probe set–preprocessed microarray data. Collapsing exon array features on the basis of correlated expression patterns reduces the number of exon array features per gene to a level similar to that of the U133 array, permitting analysis of both array platforms by a single algorithm.

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

(shown schematically in Fig. 1A). Empirically, we found that a correlation threshold of 0.7 provided the highest concordance of HuEx detected splicing events and previously known alternative splicing events annotated within the UCSC genome browser (data not shown). This resulted in a reduction of the number of HuEx platform features from 56,598 (probe sets) to 22,663 (meta-probe sets) and produced a similar distribution of the numbers of features per gene compared with the U133 platform (Fig. 1B). Therefore, features were defined as probe sets for the U133 arrays and meta-probe sets for the HuEx arrays. Within each alternatively processed gene, the most statistically significant increasing and decreasing features were identified as the "Up" and "Down" features (Fig. 1A). The "Up" and "Down" most significant features are collectively referred to as the "isoform-distinguishing features." The use of two isoform-distinguishing features provided the appropriate summarization of changes in gene structure to permit automated categorization of changes in mRNA processing as either unclassified or into one of four distinct mechanistic categories: (1) alternative 5′ initiation; (2) internal event (cassette exon, intron retention); (3) alternative 3′ terminal exon choice; or (4) use of a tandem 3′ UTR. SplicerEX was further able to differentiate and automate spatially directional changes, reported as the following: (A) lengthened versus shortened 3′-UTR choice; (B) 5′ versus 3′ located terminal exon choice; (C) 5′ versus 3′ located alternative 5′ initiation start site; and (D) inclusion versus exclusion of internal exons. (For additional details, see Materials and Methods.)

The combination of the SplicerAV core model with meta-probe set collapse, selection of primary isoform-distinguishing features, and subsequent automated event categorization and accompanying annotation tracks were combined into a single program, SplicerEX, that is capable of identifying and categorizing changes in alternative mRNA processing in both conventional U133 arrays and higher-feature-content exon arrays.

Affymetrix U133 and HuEx arrays detect nonoverlapping genes undergoing mRNA isoform changes

A total of 5375 genes were detected as being expressed in common between the U133 and HuEx array platforms in either resting human $CD19^+$ B cells or EBV-transformed LCLs (Fig. 2). Of the 5375 genes detected on both arrays, there was considerable overlap between genes that were detected as differentially expressed. A total of 364 genes were significantly induced ($p < 0.01$, fold change >2) on both array platforms, corresponding to 32% of the 1140 genes induced on the U133 and 83% of the 437 genes induced on the HuEx platform ($p < 0.0001$). Similar overlap was found among repressed genes. A total of 109 genes were significantly repressed ($p < 0.01$, fold change >2) on both array platforms, corresponding to 57% of the 191 repressed genes on the U133 and 70% of the 156 repressed genes on the HuEx platform ($p < 0.0001$).

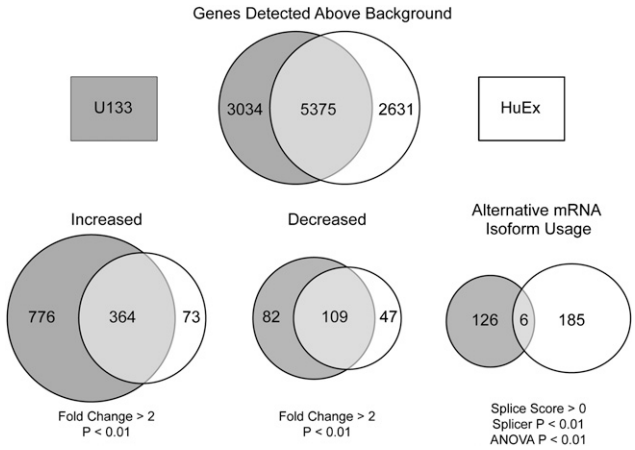

FIGURE 2. U133 and HuEx arrays detect nonoverlapping changes in alternative mRNA processing. Comparison of genes detected, increased, decreased, and alternatively processed by array platform. Comparisons of increased, decreased, and alternatively processed gene lists were limited to genes detected on both the U133 and HuEx platforms. Genes were considered to be differentially expressed with $p < 0.01$ and fold change >2 between LCL (treatment) versus B-cell (reference) samples. Genes were considered to be alternatively processed if they had a positive splice score, SplicerEX p-val <0.01, and ANOVA p-val <0.01.

In contrast, there was no significant overlap between genes that were detected to be alternatively processed by both arrays ($p = 0.33$). Only six genes were independently considered hits on both arrays (splice score >0, splicer $p < 0.01$, ANOVA $p < 0.01$). These six genes corresponded to <5% of the total genes detected by either the U133 (132 total) or HuEx (191) arrays (Fig. 2).

U133 and HuEx detect complementary classes of mRNA isoform changes

The U133 and HuEx platforms exhibited similar performance with regard to the total number of events detected and hypotheses created. In fact, testable hypotheses could be automatically generated for the majority of isoform changes on both the U133 (131/ 162, 81%) and HuEx (205/ 272, 75%) arrays (Supplemental Tables 1, 2). However, to investigate why there was such a poor overlap in alternatively processed mRNAs detected between the two microarray platforms, we plotted the location of SplicerEX-predicted changes in mRNA isoforms by array (Fig. 3). SplicerEX-assigned alternative mRNA processing categories differed significantly by platform ($p < 0.0001$). The U133 array predictions displayed a strong tendency to detect changes in mRNA processing at the 3′ ends of genes, with 93% (122/ 131) of all classified U133 events occurring at the 3′ ends of known UCSC transcripts (alternative 3′-TE choice,

