Vienna
AUSTRIA CENTER VIENNA

AS-SIG @ ISMB
Alternative Splicing - Special Interest Group meeting
July 19-20, 2007

TALKS

SHORT TALKS

POSTERS

Internationales Amtssitz- und Konferenzzentrum Vienna
Bruno-Kreisky-Platz 1
1220 Vienna

2
Content

Program, July 19th

Page 4

Program, July 20th

Page 5

Talks and abstracts

Page 6

Poster list and abstracts

Page 41

Organizing committee

Page 54

Names and e-m@ils

Page 55
Program, July 19th

REGULATORY MECHANISMS OF SPICING

09.00 - 09.45  Cellular programs of alternative splicing regulation, Juan Valcárcel
09.45 - 10.30 Splicing, Silencing and Nonsense, Christopher Smith
10.30 - 11.00 Coffee break
11.00 - 11.20 Pre-mRNA secondary structures influence exon recognition, Michael Hiller
11.20 - 11.40 Computational analysis of alternative human exons originating from tandem donors, Ralf Bortfeldt
11.40 - 12.00 Violating the splicing rules: TG dinucleotides function as alternative 3’ splice sites in U2-dependent introns, Karol Szafranski
12.00 - 13.30 Lunch break
13.30 - 14.15 Alternative splicing in C.elegans: genetic, genomic and molecular approaches to understanding regulation, Alan Zahler
14.15 - 15.00 SR protein splicing factors as targets for therapy, Jamal Tazi

EVOLUTION OF ALTERNATIVE SPICING

15.00 - 15.20 Non-coding alternative splice variants of plant-specific SR proteins are tightly regulated and highly conserved in evolution, Maria Kalyna
15.30 - 16.00 Coffee break
16.00 - 16.20 Exonization and establishment in human genes, Eduardo Eyras
16.20 - 16.40 Patterns of selection and evolution of the exon-intron structure in alternatively spliced genes of nine Drosophila species and the malarial mosquito, Ekaterina Ermakova
17.00 - 18.30 Poster session
19.00 Conference dinner
Program, July 20th

ALTERNATIVE SPLICING AND FUNCTION

09.00 - 09.45  Functional coordination and evolution of mammalian alternative splicing, Benjamin Blencowe

09.45 - 10.05  Discovery of tissue-specific exons using comprehensive human exon microarrays, Tyson Clark

10.05 - 10.25  Alternative splicing profiling and regulation in human body tissues, John Castle

10.30 - 11.00  Coffee break

11.00 - 11.20  Transcript isoform entropy: proof of extensive disruption of splicing in cancer tissues, William Ritchie

11.20 - 11.40  Alternative splicing expands the human transcriptome and proteome by one order of magnitude, Graziano Pesole

11.40 - 12.00  Comprehensive analysis of alternative splicing in the chicken, Shoba Ranganathan

12.00 - 13.30  Lunch break

GENOMICS AND BIOINFORMATICS OF SPlicing

13.30 - 14.20  Exonic splicing motifs and the design of synthetic exons, Larry Chasin

14.20 - 14.40  Murky beginnings and Alternate Ends: Transcript Diversity at 5' and 3'-ends, Uwe Ohler

14.40 - 15.00  Computational evidence for the association between transcription initiation and internal splicing, Mihaela Zavolan

15.00 - 15.20  Bootstrapping the alternative splicing annotation of newly sequenced genomes, Cheng Soon

15.30 - 16.00  Coffee break

16.00 - 16.20  Splicing isoforms in the ENCODE complement, Michael Tress

16.20 - 16.40  The Alternative Splicing and Transcript Diversity Database, Elenore Whitfield

16.40  Conclusion
How do cells deploy genetic programs that coordinate numerous alternative splicing decisions influencing cell differentiation and development? We have designed microarrays able to distinguish between alternatively spliced mRNA isoforms and used them to analyze mechanisms of alternative splicing regulation.

Sex determination in Drosophila offers a textbook example of an alternative splicing regulatory cascade affecting a handful of genes. We have investigated to which extent the fly transcriptome shows sex bias at the level of AS using splicing microarrays. We find that over 400 genes show differences in the relative abundance of mRNA isoforms between male and females flies. Interestingly, this category is enriched in factors involved in mRNA translation and pre-mRNA splicing. Searches for binding sites of sex determination factors SXL and TRA/TRA-2 revealed that only a small number of events harbor such sites, suggesting the existence of additional factors controlling sex-specific splicing in Drosophila. Evidence for alternative splicing events that show both sex- and tissue-specific splicing was derived from the analysis of tissues, cell lines and genetic manipulation of classical regulatory factors.

We have analyzed alternative splicing changes in mammalian cells in parallel with the changes in the expression of all known components of the spliceosome, splicing regulatory factors and their isoforms. We have compared a variety of biological samples, including different tissues, cell lines undergoing muscle differentiation and cell lines derived from Hodgkin lymphoma tumors at different stages of tumor progression. Surprisingly, a relatively low number of changes in the expression of splicing factors are associated with the diverse array of biological processes studied. Interestingly, some of these changes are associated with variations in the expression / alternative splicing of factors which act at late steps in the assembly of splicing complexes. Indeed, biochemical analyses have allowed us to identify factors that regulate alternative splicing by targeting late events in the spliceosome assembly pathway. For example, the protein RBM5 promotes skipping of the apoptotic receptor Fas exon 6 (leading to the production of an anti-apoptotic decoy receptor) by inhibiting incorporation of the U4/5/6 tri-snRNP to pre-spliceosomal complexes assembled on the splice sites flanking exon 6. These observations suggest that in addition to promoting or inhibiting early splice site recognition, physiological mechanisms of splicing regulation can also target subsequent steps in spliceosome assembly and splice site pairing, thus extending the range of molecular events in the splicing reaction which can be subject to regulation.
Our group is interested in the mechanisms of regulated alternative pre-mRNA splicing. Our traditional approaches have been to characterize *cis*-acting regulatory elements and *trans*-acting regulatory factors in individual model systems of regulated splicing. More recently, we have become interested in the consequences of alternative splicing, particularly in cases where regulated splicing leads to production of mRNAs that are channeled down the pathway of Nonsense Mediated Decay (NMD). We have also started to adopt more global approaches for analysis of alternative splicing, including the use of quantitative proteomics to analyze the consequences of splicing factor knockdown. By using this approach to identify targets of the splicing repressor PTB, we have uncovered a network of cross-regulation between PTB and its two tissue-restricted paralogs, nPTB and ROD1. We had previously shown that PTB expression was auto-regulated in a negative feedback loop in which PTB protein promoted PTB exon 11 skipping, thereby producing a frame-shifted RNA that is targeted by NMD. We now find similar AS-NMD switches in nPTB and ROD1. In HeLa cells the nPTB switch can be flipped by knockdown of PTB, while the ROD1 switch is flipped by knockdown of PTB and nPTB. In a second example of AS-NMD, we have analyzed a conserved pseudo-exon like sequence in the *Tpm1* gene that, if spliced into mRNA, introduces premature termination codons. Although conserved, the Nonsense Exon, is readily activated by a series of simple point mutations. Using Zhang & Chasin’s octamer data set to guide our mutagenesis, we have successfully mapped splicing regulatory elements within the Nonsense Exon. In addition to a series of splicing enhancers, a strong hnRNP H/F binding silencer was identified. We are currently investigating the potential roles, if any, of the Nonsense Exon.
Pre-mRNA secondary structures influence exon recognition

Michael Hiller¹, Zhaiyi Zhang², Rolf Backofen¹ and Stefan Stamm²

¹ Bioinformatics group, Albert-Ludwigs-University Freiburg, Georges-Koehler-Allee 106, 79110 Freiburg, Germany, ² Institute for Biochemistry, University of Erlangen, Fahrstrasse 17, 91054 Erlangen, Germany

Although mRNA does not fold into global secondary structures, its local secondary structures influence a number of processes. Several studies have shown that the regulation of alternative splicing is influenced by pre-mRNA secondary structures. For example, the tight regulation of a cluster of 48 mutually exclusive exons in the Drosophila DSCAM gene is achieved by formation of secondary structures (1). Since splicing regulatory proteins (such as SR proteins and hnRNPs) are equipped with one or more single-stranded RNA binding domains, the sequestration of their binding sites into double strands can prevent the binding of these proteins. For example, a deletion in the mouse fibronectin EDA exon leads to a shift of a critical ESE from single- into double-stranded conformation, which causes complete exon skipping (2).

In our study, we analyzed whether the structural context of binding sites for splicing regulatory proteins has a general importance. To this end, we investigated the structural context of experimentally verified splice enhancer and silencer motifs in their natural pre-mRNA. To measure the extent of base pairing within the splicing motif, we computed the probability that all bases in the motif are unpaired (denoted as the PU value) using local windows around the motif. These PU values account for all possible secondary structures and thus circumvent inaccuracies caused by considering only the optimal structure or a limited number of suboptimal structures. Furthermore, the PU values for two motifs of equal length can be directly compared.

To assess whether verified splicing motifs have a preference for single strands, we constructed several simulated data sets by changing the location of the motifs, modifying the flanking regions by dinucleotide shuffling, and choosing random motifs. Comparing the average PU values between the verified motifs and the simulated data sets, we found strong statistical support that verified motifs exhibit a noticeably higher single-strandedness.

We next tested the hypothesis that the localization of a splicing regulatory element in a single- or double-stranded RNA structure influences splice site selection in an experimental system using the SXN-minigene. We inserted known splicing enhancer and silencer motifs located in either a single-stranded hairpin loop or a double-stranded stem into the middle alternative exon. The single- and double-stranded motifs are located at equal positions in the exon to avoid positional effects. We found that the structural context of the splicing motifs has a strong influence on the splicing outcome. Enhancers lead to stronger exon inclusion when they are located in a loop compared to the location in the stem structure. Likewise, silencers located in a loop lead to stronger exon skipping compared to the loop structure. These experiments show that the pre-mRNA conformation influences the action of a splicing regulatory element.

Next, we asked whether we could detect selection on the structural context of computationally predicted splicing regulatory motifs. A classification of all hexamers into ESE, ESS, and 'splicing-neutral' motifs were taken from (3). Our strategy to detect differences in single-strandedness is to compare the single-strandedness for one hexamer between ESE-dependent and ESS-dependent regions. As ESE-dependent regions we used human exons. As ESS-dependent regions we considered (i) pseudo exons (silent intronic regions bounded by strong splice sites), (ii) 'decoy regions' (intronic regions
between an authentic and a strong intronic pseudo or 'decoy' splice site), and (iii) intronic regions adjacent to authentic splice sites.

We found that ESEs have a significantly higher single-strandedness than ESSs and splicing neutral motifs when comparing exons with pseudo exons, decoy regions, and intron flanks. Interestingly, we observed that ESSs have a significantly higher double-strandedness in exons than neutral and especially ESE motifs. These results indicate that the high ESE frequency in ESE-dependent regions is further intensified by a tendency for single-strandedness. Likewise, ESSs are rare in ESE-dependent regions and if they occur, their structural context has a tendency to be selected for double-strandedness. These results indicate a general and widespread selection pressure on the structural context of splicing motifs. Selection on secondary structures could also explain the observed constraints on synonymous codon sites and the conservation of larger intronic regions adjacent to alternative exons (4,5). We propose that a coding exon is subjected to at least three different selection pressures: (i) preserving the coding sequence, (ii) preserving the sequence of splicing motifs, and (iii) preserving an appropriate structural context for these splicing motifs.

