Netw Model Anal Health Inform Bioinforma (2012) 1:19–25
DOI 10.1007/s13721-012-0007-5

ORIGINAL ARTICLE

An exon/intron disparity framework based on the nucleotide profile of single sequence

Sing-Wu Liou · Yin-Fu Huang

Received: 30 November 2011 / Revised: 12 March 2012 / Accepted: 21 March 2012 / Published online: 11 May 2012

Abstract The RNA sequences are the major materials accessible for the nuclear splicing machinery, therefore, understanding how they are transformed into a binary decision of intron removal and exon ligation is critical in resolving the mystery of pre-mRNA splicing. This paper proposed an exon/intron discrimination framework (EIDF) to profile the intrinsic differences between exons and their immediate introns based on information of single sequence. The EIDF focuses on the frequencies of specific mono-/di-/tri-nucleotides in the individual sequence and a simple exon/intron classifier is implemented accordingly. The experimental results showed the proposed EIDF is a valuable profile of splice site sequences and the possibility of simulating the processes of splicing machinery in silico is also revealed.

1 Introduction

From the perspective of the intrinsic sequence properties, exons carry the genetic information of protein coding and the signals of mRNA localization, therefore, exons are definitely non-random. In contrast, introns are removed during the pre-mRNA maturating processes and they do not involve in the translation processes, which makes them treated as junk DNA sometimes. These biological evidences strongly suggest there exist distinguishing properties between exons and introns and many computational analysis and biological evidences showed the sequence disparities between them. The dichotomous splicing signals in flanking regions of splice sites had been identified for years (Zhang et al. 2005); it had been also observed the sequence complexity of exons is higher than introns (Orlov and Potapov 2004); significant fluctuation of nucleotide composition near exon–intron junction sites was also found (Louie et al. 2003).

The main information available to the splicing machinery is the nucleotide composition of the pre-mRNA sequence being processed, therefore, profiling the differences between exons and their *immediate* introns using only information of single sequence would bring valuable insights to the splicing research field. To clarify how the splicing machinery transforms the sequence information into a binary decision of intron removal and exon ligation, profiling the intrinsic differences between exons and introns is essential. The intrinsic differences between adjacent exonic and intronic flanks of splice sites had been revealed by investigating the distribution of tri-nucleotides in sequences (Liou and Huang 2008), which linked up the compositional heterogeneity between them with the tri-nucleotide repeats and served as single sequence-based features for exon/intron discrimination (Liou and Huang 2009). As many splicing regulatory *cis*-elements were identified, a natural next step is to integrate the available information into a predictive framework to simulate the processes of identifying exons and introns by the splicing machinery (Wang and Burge 2008). Knowledge-driven methodologies, which focus on profiling or modeling the

S.-W. Liou
Graduate School of Engineering Science and Technology,
National Yunlin University of Science and Technology,
No. 123, University Rd., Section 3, Douliu City,
Yunlin County 640, Taiwan, ROC
e-mail: g9110808@yuntech.edu.tw

Y.-F. Huang (✉)
Department of Computer Science and Information Engineering,
National Yunlin University of Science and Technology,
No. 123, University Rd., Section 3, Douliu City,
Yunlin County 640, Taiwan, ROC
e-mail: huangyf@yuntech.edu.tw

related biological evidences, are better ways to identify biologically meaningful biomarkers (Chen et al. 2009) and some efforts had been made towards achieving this goal. For example, the SpliceIT encoded the sequence features relevant to splicing events in vivo to identify the splicing signals (Malousi et al. 2010); about 200 sequence features were combined to decipher the code for alternative splicing (Barash et al. 2010).