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

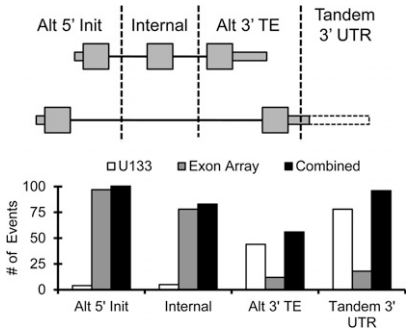

FIGURE 3. HuEx and U133 arrays detect spatially distinct mRNA isoform changes. Comparison of SplicerEX-predicted mRNA isoform changes detected using the U133 versus HuEx array platforms. Alternative 5′-transcript initiation and internal events were preferentially detected by the HuEx array. Internal events were primarily composed of cassette exons, but could also include intron retention and alternative 5′- or 3′-exon definition. Alternative 3′-TE choice and tandem 3′ UTRs were preferentially detected by the U133 array. Alternative 3′-TE choice occurred via alternative polyadenylation or alternative 3′-splice site selection. Tandem 3′ UTRs universally resulted in shortening or lengthening of mRNA 3′-UTR length.

alternative polyadenylation, or 3′-UTR length change). In contrast, HuEx arrays displayed an equally strong tendency to detect events in the 5′ and internal portions of genes, with 85% (175/205) of classified events occurring at the 5′ terminus or within an internal exon. Consistent with these findings, only 38% (50/131) of the hypotheses created by the U133 array were predicted to result in alternative ORF usage and subsequent isoform-specific protein-coding changes, while 88% (180/205) of the HuEx-derived hypotheses predicted protein-coding changes (Supplemental Tables 1, 2).

To further elucidate why U133 versus HuEx platforms detected different events, we used the SplicerEX-generated UCSC Genome Browser annotation tracks to examine the top five scoring hits that (1) corresponded to known alternative splicing events and (2) were uniquely detected as being alternatively processed by either the U133 (Fig. 4A) or HuEx (Fig. 4B) platforms. In all 10 cases, differences in detection of alternative processing events could be attributed to differential probe set coverage by the two platforms. Hits detected uniquely by the U133 platform were predominantly located toward the 3′ of transcripts (Fig. 4A), whereas hits detected uniquely by the HuEx platform were predominantly internal or located toward the 5′ end of transcripts (Fig. 4B), as suggested by the nonoverlapping distribution of automated event types shown in Figure 3.

SplicerEX predicts mRNA isoform changes validated by quantitative RT-PCR

Since our predictions of alternative mRNA isoform usage were based solely on hybridization data, we validated these events using quantitative RT-PCR (qRT-PCR). Specifically, the 20 predicted mRNA isoform changes with the highest splice scores on each platform were queried. Therefore, we analyzed 40 events (one "Up" and one "Down" for each of the 20 mRNAs) from three independent B and LCL pairs. We considered validation as the ability of the SplicerEX algorithm to accurately predict the discordance between the "Up" and "Down" exons. We labeled qRT-PCR reactions for exons predicted to be up-regulated from B to LCL in red and those predicted to be down-regulated in blue. Then, we ranked the fold changes of the 40 B-to-LCL qRT-PCR reactions for each donor and plotted these rankings as a heatmap (Fig. 5A). If all 40 events were accurately predicted, then the top half of the heatmap would be solid red, and the bottom half would be solid blue. As shown in Figure 5A, the vast majority of the "Up" qRT-PCR reactions were ranked in the top half of the heatmap, and,

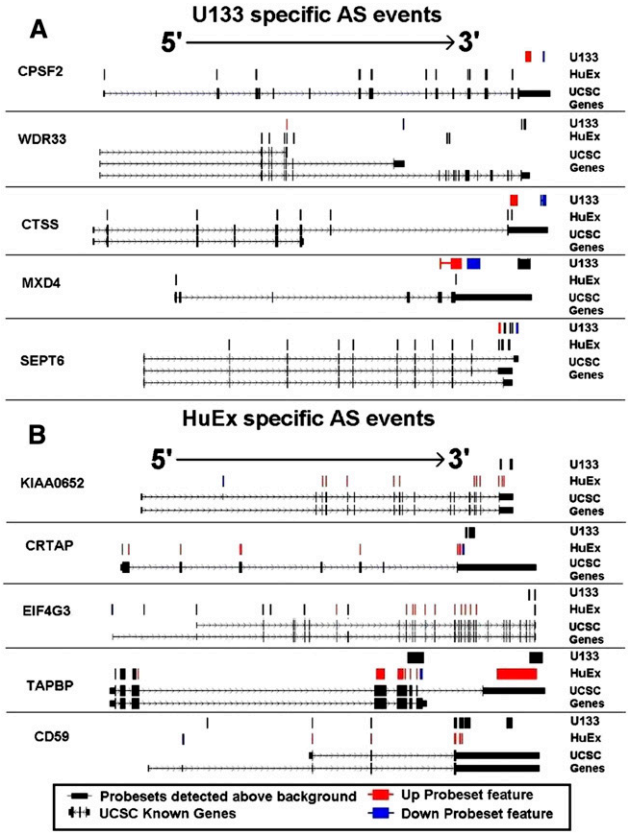

FIGURE 4. Alternative mRNA isoform changes detected by only either U133 (*A*) or HuEx (*B*) platforms. Five of the top alternative isoform usage changes detected by SplicerEX on each platform were examined by plotting the genomic location of all probe sets that overlapped known UCSC gene transcript exons. All such probe sets detected above background are shown *above* the corresponding known UCSC gene transcripts. The most significant increasing meta-probe set ("UP feature") (red); the most significant decreasing meta-probe set ("DOWN feature") (blue); all remaining probe sets (black). U133 specific events (*A*) were more often located in the 3′-UTR region, where HuEx probe sets were unable to detect transcripts. HuEx-specific events (*B*) were more often 5′ and internally located, and were not interrogated by U133 probe sets.

