Model-based prediction of *cis*-acting RNA elements regulating tissue-specifc alternative splicing

Xin Wang, Kejun Wang, Guohua Wang, Jeremy R. Sanford and Yunlong Liu

***Abstract*—Here we describe a model-based approach to predict *cis*-acting RNA elements which regulate tissue-specific alternative splicing. The model facilitates the identification of *cis*-acting elements (or CAE) and the estimation of their activities, considering the splicing variants between two different tissues as the combinatorial functions of multiple elements. We implement this model on a set of differentially expressed exons, between heart and liver, derived from Affymetrix GeneChip® Human Exon 1.0 ST Array sample data. Focusing on hexamers, we select top 15 motifs with greatest cumulative exon inclusion (EIC) scores as the potential *cis*-acting elements. Eight of the total 15 hexamers are validated based on known exonic splicing regulators (ESRs) and predicted ESRs (PESRs). Permutation test demonstrates that the predicted EIC scores are statistically significant. Based on the prediction, we propose that PTB, hnRNP-B, SRp40, as well as other unknown factors are involved in the tissue-specific alternative splicing between heart and liver.**

I. Introduction

Alternative pre-mRNA splicing is the major mechanism for expanding the protein coding capacity of human genes. Current estimates suggest that every gene produces, on average more than 5 different mRNA isoforms. Recent surveys of transcript diversity indicate that alternative splicing plays an important role in regulating tissue-specific patterns of gene expression [1, 2]. Selection of alternative splice sites is governed by the coordinated interactions of *trans*-acting RNA binding proteins with *cis*-acting RNA elements. The serine and arginine-rich family of RNA binding proteins are essential, ubiquitously expressed splicing factors whose relative expression levels can vary across tissues [3]. SR proteins function at an early stage of spliceosome assembly and are believed to play a general role as activators of splice site recognition [4]. By contrast, the expression of many RNA binding proteins is restricted to specific tissues or developmental stages. Nova proteins are the crucial factors regulating brain-specific alternative splicing to form synapse [5, 6]. The Fox-1 family and SUP-12 coordinately regulate the tissue-specific alternative splicing of fibroblast growth factor receptor gene egl-15 *in vivo* [7]. Tissue-specific RNA binding proteins may bind to specific RNA elements and regulate tissue-specific splicing [8, 9]. For example, the RNA element UGCAUG is over-represented in the proximal downstream intronic regions of many brain-specific exons, and plays a critical role in mediating tissue-specific splicing events [10].

Biological experiments are limited to a handful of model RNA transcripts and *trans*-acting RNA binding proteins, which provide important mechanistic insight but often leave the relevance to general tissue-specific splicing unanswered. It is also difficult to discover the potential factors which might be involved in tissue-specific splicing regulation. Model-based methods, integrating genome-wide microarray technologies designed for splicing variants detection, give us opportunities to discover the tissue-specific splicing code. Das and co-workers presented a computational model to identify *cis*-regulatory elements for tissue-specific alternative splicing [11]. Their model is based on the idea that the correlation of motif parameters with gene-level normalized exon expression signals can be used to identify splicing regulatory motifs. However, the combinatorial effect of different splicing factors is not considered, which may fail to model the true biological process. On the other hand, their model only focuses on upstream and downstream intronic regions. As a matter of fact, exonic splicing enhancers (ESE) and silencers (ESS) probably take part in alternative splicing regulation as well.

Here we introduce a new model to predict the potential *cis*-acting RNA elements which regulate tissue-specific splicing. This model is based on the idea that the variations of gene-level normalized exon intensities between two different tissues are controlled by the combinatorial actions of a set of tissue-specific splicing factors. The combinatorial effects of splicing factors are determined by their functional levels (or FL), which express whether and how strong the factors inhibit or enhance the expression of exons, and their corresponding occupancy frequencies in splicing regulatory regions. Platforms such as Affymetrix GeneChip® Exon Array designed for monitoring alternative splicing can be used in this model, as long as the exon-level intensities are available or can be evaluated. For a studied alternative exon, the potential regulatory regions consist of intronic region upstream of 3'ss (splicing site), exonic region downstream of 3'ss, exonic region upstream of 5'ss and intronic region downstream of 5'ss. Each candidate regulatory motif is mapped to the sequence of its regulatory region to compute the occupancy frequency. According to the occupancy frequency and factors' functional levels, a linear model is then constructed to fit the splicing index (SI), which is

Manuscript received July 5, 2008. This work was supported in part by the grants from the Indiana Genomics Initiative of Indiana University (supported in part by the Lilly Endowment, Inc.)