In this paper, the landscapes of exonic and intronic regions of splice sites are carefully surveyed, the pieces of biological evidences are integrated into an exon/intron discrimination framework (EIDF). A simple binary classifier is implemented according to the proposed EIDF, which discriminate exons from their immediate introns using only information of single sequence. Nine data sets, including sequences of manually collected from genes of six model organisms and sequences from three well-known databases, were prepared to validate the usefulness of the proposed EIDF. The results demonstrate the power of the proposed EIDF for discriminating exons from introns and some valuable insights into structure of genes are also provided.

2 Related works

The identified heterogeneity between exons and introns mainly comes from biological evidences and computational analysis. The biological evidences provide hints to discriminate them from each other, such as the basic differences between exons and introns in terms of the tri-nucleotides can be explained by the circular code theory (Arquèsa and Michel 1996) and the sharp transition at flank regions of splice sites (Zhan 1998). Many of exon/intron discrimination methods are based on analysis of sequence composition, such as the consensus sequences (Weir and Rice 2004), oligo-nucleotide frequencies (Claverie and Bougueleret 1986; Claverie et al. 1990; Solovyev et al. 1994; Louie et al. 2003), base/codon/triplet usage (Zhan 1998), the sequence determinants of splice sites (Mengeritsky and Smith 1989) and a multi-source recognition method recruiting the consensus features and statistical differences of bases usage (Nakata et al. 1985). However, none of the above methods is a single sequence-based methodology, and hence the day of fulfilling the vision of simulating the processes of splicing machinery is still awaited.

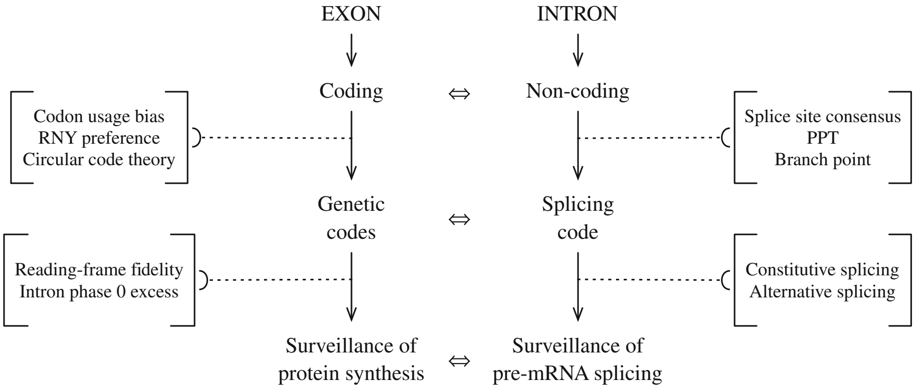
3 Methods

3.1 The exon/intron disparity framework (EIDF)

Each of the splice site sequences consist of the exons (coding sequences) and introns (noncoding sequences). A meta-analysis on them is performed, which investigates the related biological evidences from different perspectives including the concepts of genetic code, circular-code theory, reading frames of coding sequences and constitutive/alternative splicing. The blueprint of the meta-analysis is shown in Fig. 1. The most significant features characterizing exons are the genetic codes that are responsible of the surveillance of protein synthesis; and most significant features characterizing introns are the splicing signals that help the surveillance of pre-mRNA splicing. The properties of coding sequences and genetic codes are closely related to codon usage bias, RNY preference and circular code theory, which are further related to reading frame and intron phases.

The blueprint shown in Fig. 1 is systematically surveyed under bottom-up analysis, which finally leads to the exon/intron disparity framework (EIDF) as depicted in Fig. 2. The construction of EIDF starts from the basic computational concerns: what are the dominated base composition in terms of mono-/di-/tri-nucleotides. Then, the dominated base composition is identified via related biological evidences, which are further promoted to abstract concepts. The concepts are then systematically quantified via microview profiling of sequences. Finally, the disparity profiles will be formulated as equations as the concrete implementation of exon/intron disparities. EIDF represents a way to discriminate an exon from its immediate intron in terms of mono-/di-/tri-nucleotide composition.