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

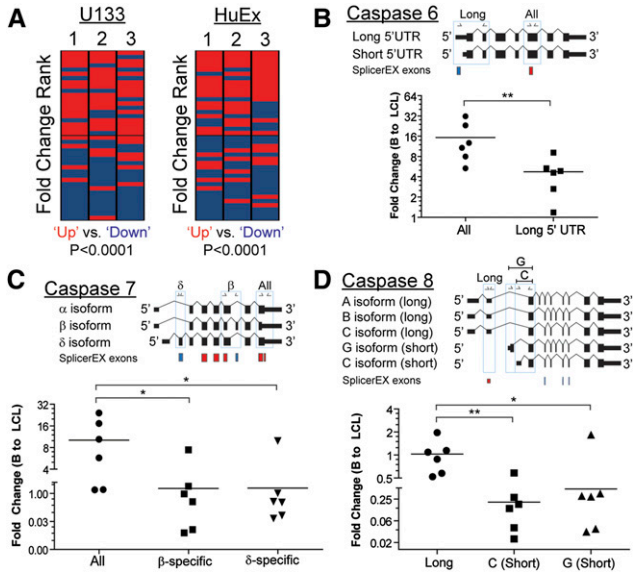

FIGURE 5. Quantitative RT-PCR validation of SplicerEX-predicted U133 and HuEx mRNA isoform changes. (*A*) The heatmaps display the fold-change rank order (from *top* to *bottom*) of 40 qRT-PCR reactions (20 "Up" and 20 "Down" for each gene) on each platform (U133 on *left* and HuEx on *right*) for three independent donors. Each reaction measures the fold change of mRNA between B cell and LCL for either the Up event (red) or the Down events (blue). If all Up reactions were ranked above the Down reactions, then the heatmap would be red on the *top* half and blue on the *bottom* half. The *P*-value shown is from the Mann-Whitney *U*-test, which was used to compare the set of Up reactions to the set of Down reactions. (*B*) A schematic diagram is shown for two caspase 6 mRNA isoforms predicted based on our SplicerEX data. The primer locations for qRT-PCR are indicated *above*, and SplicerEX meta-probe sets are displayed *below* (Up exons [red]; Down exons [blue]). *Below* the schematic is a graph plotting the fold change from B to LCL for qRT-PCR reactions characterizing the isoform change in six independent human donors. (**) $p < 0.01$ (Mann-Whitney). (*C*) As in *B*, a schematic diagram of caspase 7 mRNA isoforms is shown including primers for qRT-PCR and SplicerEX meta-probe sets. The graph indicates B-to-LCL fold-change values for reactions characterizing all isoforms, β-specific isoforms, or δ-specific isoforms. (*) $p < 0.05$ (Mann-Whitney). (*D*) As in *B* and *C*, a schematic diagram of caspase 8 mRNA isoforms is shown including primers and SplicerEX meta-probe sets. The graph plots B-to-LCL fold-change values for reactions characterizing long, C (short)–specific, and G (short)–specific isoforms. (*) $p < 0.05$; (**) $p < 0.01$ (Mann-Whitney).

similarly, nearly all of the "Down" reactions were ranked in the bottom half of the heatmap. Therefore, Splicer EX was able to successfully predict qRT-PCR measured exon-level changes of mRNAs between resting B cells and EBV-transformed LCLs (U133, $p < 0.0001$; HuEx, $p < 0.0001$; Mann-Whitney *U*-test).

EBV infection alters caspase mRNA isoform usage

We next validated SplicerEX-predicted changes for a set of biologically relevant mRNAs. Our analysis of EBV-regulated mRNA isoform changes revealed a strong enrichment in cysteine proteases and, more specifically, in caspases, which are key mediators of programmed cell death, or apoptosis. Given the known role of EBV in mitigating apoptosis (Cahir-McFarland et al. 2000) and also the important role of caspases in B-cell biology (Deming and Rathmell 2006), we further investigated the EBV-regulated caspase mRNA isoform changes. SplicerEX identified exon-level changes for caspases 6, 7, and 8 on the HuEx platform (Supplemental Table 2). We further studied these changes through exon- and isoform-specific qRT-PCR analyses.

The caspase 6 mRNA was predicted to undergo alternative 5′ initiation from resting B cells to LCLs while overall transcript levels increased (Fig. 5B). Indeed, levels of transcripts containing exons present in both isoforms were increased, while those specific to the 5′-most region of the first exon were not as strongly increased from B to LCL ($p < 0.01$). Therefore, EBV increased the overall abundance of a caspase 6 mRNA transcript with a shortened 5′ UTR presumably due to alternative 5′-transcriptional initiation.

SplicerEX predicted an increase from resting B cells to LCLs in both internal and 3′ caspase 7 exons common among mRNAs encoding the α, β, and δ isoforms (Fig. 5C). However, exons specific to the β and δ isoforms were predicted to decrease, suggesting a selective change in caspase 7 mRNA isoform usage regulated by EBV (Fig. 5C). Quantitative RT-PCR for β-specific, δ-specific, and all caspase 7 mRNA isoforms validated these predictions because neither the β-encoding nor the δ-encoding isoforms increased from B to LCL, while a 3′-UTR directed primer set detected an increase in total caspase 7 mRNA (Fig. 5C, $p < 0.05$). These data suggest a specific increase in the caspase 7 α-isoform encoding mRNA upon EBV infection of B cells.

The SplicerEX-predicted caspase 8 isoform changes from B to LCL were consistent with a decrease in expression of short isoforms (encoding protein isoforms G and C), while retaining expression of the long isoforms (encoding protein isoforms A, B, or C) (Fig. 5D). Indeed, we detected a robust EBV-mediated repression of shorter G- and C-encoding mRNA isoforms, while the long A/B/C-encoding mRNAs did not change significantly during EBV immortalization (Fig. 5D). Therefore, despite an apparent decrease in overall mRNA abundance as detected on these arrays, these qRT-PCR experiments clearly indicate decreased expression only of a set of short mRNAs with common 3′ exons (short C and G) and no change in a set of mRNAs with a common set of 5′ exons (long A/B/C). Collectively, these data suggest that SplicerEX was not only successful in predicting exon-level changes globally between resting B cells and EBV-transformed LCLs (Fig. 5A), but a specific subset of isoform-level changes in caspase mRNAs were identified by this approach and subsequently validated using qRT-PCR (Fig. 5B–D).