Xin Wang, Guohua Wang, Jeremy R. Sanford, and Yunlong Liu are with the Indiana University School of Medicine, Indianapolis, IN 46202 USA (phone: 317-278-9222; fax: 317-278-9217; e-mail: sanfordj@iupui.edu and yunliu@iupui.edu).

Xin Wang and Kejun Wang are with the Harbin Engineering University School of Automation, Harbin, Heilongjiang 150001 China

calculated by the log ratio of gene-level normalized exon intensites of two samples. The functional level parameters are estimated by linear least squares approach. This process is repeated for many times with different random sets of all candidate motifs. Cumulative Exon Inclusion (EIC) Score is introduced to define the relative contribution of each motif to the inclusion of alternative exons. Providing differentially expressed exons between two tissues, the model can predict a group of *cis*-elements with top EIC scores.

In this article, we test the model using exon array data retrieved from Affymetrix GeneChip® Human Exon 1.0 ST Array sample database [12] including a collection of 11 human tissues. Although we only focus on heart and liver in this study, it is very straightforward to compare any two of these tissues and evaluate their corresponding tissue-specific *cis*-elements. We pick out top 15 motifs with highest EIC scores from all candidate motifs as the potential *cis*-acting elements. More than half of the predicted *cis*-acting elements are validated either by known exon splicing regulators (ESRs) or predicted exon splicing regulators (PESRs). Computational validation—permutation analysis also clearly shows that the predicted EIC scores are significantly greater than randomized data. The results indicate that PTB, hnRNP-B, SRp40, as well as other predicted elements whose corresponding *trans*-acting binding proteins are not known currently, may be critical for the tissue-specific alternative splicing of heart and liver.

II. Method

Our model is designed based on a very important hypothesis that pre-mRNA splicing is regulated by the combinatorial effects of multiple *trans*-acting RNA binding proteins. Since there're different kinds of alternative splicing such as cassette exons, alternative 5' splicing, alternative 3' splicing, retained intron etc, the regulatory regions and patterns may vary in different splicing patterns. For convenience, this study only focus on cassette exons—exons skipped or retained under different regulatory conditions. However, the essential idea can be extended to other sorts of splicing very easily.

A. Identification of RNA cis-regulatory regions

Cis-acting elements consist of enhancers (enhance splicing) and silencers (inhibit splicing) which could be located within exonic or intronic regions. Most of the *cis*-acting elements exist in the splicing junctions. To model different regulatory patterns, we focus our searching of functional *cis*-acting elements in intronic region upstream of 3'ss (splicing site), exonic region downstream of 3'ss, exonic region upstream of 5'ss and intronic region downstream of 5'ss. The four regulatory regions are illustrated in Fig. 1.

It has been biologically validated that the splicing factors will absolutely lose their regulatory functions when they are placed in intronic regions more than 300bp away from splicing sites [13-15], so we focus the searching in 300bp within intronic regions and 150bp in exonic regions. If the exon is less than 150bp, the entire region will be considered. For each array-detected differentially expressed splicing variant under two different biological conditions, sequences of RNA regulatory regions, as specified in Fig. 1, will be extracted from the UCSC genome browser. Hexamers are widely used in studying exon splicing enhancers and silencers. Therefore, we select potential binding sites of splicing regulatory factors from a pool of all the 6-bp motifs. This results in a total of $4^6 = 4096$ candidates in each putative regulatory region.

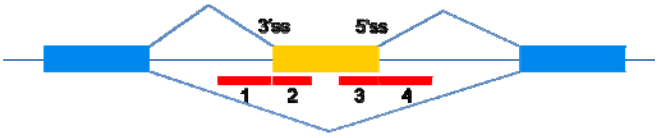

Fig. 1. Regulatory regions for cassette exon

B. Modeling the expression levels of cassette exons

We use "Splicing Index" [16] to compare the different inclusion rates of exons relative to gene expression level under two conditions. It is calculated by taking the logarithmic ratio of the normalized exon intensities (*NI*) in two samples (1).

$$SI = \log_2 \frac{NI_{sample\ 1}}{NI_{sample\ 2}} \tag{1}$$

The quantitative relationship between the expression levels of splicing variant *k* and the occurrences of binding sites of splicing regulatory factors is formulated as:

$$SI_k = \sum_{r=1}^{n_r} \sum_{i \in S_{k,r}} S_{k,i,r} x_{i,r} \tag{2}$$

where:

SI_k: logarithmic ratio of the expression levels of the pre-mRNA regions where the *k*-th splicing variants reside between two different conditions, as defined in (1);

n_r: number of regulatory regions;

$S_{k,i,r}$: number of binding sites of splicing regulatory factor *i* in the *r*-th regulatory region of the *k*-th splicing events;

$S_{k,r}$: all the functional binding sites of splicing regulatory factors having occurrences in the *r*-th regulatory region of the *k*-th splicing events;

$x_{i,r}$: functional levels of *i*-th binding sites of splicing regulatory factor in the *r*-th regulatory region

It is clearly demonstrated in (2) that the measured differential expression level of splicing variants (SI_k) is regulated by a combinatorial effect (Σ and $S_{k,i,r}$) of multiple splicing regulatory factors ($x_{i,r}$) binding in different splicing regulatory regions (n_r), as marked in Fig. 1. In this equation, SI_k can be measured using Affymetrix Exon Array data. Given a set of binding sites of splicing regulatory factors, $S_{k,i,r}$ can be calculated based on the splicing regulatory sequences retrieved from regulatory region defined in Fig. 1. Once these

two parameters are considered as known, the function of each splicing regulatory factor ($x_{i,r}$) can be estimated mathematically by fitting K equations using n_r parameters (n_r <<K).

C. Least square estimation

Given the expression variation (SI_k) and the number ($S_{k,i,r}$) of binding sites of a set of splicing factors in different regulatory regions, least-squares estimation is performed to estimate the function of each splicing factor.

$$\hat{X} = (S^T S)^{-1} S^T SI \tag{3}$$

where X, S and SI are matrix forms of $x_{i,r}$, $S_{k,i,r}$ and SI_k in (2). For a given set of splicing factors X, S and SI contain $\sum S_r \times 1$, $K \times \sum S_r$ and $1 \times K$ elements, respectively. This estimation is achieved by minimizing the sum of square error of the differences between the observed and predicted expression alteration of studied splicing variants:

$$SS = \sum_{k=1}^{K} (SI_k - \sum_{r=1}^{n_r} \sum_{i \in S_{k,r}} S_{k,i,r} x_{i,r})^2 \tag{4}$$

where K is the total number of differentially expressed exons.

D. Motif selection

The linear model and least-squares estimation (2 and 3) aim at estimating the functions of a known set of splicing regulatory factors and their binding sites that contribute to the differential splicing patterns in two different conditions. In normal practice, however, this set of splicing regulatory factors is unknown. In this step, we intend to design a procedure to evaluate the influence of each hexamers on the splicing regulation by calculating the sum of square error (4) of predicting expressional alteration of splicing variants using the hexamer being evaluated in cooperation with others. The procedure is completed in N iterations. In each iteration, n candidate motifs are selected from a pool of 4×4096 hexamers. Model errors based on the random set of selection will be calculated based on (4). Since a smaller model error implies a more influential binding site, an exon inclusion contribution score (or EIC) is assigned to each selected candidate according to the following formulations:

$$EIC_i = \sum_{c \in C_{n,r,i}} \frac{1}{SS_c^{\alpha}} \tag{5}$$

where SS is the model sum square error derived in (4); $C_{n,r,i}$ represents all the combinatorial selections (having n binding sites) that include the i-th hexamer; and α is a power factor that influences the effect of single selections ($\alpha > 1$). A larger α value usually amplifies the effect of motif selections in each iteration (here, we will use $\alpha = 5$).

Similarly, cumulative functional level of each binding site is calculated:

$$X_{i,r} = \sum_{c \in C_{n,r,i}} \hat{x}_c \tag{6}$$

where $\hat{x}$ are the estimated functional levels of selected hexamers in each iteration (3). Overall, the proposed motif selection procedure can be summarized as:

1) Randomly pick out n elements from a pool of all hexamers.
2) Calculate the predicted model sum square error SS (4).
3) Calculate the current exon inclusion contribution score for each selected binding site candidate as reciprocal to SS, and their functional levels (x).
4) Add the current contribution score and individual functional level to the cumulative exon inclusion contribution score (EIC) and functional levels (X) (6).

Repeat the procedure 1) to 4) for N times. The motif (hexamer) candidates with the highest overall exon inclusion contribution (EIC) score is considered as the putative binding sites of splicing regulatory factor that are responsible to the splicing alteration between two biological conditions. N is selected so that each motif candidate is being evaluated 10,000 times.