Our study can have implications to interpretation and design of mutagenesis experiments and the analysis of splicing-relevant SNPs. Usually an observed variation in the splicing pattern is directly interpreted as a change in splicing motifs by the mutation. However, our findings suggest that the observed splicing effects could also be explained by structural changes caused by the mutation. Such an example is a mutation in the CFTR exon 12 that causes a strong reduction in exon inclusion but does not quantitatively and qualitatively change any ESE or ESS motifs. Comparing the change in secondary structures, we found that an ESE region becomes more double-stranded, while an ESS region becomes more single-stranded, which is likely to explain the reduced exon inclusion. Moreover, secondary structures were previously found to be beneficial in RNA sequence motif finding (6) and it is conceivable that structural information increases the power of computational methods to predict splicing motifs (7).

To summarize, our computational and experimental findings suggest that pre-mRNA secondary structures are an integral part of splice site recognition and therefore important to understand the regulation of alternative splicing. Thus, they are a part of the 'mRNA splicing code' (8).

REFERENCES

Analysis of human alternative exons originating from tandem donors

Ralf Bortfeldt 1, Stefanie Schindler 2, Karol Szafranski 2, Stefan Schuster 1, and Dirk Holste 3

1 Department of Bioinformatics, Friedrich-Schiller University, Ernst-Abbe-Platz 2, 07743 Jena, Germany
2 Fritz-Lipmann Institute for Aging Research, Beutenbergstraße 11, 07745 Jena, Germany
3 Institute of Molecular Pathology, Dr-Bohr-Gasse 7, 1030, Vienna, Austria

Pre-mRNA splicing is a central mode in the regulation of eukaryotic gene expression. The recognition of authentic pre-mRNA splice sites out of many pseudo-sites, the excision of introns, and the ligation of exons to produce a correct message are catalyzed by a ribonucleoprotein complex known as the spliceosome. The processing of pre-mRNAs is often variable, producing multiple alternatively spliced (AS) mRNA isoforms. Alternative 5’ splice site exons (A5Es), which are affected in their choice of the 5’ss, are frequently observed and their regulation has been shown to play decisive roles during development (e.g., sex of the fly, developmental stages of the human CFTR gene) or human disease (e.g., in tau). A5Es are thought to be regulated by antagonistic splicing factors, which affect the choice of splice sites in a concentration-dependent manner, and by cis-regulatory elements in exons and nearby introns. While A5E extensions from the distal to the proximal splice site can typically range up to 100 and more nucleotides in mammalian exons, there is a clear bias toward shorter extensions that peaks at four nucleotides (A5EΔ4). Such subtle splice variants have been observed previously in vertebrates, invertebrates, and plants, albeit focusing on the 3’ss (A3Es), where the extension is frequently three nucleotides long. Due to the proximity of splice sites, such A5EΔ4 events may be differently regulated as more distantly spaced 5’ss. We conduct a computational analysis of A5E/A5EΔ4 events of human exons derived from transcripts of different human tissues, which were aligned to the assembled human genome to infer patterns of AS. We analyze AS patterns, transcript coverage, sequence conservation and matches to U1 snRNA, as well as known splicing regulatory motifs. We experimentally validate about a dozen of A5EΔ4 events in human cell lines, and compare A5E/A5EΔ4 events with constitutive exons and A3Es. The results suggest (1) a classification into two exon sub-populations of predominantly proximal and distal donors, (2) identify differences between rates of reading-frame preserving and destroying AS events, and (3) predict putative substrates of NMD. While the question whether a stochastic or deterministic mechanism may better explain A5EΔ4 events is currently under investigation, implications of a possible link between AS and NMD in the processing of human transcripts as well as biological interesting examples are discussed.

REFERENCES

Violating the splicing rules: TG dinucleotides function as alternative 3’ splice sites in U2-dependent introns

Karol Szafranski (1), Stefanie Schindler (1), Stefan Taudien (1), Michael Hiller (2), Klaus Huse (1), Stefan Schreiber (3), Rolf Backofen (2), Matthias Platzer (1)

1) Genome Analysis, Leibniz Institute for Age Research - Fritz Lipmann Institute, Beutenbergstr. 11, 07745 Jena, Germany, 2) Institute of Computer Science, Albert-Ludwigs-University Freiburg, Georges-Koehler-Allee 106, 79110 Freiburg, Germany, 3) Institute of Clinical Molecular Biology, Christian Albrechts University Kiel, Schittenhelmstr. 12, 24105 Kiel, Germany

1. INTRODUCTION

Despite some degeneracy of sequence signals that govern splicing of eukaryotic pre-mRNAs, it is an accepted rule that U2-dependent introns exhibit the 3’ terminal dinucleotide AG. Intrigued by anecdotal evidence for functional non-AG 3’ splice sites (3’ss) [1], we carried out a comprehensive screen of the human transcriptome.

2. RESULTS AND CONCLUSIONS

Through an analysis of available transcript data we identified TG dinucleotides functioning as 3’ss in 36 human genes. Strikingly, such non-canonical 3’ss are exclusively found as part of TG-AG 3’ss tandems that exhibit alternative splicing and result in exon length variation. The reported cases do not comprise U12-dependent introns, known for flexibility in 3’ss choice, according to the absence of characteristic sequence signatures recognized by the U12-type spliceosome [2]. We performed RT-PCR and PCR/re-sequencing of pooled genomic DNA to validate the TG-derived splice variants, with an overall success rate of 92%. Interestingly, we found that ESTs indicate non-canonical splice variants with higher sensitivity and more reliably than do GenBank mRNA entries or RefSeq transcripts. A probable explanation is that EST production involves minimal interference by researchers, resulting in stochastic error patterns, which are easily modeled. Moreover, the redundancy of EST data allows to use numerical significance as a filter criterion.

TG 3’ss and their flanking intron sequences are substantially conserved between orthologous vertebrate genes, even between human and frog, indicating functional relevance (Fig.1). Ratios of splice site choice were measured by pyrosequencing, an accurate and cost-efficient method for the analysis of DNA mixtures. In the case of Brunol4 intron 4, the ratio turned out to be highly tissue-specific (fig. 2A), corroborating the idea of tightly regulated splice site choice. The ubiquitously expressed ARS2 gene, however, showed only minor fluctuations of the relative amounts of unusual splice products (fig. 2B). Since the observed intron sequence conservation suggests regulated splicing, a stable ratio of ARS2 splice variants may be desirable for the organism. Alternatively, regulated splicing may have functional roles independent of the resulting splice variants, e.g. attenuated transcript maturation in order to control the mRNA level.

Since TGs are exclusively found as one of alternative 3’ss, the U2 spliceosome apparently accomplishes perfect specificity for 3’ AGs at an early splicing step, but may choose 3’ TGs during later steps. Given the tiny fraction of TG 3’ss compared to the vast amount of non-viable TGs, cis-acting sequence signals must significantly contribute to splice site definition. An investigation of the sequence context of TG-AG 3’ss tandems did not reveal a significant enrichment of known or novel
splice regulatory motifs. This argues in favor of a contextual and intron-specific role of cis-regulatory sequences. Thus, we consider TG-AG 3’ss tandems as promising objects for future experimental studies on 3’ss selection.

3. FIGURES

**Fig.1.** Conservation of the TG splice site found in the *RYK* gene from human to chicken. (A) Human genomic sequence and derived splice variants. Canonical AG (filled triangle) and non-canonical TG 3’ss (open triangle) are indicated. (B) Alignment of orthologous intron-exon boundary regions from several vertebrate genomes. Numbers on the right display the ratios of species-specific ESTs for the TG and AG splice sites, respectively.

**Fig.2.** Tissue-specific fractions of TG-derived splice variants. (A) *BRUNOL4* intron 6, (B) *ARS2* intron 18. Pyrosequencing assays were performed multiple times for each sample (2-4 times), and error bars depict the standard deviation of these independent measurements.

REFERENCES


Towards improving our understanding of the regulation of pre-mRNA splicing, we have been employing the nematode *Caenorhabditis elegans* as a model system. Some advantages of this system include its fully sequenced genome (along with that of several related species), ease of genetic and reverse genetic screens, three day life cycle, and the fact that homologs for the vast majority of known spliceosome-associated and splicing regulatory proteins identified in vertebrates are also found here. Our progress towards understanding alternative splicing regulation in nematodes employing several different approaches will be presented. We have used evolutionary conservation of sequences in introns flanking alternatively spliced exons to identify splicing regulatory elements. Using statistical analysis of these conserved regions, we have identified pentamer and hexamer sequence elements that are likely to function in splicing regulation. Several of these new splicing regulatory elements have been confirmed experimentally, and using biochemical approaches we have identified nuclear RNA binding proteins that interact with these regulatory sequences. We have developed a DNA microarray capable of discerning alternative splicing of 400 high-confidence exon-skipping events in these animals. Using this microarray, we have surveyed alternative splicing regulation during nematode development. We find many interesting examples of developmental stage-specific splicing regulation and we are exploring the regulatory relationship among alternative splicing events that cluster together in this analysis. We are also using the microarray to explore changes in splicing in strains carrying mutations in alternative splicing regulatory proteins in order to identify targets that have regulation in common. We have also been using genetics to identify factors that regulate cryptic 5' splice site choice when wild-type splice donors are mutated. We previously reported the identification of U1 snRNA suppressors that can suppress mutations to the canonical G nucleotide at the beginning of introns. Our current screen for cryptic 5' splice site suppressors has identified several protein coding genes capable of this type of informational suppression. This simple model system for splicing regulation is yielding important information about the overall regulation of splice site choice.
In humans the complexity of constitutive and alternative splice site recognition suggests multiple levels of regulation, each resulting from a combination of cis-elements and trans-acting factors adapting the enzyme responsible for intron excision “the spliceosome” to various situations. The significance of these observations, especially in regard to human pathologies, is that there is a considerable number of disease-causing mutations in exons and introns that disrupt previously unrecognized auxiliary cis-elements as well as the well-known classical splice sites. Furthermore, viruses like the human immunodeficiency virus (HIV) uses a combination of several alternative 5’ and 3’ splice sites to generate more than 40 different mRNAs from its single transcribed genome pre-mRNA that are essential for its life cycle. Thus, targeting either conserved constituent of the spliceosome and/or regulatory sequences or factors that bind to them, holds great promise for future therapeutic action to correct aberrant splicing caused by these mutations (1). Our lab is currently developing an entirely new approach in which splicing factors are targeted by small chemical molecules. A large screen program has been conducted with the chemical library of the Curie Institute. Several molecules were found to counter the splicing stimulatory effect of individual members of the SR protein family and have been selected for further development in pathological splicing events (2).

REFERENCES
Non-coding alternative splice variants of plant-specific SR proteins are tightly regulated and highly conserved in evolution

Maria Kalyna, Andrea Barta
Max F. Perutz Laboratories, Medical University of Vienna, Dr. Bohrgasse 9/3, A-1030, Vienna, Austria

Introns interrupt coding sequences of most eukaryotic genes. Accuracy of excision of introns from pre-mRNA relies on the precise recognition of relatively poorly conserved consensus sequences at both splice sites and the branch point region. SR (Ser/Arg) proteins govern splice site recognition and spliceosome assembly. The Arabidopsis genome encodes 19 SR proteins, which is almost twice as much as in humans. They fall into seven subfamilies (1) some of which have orthologs in metazoan (SF2/ASF, 9G8, SC35) but interestingly three of them seem to be plant-specific (RS, RS2Z, and SCL). Many of the SR protein genes are alternatively spliced both in metazoa and in Arabidopsis. It has been noted that alternative splicing occurs mainly in and around the long introns of the Arabidopsis SR genes ranging in size from about 400 to 1100 nucleotides (1), while the typical size of plant introns is less than 150 nt. Alternative splicing in Arabidopsis homologues of human SF2/ASF occurs in the long introns situated close to the 3' ends of the genes and results in the protein isoforms with the shortened for SR domains (2, 3). Members of plant-specific subfamilies are also alternatively spliced, and in some tissues alternative splice variants are the major or the only transcripts (1, 4-6). These alternative splicing events are tightly regulated. We have demonstrated that splicing of the long intron of atRSZ33, a member of the RS2Z subfamily, is autoregulated (7). This autoregulation is crucial for correct gene expression levels, because overexpression of an intronless version of atRSZ33 is lethal.