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

SplicerEX provides automated event categorization and higher concordance with known alternative processing events than comparable MiDAS analysis of HuEx data

One of the main improvements in the SplicerEX algorithm over existing methods is the ability to automatically characterize AS and other processing events, allowing researchers to quickly remove hits that are difficult to interpret biologically. At present, these "hits" compose a large part of the noise detected in AS analyses and can only be removed by tedious, manual curation and data cleaning. We therefore sought to demonstrate the ability of SplicerEX to automatically identify alternative isoform changes and describe mechanistic hypotheses and directionality as compared with a currently widely used algorithm called MiDAS (Della Beffa et al. 2008). We analyzed a publicly available HuEx data set of changes occurring following infection of lymphatic endothelial cells with a virus related to EBV, the Kaposi's sarcoma–associated herpesvirus (KSHV) (Chang et al. 2011). We analyzed the same preprocessed data as Chang and colleagues (GEO: http://www.ncbi.nlm.nih.gov/geo/, GSE26341-GPL5188 series matrix) and applied the same additional criteria that limited analysis to probe sets with an absolute expression of log70 and called splicing events using a *P*-value of <0.05. MiDAS calls alternative processing events at the level of individual exons and therefore can call multiple exons within a single gene, but does not categorize alternative processing events beyond noting whether an exon is included or excluded relative to the remaining exons in a gene. In contrast, SplicerEX performs gene-level alternative isoform analysis and will call only one event per gene, but will categorize this event with regard to mechanism.

In the same data set analyzed by MiDAS, SplicerEX was able to detect a significantly higher proportion of alternative processing events that provided specific, testable hypotheses and that were supported by known AS transcripts (Supplemental Table 3). Globally, MiDAS detected 542 genes with AS events and SplicerEX detected 426, of which 60 genes were in common (11% of MiDAS and 14% of SplicerEX hits, respectively). However, following manual categorization of the MiDAS data, no readily identified testable hypothesis was observed for 261 of the 542 genes (48%). In contrast, automated categorization by SplicerEX failed to identify testable hypotheses for only 95 of the 426 hits (22%). Furthermore, manual inspection of the top 43 SplicerEX versus top 43 MiDAS hits (the top 10% of the 426 Splicer hits) revealed that a significant and substantially higher proportion of SplicerEX hits were supported by previously annotated alternative processing events (SplicerEX, 67% [29/43] vs. MiDAS, 43% [18/43]; χ^2 test, $P = 0.03$).

SplicerEX-based characterization provided a similar distribution of events as the published, in-depth manual expert curation of events detected by MiDAS, supporting the ability of the SplicerEX algorithm to automatically provide similar data about the distribution of AS events. Characterization of the 281 MiDAS events that corresponded to known events was reported as 113 (41%) initiation or promoter usage, 160 (56%) internal events (cassette exons, intron retention, or exons with alternative 5′ or 3′ extensions), 29 (10%) 3′-terminus changes (genes were allowed to possess more than one category and added to >100%). Characterization of the 331 SplicerEX events resulted in a similar distribution of events, with 169 (51%) alternative transcription initiation events, 95 (29%) internal/cassette exon events, and 67 (20%) 3′-terminus changes. As an added benefit, SplicerEX was able to further categorize 3′-located events into tandem 3′-UTR events (15%), the products of alternative 3′ cleavage and polyadenylation, versus alternative 3′-terminal exon usage (5%). It should be noted that SplicerEX categorization also revealed a significant shift toward more upstream initiation sites (63% vs. 37%, $P < 0.0006$) in KSHV-infected cells, which was not appreciated using the MiDAS approach.

DISCUSSION

In this study, we use a novel program, SplicerEX, to characterize the ability of U133 and exon arrays to detect changes in alternative mRNA processing on identical mRNA samples. We found that both the U133 and HuEx platforms were capable of detecting comparable numbers of overall changes in differential mRNA processing, but detected almost no events in common. We show that the low overlap between the platforms is predominantly due to biases in transcript interrogation and subsequently the types of mRNA isoform changes detected by each array. We found that the U133 array was superior to the HuEx platform for detecting changes in both 3′-UTR length (78 vs. 18 events) and 3′-TE choice (44 vs. 12 events). This study suggests that the U133 2.0 Plus Array, originally designed to interrogate overall gene expression, is currently the most sensitive Affymetrix microarray for detecting differential processing of 3′-transcript regions. The U133 array design preferentially targets probe sets toward the 3′ ends of genes. Among U133 probe sets that target any known UCSC gene, we found that 90% interrogated a 3′-terminal exon. U133 arrays use cDNA prepared using oligo(dT) reverse transcription, which provides the strongest amplification of the 3′ ends of transcripts.

In contrast, a surprising finding of this analysis was the relative inability of HuEx arrays to detect differential processing at the 3′ ends of genes. There are several likely explanations as to why this might be the case. An analysis by Bemmo et al. (2008) examined the ability of HuEx arrays to detect changes in differential mRNA processing and found that a large number of false-positive events were detected at the 5′ and 3′ ends of genes. The investigators demonstrated that the signal strength of the HuEx arrays was particularly weak at the 3′ ends of genes and hypothesized that the lack of signal was due to a consequence of using

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

random primed cDNAs. A study by Robinson and Speed (2007) suggested that individual HuEx array probe set signals are less reliable than U133 probe sets as a result of smaller probe feature size and the use of fewer probes per probe set. Lastly, it is possible that HuEx array probe set target locations within the genome do not represent all 3′ UTRs. However, we found in preliminary analyses that HuEx probe sets interrogated 90% of all U133 probe set target sequences, which would not account for the fourfold differential discovery rate of 3′-located events observed in this study.

The analysis presented here suggests that researchers deciding between U133 and HuEx arrays should choose the platform based on their specific research objectives. U133 arrays focused on changes in 3′-terminal processing may be of particular interest in studying gene regulation by miRNAs, known to largely target the 3′ UTRs of most transcripts (Sandberg et al. 2008; Friedman et al. 2009). Conversely, HuEx arrays could be used to focus on changes in 5′-transcript initiation that may be of interest to those studying the use of alternative promoters. Internal processing events, composed largely of cassette exons, may be of interest to those interested in identifying splicing events that result in changes in ORFs and resultant encoded protein structure. We found that events detected using the HuEx array predicted more than three times as many protein-coding isoform changes as the U133 array.