III. RESULT

A. Identification of differentially expressed exons based on Exon Array data

Affymetrix GeneChip® Exon 1.0 ST Array is designed for the study of transcript diversity and alternative splicing. More than 1.4 million Probe Selection Regions (or PSRs) and 1 million exon clusters are integrated in a single chip to monitor the splice variants. We use exon array sample dataset (published on Affymetrix website) which contains 11 human tissues, with 3 replicates for each tissue. Without losing generality, here we only focus on the comparison of heart and liver.

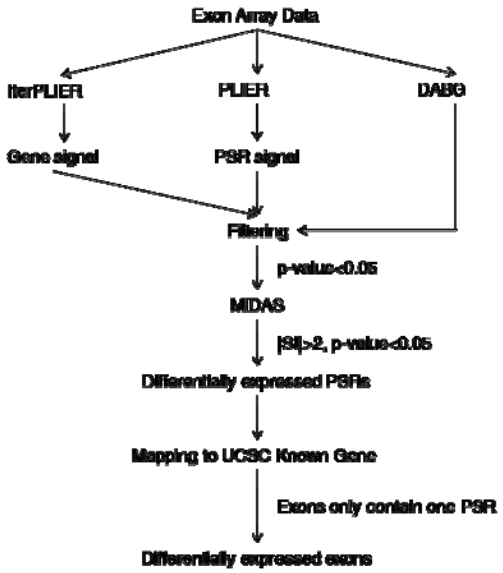

Fig. 2. Affymetrix Exon Array data processing flow chart

We firstly apply a series of processing procedures introduced in Affymetrix Technique Note [17] to find the differentially expressed PSRs. These PSRs are mapped to UCSC known gene annotation to identify their corresponding exons. As we only consider cassette exons in this study, exons containing multiple PSRs are filtered out. The data processing flow chart is summarized in Fig. 2.

TABLE I
Exon array analysis

	Number
PSR	1404903
Core gene	17800
Present PSR	489804
Present core gene	4160
Present PSR in present core gene	98839
Differentially expressed PSR (MiDAS p-value<0.05 and $\|SI\|$>=2)	998
Exons containing only 1 PSR	326

Three versions of transcript annotations are supplied by Affymetrix—core (RefSeq transcripts and full-length mRNAs), extended (Core + cDNA-based annotations) and full (Extended + ab-initio gene predictions) [17]. We use core transcript annotation which is claimed to be much more significant and important compared with the other two. Only 7% of the total 1.4 million PSRs are identified as present (p-value<0.05). MiDAS [18] calculate the gene-level normalized PSR intensities and test the hypothesis that the gene-level normalized expression levels of the PSRs of heart and liver are the same. We filter the PSRs based on the strategy that p-value<0.05 and $|SI|$>2. Data in each processing procedure are listed in TABLE I.

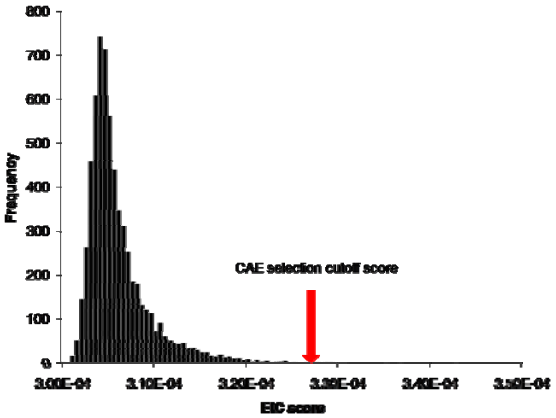

Fig. 3. EIC Score Distribution

B. *Modeling results for 6-bp motif prediction*

Since we focus on 6-bp RNA motifs in this study, there are $4 \times 4^6 = 16384$ potential motifs which could be the *cis*-acting elements theoretically. Considering that some of the motifs seldom appear in the differentially expressed exons, we do a filtering of these motifs to speed up the searching of binding sites. Exon occupancy frequency (EOF) is calculated for each motif. Only 6389 retained when we filtered out the motifs with EOF less than 5%. Then, we implement the model using these motifs and their corresponding *SI* values.

The EIC score distribution is illustrated in Fig. 3. We selected top 15 motifs with the highest EIC scores as the regulatory elements which contribute to the tissue-specific splicing of exons in heart and liver (TABLE II). Interestingly, about half of the motifs (Fig. 4, 7 motifs with red diamonds) enhance the expression levels of exons in heart. Another half of the motifs (8 motifs with blue diamonds) prevent the expression of heart exons and favor liver exons. Only one of the top 15 motifs is located within exon, while other motifs are intronic regulators.