Interestingly, in the plant-specific SR protein subfamilies alternatively spliced intron is located between the RNP2 and RNP1 motifs of their N-terminal RRM (RNA recognition motif). Analysis of the alternative splice forms in all three subfamilies shows that they can encode only extremely truncated proteins which would contain only a part of the RRM due to premature termination codons (PTC) generated by the inclusion of intronic sequences (1, 4-6). These proteins, if produced, should have no influence on splicing activity as they lack both RNA and protein interaction domains. We have shown that alternative splice variants of atRSp31, a member of the RS subfamily, do not produce predicted truncated protein (8). In addition atRSp31 is strongly regulated both on the transcriptional and post-transcriptional levels at different developmental stages, stress conditions and in a tissue-specific manner. However, splicing events in atRSp31 are not regulated by atRSp31 itself, but by atRSZ33. Furthermore, our results indicate that major alternative splice variants are not subjected to nonsense-mediated decay in spite presence of PTC, in contrast to splice variants of human SR genes (9, 10).

To investigate the significance of these alternative splicing events we have analyzed splicing profiles of orthologous genes of RS and RS2Z subfamilies in different species. First, we have identified orthologues of Arabidopsis SR proteins in monocots, gymnosperms, mosses and green algae. Interestingly, our search revealed presence of genes coding for SR proteins and their alternative splicing in the unicellular organism, green algae Chlamydomonas reinhardtii. Next, we have shown that all orthologues contain the intron in the conserved position between RNP-2 and RNP-1 of their N-terminal RRM. Furthermore, in all analyzed species, existing EST and cDNA data support presence of
alternative splice variants that involve this intron and contain PTC. In addition, alternative splice sites and adjacent sequences are highly conserved in different lineages. However, neither included alternative exons, nor positions of PTCs are conserved.

Evolutionary preservation of splice forms, which are not translated into protein together with highly, conserved intron features suggest other functions hidden in genes encoding plant-specific SR proteins.

REFERENCES


1. INTRODUCTION

It is fully accepted that Alternative Splicing (AS) is a widespread mechanism responsible for increased protein diversity and complexity among eukaryotes. The importance of this mechanism in the regulation of gene function has raised the question of its role in the context of evolution. There is recent evidence suggesting that AS can provide an organism the possibility to explore new protein functions while not disrupting the fitness of the original protein. In this way, it contributes to an increased rate of intron-exon structure evolution by reducing negative selective pressure against events such as exon creation (1). It has also been demonstrated that Alu elements play a key role in shaping intron-exon structures in primates. They contribute to exon creation due to the fact that they present, as part of their consensus sequence, motifs resembling canonical splice sites (2). However, few studies have focused on the distribution of cis-Regulatory Elements (cisRE) along them. In order to understand the requirements for de novo exonization, and how splicing regulation changes with time, three groups of human exons with different evolutionary ages were defined: Primate Specific (PS), Mammalian Specific (MS) and Vertebrate or Older (VO), and parameters such as EST inclusion level, splice site score and cis-RE (3,4,5,6) densities were determined.

In accordance with previously published results, PS exons present very low EST inclusion levels, while MS and VO exons are highly included (7). Our analyses suggest that splicing decisions might greatly depend on the exon/intron density ratio for cisRE (Figure 1A), and that splicing regulation changes throughout an exon's lifespan. Older exons presented higher exon/intron density ratios for Exonic Splicing Enhancers (ESEs) and lower for Exonic Splicing Silencers (ESSs), leading to higher inclusion levels. The obtained results for mobile element density and overlap show that MS and VO exons present low density of mobile elements in the adjacent intronic regions and negligible overlap, while PS exons greatly overlap with such elements, in particular with Alu sequences. Our results suggest that Alu exonization might not only depend on the presence of splice sites but also on the fact that between these, there is a desert of ESSs creating the correct environment for exonization (Figure 1B). This property seems to be Alu specific and might explain why Alu elements broadly serve as substrate for exon creation events as well as alternative splice site creation. We also suggest that a lack of ESEs might be the reason why thousands of Alu sequences containing optimal splice sites do not get exonized.
2. FIGURES

![Diagram showing exonic densities and motif exon/intron density ratios for VO, pseudo exons (A) and PS exons overlapping and not overlapping Alu elements (B). Exonic densities are given by the proportion of bases overlapped by motifs belonging to a particular set. Motif exon/intron density ratios are given by this formula: \( \ln \left( \frac{\text{exonic density}}{\text{adjacent intronic density}} \right) \).]

REFERENCES

1. INTRODUCTION

The evolution of alternative splicing in insects is not as well studied as in mammals: the alternative regions are shorter and there are less EST data. The patterns of selection in constitutive and alternative regions in flies and in mammals seem to differ [1]; however, the rates of nucleotide substitutions and selection patterns in alternatively spliced genes in the latter are subject of controversy [2-4].

In this study we consider patterns of selection and evolution of the exon-intron structure in alternatively spliced genes of nine Drosophila species: (see Figure 1 for the phylogenetic tree) and malarial mosquito Anopheles gambiae (Agam)

1. D. melanogaster (Dmel)
2. D.simulans (Dsim)
3. D.yakuba (Dyak)
4. D.erecta (Dere)
5. D. ananassae (Dana),
6. D. pseudoobscura (Dpse)
7. D. mojavensis (Dmoj),
8. D. virilis (Dvir)
9. D. grimshawi (Dgri)

Figure 1. Phylogenetic tree of nine considered Drosophila species.

2. RESULTS

Like in mammals, alternative regions in insects absorb more nucleotide substitutions than constitutive ones and negative selection in alternative regions is weaker and/or positive selection is stronger than in constitutive regions. Unlike mammals where the rates of nucleotide substitutions are higher and the dS/dN ratio is greater in C-terminal alternatives, terminal alternative regions in flies evolve slower than
internal ones. There is strong evidence for positive selection in internal alternative regions in flies (dN/dS=1.43, Dmel vs Dpse, the Ina method I (Ina 1995)).

The patterns of intron gains and losses are different in the melanogaster subgroup (Dsim, Dyak, Dere), more “distant” flies of the Sophophora subgenus (Dana and Dpse), and the species of the Drosophila subgenus (Dmoj, Dvir, Dgri). Taking the Dmel structure as the reference, we observe that Dana and Dpse have similar counts of lost and gained introns, whereas in other species, intron gains prevail over losses (Dyak fourfold, Dsim sixfold, more distant ones, about fifty percent more, see Table 1).

<table>
<thead>
<tr>
<th></th>
<th>Conserved</th>
<th>Lost</th>
<th>Gained</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dmel</td>
<td>20954</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Dsim</td>
<td>18754</td>
<td>23</td>
<td>162</td>
</tr>
<tr>
<td>Dyak</td>
<td>19357</td>
<td>33</td>
<td>122</td>
</tr>
<tr>
<td>Dere</td>
<td>19476</td>
<td>19</td>
<td>50</td>
</tr>
<tr>
<td>Dana</td>
<td>18399</td>
<td>222</td>
<td>220</td>
</tr>
<tr>
<td>Dpse</td>
<td>17878</td>
<td>294</td>
<td>290</td>
</tr>
<tr>
<td>Dmoj</td>
<td>16718</td>
<td>351</td>
<td>560</td>
</tr>
<tr>
<td>Dvir</td>
<td>16891</td>
<td>360</td>
<td>481</td>
</tr>
<tr>
<td>Dgri</td>
<td>16841</td>
<td>366</td>
<td>580</td>
</tr>
<tr>
<td>Agam</td>
<td>5542</td>
<td>2736</td>
<td>1410</td>
</tr>
</tbody>
</table>

Table 1. Counts of conserved, lost, and inserted introns using Dmel as the reference genome. Only introns in genes having orthologs in the respective genome are considered in each line.

We also compared conservation of alternative and constitutive Dmel genome segments. We considered constitutive exons, cassette exons, mutually exclusive exons, constitutive parts of exons with alternative fragments, donor and acceptor extensions, and retained introns separately. Constitutive parts of exons with alternative fragments were more conserved than alternative segments but less conserved than constitutive exons (in agreement with results of [5] for mammals), donor site extensions were more conserved than acceptor site extensions, and mutually exclusive exons — more conserved than cassette ones.

REFERENCES

Recent large-scale analyses of transcript sequence and microarray data have begun to provide fundamental insights into the global properties of alternative splicing in metazoans. We have developed custom microarrays and computational tools for the global-quantitative analysis of alternative splicing patterns in mammalian cells and tissues. From profiling a diverse range of cell and tissue types, as well as cells grown under different physiological conditions, we have uncovered sets of regulated alternative exons that are concentrated in genes, which operate in specific biological processes and pathways. A reoccurring observation is that primarily non-overlapping subsets of functionally associated genes display cell/tissue-type or condition-dependent regulation at the levels of alternative splicing and transcription. The data from these studies is facilitating the identification of elements of the “code” underlying regulated alternative splicing patterns. We have also applied our quantitative profiling system to compare alternative splicing patterns of orthologous exons in different mammalian species, including human and chimpanzee. These studies are providing among the first insights into the sets of genes, which display patterns of alternative splicing that have diverged within relatively short evolutionary time intervals. Results from the above areas of investigation will be presented.
Discovery of tissue-specific exons using comprehensive human exon microarrays

Tyson A. Clark, Anthony C. Schweitzer, Michelle K. Staples, Tina X. Chen, John E. Blume

Expression Research, Affymetrix, Inc.
3420 Central Expressway, Santa Clara, CA 95051

Higher eukaryotes express a diverse population of messenger RNAs generated by alternative splicing. Large-scale methods for monitoring gene expression must adapt in order to accurately detect the transcript variation generated by this splicing.

We have designed a high-density oligonucleotide microarray with probesets for more than one million annotated and predicted exons in the human genome. Using these arrays and a simple algorithm that normalizes exon signal to signal from the gene as a whole, we have identified tissue-specific exons from a panel of 16 different normal adult tissues. RT-PCR validation confirms approximately 86% of predicted tissue-enriched probesets. Pair-wise comparisons between the tissues suggest that as many as 73% of detected genes are differentially alternatively spliced. We also demonstrate how an inclusive exon microarray can be used to discover novel alternative splicing events. As examples, 17 new tissue-specific exons from 11 genes were validated by RT-PCR and sequencing.

In conjunction with a conceptually simple algorithm, comprehensive exon microarrays can detect tissue-specific alternative splicing events. Our data suggest significant expression outside of known exons and well annotated genes and a high frequency of alternative splicing events. In addition, we identified and validated a number of novel exons with tissue-specific splicing patterns. The tissue map data will likely serve as a valuable source of information on the regulation alternative splicing (1).

REFERENCES

Alternative splicing profiling and regulation in human body tissues

John Castle, Chris Armour, Jyoti Shah, Chris Raymond, Jason Johnson
Rosetta Inpharmatics, LLC (Merck & Co.), Seattle, Washington 98109, USA

We computationally identified over 20,000 RNA alternative splicing events and 200,000 individual exons. We generated microarray patterns to monitor these splicing events and individual exons in many human tissues and cell lines. After comparing microarray patterns and analysis techniques, we conclude that for alternative splice form monitoring, the use of both exon and junction probes is superior to either probe type individually, that analysis algorithms that incorporate multiple probes per event outperform algorithms that use single probes/probesets per event, and present a novel analysis algorithm. These profiling results include many tissue-enriched splicing events, including events specific to tumor and cell lines, and both known and novel splice form expression patterns. Using a novel motif-recognition protocol, these data clearly and de novo identify known alternative splicing motifs in the correct tissues and identify many more putative tissue-specific alternative splicing regulatory motifs. While samples are similarly clustered using splice-event and gene expression, the genes showing changes in alternative splicing are distinct from the genes showing overall expression changes and the two gene sets exhibit distinct pathway enrichment, as expected from two distinct regulatory mechanisms. Finally, onco-related expression changes are more apparent from alternative splice expression data, versus from gene expression data.
The Alternative Splicing and Transcript Diversity (ASTD) database is a new database available at the European Bioinformatics Institute (1). This database is a result of the merging, continuation and improvement of the Alternative Splicing Database (ASD) (2) and the Alternative Transcript Diversity database (ATD) (3), previously maintained at the EBI.

