V26 Regular vs. alternative splicing

• Regular splicing
  
  mechanistic steps

  recognition of splice sites

• Alternative splicing
  
  different mechanisms

  how frequent is alternative splicing?

• Effect of alternative splicing on protein – protein interactions

• Interplay of alternative splicing and epigenetic modifications
Our contact with splicing:
MBD2 recognizes methylated cytosines

MBD2 is alternatively spliced and then plays a role for maintenance of pluripotency.
STIM: Orai channels

STIM proteins regulate **store-operated calcium entry** (SOCE) by sensing $\text{Ca}^{2+}$ concentration in the ER and forming oligomers to trigger $\text{Ca}^{2+}$ entry through plasma membrane-localized Orai1 channels.

Alternative splicing may affect PP interactions: STIM2 splice variant

Niemeyer and co-workers characterized a STIM2 splice variant which retains an additional 8-amino acid exon in the region encoding the channel-activating domain.

STIM2.1 knockdown increases SOCE in naive CD4⁺ T cells, whereas knockdown of STIM2.2 decreases SOCE.

Overexpression of STIM2.1, but not STIM2.2, decreases SOCE.

**STIM2.1 interaction with Orai1** is impaired and prevents Orai1 activation.

Effect of AS on protein domain architecture

(left) fraction of proteins where the domain architecture (DA) is altered as a result of splicing (based on Swissprot transcripts)
(right) number of isoforms for 3 databases; Ensembl, Vega/ Havana and Swissprot.

small nuclear RNA (snRNA) genes are part of the spliceosome. Shown are cis-acting elements and trans-acting factors involved in the expression of snRNA genes.

DSE: distal sequence element, and PSE: proximal sequence element TSS, transcription start site.

snRNA promoters recruit the little elongation complex (LEC).

Initiation of snRNA transcription requires general transcription factors (GTFs), as well as the snRNA-activating protein complex (SNAPc).
Shown in (b) are *cis*-acting elements and *trans*-acting factors involved in the expression of protein-coding mRNA genes. DSE and PSE of snRNAs are roughly equivalent to the enhancer and TATA box elements, respectively, of mRNA genes. Ex, exon pA, polyA signal ss, splice site.

While snRNA promoters recruit the LEC, mRNA promoters recruit the super elongation complex (SEC).

**Integrator subunit 11 (INTS11) and INTS9** have sequence similarities to the mRNA 3´-processing factors cleavage and polyadenylation specificity factor 73 kDa subunit (CPSF73) and CPSF100, respectively.
Assembly of the splicesome + splicing steps of pre-mRNA

Spliceosome assembly takes place at sites of transcription.

The U1 and U2 small nuclear ribonucleoproteins (snRNPs) assemble onto the pre-mRNA in a co-transcriptional manner through recognition of the 5′ splice site (5′ss) and 3′ss.

Recognition is mediated by the carboxy-terminal domain (CTD) of polymerase II.

The U1 and U2 snRNPs of different exons then interact with each other to form the pre-spliceosome (complex A).

This process is dependent on DExD/H helicases pre-mRNA-processing 5 (Prp5) and Sub2.

Complex B* then carries out the first catalytic step of splicing, generating complex C, which contains free exon 1 (Ex1) and the intron–exon 2 “lariat intermediate”.

Assembly of the splicesome + splicing steps

In a subsequent reaction catalysed by Prp28, the preassembled tri-snRNP U4–U6•U5 is recruited to form complex B.

The resulting complex B undergoes a series of rearrangements to form a catalytically active complex B (complex B*), which requires multiple RNA helicases (Brr2, Snu114, Prp2) and results in the release of U4 and U1 snRNPs.

Complex C undergoes additional rearrangements and then carries out the second catalytic step, resulting in a post-spliceosomal complex that contains the lariat intron and spliced exons.

Finally, the U2, U5 and U6 snRNPs are released from the mRNP particle and recycled for additional rounds of splicing. Release of the spliced product from the spliceosome is catalysed by the DExD/H helicase Prp22.

During splicing, RNA–RNA interactions are rearranged in a stepwise manner to create the **catalytic centre** of the spliceosome.

- Initially, U1 and U2 small nuclear RNA (snRNA) pair with the 5´ss and the branch point sequence within complex A (the branch point adenosine is indicated by the letter A).
- Subsequently, complex A associates with the U4–U6•U5 tri-snRNP, leading to new base pairs between U2 and U6 snRNA and between U5 snRNA and exonic sequences near the 5´ss.
- The U4 snRNA is disassociated from U6 to expose the 5´ end of U6, which then base pairs with the 5´ss to displace U1 snRNA.
- In the end, an extensive network of base-pairing interactions is formed between U6 and U2, **juxtaposing** the 5´ss and branch-point adenosine for the first catalytic step of splicing. The central region of U6 snRNA forms an intramolecular stem-loop (the U6-ISL), which is essential for splicing catalysis.