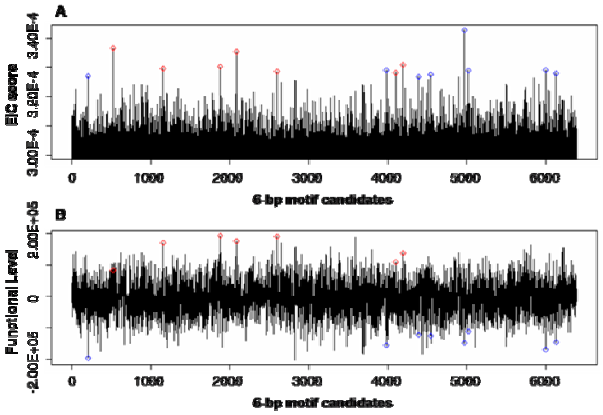

Fig. 4. (A) Spectrum of EIC scores; (B) Spectrum of Motif functional level scores. (Red diamonds stand for selected CAEs with FL>0, while blue ones are selected CAEs with FL<0)

TABLE II
Selected top CAEs and validation

Regulatory region	CAE	EIC	Functional level	Validation
4	CUCUUG	3.43E-04	-1.47E+05	hnRNP-I
1	AUAUUU	3.37E-04	8.35E+04	PESR
1	UGAUAU	3.35E-04	1.75E+05	
4	ACUGCA	3.31E-04	1.38E+05	SRp40
1	UAUGGA	3.30E-04	1.92E+05	PESR
1	GAAACA	3.30E-04	1.71E+05	PESR
4	UGAACC	3.29E-04	-1.69E+05	
4	AACCAG	3.29E-04	-1.55E+05	PESR
4	CUGUCU	3.29E-04	-1.11E+05	
2	AUAAAA	3.29E-04	1.90E+05	
4	AAUUUA	3.28E-04	1.09E+05	hnRNP-B
4	UGGUCU	3.28E-04	-1.45E+05	
4	AUUCCU	3.28E-04	-1.25E+05	
1	ACAGCU	3.27E-04	-1.96E+05	SRp40
4	AGUGGG	3.27E-04	-1.22E+05	

To elucidate the regulatory interactions of selected motifs and the exons, we plotted a regulatory network (Fig. 5) using Cytoscape [19]. The regulatory relationship between motif and exon is called “up-regulate” if the motif’s functional level is greater than 0; otherwise, we call it “down-regulate”. Very clearly, most of the exons regulated by motifs with FL>0 have positive SI scores, while exons regulated by motifs with FL<0 have negative SI scores.

C. *Validation of selected motifs*

1) *Computational validation—Permutation analysis*

Permutation analysis is performed to test the significance of the results of modeling. Basically, we shuffle the SI scores and rerun the model. Significant lower EIC scores are obtained for the randomized SI scores (Fig. 6). We also did wilcoxon test between EIC scores of experimentally-determined and randomized data, and we saw a very significant result (p-value<2.2e-16). The permutation analysis tells us that the modeling results are statistically significant.

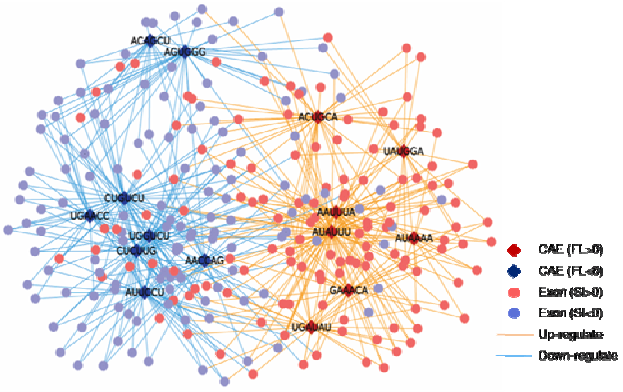

Fig. 5. Predicted CAE regulation network

2) *Validation using known ESRs and PESRs*

Due to the complexity of alternative splicing and the limited number of known *cis*-regulatory elements, we are not able to know all of the *cis*-acting RNA motifs related to their corresponding proteins at current stage. So it is very challenging for us to validate our predicted top CAEs based on biologically validated *cis*-acting RNA elements systematically. Fortunately, we still can evaluate the model's results using the collected known ESRs and PESRs provided by some research groups [20-23]. Surprisingly, of all the top 15 selected CAEs, 4 motifs are mapped to known ESRs, and 5 motifs are hit by PESRs. In total, a little more than half of seleted motifs can be validated.