The ASTD database has genome wide alternative splicing data for human, mouse and rat. The aim of the database is to predict full-length transcripts for all three species using all publicly available EST and mRNA data. Pair wise comparison of transcripts for each gene, allows us to display alternative splice events within the transcriptome. Transcripts are annotated for alternative transcription start sites, alternative polyadenylation sites, splice events, SNPs and splice site conservation information between orthologous species. eVOC annotations for each transcript also allows us to display potential expression patterns. Translations are derived for each transcript where possible, and, if relevant, are tagged as potential candidates for nonsense-mediated RNA decay (4). More species will be added in future releases of ASTD.

ASTD consistently uses community acknowledged ontologies in the database: Gene Ontology (GO) (5), Sequence Ontology (SO) (6) and eVOC (7) terms have been implemented. GO aims to provide a structured controlled vocabulary for annotating genes, gene products and sequences, SO is developing an ontology to annotate the nucleic acid sequence and eVOC is a set of orthogonal controlled vocabularies that unify gene expression data by facilitating a link between the genome sequence and expression phenotype information. The eVOC, GO and SO data are increasingly being accepted as a standard for describing gene expression and annotations within genomic databases. Researchers developing a “community approved” splice event ontology have taken the descriptions of splice events used with ASTD into consideration. Once approved, this ontology will also be incorporated.

ASTD has been designed to include curated information available through consortium members, currently EURASNET NoE members (8) and formerly the ASD and ATD (9) consortium members. For a subset of genes, experimental confirmation of alternative poly(A) sites and transcript structure in normal and disease tissues is available. We hope to incorporate microarray data in the future, thereby adding evidence to exons that are known to exist within a certain transcript structure. The database can also be extended to link features from other Alternative Splicing data sources (Ensembl (10), VEGA (11), etc.). At present, Ensembl features are loaded to compare our alternative splicing predictions with Ensembl transcripts annotations.

The new database schema design has allowed the web interfaces to have the pleasure of a new face. They have been designed to:

i. Provide an intuitive and user-friendly interface for a wide variety of users to easily navigate the data;
ii. Speed up the retrieval and display of the information.
In particular, a graphical genome browser has been developed to navigate on the genome. An Internet search engine-like page enables users to access quickly the information they need. All the development is based on state of the art technologies (ORM, Text search engine, AJAX, Web Services).

The ASTD data is available in GFF3 / GTF 2.2 exchange format. The file will contain all features available in ASTD, including coding sequence (CDS) and orthology information. It is being extended depending on user requirements.

The ASTD project at EBI is supported by a grant from the EC: Eurasnet Network of Excellence (LSHG-CT-2005-518238). It was supported by the ASD grant from the EC (QLRT-CT-2001-02062) until November 2005 and by ATD consortium (LSHG-CT-2003-503329) until March 2007.

REFERENCES


[11] Loveland J.2005 VEGA, the genome browser with a difference. Brief Bioinformatics 6(2):189-93
Alternative splicing expands the human transcriptome and proteome by one order of magnitude: a benchmark analysis of ENCODE genes through the ASPIC software

Ernesto Picardi¹, Marina Mangiulli¹, Annamaria D’Erchia¹, Tiziana Castrignanò², Raffaella Rizzi³, Anna Anselmo¹, Paola Bonizzoni², Flavio Mignone¹ and Graziano Pesole¹*

¹ University of Bari, Dipartimento di Biochimica e Biologia Molecolare "E. Quagliariello", Bari 70126, Italy, ² Consorzio Interuniversitario per le Applicazioni di Supercalcolo per Universita’ e Ricerca, CASPUR, Rome, Italy, ³ DISCo, University of Milan Bicocca, Milan, 20135, Italy, ⁴ University of Milan, Dipartimento di Scienze Biomolecolari e Biotecnologie, Milan 20133, Italy

Alternative splicing is now emerging as a major mechanism contributing to the expansion of the transcriptome and proteome complexity of multicellular organisms. The fact that a single gene locus may give rise to multiple mRNAs and protein isoforms, showing both major and subtle structural variations, is an exceptionally versatile tool in the optimization of the coding capacity of the eukaryotic genome. The huge and continuously increasing number of genome and transcript sequences provides an essential information source for the computational detection of genes alternative splicing pattern. However, much of this information is not optimally or comprehensively used in gene annotation by current genome annotation pipelines.

We recently developed ASPIC, a novel algorithm for the investigation of alternative splicing of user submitted genes and a related web resource, based on comparative analysis of available transcript and genome data from a variety of species. The ASPIC web resource provides graphical and tabular views of the splicing patterns of all full-length mRNA isoforms compatible with the detected splice sites of genes under investigation as well as relevant structural and functional annotation. The ASPIC web resource - available at http://www.caspur.it/ASPIC/ - is dynamically interconnected with the Ensembl and Unigene databases and also implements an upload facility.

The ASPIC software has been applied to a set of 300 human genes included in the 44 regions included in the ENCODE project and accounting for 1% of the entire human genome. The results we obtained confirm the effectiveness and reliability of the ASPIC tool and show that about 90% of human genes undergo alternative splicing and that on average more than ten different transcript are generated by each gene. This implies that the human transcriptome and proteome complexity is expanded of at least one order of magnitude with respect to its gene complement.

REFERENCES

Comprehensive analysis of alternative splicing in chicken using a splicing graph approach

Durgaprasad Bollina¹, Shoba Ranganathan¹².

¹Department of Chemistry and Biomolecular Sciences & Biotechnology Research Institute, Macquarie University, Sydney, NSW 2109, Australia
²Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 119260

Alternative transcript diversity manifests itself a prime cause of complexity in higher eukaryotes. Alternative splicing is a widespread phenomenon. Recently, transcript diversity studies have suggested that 60-80% of human genes [1] are alternatively spliced. Although several alternative splicing databases [2] exist, they provide neither visualization nor standardized classification of alternative splicing events. We have used a splicing graph approach [3-4] to represent all transcripts from a single gene pattern approach, providing compact visualization as well as automated rule-based event classification. Here we report an analysis of 9238 orthologus genes in human and mouse and 556 orthologus genes in human, mouse and chicken [5] and we have described nine alternative splicing patterns derived from four novel splicing graph patterns and applied to human, mouse and chicken alternatively spliced genes. While alternative splicing has the potential to create many RNA isoforms from a single locus, the majority of loci generate only two or three isoforms and transcript support indicates that these isoforms are generally not rare events. Most alternative splicing variations are localized to the protein coding sequence and are predicted to significantly alter the coding sequence [5].

REFERENCES

Three aspects of the identification and behavior of exonic splicing motifs will be discussed. The first will be a study of the recent evolution of exonic splicing regulatory motifs (ESRs) in primates. The results show a purifying selection against the disruption of hexameric exonic splicing enhancer (ESE) motifs and against the creation of exonic splicing silencer (ESS) motifs. As well, positive selection is acting in favor of ESE creation and ESS disruption. These trends are detectable even between chimpanzees and humans, suggesting a dynamic state in the maintenance of this splicing information. Second, the apparent cornucopia of ESRs will be discussed, combining the lists from different laboratories. The results imply that any RNA sequence will be densely populated with such motifs. Third, the construction of synthetic exons will be described. These exons have been created using only three prototype motifs, one ESE, one ESS and one neutral sequence. These 8-mers have been placed between a pair of natural 3’ and 5’ splice sites to form the central exon of a 3-exon minigene. The objective here is to simplify the splicing regulatory information so as to better understand how the splicing code works. Transient transfection analysis showed ESEs are required for efficient splicing and ESSs prevent splicing. However, the combination of elements in various permutations still presents a complex picture.
**Murky beginnings and alternate endings: transcript diversity on the 5' and 3' ends**

**Uwe Ohler**, Elizabeth Rach, William Majoros

Institute for Genome Sciences and Policy, Duke University, Durham NC 27708, US

1. INTRODUCTION

Alternative splicing can generate a large number of transcript isoforms form a single locus, by means of post-transcriptional regulation on the level of the pre-messenger RNA. Additional transcript variation can be introduced by the usage of alternative transcription start sites (1), leading to longer or alternative first exons, and alternative cleavage/poly-adenylation, leading to longer or alternative terminal exons.

2. ALTERNATIVE PROMOTERS/TRANSCRIPTION INITIATION

Similar to the arrival of large EST libraries which led to an explosion in the number of observed splice isoforms, recent high-throughput 5' end sequencing has seen a large increase in the number of genes with alternative start sites (2). As in splicing, the question is how much of this variation is due to (a) experimental noise; (b) biological noise; and (c) condition-specific functional variability. We have built a pipeline to infer alternative transcription start sites in *D. melanogaster*, and will present results of our ongoing analysis. Among the questions we currently address are: Are alternative TSS utilized under different conditions? What are the sequence-specific determinants and characteristics of alternative TSS? What are the consequences on the isoform, i.e. do alternative TSS lead to changes in the protein and/or the 5' UTR sequence?

3. ALTERNATIVE POLY-ADENYLATION

Gene expression is regulated heavily on the post-transcriptional level, and a number of mechanisms have been discovered which control the fate of a spliced message once it has reached the cytoplasm. In addition to RNA-binding proteins, the regulation by non-coding RNAs, with microRNAs as the paramount example, has gained a lot of attention. An open and interesting question is the interplay of these different post-transcriptional mechanisms.

Animal miRNAs preferentially target complementary sites located in the 3'UTR of specific mRNAs. Starting from the polyA_DB database of poly-adenylation sites (3), we have constructed a set of reliable alternative 3'UTR regions in the human genome (4). This allowed us to study the distribution of miRNA target sites within (alternative) UTRs, with the result that more than 40% of target sites lie in alternative UTR segments which will not always be part of the mature transcript. This means that the miRNA-target gene relationship is dynamic in nature, and that it depends on the condition-specific presence of a particular isoform of its target gene. In addition, miRNA target sites are clustered around poly-adenylation sites in long UTRs, possibly due to their increased potential to form secondary structure and prevent the miRNAs from finding their complementary target sites.

To gain a full understanding of the function of a particular gene, one must know the complete compendium of its transcript diversity. Taken together, our studies provide an overview of variation at
the beginning and end of transcripts, and our analysis led to initial results regarding the functional consequences of such variation. Variation at the beginning and end has also been proposed to be functionally correlated with alternative splicing events: One model for alternative splicing sees functional splicing factors loaded at the stage of transcription, and alternative promoters would provide a convenient condition-specific mechanism for this to happen. And for the correct processing of transcripts utilizing a more downstream terminal exon, splicing to the upstream alternatives must be suppressed. Our data sets provide excellent starting points to further study the correlation of this kind of variation with alternative splicing.

REFERENCES


Computational evidence for the association between transcription initiation and internal splicing

Mihaela Zavolan, Erik van Nimwegen, Tzu-Ming Chern

Biozentrum, University of Basel, Klingelbergstrasse 50-70, Basel 4056, Switzerland

1. INTRODUCTION

The most common form of splice variation is the inclusion of an exon in some but not all of the transcripts of a gene (1,2). Numerous studies have been dedicated to specific instances of such exons, which are known by various names, such as "cassette", "alternative", "skipped" and "cryptic" exons. The regulatory signals that generally lead to the inclusion or exclusion of an exon have also been widely studied. Computational studies are converging towards the view that cassette exons are generally less recognizable to the splicing machinery than constitutive exons due to their shorter length (3), lower strength of splice sites (4) and poorer representation of general splice enhancers (1,5). The tissue-specific inclusion of these exons appears to be dependent upon specific regulatory elements, at least some of which are located in the strongly conserved intronic regions that flank the cassette exons (6, 2, 7, 8).