Several programs have been developed to analyze Affymetrix exon array data for changes in alternative mRNA processing. Of these programs, PLATA (Sandberg et al. 2008) has previously been used to calculate changes in tandem 3′-UTR length within proliferating cell types, but is not capable of otherwise characterizing differential processing events. Previous algorithms have been used to deconvolute the relative abundance of specific splice variants (Li and Wong 2001; Wang et al. 2003). However, these algorithms have only been applied to a handful of well-characterized genes and are not applicable to commercially available microarrays. SplicerEX remains the only program available capable of automatically categorizing differential mRNA processing events using commercially available microarrays. We have shown in this analysis a 67% versus 42% correspondence of SplicerEX versus MiDAS predicted events that can be directly supported by previously annotated alternative processing events, and found a similar advantage in initial investigations of SplicerAV applied to exon arrays compared with X-Ray (62% vs. 20%) (Miller et al. 2011). These observations suggest that SplicerEX is a robust algorithm for the identification of known as well as previously unidentified mRNA isoform changes. Importantly, the automated characterization of exon-level changes provides specific hypotheses that can be experimentally validated.

We used a qRT-PCR approach to both broadly validate SplicerEX-generated hits and specifically confirm hypotheses in this newly identified set of EBV-regulated mRNA isoform changes in primary B cells. We found that, collectively, EBV-regulated "Up" exons were significantly different in expression level from "Down" exons. To specifically address hypotheses predicted by SplicerEX, we interrogated isoform-level changes in caspases 6, 7, and 8. For each gene, we validated the predicted hypotheses and thereby identified a set of unique isoform-level changes in these critical cell death-regulatory proteins. The biological significance of these changes will be addressed in the future, but these findings suggest that EBV exerts a previously unrecognized level of regulation on the apoptotic pathway.

To our knowledge, the SplicerAV/EX set of programs remains the only programs currently available to analyze differential mRNA processing on U133 arrays at the level of the probe set (Robinson et al. 2010). Future work to automate and facilitate analysis of complex changes in alternative mRNA processing remains a promising and active area of research and holds the potential to improve our understanding of the regulation of alternative mRNA processing and its role in human disease. SplicerEX is freely available upon request.

MATERIALS AND METHODS

Preprocessing and implementation details

SplicerEX takes normalized probe set intensities as input, which can be generated using several existing software options. Readily available options for probe set level normalization of U133 and exon array data include the Affymetrix Expression Console, bioconductor R packages, Partek, XRAY, and others. For our analysis, we used RMA sketch quantile normalization within the freely available Affymetrix Power Tools software suite. We found that the correspondence between predicted and known alternative mRNA isoform changes was best when limiting U133 data to probe sets with a mean $\log_2$ expression >6 and HuEx probe sets with a mean $\log_2$ expression >8. HuEx genes were additionally required to have at least two expressed probe sets to be included in the analysis.

To improve the quality of hypotheses generated by SplicerEX, the program can limit the analysis to a user-specified set of probe sets. We have generated default probe set lists for both the U133 and HuEx platforms that limit analysis to probe sets that overlap one or more UCSC gene transcripts. Overlap was determined using Affymetrix annotated target probe set sequence coordinates, comparing them to the March 2006 (hg18) version of the human genome, and checking for overlap with known UCSC genes. Genomic coordinates of U133 probe set target coordinates were not publicly available from Affymetrix or from the UCSC Genome Browser and were generated using BLAT to align Affymetrix probe set sequences against the March 2006 genome; they are provided as Supplemental Table 4.

SplicerEX was implemented in Perl for use on a personal computer (PC). Typical run times are 3–5 min for U133 2.0 Plus data and 10–15 min for Exon Array data using meta-probe set features. Perl is a freely available programming language that is available for most operating systems.

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

SplicerAV base model and generation of splice score

SplicerEX is based on the underlying statistical model provided by SplicerAV, described in detail previously, and should be referenced for further details (Robinson et al. 2010). In brief, SplicerEX uses a maximum likelihood ratio (MLR) to compare the relative likelihood that changes in probe set expression levels are described by alternatively processing versus overall transcription level changes, which are respectively calculated using a one versus two Gaussian mixture model. This base MLR is then modified by user set parameters to generate the final splice score: a multiple probe set correction to adjust for total possible paired groupings of probe sets, an expression cutoff modifier to specify the minimum change required between isoforms, and a centering modifier to preferentially rank genes whose probe set expression levels change in opposite directions. The multiple probe set correction allows for the user to focus on less complicated AS events, the expression cutoff allows the user to focus on events with a prespecified effect size, and the centering modifier provides a means of reducing false positives caused by attenuation of probe set expression levels due to either signal saturation or 5′- and 3′-located detection sensitivity (Bemmo et al. 2008).

Meta-probe set implementation and selection of correlation threshold

Meta-probe set feature selection was accomplished in several steps. First, pairwise Pearson correlations were calculated between all probe sets targeting the same gene. If this correlation exceeded a user set threshold, these two probe sets would be joined together into a single probe set. Correlations between the joined probe sets and remaining probe sets would then be averaged to create a new pairwise correlation matrix. The process was then repeated until no remaining features were correlated above the set threshold. The resulting features, each made up of one or multiple probe sets, constituted a meta-probe set. Meta-probe set collapse using an empirically derived Pearson correlation threshold of 0.7 maximized our ability to accurately categorize mRNA processing events and greatly reduced the average number of features per gene (Fig. 1B).

SplicerEX categorization decision tree

We devised an algorithm to categorize SplicerEX-predicted changes in alternative mRNA processing into distinct mechanistic and directional categories. This algorithm was created with two main goals: (1) to provide biologically useful distinct categories and (2) to provide an algorithm that was as simple as possible to promote transparency of the method. To categorize AS events, the algorithm uses two sources of input for each gene: (1) the genomic coordinates of the two features used to generate the isoform ratio (see Robinson et al. 2010) and (2) known UCSC gene transcript genomic coordinates. The two features used to generate isoform ratios come from the single most significant probe sets/meta-probe sets found in the A and B probe set groups. The program uses single probe sets as features for the U133 array and single meta-probe sets for the HuEx array. The UCSC known gene list contains predictions based on data from RefSeq, GenBank, CCDS, and UniProt and contains ~10% more protein-coding genes than RefSeq and about twice as many splice variants (Kent et al. 2002).