Fig. 1 The disparities between exons and introns



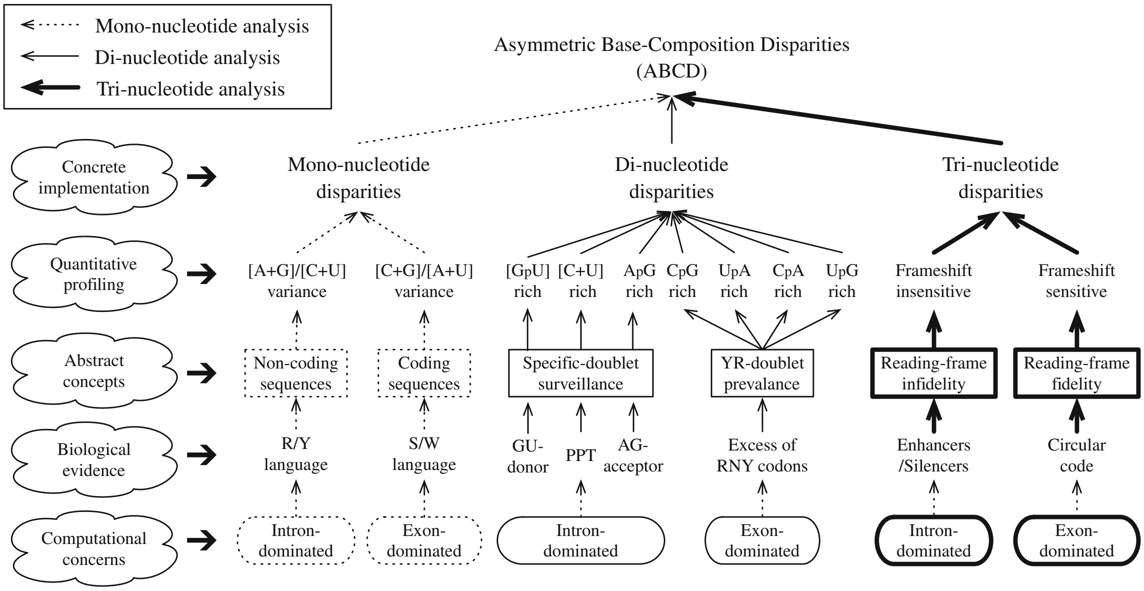


Fig. 2 The exon/intron disparity framework (EIDF)

3.2 Meta-analysis of mono-nucleotide composition

In general, any genomic sequence can be uniquely described by two independent RNA languages: the strong/weak (S/W, where S=G/C and W=A/U) bases and the purine/pyrimidine (R/Y, where R=A/G and Y=C/U) bases. The two distributions are invariant under the transforms of the RNA group in some sense, which indicates they are inherent for the sequences (Zhan 1997), moreover, the coding sequences prefer the S/W and the noncoding regions (e.g., introns, 5′-UTR, 3′-UTR) are often full of R/Y bases (Luo and Ji 1997). The variance of S/W codes and R/Y codes, denoted by M_{SW} and M_{RY} (M stands for mono-nucleotide), are defined in Eqs. 1 and 2, where μ is the average number of A/C/G/U in the specific single sequence and N_x is the number of base x (i.e.., A, C, G or U) in the sequence. The disparity in terms of mono-nucleotide composition, D_1, is the mean of variance M_{SW} and M_{RY}, which is defined in Eq. 3.