One attractive hypothesis concerning the mechanism of tissue-specific inclusion of cassette exons involves the coupling between (tissue-dependent) transcription and splicing: it has been shown that the promoter from which transcription is initiated can affect the inclusion of downstream exons through the recruitment of transcription factors and co-activators that modulate the elongation rate of RNA polymerase II (9), or through interactions with splicing factors (10). Here we focus on the connection between transcription initiation and internal splicing. We analyzed the FANTOM3 dataset of full-length mouse cDNAs (11) and the H-Invitational dataset of full-length human cDNAs (12), using the following Bayesian model. We first identify all cassette exons in our datasets, and compute the number of times each of these exons was included and skipped, respectively, when each particular transcription start site (TSS) was used. Then, we calculated for each exon the likelihoods of the counts of inclusion/TSS combinations under two models. The first model assumes that the probability of exon inclusion is independent of the TSS used to transcribe the pre-mRNA. The second model assumes that there is a TSS-associated probability of exon inclusion, and that these probabilities may differ between the different TSSs. Using these likelihoods we calculated the posterior probability \( P(f|\text{Data}) \) that the inclusion of a fraction \( f \) of the exons is dependent on the TSSs used to initiate the transcripts. To test the statistical significance of this result we further created a randomized dataset by permuting, for each cassette exon, the inclusions and exclusions among the TSSs in such a way that the total number of times each TSS was used, and the total number of times the exon was included remained unchanged.

On the randomized dataset, our Bayesian procedure infers, as expected, that the inclusion of less that 5% of exons is TSS-dependent. In contrast, on the real dataset, we infer that the inclusion of 20% of cassette exons depends on TSS. The exon with the highest posterior probability of TSS-dependence in the mouse dataset is shown in the figure, indicated by the black frame. The exon is predominantly skipped when the upstream TSSs (numbered 1 and 2) are used, and it is predominantly included when the downstream TSS (numbered 3) is used. The number of times each particular splicing pattern was present among the full-length cDNA sequences is indicated on the right side of each splicing pattern. Exons are shown as boxes, and introns as lines.
Further analyzing the properties of TSS-associated exons, we find that close to 90% of the top 10% exons in the order of their posterior probability of TSS-association are conserved at least to some extent in the homologous species (human for mouse and mouse for human). This, as well as the bias for a length that is a multiple of 3, indicates that these exons are functionally selected. Finally, we found two exons in our dataset that have a relatively high probability of TSS-association in both human and mouse.

Our analysis thus suggests that events that take place during transcription initiation have an important modulatory role on the splicing of internal exons, and provides a large dataset of exons on which the mechanism underlying this association can be studied.

2. FIGURES

**Figure 1.** The exon with highest probability of TSS-association in the mouse dataset.

![Exon Multiplicity Diagram](image)

**REFERENCES**

Bootstrapping the alternative splicing annotation of newly sequenced genomes

Cheng Soon Ong\textsuperscript{1,2}, Lisa Hartmann\textsuperscript{1}, Regina Bohnert\textsuperscript{1}, Nina Krüger\textsuperscript{1}, Anja Bohlen\textsuperscript{1}, Gunnar Rätsch\textsuperscript{1}

\textsuperscript{1}Friedrich Miescher Laboratory, Max Planck Society, Spemannstrasse 39, 72076 Tübingen, Germany
\textsuperscript{2}Max Planck Institute for Biological Cybernetics, Spemannstrasse 38, 72076 Tübingen, Germany

1. INTRODUCTION

While increasingly more genomes are becoming available every year, many genomes have rather poor EST coverage. The small number of ESTs not only means that the identification of genes has to rely mostly on computational gene finders, but also that almost no alternative splicing events can be identified as most introns are only covered by a single EST sequence. Hence, the initial alternative splicing annotation is a particularly challenging problem that we will consider in this work. We discuss several methods of identifying candidate regions that are highly likely to contain alternative splicing events for further experimental analysis. We compare two approaches for discovery of exon skipping events in \textit{Caenorhabditis remanei} using known events in \textit{C. elegans}, a closely related and well studied organism: (a) by finding close homologs in \textit{C. remanei} of alternatively spliced genes in \textit{C. elegans} and (b) by learning about discriminative characteristics of alternatively spliced genes in \textit{C. elegans} using Support Vector Machines (SVMs) \cite{1}, and using the trained SVM to predict on \textit{C. remanei}. We evaluate the success of the methods by experimentally analyzing considered regions (by RT-PCR and sequencing). After having obtained a reasonably sized set of known alternative splicing events we propose a method called \textit{active learning} as similarly used in drug discovery \cite{2}. The idea is to iteratively predict alternative splicing and to perform a few biological validation experiments. In a simulation study on \textit{C. elegans} we show that using this method we can significantly reduce the number of experiments needed to identify a reasonably large number of confirmed alternative splicing events.

2. PREDICTION OF ALTERNATIVELY SPliced EXONS IN \textit{C. REMANEI}

We started by aligning ESTs from dbEST and available cDNAs against the genomes using \textit{blat} with proper post-processing and quality control \cite{3}. From this transcript data we obtained lists of introns and exons, which were first clustered and then combined into splicegraphs. We detected exon skipping events from these splicegraphs and used them as positive examples for training and testing our algorithms. Negative examples were generated from sites with no evidence of alternative splicing additionally requiring that every splice site is confirmed at least five times. This leads to 413 and 8 positive as well as 14081 and 258 negative examples for \textit{C. elegans} and \textit{C. remanei}, respectively. Regions covered by only one EST or cDNA sequence, which are therefore good candidates for yet unknown alternative splicing events, were used as an unlabeled set that we would like to annotate using our method (5341 cases in \textit{C. remanei}).

\textit{Discovering new events using conservation:} We looked for homology on the mRNA and DNA level. For exon skipping events, which were at least three times EST confirmed in \textit{C. elegans} (101 cases), we used \textit{blat} to compare the mRNA and DNA sequences of skipped exons including their flanking exons with the genome of \textit{C. remanei}. If we find more than 50% identity and an EST covering the region in \textit{C. remanei}, we call the exon conserved and predict that the exon is alternatively spliced in \textit{C. remanei}.

\textit{SVM based predictions:} We use SVMs to learn to distinguish alternatively from constitutively spliced exons as similarly described in \cite{1}. Our method learns to recognize sequence elements at loosely defined positions relative to the splice sites in the training sequences using so-called string
kernels. We train this method on *C. elegans* and apply it to yet uncharacterized exons in *C. remanei*. We predict an exon in *C. remanei* to be alternatively spliced if the prediction score is above a threshold corresponding to a 0.25% false positive rate in *C. elegans* leading to 32 cases. Four out of the eight known exon skips in *C. remanei* were predicted to be positive, two of the remaining cases were above the 5% threshold and two cases could not be identified as alternatively spliced exons. This indicates that the method performs quite well when applied between *C. elegans* and *C. remanei*, and similar sequence elements are important for alternative splicing.

<table>
<thead>
<tr>
<th>Prediction method</th>
<th>Candidates</th>
<th>Confirmed</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>mRNA homology</td>
<td>9</td>
<td>4</td>
<td>44%</td>
</tr>
<tr>
<td>DNA homology</td>
<td>4</td>
<td>1</td>
<td>25%</td>
</tr>
<tr>
<td>SVMs</td>
<td>25</td>
<td>8</td>
<td>32%</td>
</tr>
</tbody>
</table>

**Tab. 1:** Number of exon skipping events confirmed by RT-PCR and sequencing.

3. WETLAB CONFIRMATION

For every predicted exon skipping event, primer pairs were designed using *primer3* with default parameters, at least 60nt away from the splice site of the flanking exons. We had to exclude a few cases where the primer design was not successful. Total RNA was extracted from wild type *C. remanei*, and converted to cDNA by RT-PCR. The cDNA was used with the designed primers in touchdown PCR, and all bands on the agarose gel were cut out and sequenced after gel extraction. In total we analyzed 31 cases and found 10 positive events, (cf. Tab. 1; some events are counted more than once). The success rate for all methods were rather similar, but the number of events predicted by SVMs were considerably larger. Hence SVM predictions have a larger potential to identify many alternative splicing events.

4. ACTIVE LEARNING

After this initial analysis we have a set of 18 confirmed alternatively spliced exons in *C. remanei*—eight previously known
ones and 10 newly confirmed exon skips. This set can serve as a starting point for applying a technique called active learning as similarly used in drug discovery (see e.g. [2]). The idea is to use the confirmed positives and a set of very likely negatives to train the SVM classifier on C. remanei exons. Applying this classifier we can rank all remaining unlabeled exons and choose a reasonable number of the most promising candidates for experimental analysis. The resulting newly labeled examples are used in the next iteration for retraining the SVM classifier together with the previous examples. This way the classifier continually improves due to the growing training set. An alternative would be to use the same classifier to label all unlabeled exons and decide only once which exons to test. This approach is less accurate as the classifier does not improve along the way. We illustrate the advantage of our approach in a simulation on C. elegans (cf. Fig. 2). The red curve represents results always obtained from the same classifier: at each iteration, 20 of the most likely remaining candidates are “analyzed”, resulting in 155 confirmed exon skips after analyzing 300 exons. The blue curve represents results using active learning; at each iteration, the results of the experiments of the previous iteration are included into the training set. This method finds 224 confirmed exon skips by testing the same number of exons, which is an improvement of 45%.

5. DISCUSSION

It has been claimed that only a very small fraction of alternative splicing events may be conserved among closely related species (only 11% estimated in [4]). Other studies have shown that alternative splicing is highly conserved (83%, see [5]), in a few selected cases even quantitatively [6]. Our preliminary data indicate that the conservation of alternative splicing between C. remanei and C. elegans is relatively low (25%-44%, assuming the experiments would always lead to the correct label). Additionally, we illustrated that SVM based predictions are almost as accurate as homology based predictions—with the additional benefit that we also find non-conserved alternative splicing events. Furthermore, in a simulation study for C. elegans we have shown that SVMs excel when used in an active learning setting, which would not be possible for homology based predictions. Finally, we demonstrated that with relatively little experimental effort we can significantly improve the alternative splicing annotation in C. remanei.

REFERENCES

1. INTRODUCTION

Alternative splicing is a widespread phenomenon in mammals through which multiple transcripts can be formed from the same gene. Disruption of alternative splicing affects the balance in transcript levels and has often been associated to cancer. By comparing the expression levels of alternative isoforms between normal and cancer tissues, several studies have implied that specific isoforms expressed at abnormal levels actively contribute to the transformed cancer state. However, no study has addressed the relationship between cancer and the alternative splicing activity in general. To this end we considered the distribution of all isoforms in a given tissue instead of focusing on specific isoforms that are differentially expressed in cancer.

2. RESULTS AND DISCUSSION

We studied the level of disorder in alternative transcription by using a measure of uncertainty called Shannon’s entropy, applied to all isoforms of a given gene in a given tissue (Figure 1). We measured the difference in entropy of alternative transcription between normal and cancer tissues from the same anatomical site. Figure 2 shows the results of this study for alternative initiation, alternative polyadenylation and alternative splicing.

*Figure 1:* Example of Shannon’s entropy calculation for a gene with four splicing isoforms SP1..SP4. EST counts are provided for each isoform in a normal and cancer tissue. In this example, isoform entropy is higher in the cancer tissue (1.38 vs 1.16 bits).