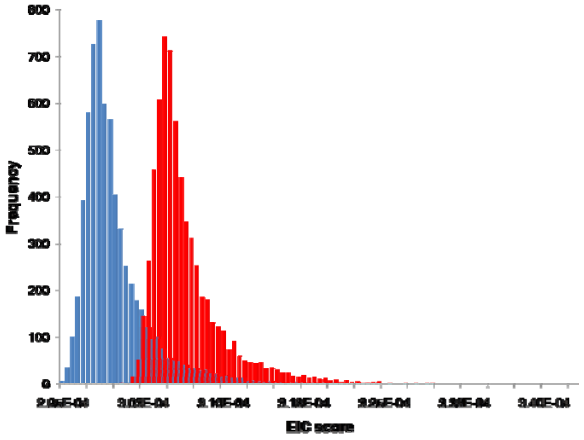

Fig. 6. Histograms of EIC scores of predicted 6-bp motifs based on *SI* with original orders (red histogram) and randomized orders (blue histogram)

IV. Discussion

This article presents a new model-based approach to identify *cis*-acting elements for tissue-specific alternative splicing. The model is based on the hypothesis that the splicing variants between two tissue samples are determined by the combinatorial activities of *cis*-acting elements. Using heart and liver sample data from Affymetrix GeneChip® Human Exon 1.0 ST Array database, we demonstrate that the predicted EIC scores of 6-bp motifs are of statistical significance. Eight of the selected top 15 motifs are validated by known ESRs and PESRs, although the other seven motifs have no corresponding targets. As the research on RNA binding proteins goes further, the *cis*-acting RNA element database will become more established (like today's transcription factor database), and hopefully we will be able to validate more predicted motifs.

From the predicted top 15 hexamers listed in TABLE II, we notice that 14 of them are located in upstream or downstream intron. This result is consistent with previous conclusion that tissue-specific *cis*-elements are enriched in the flanking intronic regions [10, 24-26]. We identified 3 known splicing factors: hnRNP-I, SRp40, and hnRNP-B. HnRNP-I protein, also known as PTB, is expressed throughout heart development, and both CELF and PTB activities are required for appropriate splicing in cardiomyocytes [27]. The predicted functional level of PTB is positive, indicating that PTB is a very important splicing factor contributing more exon inclusion in heart than in liver. We also found that SRp40 are targeted twice by the top 15 hexamers, but the functional levels are opposite, which seems to be contradictory. However, some articles have proved that SR proteins usually function as enhancers, but interestingly, can also be inhibitors in some splicing systems [28-32]. SRp40 can not only bind on intronic splicing enhancer (ISE) [33], but also intronic splicing silencer (ISS) [34].

SR proteins and hnRNP proteins are two of the most important and largest RNA binding protein families which regulate alternative splicing. Previous studies show that both the alternative splice site choice and the inclusion/exclusion ratio of selected alternative exons can be controlled by the changes in the relative amounts of hnRNP A/B proteins and SR proteins [35-39]. Consistent with this point, our results demonstrate that hnRNP proteins, SR proteins, as well as other unknown factors contribute to the splice variation between heart and liver. Understanding the combinatorial effect of splicing factors systematically probably will be very difficult at current stage. Computational models, however, can help us discover more potential factors, their functions and combinatorial regulations on tissue-specific alternative splicing.

Despite the success of our approach, we feel the model is still in an early stage of development. Our future work will be focused on improving this model by adding two additional parameters—RNA secondary structure and sequence degeneracy of candidate motifs. We believe these considerations will increase the accuracy and specificity of

our predictions and ultimately facilitate deciphering of the splicing code.

REFERENCES

[1] Q. Pan, O. Shai, C. Misquitta, W. Zhang, A.L. Saltzman, N. Mohammad, T. Babak, H. Siu, T.R. Hughes, Q.D. Morris, B.J. Frey, and B.J. Blencowe, "Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform," *Mol Cell*, vol. 16, (no. 6), pp. 929-41, Dec 22 2004.

[2] G. Yeo, D. Holste, G. Kreiman, and C.B. Burge, "Variation in alternative splicing across human tissues," *Genome Biol*, vol. 5, (no. 10), pp. R74, 2004.

[3] A. Hanamura, J.F. Caceres, A. Mayeda, B.R. Franza, Jr., and A.R. Krainer, "Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors," *RNA*, vol. 4, (no. 4), pp. 430-44, Apr 1998.

[4] O. Ram and G. Ast, "SR proteins: a foot on the exon before the transition from intron to exon definition," *Trends Genet*, vol. 23, (no. 1), pp. 5-7, Jan 2007.