$$M_{\mathrm{SW}} = \frac{1}{2}((N_{\mathrm{C}} + N_{\mathrm{G}}) - \mu)^2 + ((N_{\mathrm{A}} + N_{\mathrm{U}}) - \mu)^2) \qquad (1)$$

$$M_{\mathrm{RY}} = \frac{1}{2}((N_{\mathrm{A}} + N_{\mathrm{G}}) - \mu)^2 + (N_{\mathrm{C}} + N_{\mathrm{U}}) - \mu)^2) \qquad (2)$$

$$D_1 = \frac{(M_{\mathrm{SW}} + M_{\mathrm{RY}})}{2} \qquad (3)$$

3.3 Meta-analysis of di-nucleotide composition

Many significant di-nucleotides in exonic and intronic regions are derived according to the properties of coding and noncoding sequences. The RNY preference (Nikolaou and Almirantis 2004) in coding sequences suggests the YR-doublet (Y stands for C/U and R for A/G) will be abundant in exons (i.e., RNYRNYRNYRNY...), and thus the CA/UA/CG/UG are the candidates of frequent di-nucleotides. In DNA sequences, there is marked variability among genes in the frequency of the di-nucleotide CpG (Bulmer 1987). Introns are similar to the intergenic regions, and thus the frequencies of CpG islands in exons (coding sequences) and introns (noncoding sequences) are expected to be very different. Therefore, the CpG is added to the candidate set of distinguishing di-nucleotides. From the perspective of constitutive splicing, it is anticipated that natural selection decreases frequency of GpU and ApG near the 5′ and 3′ ends of exons to prevent the appearance of cryptic splicing sites, this is so called the site avoidance (in exons) (Eskesen et al. 2004). While from the perspective of alternative splicing, a clear preference of isoforms was identified (Bortfeldt et al. 2008), which showed the potential tandem repeats of GpU/ApG di-nucleotide in introns. Thus, the GpU and ApG are recruited to the di-nucleotide candidate set. The intronic regions near acceptor site is a C/U rich region, which implies the occurrence of CpC/CpU/UpC/UpU di-nucleotides will be more frequent than other regions, and thus all of the four di-nucleotides are added into the candidate set. By performing the sensitivity analysis, some of the above-mentioned di-nucleotide candidates are filtered out and only the di-nucleotides with significant discrimination power are reserved. The significant exonic dimers, denoted as D_2^E shown in Eq. 4, are {CA, CG}/{CG, GA}; whereas the intronic dimers, denoted as D_2^I shown in Eq. 5, are {GU, UA}/{AG, UA, UC, CU, UU}. The dimer disparity D_2^* is defined as the difference between exonic dimers and intronic dimers as shown in Eq. 6.

$$D_2^E = \begin{cases} N_{\mathrm{CA}} + N_{CG}, & \text{if Site} = \text{Donor} \\ N_{\mathrm{CG}} + N_{\mathrm{GA}}, & \text{if Site} = \text{Acceptor} \end{cases} \tag{4}$$

$$D_2^I = \begin{cases} N_{\mathrm{GU}} + N_{\mathrm{UA}}, & \text{if Site} = \text{Donor} \\ N_{\mathrm{AG}} + N_{\mathrm{UA}} + N_{\mathrm{UC}} + N_{\mathrm{CU}} + N_{\mathrm{UU}}, & \text{if Site} = \text{Acceptor} \end{cases} \tag{5}$$

$$D_2^* = (D_2^E - D_2^I)/(L-1), \quad \text{where } L \text{ is sequence length.} \tag{6}$$

3.4 Meta-analysis of tri-nucleotide composition

The tri-nucleotide preference has great implications in analyzing pre-mRNA sequences and characterizing splice sites (Staden and McLachlan 1982; Nikolaou and Almirantis 2004; Willie and Majewski 2004). Tri-nucleotide repeats were also shown to be closely related with splicing regulation (Parmley and Hurst 2007), which are usually tandem repeats around splice sites (Zhuo et al. 2007); mutually symmetric and complementary triplets provide hints to distinguish coding sequences from noncoding ones (Nikolaou and Almirantis 2003). The tri-nucleotide disparity D_3 is derived via circular code-based analysis, which is described in Algorithm 1. From line 1 to 3, the function *codoncount* will find the frequencies of all the 64 codons in the sequences; the codons that do not appear in any of the three reading frame will be excluded by the function DeleteEntry (line 4). *CircularCodes* are the sets of codons that appear in the same reading frame (line 6). If the reading frame preference is observed, the specific tri-nucleotide gets award (the frequency is increased by half of the original frequency, it is the meaning of the $\frac{3}{2}$ in line 7). The D_3 is the weighted sum of the frequencies of tri-nucleotides in each of the reading frames, the $+\frac{1}{2}$ in line 8 stands for the appearing in another reading frame is also acceptable but it is less important, while $-\frac{1}{4}$ for the appearance in the third reading frame is unacceptable and it is treated as a penalty.