Whereas alternative initiation and polyadenylation show no significant gain or loss of entropy between normal and cancer tissues, alternative splicing shows a significant increase in entropy for 13 of the 29 tissues studied. This gain of entropy results from a disordered isoform expression in many different genes, suggesting that splicing perturbation is caused by factors that regulate multiple genes at the same time. It is worth noting that the three non-cancer diseases studied...
(ascites, schizophrenia, arthritis) showed no significant gain or loss of entropy, indicating that the gain of entropy is specific to cancer.

The genes that present the highest gain in entropy are enriched in GO terms that designate splicing factors. This suggests that the observed gain of entropy is linked to an impairment of the splicing machinery. Genes that code for splicing factors do generally present a level of alternative splicing that is higher than average, which could explain the enrichment in GO terms linked to alternative splicing. However, only the group of genes that show a high level of entropy are enriched in splicing factors, indicating that there is a genuine bias towards splicing alteration of splice factors.

Figure 2. Average isoform entropy per gene in cancer tissues/normal tissues for: a) alternative initiation sites, b) alternative polyadenylation sites, c) alternative splice forms. The ratio of entropy between cancer and normal tissues is represented by the horizontal bars and statistically significant loss and gain of entropy boundaries are represented by vertical dots and dashes respectively.

Finally, we observed that the gain in entropy of alternative splice forms between normal and cancer tissues is correlated to high levels of cellular proliferation. Proliferation signatures for each cancer type were obtained by analysing the expression levels of proliferation-associated genes across cancer-related microarray experiments. High proliferation signatures were found to be correlated with a high gain of entropy (not shown).

The observed gain of entropy between normal and cancer tissues most likely reflects a widespread deregulation of the splicing machinery that has a major impact on transcription levels of many genes. This disruption may develop in part from splicing alteration in splice factors and is correlated to proliferation levels. This generalized disruption, which prevails in about half of the cancers we studied, could mark a level of deregulation such that therapy based on the specific targeting of specific isoforms would be ineffective. Studies seeking alternative splicing isoforms as cancer signatures should also be aware of this context of general splicing alteration

REFERENCES

Alternative messenger RNA (mRNA) splicing, the generation of a diverse range of mature RNAs, has considerable potential to expand the cellular protein repertoire (1,2). Splicing events that occur within protein coding regions have the potential to alter the biological function of the expressed protein and even to create new protein functions. Recent studies have estimated that 40–80% of multi-exon human genes can produce differently spliced mRNAs (3,4). The importance of alternative splicing in processes such as development (5) has long been recognized and proteins coded by alternatively spliced transcripts have been implicated in a number of cellular processes. The extent of alternative splicing in eukaryotic genomes has lead to suggestions that alternative splicing is key to understanding how human complexity can be encoded from so few genes. The pilot project of the Encyclopaedia of DNA Elements (ENCODE) project (6), which aims to identify all the functional elements in the human genome, has undertaken a comprehensive analysis of 44 selected regions that make up 1% of the human genome. One valuable element of the project has been the detailing of a reference set of manually annotated splice variants by the GENCODE consortium (7). The annotation by the GENCODE consortium is an extension of the manually curated annotation by the Havana team at the Sanger Institute. We have carried out a detailed study of the alternatively spliced gene products annotated in the ENCODE pilot project (8). We found that alternative splicing in human genes is likely to be even more frequent than has commonly been suggested and we demonstrate that many of the potential alternative gene products will have markedly different structure and function from their constitutively spliced counterparts. For the majority of alternative isoforms we found little evidence to suggest they have a role as functional proteins and it seems unlikely that alternative splicing will meaningfully extend the cellular repertoire. As part of the work we carried out on the proteins in the ENCODE set we were able to demonstrate that a large proportion of the human proteins recorded as the constitutive variants in the SwissProt database (9) were likely to be alternative variants. In many cases we were able to predict the probable constitutive variant.

REFERENCES

<table>
<thead>
<tr>
<th>Name</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richard Dixon</td>
<td>Complementary intron sequence motifs associated with human exon repetition</td>
</tr>
<tr>
<td>Ivana Vukusic</td>
<td>Intron Retention: an alternative path to exonization?</td>
</tr>
<tr>
<td>Axel Rasche</td>
<td>Comparison of exon array analysis methods for detecting splice variants</td>
</tr>
<tr>
<td>Alphonse Thanaraj</td>
<td>SpliVaP: A Data Resource on Splice-Mediated Changes in Isoform Protein Sequences</td>
</tr>
<tr>
<td>Jose Muino</td>
<td>Coupling alternative splicing with non-mediated mRNA decay to regulate flowering in Arabidopsis</td>
</tr>
<tr>
<td>Vanessa Kane</td>
<td>Alternative splicing regulation for the tumor associated NADH oxidase tNOX</td>
</tr>
<tr>
<td>Meelis Kull</td>
<td>DiDASE - Discovery of Differential Alternative Splicing Events</td>
</tr>
<tr>
<td>Ana Rita Grosso</td>
<td>Tissue-specific splicing factor gene expression signatures</td>
</tr>
<tr>
<td>Michael Hiller</td>
<td>Pre-mRNA secondary structures influence exon recognition</td>
</tr>
<tr>
<td>Ralf Bortfeldt</td>
<td>Computational analysis of alternative human exons originating from tandem donors</td>
</tr>
<tr>
<td>Karol Szafranski</td>
<td>Violating the splicing rules: TG dinucleotides function as alternative 3’ splice sites in U2-dependent introns</td>
</tr>
<tr>
<td>Maria Kalyna</td>
<td>Non-coding alternative splice variants of plant-specific SR proteins are tightly regulated and highly conserved in evolution</td>
</tr>
<tr>
<td>Eduardo Eyras</td>
<td>Exonization and establishment in human genes</td>
</tr>
<tr>
<td>Ekaterina Ermakova</td>
<td>Patterns of selection and evolution of the exon-intron structure in alternatively spliced genes</td>
</tr>
<tr>
<td>Tyson Clark</td>
<td>Discovery of tissue-specific exons using comprehensive human exon microarrays</td>
</tr>
<tr>
<td>John Castle</td>
<td>Alternative splicing profiling and regulation in human body tissues</td>
</tr>
<tr>
<td>Name</td>
<td>Title</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Eleanor Whitfield</td>
<td><em>The Alternative Splicing and Transcript Diversity Database</em></td>
</tr>
<tr>
<td>Graziano Pesole</td>
<td>Alternative splicing expands the human transcriptome and proteome by one order of magnitude</td>
</tr>
<tr>
<td>Shoba Ranganathan</td>
<td>Alternative splicing expands the human transcriptome and proteome by one order of magnitude</td>
</tr>
<tr>
<td>Uwe Ohler</td>
<td><em>Murky beginnings and Alternate Ends: Transcript Diversity at 5'- and 3'-ends</em></td>
</tr>
<tr>
<td>Mihaela Zavolan</td>
<td>Computational evidence for the association between transcription initiation and internal splicing</td>
</tr>
<tr>
<td>Cheng Soon</td>
<td><em>Bootstrapping the alternative splicing annotation of newly sequenced genomes</em></td>
</tr>
<tr>
<td>William Ritchie</td>
<td><em>Transcript isoform entropy: proof of extensive disruption of splicing in cancer tissues</em></td>
</tr>
<tr>
<td>Michael Tress</td>
<td>Splicing isoforms in the ENCODE complement</td>
</tr>
<tr>
<td>Dorothy M. Morre</td>
<td>Alternative splicing as the basis for specific localization of tNOX, a tumor associated hydroquinone (NADH) oxidase, to the cancer cell surface</td>
</tr>
<tr>
<td>D. James Morre</td>
<td>Cancer-specific expression of a tumor-associated NADH oxidase (tNOX) exon4-minus splice variant mRNA in a model system</td>
</tr>
</tbody>
</table>
**Poster abstracts**

**Complementary intron sequence motifs associated with human exon repetition**

Dixon RJ\(^1\), Eperon IC\(^2\) and Samani NJ\(^1\)

(1) Department of Cardiovascular Sciences; (2) Department of Biochemistry, University of Leicester, Leicester, U.K., LE3 9QP

Exon repetition (non-linear mRNA) is a phenomenon in which tandemly repeated exons are present in mRNA in the absence of duplications in the genome. We have previously undertaken the first genome wide survey of candidate exon repetition events in expressed sequences from multiple species and found that this phenomenon is more common than previously thought\(^1\). The regulation of this process is not fully understood. We therefore investigated the entire flanking intronic sequences of human exons involved in exon repetition for common sequence elements. A computational analysis of 48 human single exon repetition events identified two common sequence motifs\(^2\). One of these motifs is pyrimidine-rich and is more common in the upstream intron, whilst the other motif is highly enriched in purines and is more common in the downstream intron. As the two motifs are complementary to each other, they support a model by which exon repetition occurs as a result of trans-splicing between separate pre-mRNA transcripts from the same gene that are brought together during transcription by complementary intronic sequences. The majority of the motif instances overlap with the locations of mobile elements such as Alu elements. We explore the potential importance of complementary intron sequences in an rat gene that undertakes natural exon repetition in a strain specific manner. The possibility that distant complementary sequences can stimulate inter-transcript splicing during transcription suggests an unsuspected new role for potential secondary structures in endogenous genes.

**REFERENCES**


Recently, we provided a classifier based on the principles of GP (Genetic Programming), which is well suited to detect retained introns (SIRs) (1). The classifier primarily takes advantage of the characteristic resemblance of retained introns to exons; making them easily distinguishable from “typical” introns due to several well known features. In contrast to the findings by Sakabe and de Souza (2) the intronic splicing enhancer GGG was not picked up by the system very frequently (Fig. 1a), rather, the most important feature considered was the number of ESS (exonic splicing silencers), defined by Zhang and Chasin (3).

We generated a dataset of pseudo SIRs by cutting out intron-like sequences from long constitutive exons and analyzed the densities of the certain ESE and ESS motifs, defined by different groups (3-5). We found out that SIRs are sometimes more similar to exons, but when considering another dataset of motifs, they become more similar to introns (Fig. 2). This picture changed as soon as we took the context of the flanking regions into consideration: By calculating the ratio of the motif
densities for each flanking exon and dividing it by the density found in the corresponding intron, we found that the ratios are in average positive for ESE and negative for ESS in SIR (Fig. not shown). In pseudo SIR this ratio tends to be around 0, suggesting that the overall density is not as important as the local sequence context. It would be interesting to find out whether the retained introns are in most cases introns that are changing their properties to become new exons. Our preliminary results point to this direction, as almost no SIR introns are exonized in mouse.

**Figure 2:** ESE and ESS densities calculated in a) pseudo SIR, b) SIR, c) introns and d) constitutive exons. For each of the four classes, we normalized the sequence length to 100bp by taking 50bp from the donor- and acceptor-splice site. We removed 10bp from the donor splice site (-3:+7) and 23bp from acceptor splice site (-20:+3). For each base pair, we calculated the percentage, covered by an ESS/ESE motif.

**REFERENCES**


Comparison of exon array analysis methods for detecting splice variants

Axel Rasche, Ralf Herwig

Max-Planck-Institute for Molecular Genetics. Vertebrate Genomics, Bioinformatics Group
Ihnestr. 63-73, 14195 Berlin

1. INTRODUCTION

Recently, Affymetrix released their newest chip technology based on exon arrays [1] that are supposed to yield a higher resolution picture of the transcriptome compared to standard 3’ gene expression arrays. The expression of a gene can now be dissected in the expression of its different transcripts. On the one hand, the known transcripts can be analysed on the other hand new transcripts may be detected. An important prerequisite for data analysis is the question which splicing events can be detected. In the figure included in the document a list of possible splicing events is displayed. Theoretically, the microarrays are able to identify the most interesting splicing events, however, not all possible splicing can be detected. In this work we highlight the possible splice variants that are detectable and conduct a comparative study of existing methods on a defined set of experiments and validation resources.