### Composition of spliceosomal snRNPs

<table>
<thead>
<tr>
<th>snRNP</th>
<th>RNA secondary structure</th>
<th>Sm proteins</th>
<th>Other core proteins associated with snRNA</th>
<th>Associated proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>U1</td>
<td>568 nt</td>
<td>B, D3, G, E, F, D2 and D1</td>
<td>Snp1 (U1–70K), Mud1 (U1A) and Yhc (U1C)</td>
<td>Prp39, Prp40, Prp42, Snu71, Nam8, Snu56 and Urn1</td>
</tr>
<tr>
<td>U2</td>
<td>1,175 nt</td>
<td>B, D3, G, E, F, D2 and D1</td>
<td>Lea1 (U2A), Msl1( U2B’), Prp9 (SF3B60), Prp11 (SF3B66), Prp21 (SF3A120), Rds3 (SF3B14B), Snu17 (SF3B14A; also known as p14), Hsh155 (SF3B155), Cus1 (SF3B145), Rse1 (SF3B130), Hsh49 (SF3B49) and Ysf3 (SF3B10)</td>
<td>U2AF35, Mud2 (U2AF65) and Msl5 (SF1; also known as BBP)</td>
</tr>
<tr>
<td>U4–U6</td>
<td>160 nt and 112 nt, respectively</td>
<td>U4: B, D3, G, E, F, D2 and D1; U6: Lsm2–8</td>
<td>Prp3, Prp31, Prp4 and Snu13</td>
<td></td>
</tr>
<tr>
<td>U5</td>
<td>179 nt for short form; 214 nt for long form</td>
<td>B, D3, G, E, F, D2 and D1</td>
<td>Prp8, Prp6, Prp28, Brr2, Snu114, U5-40K and Dib1</td>
<td>Snu23, Prp38, Prp2, Spp2, Yju2 and Cbc2 (52K)</td>
</tr>
<tr>
<td>U4–U6•U5</td>
<td>U4 and U5: B, D3, G, E, F, D2 and D1 (i.e. two sets); U6: Lsm2–8</td>
<td>Prp3, Prp31, Prp4, Snu13/15.5K, Prp8, Prp6, Prp28, Brr2, Snu114, U5-40K, snRNP27 and Dib1</td>
<td>Snu23, Prp38, Prp2, Spp2, Yju2, Snu66 and Sad1</td>
<td></td>
</tr>
</tbody>
</table>

*Cus1, cold-sensitive U2 snRNA suppressor 1; Lea1, looks exceptionally like U2A 1; Lsm, Sm-like; Msl, Male-specific lethal homologue; nt, nucleotides; Prp, pre-mRNA-processing; Rds3, regulator of drug sensitivity 3; Rse1, RNA splicing and ER to Golgi transport factor 1; Sad1, snRNP assembly-defective 1; SF, splicing factor; Spp2, suppressor of PRP; snRNA, small nuclear RNA; snRNP, small nuclear ribonucleoprotein; Snu, US small nuclear ribonucleoprotein component; U2AF, U2 auxiliary factor; Urn1, U2-US-U6 snRNP; RES complex and NTC-interacting pre-mRNA-splicing factor 1. *The protein composition is primarily based on a mass spectrometric analysis of the yeast spliceosome[49]; certain regulatory factors that are closely associated with the core spliceosome (such as SR proteins) are not included. Proteins are listed using the budding yeast nomenclature unless there is no known yeast homologue. In certain cases, the common name of a metazoan homologue is also included in brackets. The snRNA lengths are based on yeast transcripts.
The splicing starts with an **AG site** and is preceded with a non-AG (pyrimidine rich) region preceded by the branch point that includes an **Adenosine** residue.

The 5’ end of the intron contains an almost invariant **GU sequence**.
Mechanisms of alternative splicing

- **Gray boxes**: exons
- **White boxes**: introns

The (gray) protein coding regions are excluded/included in different transcripts.

---


26. lecture SS 2018

Bioinformatics III
Splice site choice is regulated through \textit{cis}-acting splicing regulatory elements (SREs) and \textit{trans}-acting splicing factors.\medskip

Shown sequence motifs are the consensus motifs of splice sites.\medskip

The height of each letter represents the nucleotide frequency in each position.\medskip

The dashed arrow represents the formation of the exon definition complex.

On the basis of their relative locations and activities, splicing regulatory elements are classified as
- exonic splicing enhancers (ESEs),
- intronic splicing enhancers (ISEs),
- exonic splicing silencers (ESSs) or
- intronic splicing silencers (ISSs).