Algorithm 1 Circular-code based tri-nucleotide disparity

```
1. $F(:,1) = codoncount(Seq,'ReadingFrame1');$
2. $F(:,2) = codoncount(Seq,'ReadingFrame2');$
3. $F(:,3) = codoncount(Seq,'ReadingFrame3');$
4. $F_{sorted} = sort(F);$
5. $F'_{sorted} = DeleteEntry(find(F(:,1) == 0\&\&F(:,2) == 0\&\&F(:,3) == 0));$
6. Let $CircularCodes = find(F'_{sorted}(:,2) == 0);$
7. $CircularCodes(i,1) = CircularCodes(i,1) \times \frac{3}{2};$
8. $D_3 = sum\left(F_{sorted}(:,1) + \frac{1}{2}F_{sorted}(:,2) - \frac{1}{4}F_{sorted}(:,3)\right);$
```

3.5 The simple EIDF classifier

The defined D_3 and D_2 are exon-dominated factors and the D_1 is intron-dominated factor. After formulating the disparities of mono-/di-/tri-nucleotide composition, a simple EIDF classifier is implemented to perform a bipartite comparison between the exon and its immediate intron within the single sequence. The discrimination function aims at maximizing the $EIDF_{exon}$ (i.e., minimizing the $EIDF_{intron}$), therefore, the D_3 and D_2 are treated as positive factors and D_1 is treated as a negative factor. Accordingly, the EIDF classifier is defined by an easy addition model as shown Eq. 7. The classifier performs bipartite comparison between exons and their immediate introns based on Eq. 8. Generally, the value of $EIDF^i_{exon}$ (the ith exon) is expected larger than the corresponding one of $EIDF^i_{intron}$ (the immediate intron of ith exon). The comparison result is either 1 or 0, thus, for a data set comprising n exons and introns, the total number of positive results is defined as the *support* of tests as defined in Eq. 9.

$$EIDF = D_3 + D_2 - D_1 \tag{7}$$

$$Support_i = \begin{cases} 1, & \text{if } EIDF^i_{exon} > EIDF^i_{intron} \\ 0, & \text{else} \end{cases}. \tag{8}$$

$$Support = \frac{\sum_{i=1}^{n} Support_i}{n} \tag{9}$$

4 Results

4.1 Data sets

There were nine different data sets used to test performance of the EIDF classifier. Firstly, the data sets were collected from the UCI repositories (Arthur and Newman 2007), SpliceDB (Burset et al. 2001) and HS3D (Pollastro and Rampone 2002). The number of experimental sequences in the three data sets is listed in Table 1 and they were used to validate the basic performance of the proposed EIDF classifier. Other data sets were manually extracted from the Xpro database (Gopalan et al. 2004), including the higher eukaryotes (human, mouse, rat) and lower eukaryotes (*C. elegans*, *Drosophila*, *Arabidopsis*), they were prepared to investigate the regional disparities between exons and

Table 1 Experimental sequences collected from three public databases

Data sets	UCI	SpliceDB	HS3D
#Exon–DONOR–Intron sequences	762	19,073	2,796
#Intron–ACCEPTOR–Exon sequences	765	19,161	2,880
Length of exonic/intronic regions	30	40	70

Table 2 Experimental sequences manually extracted from the Xpro database

Species	Human	Mouse	Rat	*C. elegans*	*Arabidopsis*	*Drosophila*
#Exon–DONOR–Intron sequences	29,795	6,749	811	44,204	51,522	28,505
#Intron–ACCEPTOR–Exon sequences	30,592	7,151	847	53,531	52,982	38,631
Length of exonic/intronic regions	70	70	70	70	70	70

introns. The number of experimental sequences in the six data sets is listed in Table 2.