The categorizer works by stepping through a binary decision tree that asks questions about the two distinguishing probe set features (probe sets for U133 and meta-probe sets for HuEx) and their overlap with known UCSC genes (Fig. 6; Table 1). The categorizer sequentially asks a set of five questions. If the answer to any of these questions is no, the program assigns a category and then terminates. If no category has been assigned after these five questions, the common transcript inference algorithm is used to assign a category. Events placed in the common transcript inference (CTI) category are categorized based on each probe set's exon count and location within the common transcript. CTI events can be categorized as internal events, alternative 5′ initiation, or uncategorized (Table 2). Using the relative 5′ versus 3′ position of the "Up" and "Down" features on known UCSC genes, SplicerEX further differentiates (1) lengthened versus shortened 3′-UTR choice, (2) 5′- versus 3′-located terminal exon choice, (3) 5′- versus 3′-located alternative 5′-initiation start site, and (4) inclusion versus exclusion of internal exons.

Of events not assigned hypotheses, only a handful of events had reasonably clear hypotheses that could be assessed subjectively by the researchers. These additional events could be accurately categorized by SplicerEX with additional modifications to the categorization algorithm. However, it was felt that the increased additional complexity required to categorize these few events was not worth the loss in transparency and ease of implementation. These modifications were not included in the final SplicerEX program.

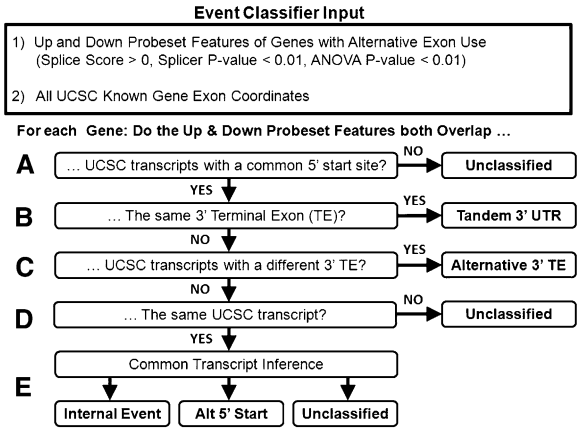

FIGURE 6. SplicerEX decision tree used to categorize alternative mRNA processing events into discrete categories. SplicerEX assigned mechanistic categories of mRNA isoform changes by comparing the location of the two primary distinguishing probe set features (probe sets for the U133 array and meta-probe sets for the HuEx arrays) with known UCSC gene transcripts. The algorithm first removes a gene if either feature does not overlap a known UCSC transcript, resulting in an unclassified event. Second, the event is recorded as unclassified if neither feature overlaps transcripts with a common 5′ start site (*A*). If both features overlap the same 3′-terminal exon, then the event reflects a change in 3′-UTR length, indicating the presence of tandem 3′ UTRs (*B*). If each feature targets a different 3′ terminal exon, then the gene is categorized as undergoing alternative 3′ terminal exon choice (*C*). If the event has not been categorized at this point, but both features target at least one common UCSC transcript (*D*), then the event is categorized as an internal event, an alternative 5′ initiation event, or unclassified using the common transcript inference (CTI) algorithm (*E*) (Table 2).

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

TABLE 1. Categories of mRNA processing events assigned by SplicerEX

Category	Description	Directional subtypes	Optimal detection platform
5′ Initiation	Change in 5′ transcription initiation site Considered to result in coding changes if 5′ isoform possesses three or more exons before the start of the 3′ isoform	Relative increase in 5′-most initiation site (more 5′) Relative decrease in 5′-most initiation site (more 3′)	HuEx
Internal event	Changes in internal exon content Includes primarily cassette exons, alt 5′ SS, alt 3′ SS, and intron retention	Inclusion Exclusion	HuEx
Tandem 3′ terminal exons	Change in length of 3′ TE Almost universally results in noncoding changes in 3′-UTR length	Shorten Lengthen	U133
3′ Terminal exon choice	Change in choice of 3′ TE Almost universally results in protein coding changes	More 5′ More 3′	U133
Unclassified	Unable to be categorized using above	N/A	

Calculations of differential gene expression, differentially processed genes, and gene list overlap

To provide an equitable comparison of the ability of U133 and HuEx arrays to detect changes in both whole-gene and isoform-specific mRNA changes, all analyses directly comparing the two platforms were limited to genes detected above background by both platforms. For the U133 array, differential gene expression was reported using the probe set that reported the most statistically significant change in expression by a simple *t*-test of $\log_2$ expression values between treatments versus controls. For the HuEx array, differential gene expression fold change and significance was calculated by a two-tailed, homoskedastic *t*-test of all of a gene's $\log_2$ probe set expression values in treatments versus controls. Genes were classified as differentially expressed with a *P*-value <0.01 and fold change >2. Genes were classified as differentially processed if they had a positive splice score, SplicerEX *P*-value <0.01, and ANOVA *P*-value <0.01. These criteria used to detect AS events were established and described previously in work involving SplicerEX's precursor, the SplicerAV program (Pearson et al. 2008; Robinson et al. 2010). Comparisons of differential frequencies in mechanistic categories and gene list overlap were made using a two-tailed Fisher's exact test. Enrichment of directional splicing changes (3′-UTR length, internal exon inclusion/exclusion) was performed using a binomial distribution to test the null hypothesis that $p = 0.5$.