**Figure 1.** Detectable splicing events: Above we see a systematic of possible splicing events. Beside the constitutive splicing we see on the right side splicing events detectable with Affymetrix exon arrays.
2. DATA ANALYSIS

Goal of our work is to assess performance of different methods for the identification of alternative splicing from exon microarrays. Simple approaches like the computation of Splicing Indices or PAC calculate a numerical value from the expression means. Alternatively, ANOVA has also been proposed in diverse variants as for example ANOSVA [2-4]. The major drawback in developing and comparing new methods is the lack of controlled data sets like Spike-In data set or similar data sets as available for the standard arrays [2]. Affymetrix publishes several datasets on its homepage [1]. With the accompanying publications [3, 4] at least some validated genes are at hand.

In a first phase a pipeline was implemented in the R/BioC framework covering the quality control, preprocessing and analysis on exon/gene level [5, 6]. The different results of the alternative splicing evaluation methods are compared on the same data sets and with the same preprocessing. Where available, true positives were used for assessment, e.g. with ROC curve analysis. Validation using external data is a crucial element of the work. We used, for example, EST data for validation.

REFERENCES

SpliVaP: a data resource on splice-mediated changes in isoform protein sequences

Massimiliano Orsini, Matteo Floris, Anna Tramontano, Alphonse Thanaraj Thangavel

CRS4-Bioinformatica, Parco Scientifico e Technologico, POLARIS, Edificio 3, 09010 PULA (CA), Sardinia, Italy.

It is often the case that mammalian genes are alternatively spliced; most of the alternate transcripts resulting from a gene are translated in the cell as isoform protein sequences. We have developed a computational pipeline that characterizes splice-mediated protein isoform sequences for functional differences. Individual isoform sequences are annotated for presence of signatures that characterize functional properties of the proteins. Such signatures include domains, fingerprints, and prosite motifs. The pipeline then delineates differences in such signatures among the isoform protein sequences and reports splice-mediated protein functional changes. It is seen that between two isoform protein sequences, signatures are often included/excluded, alternated, and truncated in length. Examination of the structural data (where available) reveal that the affected regions are often involved in dimerization, ligand binding, and protein-protein interactions. Effects of changes on « surface patterns » (that are made of distant residues and that are exposed on the protein surface) are examined. Major classes of proteins that are affected due to pre-mRNA splicing are receptors, hydrolases, transferases, other transport proteins, RNA/DNA associated proteins, structural proteins, cytokines and growth factors & inhibitors. We also see that a number of genes, which are alternatively spliced to code for protein isoforms with functional differences are associated with diseases. The generated data resource (SpliVaP) on isoform protein sequences with value-added annotation is presented through user-friendly query interfaces as web resources from http://bioinfo15.crs4.org/splicing/.
Coupling alternative splicing with non-mediated mRNA decay to regulate flowering in Arabidopsis

Jose M. Muino, Pawe Krajewski
Institute of Plant Genetics-PAN, Strzeszynska 34, 60-479 Poznan, Poland

Alternative splicing is an essential biological process that generates multiple different transcripts from the same precursor mRNA. Sometimes, the splice event will cause a reading frameshift which will generate a premature stop codon (PCT) on the transcript, in humans it is estimated that roughly a third of the inferred splicing events results in a mRNA isoform harboring a PCT. These isoforms containing a PCT will be candidates for degradation due the mechanism so-called non-mediated mRNA decay. In such a way, the cell can regulate the abundance of some transcripts at the translational level coupling alternative splicing with non-mediated mRNA decay (1).

It is estimated that at least 35% of human genes undergo alternative splicing during development, cellular differentiation and other cellular processes and this is a tissue-dependent process. Recently, it was shown that one of the factors of non-mediated mRNA decay (upf1) has an important role in the development of Drosophila (2), and also in growth and size control in Arabidopsis (3). Therefore, it is of some interest to study how this kind of regulation can have some effect to the regulation of flowering in Arabidopsis.

Because there is not exon array chips for Arabidopsis to study this mechanism, and also because it is desirable to use the big amount of previous expression array experiments, we present a method to study this kind of regulation using standard Affymetrix DNA chips (eg: ATH1). In short, each probe is located in the genome and grouped depending which exon is targeting and/or if they are targeting a region of the gene before or after the PCT at study. In such a way, for a significant number of genes it is possible to detect exon-skipping events, to check if a predicted PCT have a significant effect in the expression pattern of the transcript and even, in some very particular cases, it can be detected intron retention events. This last type of alternative splicing seems to be the most prevalent type in Arabidopsis (about 56%), in contrast with humans where exon skipping is the most common phenomenon (4). The method will be exemplified studying some experiments about flowering.

REFERENCES


Pre-mRNAs to be alternatively spliced contain conserved sequences in the exons to direct the spliceosome to either include or exclude said exon in the final product. These sequences have been classified as Exonic Splicing Enhancers (ESEs) which include the exon, or Exonic Splicing Silencers (ESSs) which exclude the exon (1). Serine/Arginine rich proteins (SR proteins) and heteronuclear ribonucleoproteins (hnRNPs) bind the ESE and ESS sequences, respectively (2). Our laboratory has demonstrated that exon 4 is not present in the final mRNA product of Tumor-Associated NADH Oxidase (tNOX) in cancer cells (3). In this study, the gene sequence for tNOX, located on the X chromosome, was analyzed via Genscan (4) and compared with previously identified ESE and ESS sequences (2, 5, 6).

The ESE sequences are evenly distributed throughout exon 4 while the ESS sequences are mostly present in the first two thirds of the exon. Only one ESS sequence is located at the 3' end of exon 4 among the sequences differentiating the genome sequence as an intron (ACE-in) (7). The sequences indicating the genome sequence as an exon (ACE-ex) are exclusively present in the first third of the exon. Should an hnRNP bind the 3' end ESS, the surrounding visible sequences will be the intron sequences (ACE-in). This may cause the spliceosome to interpret exon 4 as an intron, leaving it out of the final mRNA product.

To isolate and characterize the splicing factor binding the above-mentioned ESS sequences, we are designing an RNA affinity column anchoring exon 4 RNA to a glutathione bead support (Figure 2) (8, 9). HeLa nuclear extract (10) will be incubated with the column and the bound factors isolated. A minigene (Figure 1) consisting of two halves of an enhanced green fluorescent protein (EGFP) (11) with exon 4 between the EGFP halves will be used in a HeLa S100 splicing deficient extract supplemented with the factors isolated from the column (10). To do this, in vitro transcription, translation, and RNA purification kits (Ambion) will be utilized. If the splicing factor responsible for silencing exon 4 from the final RNA product is isolated, a fully functional EGFP will be formed, and fluorescence quantifiable.
To determine if this splicing factor is cancer specific, a transfection experiment using the EGFP-exon 4 minigene construct will be performed. MCF10A and BT20 cells will be transfected using Lipofectamine 2000 (Invitrogen). The EGFP-exon 4 minigene, EGFP minigene backbone without exon 4, and a vehicle control will be used. 24 hours post-transfection, the fluorescence will be quantified as a ratio to vehicle control. Once the splicing factor has been isolated and determined to be cancer specific or non-cancer specific, the isolated factors purified from the column will be characterized via N-terminal sequencing.

REFERENCES

DiDASE - Discovery of Differential Alternative Splicing Events

Meelis Kull, Jaak Vilo

Estonian Biocentre, Riia 23b, Tartu 51010, Estonia
University of Tartu, Institute of Computer Science, Liivi 2-306, Tartu 50409, Estonia

Alternative splicing has been shown to act differentially in distinct anatomical sites, developmental stages, and pathological states. Such condition-specificity of splice variants has been studied using expressed sequence tag (EST) libraries and splicing-specific microarrays. The microarrays give more throughput, but require a lot of resources. EST libraries form a huge existing source of information and can be studied by bioinformatic means. We have worked with EST sequence data and produced a software tool DiDASE - Discovery of Differential Alternative Splicing Events. The novelty of the tool lies in two aspects. Firstly, DiDASE exhaustively studies all combinations of anatomical site, developmental stage, and pathological state, while previously these have been studied only separately in a genome-wide manner. Secondly, it uses a new yet simple method for assessing the evidence that splicing indeed is differential.

Input of DiDASE can be any set of spliced EST sequences, together with their alignment on the genome and annotations telling which anatomical site, developmental stage, and pathological state the sample was taken from. DiDASE splits all genes into regions at splice sites that are used in more than one library. Each region is studied separately to discover conditions where this region is almost never exonic (or intronic) while it is often exonic (or intronic) in another condition. Each discovery is given a reliability score by which the results are ranked.

DiDASE is capable of exhaustively covering all combinations of anatomy, development, and pathology, such as "normal fetal brain", and comparing these with the three opposite conditions, which in this case are "non-normal fetal brain", "normal non-fetal brain" and "normal fetal non-brain". For instance, the second comparison would find differences of splicing in fetal and non-fetal normal brain. Many of such combinations do not have any representative cDNA libraries and are thus not of interest. However, quite a few interesting cases can be discovered this way. Usage of a wildcard allows for comparisons such as "normal any-development brain" vs "normal any-development non-brain" which is the usual tissue-specific splicing. Similarly, disease-specific and development-specific splicing can be discovered.

The visualisation of results uses bars with varying width to indicate different percentage of exons and introns for a region. This visualisation technique is going to be used while integrating DiDASE with ASTD database run by EBI (1).

Discoveries of DiDASE can easily be experimentally tested and give a good basis for studying the regulation of alternative splicing. We are currently validating experimentally the results on human EST and cDNA data from ASTD and annotations from eVOC (2).

REFERENCES

**Tissue-specific splicing factor gene expression signatures**

Ana Rita Grosso\(^1,2\), Anita Gomes\(^1\), Nuno Morais\(^2\), Sandra Caldeira\(^1\), Natalie Thorne\(^2\), Maria Carmo-Fonseca\(^1\)

\(^1\)Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, Portugal
\(^2\)University of Cambridge, Department of Oncology, Hutchison-MRC Research Centre, Cambridge CB2 2XZ, UK

The alternative splicing code that controls and coordinates the transcriptome in metazoan organisms remains poorly understood. It has long been argued that regulation of alternative splicing relies on combinatorial interactions between multiple proteins, and that tissue-specific splicing decisions most likely result from differences in the concentration and/or activity of these proteins. However, large-scale data to systematically address this issue have just recently started to become available.

Here we show that splicing factor gene expression signatures can be identified that reflect cell type and tissue specific patterns of alternative splicing. We developed a computational analysis tool that was applied to microarray-based gene expression profiles of splicing factors from mouse, chimpanzee and human tissues. Our results reveal that brain and testis, the two tissues with highest levels of exon skipping, have the highest number of splicing factor signatures. From a total of over 100 tissue-specific signatures we identified known factors, as well as novel candidate splicing regulators that include SR protein kinases, DEAD-box RNA helicases, and snRNP proteins. These results indicate the power of generating signature-based predictions as an initial computational approach into a global view of tissue-specific alternative splicing regulation.