4.2 Basic performance of EIDF classifier

The EIDF classifier is firstly tested using UCI, SpliceDB and HS3D data sets, and the performances are listed in in Table 3. The average supports reached 0.89 and 0.92, therefore, it is obvious that there exists significant heterogeneous base composition between exonic and intronic flank regions of splice sites. The strongly negative Spearman rank correlations between intronic EIDF classifier values and the difference of exon/intron EIDF classifier values provide another insight into exon/intron discrimination. Taking the HS3D experiments as an example, the EIDF classifier values of true exons and true introns are shown in the left part of Fig. 3.

4.3 The regional disparity

To identify the most significant regions discriminating exons from introns, the sequences were investigated under restricted lengths. The discrimination analysis on restricted exonic/intronic regions flanking splice sites of six model organisms (the data sets are listed in Table 2) was performed with restricted lengths ranging from 11 to 68 bp (considering the calculation of tri-nucleotide disparity, the sequence length is set to be $3k+2$ base pairs). As shown in the left part of Fig. 4, the degree of disparity increases as the sequence length gets longer for the donor site, but the exons and introns flanking acceptor site showed significant disparity (the left part of Fig. 4) at the range of about 30 bases. The regional disparities revealed that there exist different sequence properties between the exonic regions and their immediate intronic regions.

4.4 Robustness of EIDF classifier

For validating the robustness of the proposed EIDF classifier, an extended data set is prepared which consists of genes with large number of exons. As listed in Table 4, the complicated exon/intron organization provides a chance to test the robustness of EIDF classifier, where the *GID* is the gi number of the selected gene in NCBI, *nExon* is the number of exons in the specified gene; *Seq_len* is the sequence length; and *Window_Size* is the size of sliding window (in terms of number of base pairs) determined by two-thirds of the length of the shortest intron in that gene. The supports are shown in Table 5, the high correction ratios demonstrated the robustness of EIDF classifier.

Table 3 EIDF classifier on three public data sets

Data set	UCI	SpliceDB	HS3D
Exon–Donor–Intron	0.91	0.85	0.92
Intron–Acceptor–Exon	0.91	0.92	0.93

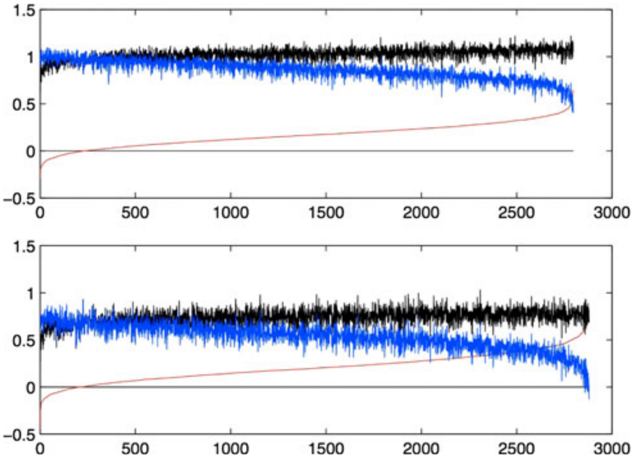


Fig. 3 The EIDF values of true Exon–Donor–Intron sequences (*top*) and the EIDF values of true Intron–Acceptor–Exon sequences (*bottom*)

Table 4 Example genes with complicated exon/intron organization

Species	GID	nExons	Seq_len	Window_Size
Human	11878411	105	29,713	41
Mouse	13517499	87	97,586	83
Rat	204099	42	15,360	70
C. elegans	24620453	65	80,941	42
Arabidopsis thaliana	4204276	68	25,800	47
Drosophila melanogaster	28380266	77	94,800	53