[5] J. Ule, K.B. Jensen, M. Ruggiu, A. Mele, A. Ule, and R.B. Darnell, "CLIP identifies Nova-regulated RNA networks in the brain," *Science*, vol. 302, (no. 5648), pp. 1212-5, Nov 14 2003.

[6] J. Ule, A. Ule, J. Spencer, A. Williams, J.S. Hu, M. Cline, H. Wang, T. Clark, C. Fraser, M. Ruggiu, B.R. Zeeberg, D. Kane, J.N. Weinstein, J. Blume, and R.B. Darnell, "Nova regulates brain-specific splicing to shape the synapse," *Nat Genet*, vol. 37, (no. 8), pp. 844-52, Aug 2005.

[7] H. Kuroyanagi, G. Ohno, S. Mitani, and M. Hagiwara, "The Fox-1 family and SUP-12 coordinately regulate tissue-specific alternative splicing in vivo," *Mol Cell Biol*, vol. 27, (no. 24), pp. 8612-21, Dec 2007.

[8] J. Qi, S. Su, M.E. McGuffin, and W. Mattox, "Concentration dependent selection of targets by an SR splicing regulator results in tissue-specific RNA processing," *Nucleic Acids Res*, vol. 34, (no. 21), pp. 6256-63, 2006.

[9] X. Xu, D. Yang, J.H. Ding, W. Wang, P.H. Chu, N.D. Dalton, H.Y. Wang, J.R. Bermingham, Jr., Z. Ye, F. Liu, M.G. Rosenfeld, J.L. Manley, J. Ross, Jr., J. Chen, R.P. Xiao, H. Cheng, and X.D. Fu, "ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle," *Cell*, vol. 120, (no. 1), pp. 59-72, Jan 14 2005.

[10] S. Minovitsky, S.L. Gee, S. Schokrpur, I. Dubchak, and J.G. Conboy, "The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons," *Nucleic Acids Res*, vol. 33, (no. 2), pp. 714-24, 2005.

[11] D. Das, T.A. Clark, A. Schweitzer, M. Yamamoto, H. Marr, J. Arribere, S. Minovitsky, A. Poliakov, I. Dubchak, J.E. Blume, and J.G. Conboy, "A correlation with exon expression approach to identify cis-regulatory elements for tissue-specific alternative splicing," *Nucleic Acids Res*, vol. 35, (no. 14), pp. 4845-57, 2007.

[12] http://www.affymetrix.com/support/technical/sample_data/exon_array_data.affx.

[13] B.R. Graveley, K.J. Hertel, and T. Maniatis, "A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers," *EMBO J*, vol. 17, (no. 22), pp. 6747-56, Nov 16 1998.

[14] A. Lavigueur, H. La Branche, A.R. Kornblihtt, and B. Chabot, "A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding," *Genes Dev*, vol. 7, (no. 12A), pp. 2405-17, Dec 1993.

[15] M. Tian and T. Maniatis, "A splicing enhancer exhibits both constitutive and regulated activities," *Genes Dev*, vol. 8, (no. 14), pp. 1703-12, Jul 15 1994.

[16] K. Srinivasan, L. Shiue, J.D. Hayes, R. Centers, S. Fitzwater, R. Loewen, L.R. Edmondson, J. Bryant, M. Smith, C. Rommelfanger, V. Welch, T.A. Clark, C.W. Sugnet, K.J. Howe, Y. Mandel-Gutfreund, and M. Ares, Jr., "Detection and measurement of alternative splicing using splicing-sensitive microarrays," *Methods*, vol. 37, (no. 4), pp. 345-59, Dec 2005.

[17] Affymetrix, "Identifying and Validating Alternative Splicing Events," *Affymetrix Technote*, 2006.

[18] Affymetrix, "Alternative Transcript Analysis Methods for Exon Arrays," *Affymetrix Technote*, 2006.

[19] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker, "Cytoscape: a software environment for integrated models of biomolecular interaction networks," *Genome Res*, vol. 13, (no. 11), pp. 2498-504, Nov 2003.

[20] A. Goren, O. Ram, M. Amit, H. Keren, G. Lev-Maor, I. Vig, T. Pupko, and G. Ast, "Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers," *Mol Cell*, vol. 22, (no. 6), pp. 769-81, Jun 23 2006.

[21] W.G. Fairbrother, R.F. Yeh, P.A. Sharp, and C.B. Burge, "Predictive identification of exonic splicing enhancers in human genes," *Science*, vol. 297, (no. 5583), pp. 1007-13, Aug 9 2002.