Sequence motifs cannot exert their effects directly → these SREs specifically recruit splicing factors to promote or inhibit recognition of nearby splice sites:
- SR proteins recognize ESEs to promote splicing,
- heterogeneous nuclear ribonucleoproteins (hnRNPs) typically recognize ESSs to inhibit splicing.

Matera & Wang,
Nature Rev Mol Cell Biol
15, 108–121 (2014)
Activity of splicing factors and SREs

The activity of splicing factors and cis-acting SREs is context-dependent.

**Oligo-G tracts** are recognized by hnRNP H.

(Top) When the oligo-G tracts are located inside an intron, they function as intronic splicing enhancers (ISE) to promote splicing.

(Bottom) When they are located within exons, they function as exonic splicing silencers (ESSs).

Matera & Wang,
Nature Rev Mol Cell Biol
15, 108–121 (2014)
YCAY motifs are recognized by neuro-oncological ventral antigen (NOVA).

Y stands for pyrimidine (C/T).

(Top) When YCAY motifs are located inside an exon, they act as ESEs,

(Middle) When they are located in the upstream intron of an alternative exon, they act as ISSs,

(Bottom) When they are located inside an intron, they act as ISEs.

Matera & Wang,
Nature Rev Mol Cell Biol
15, 108–121 (2014)
Activity of splicing factors and SREs

Binding sites for SR proteins and hnRNP A1 also have distinct activities when located at different regions on the pre-mRNA.

Matera & Wang,
Nature Rev Mol Cell Biol
15, 108–121 (2014)
Define 3 regions C1/I1(5')/I1(3') before an „alternative exon“ (A) and 3 regions I2(5')/I2(3')/C2 behind the alternative exon.
Approach to extract RNA features

Barash et al. Nature
465, 53- (2010)
Bars without black line denote feature enrichment, bars with **black line** feature depletion.

Bar size conveys enrichment P-value; P < 0.005 in all cases.

Column “Feature” lists sequence motifs.

E.g. the CUG-rich motif of the second row recruits the binding protein Cugbp.

Each table cell contains 5 bars. They show the region-specific activity of each feature for increased exon inclusion (**red bar**) or exclusion (**blue bar**) in 5 different mouse tissues:

- CNS (C),
- muscle (M),
- embryo (E) and
- digestive (D) tissues, plus a tissue-independent mixture (I).

<table>
<thead>
<tr>
<th>Feature</th>
<th>C1</th>
<th>I1(5')</th>
<th>I1(3')</th>
<th>A</th>
<th>I2(5')</th>
<th>I2(3')</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-rich</td>
<td>IC MED</td>
<td>IC MED</td>
<td>IC MED</td>
<td>IC MED</td>
<td>IC MED</td>
<td>IC MED</td>
<td>IC MED</td>
</tr>
<tr>
<td>CUG-rich(Cugbp)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# The splicing code


<table>
<thead>
<tr>
<th>Feature</th>
<th>C1 CMED</th>
<th>I1(5’) CMED</th>
<th>I1(3’) CMED</th>
<th>A CMED</th>
<th>I2(5’) CMED</th>
<th>I2(3’) CMED</th>
<th>C2 CMED</th>
</tr>
</thead>
<tbody>
<tr>
<td>U-rich</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CUG-rich(Cugbp)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[U]GCAUG(Fox)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACUAAY(Qk)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CU-rich(consPTEB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YCAVY(Nova)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YGCUCY(MbnI)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GRYYcSYR(SC35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YRCYKM(SRp55)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YYACWSS(SRp40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UGAUUUU]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UAUAUG]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UUGGUU]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UUUAUAA]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[GUAAG]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UGCUU]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UCAUUC]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UUAGAA]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[GUUU]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UGGCUU]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[AGUC]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[UUUGAG]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mot[AUUAAAU]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTCinc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTCexc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frameshift</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junction score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First[AG]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

26. lecture SS 2018       | Bioinformatics III |
Tissue-specific alternative splicing

Shown are mRNA-Seq reads mapping to a portion of the SLC25A3 gene locus. SLC25a3 is a mitochondrial phosphate carrier.

The number of mapped reads starting at each nucleotide position is displayed ($\log_{10}$) for the tissues listed at the right.

Bottom: exon/intron structures of representative transcripts containing **mutually exclusive exons** 3A and 3B (GenBank accession numbers AK074759 and AK092689).