EBV-induced B-cell transformation and mRNA preparation

Human B cells were obtained from normal donor buffy coats through the Carolina Red Cross, and peripheral blood mononuclear cells (PBMC) were isolated by Ficoll Hystopaque gradient (Sigma-Aldrich #H8889). $CD19^+$ B cells were purified from PBMC using the BD iMag Negative Isolation Kit (BD, cat. #558007). Purity was routinely >90% as determined by flow cytometry. Total mRNA was prepared from four normal donors in two conditions: (1) uninfected purified $CD19^+$ B cells and (2) monoclonal LCL derived by limiting B95-8 virus dilution on PBMC. B95-8 virus was produced from the B95-8 Z-HT cell line as previously described (Johannsen et al. 2004). Three sets of lymphoblastoid cell lines (LCL) were created from matching B-cell donors, and one set of unmatched LCLs and B cells were also analyzed. RNA from all four LCL samples and all four resting B-cell samples were successfully hybridized to both the U133 2.0 Plus and Human Exon Arrays for a total of 16 independent hybridizations (four LCLs and four B cells on the U133 2.0 Plus Array and four LCLs and four B cells on the Exon Array).

cDNA preparation, labeling, and fragmentation were performed using the GeneChip Wild-Type (WT) cDNA Synthesis and Amplification Kit (Affymetrix cat. #900673) and Exon Array Labeling Kits (Affymetrix cat. #900671). Eight samples (four of

TABLE 2. Common transcript inference (CTI) categories

Category	Criteria	Subtypes	Criteria
Internal event	(1) Probe set 1: Contained internally within probe set 2 OR (2) Probe set 1: Single exon, internal Probe set 2: Multiple exons	Inclusion Exclusion	Probe set 1 is from Group A Probe set 1 is from Group B
5′ Alt initiation	(1) Probe set 1: Single 5′ located exon OR (2) Probe sets 1 and 2 each target three or more exons and do not overlap	More 5′ More 3′	Group A probe set is more 5′ Group A probe set is more 3′
Unclassified/no class	Unable to be categorized using above	N/A	

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

each condition) were hybridized to HuEx 1.0ST Exon Arrays (Affymetrix cat. #900650), and the chips were scanned in the Duke Microarray Facility.

Primer design and quantitative RT-PCR validation of SplicerEX hits

The top 20 SplicerEX hits by splice score were validated for each platform. We designed PCR primers to specifically amplify the exons of the mRNA labeled as the "Up" and "Down" exons by the algorithm (see example in Fig. 4). SplicerEX was able to automatically extract exon coordinate information for the probe sets used on each array. With this information, the differentially detected probe sets between the B and LCL states could be extrapolated to differentially expressed exons, which could then be queried by qRT-PCR. Two sets of primers were designed using IDT PrimerQuest for each "Up" and "Down" exon based on exon–exon junctions (where possible, e.g., cassette exons, but not tandem 3′ UTRs) or simply to specifically amplify the "Up" or "Down" exon only (primer sequences are available upon request). We used total mRNA isolation followed by random hexamer priming to generate cDNA (High Capacity cDNA Kit, Applied Biosystems) from the B and LCL samples for three independent human donors. RT-PCR reactions were performed using SYBR Green (Quanta Biosciences) on the StepOne Plus QPCR instrument (Applied Biosystems) in technical duplicate with the first primer set for each Up and Down exon. If technical replicates achieved a standard deviation (SD) of <0.5 cycle, then the $\Delta\Delta C_T$ method of relative abundance changes was used to determine the fold change of the individual exon-containing mRNAs from B to LCL (reactions were normalized to total RNA concentration). If the SD between technical replicates was ≥0.5 cycle, then we used the second primer set. For all 240 qRT-PCR reactions (20 genes/array × 2 arrays × 2 [1 Up + 1 Down exon] × 3 donors), we were able to generate fold-change values with either the first or second primer set. Thus, our technical measurements and ability to predict expressed exons from probe set information was quite robust. The fold-change values were then used to perform statistical analysis described in the Results section. Briefly, given the heterogeneity between human donors for individual mRNA expression levels, we used a nonparametric test to assess the validation of SplicerEX predictions for the top 20 hits on each platform. For each donor, we asked whether the group of predicted Up exons differed significantly from the group of Down exons using the Mann-Whitney *U*-test. The Mann-Whitney test provided a conservative, nonparametric approach that confirmed that SplicerEX predicted robust and directional (increasing vs. decreasing) mRNA isoform-specific changes within individual donors.

DATA DEPOSITION

Microarray data are available for download from the Gene Expression Omnibus (GEO) under accession number GSE29301.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Uwe Ohler, Sandeep Dave, Erich Huang, and Joe Lucas for useful discussions during the development and application of SplicerEX. We acknowledge funding from the NIH grants 5R01-GM63090 (to M.G.B.) and 1R01-CA127727 (to M.G.B.), 1R01-CA140337 (to M.A.L.), the DOD grant GRANT00412169 (Predoctoral Traineeship Award to T.J.R.), and pilot grants from the American Cancer Society and Duke Center for AIDS Research (P30-AI064518 to M.A.L.).

Received April 2, 2012; accepted May 4, 2012.

REFERENCES

Affymetrix. 2005. *Alternative transcript analysis methods for exon arrays*. Affymetrix Whitepaper 2005. http://www.affymetrix.com/support/technical/whitepapers.affx.

Bemmo A, Benovoy D, Kwan T, Gaffney DJ, Jensen RV, Majewski J. 2008. Gene expression and isoform variation analysis using Affymetrix Exon Arrays. *BMC Genomics* **9:** 529. doi: 10.1186/1471-2164-9-529.

Blencowe BJ. 2006. Alternative splicing: New insights from global analyses. *Cell* **126:** 37–47.

Cahir-McFarland ED, Davidson DM, Schauer SL, Duong J, Kieff E. 2000. NF-κB inhibition causes spontaneous apoptosis in Epstein–Barr virus-transformed lymphoblastoid cells. *Proc Natl Acad Sci* **97:** 6055–6060.

Chang TY, Wu YH, Cheng CC, Wang HW. 2011. Differentially regulated splice variants and systems biology analysis of Kaposi's sarcoma-associated herpesvirus-infected lymphatic endothelial cells. *Nucleic Acids Res* **39:** 6970–6985.

Cheung HC, Baggerly KA, Tsavachidis S, Bachinski LL, Neubauer VL, Nixon TJ, Aldape KD, Cote GJ, Krahe R. 2008. Global analysis of aberrant pre-mRNA splicing in glioblastoma using exon expression arrays. *BMC Genomics* **9:** 216. doi: 10.1186/1471-2164-9-216.

Cline MS, Blume J, Cawley S, Clark TA, Hu JS, Lu G, Salomonis N, Wang H, Williams A. 2005. ANOSVA: A statistical method for detecting splice variation from expression data. *Bioinformatics* (Suppl 1) **21:** i107–i115.

Cooper TA, Wan L, Dreyfuss G. 2009. RNA and disease. *Cell* **136:** 777–793.