This work was supported by FCT in Portugal (“Fundação para a Ciência e Tecnologia”: SFRH/BD/22825/2005 and POCTI/MGI/49430/2002) and EURASNET (European Alternative Splicing Network).
Alternative splicing as the basis for specific localization of tNOX, a tumor-associated hydroquinone (NADH) oxidase, to the cancer cell surface

Xiaoyu Tang\textsuperscript{1}, Zengsui Tian\textsuperscript{2}, Pin-Ju Chueh\textsuperscript{2}, Ssuhen Chen\textsuperscript{2}, D. James Morré\textsuperscript{2}*, Dorothy M. Morré\textsuperscript{1}

\textsuperscript{1}Department of Foods and Nutrition, Purdue University, 700 West State Street, West Lafayette, IN 47907, USA
\textsuperscript{2}Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, 201 South University Street, West Lafayette, IN 47907, USA

A tumor-associated hydroquinone and NADH oxidase with protein disulfide-thiol interchange activity (designated ENOX2 or tNOX) is exclusively associated with the outer leaflet of the plasma membrane at the surface of cancer cells and in sera of cancer patients. It is absent from the surface of non-cancer cells and from sera from healthy individuals. Full length tNOX mRNA has approximately the same abundance in both normal and tumor cells but appears not to be expressed in either. We will present evidence that the cancer specificity of tNOX expression at the cell surface is the result of alternative splicing. An exon 4 minus form is present in cancer cell lines and is absent in non-cancer cell lines. In contrast to full length tNOX cDNA, transfection of COS cells with tNOX exon 4 minus cDNA resulted in overexpression of mature 34 kDa tNOX protein at the plasma membrane. The exon 4 minus form results in initiation of translation at a downstream M231 initiation site distinct from full length mRNA. Site-directed mutagenesis of M231 prevented the translation of exon 4 minus cDNA and cell surface expression of 34 kDa mature tNOX. The unprocessed molecular weight of 47 kDa of the exon 4 minus cDNA translated from methionine 231 corresponded to that of the principal native tNOX form of the endoplasmic reticulum. Taken together, the molecular basis of cancer cell-specific expression of tNOX appears to reside in the cancer-specific expression of exon 4 minus splice variant mRNA.
Cancer-specific expression of a tumor-associated NADH oxidase (tNOX) exon4-minus splice variant mRNA in a model system

†C. P. Kuntz, †X. Tang, †V. C. Kane, †D. M. Morré, †D. J. Morré

†Department of Foods and Nutrition, Purdue University, 700 West State Street, West Lafayette, IN 47907, USA and †Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, 201 South University Street, West Lafayette, IN 47907, USA

A cancer-specific cell surface NADH oxidase with protein disulfide-thiol interchange activity designated as tNOX, is a member of a larger family of related cell surface NADH oxidases or ECTO-NOX proteins. Expression of tNOX at the cell surface results from an exon4-minus splice variant mRNA transcript. To investigate the mRNA-processing events leading to the splice variant, 3T3 (mouse) and HUVEC (human vascular endothelial) cells were inoculated with SV-40 (ATCC pm1-2) and characterized post-infection for tNOX presence at the cell surface by Western blot analysis and presence of drug- (capsaicin) inhibited NADH oxidase activity, a defining characteristic of tNOX protein. With SV-40-infected 3T3 cells, tNOX was absent on day 3 post-infection but strongly present on day 6 post-infection. With SV-40-infected HUVEC cells, the kinetics of tNOX appearance were similar with 32 kDa and 34 kDa bands as indicated by Western blot analysis being absent on day 3 post-infection and increasingly evident on days 5 and 7 post-infection. The SV-40-infected HUVEC cells exhibited proteins cross-reactive with anti-tNOX antibodies at 32 and 34 kDa plus a protein, also cross-reactive with anti-tNOX antibody, corresponding to a previously described tNOX-specific protein with a M₆ of 52 kDa (1). Control (uninfected) 3T3 and HUVEC cells lacked tNOX presence. The findings suggest that either model system, SV-40-infected 3T3 cells or SV-40-infected HUVEC cells, offer a relatively narrow window post-infection (between day 3 post-infection and days 5 to 7 post-infection) for exon4-minus mRNA initiation and translation and delivery of tNOX to the cell surface. The 3T3 system is presently under investigation to test the involvement of exonic splicing silencer sequence-binding proteins as initiators of the cancer-specific expression of the tNOX protein.

REFERENCE

Organizing committee

Shoba Ranganathan, Professor
Macquarie University
shoba@cbms.mq.edu.au
Department of Chemistry and Biomolecular Sciences
Macquarie University NSW 2109, Sydney, Australia
Telephone +612 9850 6262
FAX +612 9850 8313
http://biolinfo.org

Dirk Holste
Institute of Molecular Pathology
holste@imp.ac.at
Institute of Molecular Pathology, Dr-Bohr-Gasse 7, A-1030, Vienna, Austria
Telephone +43-1-79044 4508
FAX +43-1-79044 110
http://www.mit.edu/people/holste
http://www.vdrc.at/holste.htm

Eduardo Eyras
Pompeu Fabra University
eduardo.eyras@upf.edu
Biomedical Informatics Unit
Pompeu Fabra University, PRBB, Dr. Aiguader 88, E8003 Barcelona, Spain
Telephone +34 933160502
FAX +34 933160550
http://genomics.imim.es/regulation

Andrea Barta, Professor
University of Vienna
andrea.barta@meduniwien.ac.at
Max-F-Perutz Laboratories, Dr-Bohr-Gasse 9/3, A-1030, Vienna, Austria
Telephone +43-1-427761640
FAX +43-1-42779616
http://www.mfpl.ac.at
<table>
<thead>
<tr>
<th>Names</th>
<th>Emails</th>
</tr>
</thead>
<tbody>
<tr>
<td>Irena I. Artamonova</td>
<td><a href="mailto:irenart@gmail.com">irenart@gmail.com</a></td>
</tr>
<tr>
<td>Rolf Backofen</td>
<td><a href="mailto:backofen@informatik.uni-freiburg.de">backofen@informatik.uni-freiburg.de</a></td>
</tr>
<tr>
<td>Andrea Barta</td>
<td><a href="mailto:andrea.barta@meduniwien.ac.at">andrea.barta@meduniwien.ac.at</a></td>
</tr>
<tr>
<td>Benjamin J. Blencowe</td>
<td><a href="mailto:b.blencowe@utoronto.ca">b.blencowe@utoronto.ca</a></td>
</tr>
<tr>
<td>Ralf H. Bortfeldt</td>
<td><a href="mailto:bortfeldt@minet.uni-jena.de">bortfeldt@minet.uni-jena.de</a></td>
</tr>
<tr>
<td>John Calley</td>
<td><a href="mailto:calley_john_n@lilly.com">calley_john_n@lilly.com</a></td>
</tr>
<tr>
<td>Doina Caragea</td>
<td><a href="mailto:dcaragea@ksu.edu">dcaragea@ksu.edu</a></td>
</tr>
<tr>
<td>John Castle</td>
<td><a href="mailto:john_castle@merck.com">john_castle@merck.com</a></td>
</tr>
<tr>
<td>Lawrence Chasin</td>
<td><a href="mailto:lac2@columbia.edu">lac2@columbia.edu</a></td>
</tr>
<tr>
<td>Tyson Clark</td>
<td><a href="mailto:tyson_clark@affymetrix.com">tyson_clark@affymetrix.com</a></td>
</tr>
<tr>
<td>Melanie Courtot</td>
<td><a href="mailto:courtot@ebi.ac.uk">courtot@ebi.ac.uk</a></td>
</tr>
<tr>
<td>Anneleen Daemen</td>
<td><a href="mailto:Anneleen.Daemen@esat.kuleuven.be">Anneleen.Daemen@esat.kuleuven.be</a></td>
</tr>
<tr>
<td>Lawrence D'Antonio</td>
<td><a href="mailto:ldant@ramapo.edu">ldant@ramapo.edu</a></td>
</tr>
<tr>
<td>Adam Dawe</td>
<td><a href="mailto:adam@sanbi.ac.za">adam@sanbi.ac.za</a></td>
</tr>
<tr>
<td>André De Sousa Corvelo</td>
<td><a href="mailto:acorvelo@imim.es">acorvelo@imim.es</a></td>
</tr>
<tr>
<td>Sandro De Souza</td>
<td><a href="mailto:sandro@ludwig.org.br">sandro@ludwig.org.br</a></td>
</tr>
<tr>
<td>Richard Dixon</td>
<td><a href="mailto:rd67@le.ac.uk">rd67@le.ac.uk</a></td>
</tr>
<tr>
<td>Xianjun Dong</td>
<td><a href="mailto:xianjund@ii.uib.no">xianjund@ii.uib.no</a></td>
</tr>
<tr>
<td>Ekaterina Ermakova</td>
<td><a href="mailto:ermakova@iitp.ru">ermakova@iitp.ru</a></td>
</tr>
<tr>
<td>Eduardo Eyras</td>
<td><a href="mailto:eeyras@imim.es">eeyras@imim.es</a></td>
</tr>
<tr>
<td>Elisa Napolitano Ferreira</td>
<td><a href="mailto:eferreira@ludwig.org.br">eferreira@ludwig.org.br</a></td>
</tr>
<tr>
<td>Daniel Gautheret</td>
<td><a href="mailto:daniel.gautheret@u-psud.fr">daniel.gautheret@u-psud.fr</a></td>
</tr>
<tr>
<td>Mikhail Gelfand</td>
<td><a href="mailto:gelfand@iitp.ru">gelfand@iitp.ru</a></td>
</tr>
<tr>
<td>Ivo Grossesse</td>
<td><a href="mailto:grossesse@ipk-gatersleben.de">grossesse@ipk-gatersleben.de</a></td>
</tr>
<tr>
<td>Ana Rita Grosso</td>
<td><a href="mailto:agrosso@fm.ul.pt">agrosso@fm.ul.pt</a></td>
</tr>
<tr>
<td>Hedí Hegyi</td>
<td><a href="mailto:hegyi@enzim.hu">hegyi@enzim.hu</a></td>
</tr>
<tr>
<td>Ines Hellmann</td>
<td><a href="mailto:IHeermann@bi.ku.dk">IHeermann@bi.ku.dk</a></td>
</tr>
<tr>
<td>Michael Hiller</td>
<td><a href="mailto:hiller@informatik.uni-freiburg.de">hiller@informatik.uni-freiburg.de</a></td>
</tr>
<tr>
<td>Dirk Holste</td>
<td><a href="mailto:holste@alum.mit.edu">holste@alum.mit.edu</a></td>
</tr>
<tr>
<td>Agnes Hotz-Wagenblatt</td>
<td><a href="mailto:hotz-wagenblatt@dkfz.de">hotz-wagenblatt@dkfz.de</a></td>
</tr>
<tr>
<td>Maria Kalyna</td>
<td><a href="mailto:mariya.kalyna@univie.ac.at">mariya.kalyna@univie.ac.at</a></td>
</tr>
<tr>
<td>Jens Keilwagen</td>
<td><a href="mailto:Jens.Keilwagen@ipk-gatersleben.de">Jens.Keilwagen@ipk-gatersleben.de</a></td>
</tr>
<tr>
<td>Nam Shin Kim</td>
<td><a href="mailto:deepreds@gmail.com">deepreds@gmail.com</a></td>
</tr>
<tr>
<td>Gautier Koscieln</td>
<td><a href="mailto:koscieln@ebi.ac.uk">koscieln@ebi.ac.uk</a></td>
</tr>
<tr>
<td>Meelis Kull</td>
<td><a href="mailto:meelisk@ut.ee">meelisk@ut.ee</a></td>
</tr>
<tr>
<td>Min-Seok Kwon</td>
<td><a href="mailto:mskwon@yonsei.ac.kr">mskwon@yonsei.ac.kr</a></td>
</tr>
<tr>
<td>Wolfram Langer</td>
<td><a href="mailto:wolfram.langer@bayerhealthcare.com">wolfram.langer@bayerhealthcare.com</a></td>
</tr>
<tr>
<td>Jacek Majewski</td>
<td><a href="mailto:jacek.majewski@mcgill.ca">jacek.majewski@mcgill.ca</a></td>
</tr>
<tr>
<td>Marina Mangiulli</td>
<td><a href="mailto:marina.mangiulli@gmail.com">marina.mangiulli@gmail.com</a></td>
</tr>
<tr>
<td>Ines Mollet</td>
<td><a href="mailto:inesmollet@yahoo.com">inesmollet@yahoo.com</a></td>
</tr>
</tbody>
</table>