Table 5 Supports of EIDF classifier on complicated genes

Data set	Human	Mouse	Rat	*C. elegans*	*Arabidopsis*	*Drosophila*
Exon–Donor–Intron	0.86	0.92	0.94	0.90	0.87	0.89
Intron–Acceptor–Exon	0.95	0.96	0.89	0.98	0.89	0.96

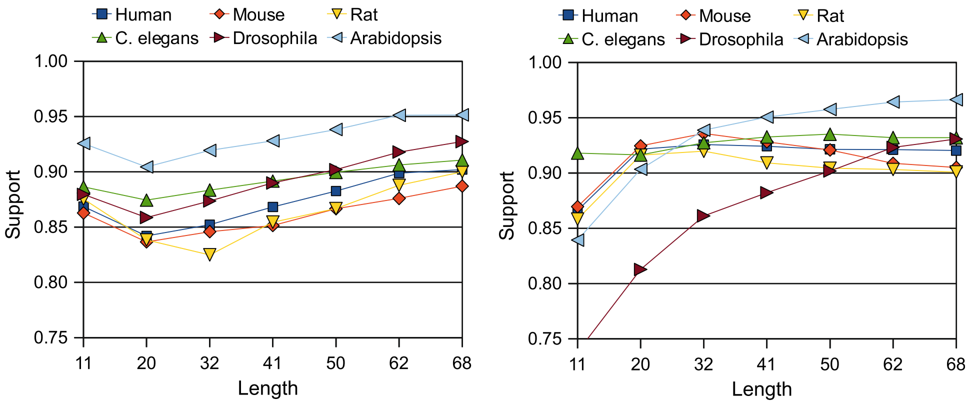


Fig. 4 The regional disparity between exonic and intronic regions flanking splice sites. The *y* axis is the support of EIDF classifier and the *x* axis is the restricted sequence length starting from the splice sites

5 Discussion and conclusions

The precise processes of intron removal and exon ligation during pre-mRNA splicing remain far beyond well understood; most of related works were devoted to discovering sequence features or trying to identify discriminative motifs based on results of sequence analysis. In this paper, an EIDF is proposed, which recruits the diverse base disparities between exons and introns. The EIDF is implemented as a simple binary classifier; the experimental results showed that it is effective and reliable in distinguishing exons from introns using only information of single sequence. The results suggest the proposed EIDF is a valuable profile of splice site sequences and it can serve as a new basis for investigating the splice site sequences. The proposed EIDF demonstrated a new paradigm of methodology in exon/intron discrimination, which provides new insights into the pre-mRNA splicing events and the possibility of simulating the processes of splicing machinery in silico is also revealed.

References

Arquèsa DG, Michel CJ (1996) A complementary circular code in the protein coding genes. J Theor Biol 182(1):45–58

Asuncion A, Newman D (2007) UCI machine learning repository

Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ (2010) Deciphering the splicing code. Nature 465(7294):53–59

Bortfeldt R, Schindler S, Szafranski K, Schuster S, Holste D (2008) Comparative analysis of sequence features involved in the recognition of tandem splice sites. BMC Genomics 9:202

Bulmer M (1987) A statistical analysis of nucleotide sequences of introns and exons in human genes. Mol Biol Evol 4(4):395–405

Burset M, Seledtsov IA, Solovyev VV (2001) SpliceDB: database of canonical and non-canonical mammalian splice sites. Nucleic Acids Res 29(1):255–259

Chen L, Xuan J, Wang C, Wang Y, Shih I-M, Wang T-L, Zhang Z, Clarke R, Hoffman EP (2009) Biomarker identification by knowledge-driven multilevel ica and motif analysis. Int J Data Min Bioinform 3(4):365–381