[22] X.H. Zhang and L.A. Chasin, "Computational definition of sequence motifs governing constitutive exon splicing," *Genes Dev*, vol. 18, (no. 11), pp. 1241-50, Jun 1 2004.

[23] Z. Wang, M.E. Rolish, G. Yeo, V. Tung, M. Mawson, and C.B. Burge, "Systematic identification and analysis of exonic splicing silencers," *Cell*, vol. 119, (no. 6), pp. 831-45, Dec 17 2004.

[24] C.W. Sugnet, K. Srinivasan, T.A. Clark, G. O'Brien, M.S. Cline, H. Wang, A. Williams, D. Kulp, J.E. Blume, D. Haussler, and M. Ares, Jr., "Unusual intron conservation near tissue-regulated exons found by splicing microarrays," *PLoS Comput Biol*, vol. 2, (no. 1), pp. e4, Jan 2006.

[25] M. Brudno, M.S. Gelfand, S. Spengler, M. Zorn, I. Dubchak, and J.G. Conboy, "Computational analysis of candidate intron regulatory elements for tissue-specific alternative pre-mRNA splicing," *Nucleic Acids Res*, vol. 29, (no. 11), pp. 2338-48, Jun 1 2001.

[26] B.J. Blencowe, "Alternative splicing: new insights from global analyses," *Cell*, vol. 126, (no. 1), pp. 37-47, Jul 14 2006.

[27] A.N. Ladd, M.G. Stenberg, M.S. Swanson, and T.A. Cooper, "Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development," *Dev Dyn*, vol. 233, (no. 3), pp. 783-93, Jul 2005.

[28] A. Kanopka, O. Muhlemann, and G. Akusjarvi, "Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA," *Nature*, vol. 381, (no. 6582), pp. 535-8, Jun 6 1996.

[29] L.M. McNally and M.T. McNally, "An RNA splicing enhancer-like sequence is a component of a splicing inhibitor element from Rous sarcoma virus," *Mol Cell Biol*, vol. 18, (no. 6), pp. 3103-11, Jun 1998.

[30] L.M. McNally and M.T. McNally, "U1 small nuclear ribonucleoprotein and splicing inhibition by the rous sarcoma virus negative regulator of splicing element," *J Virol*, vol. 73, (no. 3), pp. 2385-93, Mar 1999.

[31] M.J. Simard and B. Chabot, "SRp30c is a repressor of 3' splice site utilization," *Mol Cell Biol*, vol. 22, (no. 12), pp. 4001-10, Jun 2002.

[32] Y. Wang, J. Wang, L. Gao, R. Lafyatis, S. Stamm, and A. Andreadis, "Tau exons 2 and 10, which are misregulated in neurodegenerative diseases, are partly regulated by silencers which bind a SRp30c.SRp55 complex that either recruits or antagonizes htra2beta1," *J Biol Chem*, vol. 280, (no. 14), pp. 14230-9, Apr 8 2005.

[33] A. Srebrow and A.R. Kornblihtt, "The connection between splicing and cancer," *J Cell Sci*, vol. 119, (no. Pt 13), pp. 2635-41, Jul 1 2006.

[34] E. Buratti, C. Stuani, G. De Prato, and F.E. Baralle, "SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer," *Nucleic Acids Res*, vol. 35, (no. 13), pp. 4359-68, 2007.

[35] A. Mayeda and A.R. Krainer, "Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2," *Cell*, vol. 68, (no. 2), pp. 365-75, Jan 24 1992.

[36] A. Mayeda, D.M. Helfman, and A.R. Krainer, "Modulation of exon skipping and inclusion by heterogeneous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF," *Mol Cell Biol*, vol. 13, (no. 5), pp. 2993-3001, May 1993.

[37] J.F. Caceres, S. Stamm, D.M. Helfman, and A.R. Krainer, "Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors," *Science*, vol. 265, (no. 5179), pp. 1706-9, Sep 16 1994.

[38] Y. Bai, D. Lee, T. Yu, and L.A. Chasin, "Control of 3' splice site choice in vivo by ASF/SF2 and hnRNP A1," *Nucleic Acids Res*, vol. 27, (no. 4), pp. 1126-34, Feb 15 1999.

[39] M. Blanchette, R.E. Green, S.E. Brenner, and D.C. Rio, "Global analysis of positive and negative pre-mRNA splicing regulators in Drosophila," *Genes Dev*, vol. 19, (no. 11), pp. 1306-14, Jun 1 2005.