Della Beffa C, Cordero F, Calogero RA. 2008. Dissecting an alternative splicing analysis workflow for GeneChip® Exon 1.0 ST Affymetrix arrays. *BMC Genomics* **9:** 571. doi: 10.1186/1471-2164-9-571.

Deming PB, Rathmell JC. 2006. Mitochondria, cell death, and B cell tolerance. *Curr Dir Autoimmun* **9:** 95–119.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19:** 92–105.

Garcia-Blanco MA, Baraniak AP, Lasda EL. 2004. Alternative splicing in disease and therapy. *Nat Biotechnol* **22:** 535–546.

Gardina PJ, Clark TA, Shimada B, Staples MK, Yang Q, Veitch J, Schweitzer A, Awad T, Sugnet C, Dee S, et al. 2006. Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7:** 325. doi: 10.1186/1471-2164-7-325.

Johannsen E, Luftig M, Chase MR, Weicksel S, Cahir-McFarland E, Illanes D, Sarracino D, Kieff E. 2004. Proteins of purified Epstein–Barr virus. *Proc Natl Acad Sci* **101:** 16286–16291.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The Human Genome Browser at UCSC. *Genome Res* **12:** 996–1006.

Krawczak M, Reiss J, Cooper DN. 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: Causes and consequences. *Hum Genet* **90:** 41–54.

Laajala E, Aittokallio T, Lahesmaa R, Elo LL. 2009. Probe-level estimation improves the detection of differential splicing in Affymetrix exon array studies. *Genome Biol* **10:** R77. doi: 10.1186/gb-2009-10-7-r77.

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

Li C, Wong WH. 2001. Model-based analysis of oligonucleotide arrays: Expression index computation and outlier detection. *Proc Natl Acad Sci* **98:** 31–36.

Lopez-Bigas N, Blencowe BJ, Ouzounis CA. 2006. Highly consistent patterns for inherited human diseases at the molecular level. *Bioinformatics* **22:** 269–277.

Miller HB, Robinson TJ, Gordan R, Hartemink AJ, Garcia-Blanco MA. 2011. Identification of Tat-SF1 cellular targets by exon array analysis reveals dual roles in transcription and splicing. *RNA* **17:** 665–674.

Pearson JL, Robinson TJ, Munoz MJ, Kornblihtt AR, Garcia-Blanco MA. 2008. Identification of the cellular targets of the transcription factor TCERG1 reveals a prevalent role in mRNA processing. *J Biol Chem* **283:** 7949–7961.

Pickrell JK, Marioni JC, Pai AA, Degner JF, Engelhardt BE, Nkadori E, Veyrieras JB, Stephens M, Gilad Y, Pritchard JK. 2010. Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* **464:** 768–772.

Robinson MD, Speed TP. 2007. A comparison of Affymetrix gene expression arrays. *BMC Bioinformatics* **8:** 449. doi: 10.1186/1471-2105-8-449.

Robinson TJ, Dinan MA, Dewhirst M, Garcia-Blanco MA, Pearson JL. 2010. SplicerAV: A tool for mining microarray expression data for changes in RNA processing. *BMC Bioinformatics* **11:** 108. doi: 10.1186/1471-2105-11-108.

Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. 2008. Proliferating cells express mRNAs with shortened 3′ untranslated regions and fewer microRNA target sites. *Science* **320:** 1643–1647.

Takeda J, Suzuki Y, Nakao M, Barrero RA, Koyanagi KO, Jin L, Motono C, Hata H, Isogai T, Nagai K, et al. 2006. Large-scale identification and characterization of alternative splicing variants of human gene transcripts using 56,419 completely sequenced and manually annotated full-length cDNAs. *Nucleic Acids Res* **34:** 3917–3928.

Thorsen K, Sorensen KD, Brems-Eskildsen AS, Modin C, Gaustadnes M, Hein AM, Kruhoffer M, Laurberg S, Borre M, Wang K, et al. 2008. Alternative splicing in colon, bladder, and prostate cancer identified by exon array analysis. *Mol Cell Proteomics* **7:** 1214–1224.

Venables JP. 2006. Unbalanced alternative splicing and its significance in cancer. *Bioessays* **28:** 378–386.

Venables JP, Klinck R, Koh C, Gervais-Bird J, Bramard A, Inkel L, Durand M, Couture S, Froehlich U, Lapointe E, et al. 2009. Cancer-associated regulation of alternative splicing. *Nat Struct Mol Biol* **16:** 670–676.

Wang H, Hubbell E, Hu JS, Mei G, Cline M, Lu G, Clark T, Siani-Rose MA, Ares M, Kulp DC, et al. 2003. Gene structure-based splice variant deconvolution using a microarray platform. *Bioinformatics* (Suppl 1) **19:** i315–i322.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476.

Xi L, Feber A, Gupta V, Wu M, Bergemann AD, Landreneau RJ, Litle VR, Pennathur A, Luketich JD, Godfrey TE. 2008. Whole genome exon arrays identify differential expression of alternatively spliced, cancer-related genes in lung cancer. *Nucleic Acids Res* **36:** 6535–6547.

Xing Y, Stoilov P, Kapur K, Han A, Jiang H, Shen S, Black DL, Wong WH. 2008. MADS: A new and improved method for analysis of differential alternative splicing by exon-tiling microarrays. *RNA* **14:** 1470–1479.

Downloaded from rnajournal.cshlp.org on March 28, 2022 - Published by Cold Spring Harbor Laboratory Press

SplicerEX: A tool for the automated detection and classification of mRNA changes from conventional and splice-sensitive microarray expression data

Timothy J. Robinson, Eleonora Forte, Raul E. Salinas, et al.

RNA 2012 18: 1435-1445 originally published online June 26, 2012
Access the most recent version at doi:10.1261/rna.033621.112

Supplemental Material http://rnajournal.cshlp.org/content/suppl/2012/06/05/rna.033621.112.DC1

References This article cites 32 articles, 10 of which can be accessed free at:
http://rnajournal.cshlp.org/content/18/8/1435.full.html#ref-list-1

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

To subscribe to *RNA* go to:
http://rnajournal.cshlp.org/subscriptions

Copyright © 2012 RNA Society