Claverie J-M, Bougueleret L (1986) Heuristic informational analysis of sequences. Nucleic Acids Res 14(1):179–196

Claverie J-M, Sauvaget I, Bougueleret L (1990) K-tuple frequency analysis: from intron/exon discrimination to t-cell epitope mapping. Methods Enzymol 183:237–252

Eskesen ST, Eskesen FN, Ruvinsky A (2004) Natural selection affects frequencies of ag and gt dinucleotides at the 5′ and 3′ ends of exons. Genetics 167(1):543–550

Gopalan V, Tan TW, Lee BTK, Ranganathan S (2004) Xpro: database of eukaryotic protein-encoding genes. Nucleic Acids Res 32(Database issue):D59–63

Liou S-W, Huang Y-F (2008) Investigating the intrinsic differences in flank regions of exon–intron junction sites. International Conference on biomedical engineering and informatics (BMEI'08), vol 2, pp 96–101

Liou S-W, Huang Y-F (2009) Identifying new sequence features for exon–intron discrimination by rescaled-range frameshift analysis. International Conference on computational and systems biology (ICCSB'09), vol 37, pp 349–353

Louie E, Ott J, Majewski J (2003) Nucleotide frequency variation across human genes. Genome Res 13(12):2594–2601

Luo L, Ji F (1997) The preferential mode analysis of DNA sequence. J Theor Biol 188(3):343–353

Malousi A, Chouvarda I, Koutkias V, Kouidou S, Maglaveras N (2010) SpliceIT: a hybrid method for splice signal identification based on probabilistic and biological inference. J Biomed Inform 43(2):208–217

Mengeritsky G, Smith TF (1989) New analytical tool for analysis of splice site sequence determinants. Comput Appl Biosci 5(2):97–100

Nakata K, Kanehisa M, DeLisi C (1985) Prediction of splice junctions in mRNA sequences. Nucleic Acids Res 13(14):5327–5340

Nikolaou C, Almirantis Y (2003) Mutually symmetric and complementary triplets: differences in their use distinguish systematically between coding and non-coding genomic sequences. J Theor Biol 223(4):477–487

Nikolaou C, Almirantis Y (2004) Measuring the coding potential of genomic sequences through a combination of triplet occurrence patterns and RNY preference. J Mol Evol 59(3):309–316

Orlov YL, Potapov VN (2004) Complexity: an internet resource for analysis of DNA sequence complexity. Nucleic Acids Res 32(Web Server issue):W628–W633

Parmley JL, Hurst LD (2007) Exonic splicing regulatory elements skew synonymous codon usage near intron–exon boundaries in mammals. Mol Biol Evol 24(8):1600–1603

Pollastro P, Rampone S (2002) HS3D: *Homo sapiens* splice site data set. Nucleic Acids Res (Annual Database Issue)

Solovyev VV, Salamov AA, Lawrence CB (1994) Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. Nucleic Acids Res 22(24):5156–5163

Staden R, McLachlan AD (1982) Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucl Acids Res 10(1):141–156

Wang Z, Burge CB (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14(5):802–813

Weir M, Rice M (2004) Ordered partitioning reveals extended splice-site consensus information. Genome Res 14(1):67–78

Willie E, Majewski J (2004) Evidence for codon bias selection at the pre-mRNA level in eukaryotes. Trends Genet 20(11):534–538

Zhang C-T (1997) A symmetrical theory of DNA sequences and its applications. J Theor Biol 187(3):297–306

Zhang MQ (1998) Statistical features of human exons and their flanking regions. Hum Mol Genet 7(5):919–932

Zhang XH-F, Leslie CS, Chasin LA (2005) Dichotomous splicing signals in exon flanks. Genome Res 15(6):768–779

Zhuo D, Madden R, Elela SA, Chabot B (2007) Modern origin of numerous alternatively spliced human introns from tandem arrays. Proc Natl Acad Sci USA 104(3):882–886