**tissue-specific regulation of alternative mRNA isoforms**

<table>
<thead>
<tr>
<th>Alternative transcript events</th>
<th>Total events ($\times 10^3$)</th>
<th>Number detected ($\times 10^3$)</th>
<th>Both isoforms</th>
<th>Number tissue-regulated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skipped exon</td>
<td>37</td>
<td>35</td>
<td>10,436</td>
<td>6,822</td>
</tr>
<tr>
<td>Retained intron</td>
<td>1</td>
<td>1</td>
<td>167</td>
<td>96</td>
</tr>
<tr>
<td>Alternative 5’ splice site (A5SS)</td>
<td>15</td>
<td>15</td>
<td>2,168</td>
<td>1,386</td>
</tr>
<tr>
<td>Alternative 3’ splice site (A3SS)</td>
<td>17</td>
<td>16</td>
<td>4,181</td>
<td>2,655</td>
</tr>
<tr>
<td>Mutually exclusive exon (MXE)</td>
<td>4</td>
<td>4</td>
<td>167</td>
<td>95</td>
</tr>
<tr>
<td>Alternative first exon (AFE)</td>
<td>14</td>
<td>13</td>
<td>10,281</td>
<td>5,311</td>
</tr>
<tr>
<td>Alternative last exon (ALE)</td>
<td>9</td>
<td>8</td>
<td>5,246</td>
<td>2,491</td>
</tr>
<tr>
<td>Tandem 3’ UTRs</td>
<td>7</td>
<td>7</td>
<td>5,136</td>
<td>3,801</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>105</strong></td>
<td><strong>100</strong></td>
<td><strong>37,782</strong></td>
<td><strong>22,657</strong></td>
</tr>
</tbody>
</table>

**Blue, red, grey**: mapped reads supporting expression of upper isoform, lower isoform or both isoforms.

The four columns to the right show the numbers of events of each type:
1. supported by cDNA and/or EST data;
2. 1 isoform supported by mRNA-Seq reads;
3. both isoforms supported by reads;
4. events detected as tissue-regulated (difference significant under Fisher's exact test).

Review from V3: Not considered yet: alternative splicing

DNA

---

exon 1 | exon 2 | exon 3 | exon 4

5' | | | 3'

3' | | | 5'

primary RNA transcript

---

transcription

---

alternative splicing (~95% of human multi-exon genes)

mRNAs

---

translation

---

translation

---

translation

protein isoforms

---

AS affects ability of proteins to interact with other proteins
Differential PPI wiring analysis

112 matched normal tissues (TCGA)  112 breast cancer tissues (TCGA)

comparison 1: $d_1$

comparison 2: $d_2$

comparison 3: $d_3$  $\sum d_i$

one-tailed binomial test + BH/FDR ($<0.05$)

Check whether rewiring of a particular PP interaction occurs in a significantly large number of patients compared to what is expected by chance rewiring events.
PPIXpress method

I. mapping
reference: principal protein isoforms

II. instantiation
built using most abundant protein isoforms
The enriched terms that are exclusively found by the transcript-level method (right) are closely linked to carcinogenetic processes.

Hardly any significant terms are exclusively found at the gene level (left).

Table S16: Comparison of rewiring results between the gene-based construction and a transcript-based construction method for the BioGRID network. Here, the top five enriched terms and their $p$-values are shown for the proteins affected by interactions exclusively found by the transcript-based method using the ALL-DDI dataset or the gene-based approach, respectively. Enrichment in KEGG pathways and GO biological processes was determined using DAVID [2] where we used the proteins included in the corresponding input network as the background. Enrichment was defined as $p < 0.05$ (Bonferroni-adjusted).
Hematopoiesis (development of blood cells)

Hematopoietic stem cells (in bone marrow)

progenitor cells

terminally differentiated cells

Bone marrow

HSC

MPP

CMP

CLP

LMPP

MEP

EB

MK

GMP

Blood

N

M

CD4
PPICmpare workflow

A

group 1

all pairs

group 2

pairwise comparison i

P_{rew,i} = J_d(a_i,b_i)

d_i

mean(P_{rew,i})

\sum d_i

B

P

n

+ FDR

C

P3 & P2

P2

P3

P4
Calibrating RNAseq on/off threshold against proteome data (MS)

![Graph showing the relationship between threshold [TPM] and proteins in network for different cell types. The graph highlights the best agreement at a specific threshold.]
How many RNAseq samples are needed?

X-axis:
Total number of samples in both groups

Blue -> violet:
Imbalance reduces

**Subsampling** shows that reasonable results are obtained for ≥ 3 samples
Rewiring is due to …

Different types of alterations can cause same rewiring event

DE: differential expression
DE/DE: both proteins DE

AS: alternative splicing
DE/AS: one protein affected by DE, the other by AS

Simultaneous deregulation of interaction partners is often biologically meaningful.

- more likely to participate in known protein complexes
- lower betweenness in interaction network
- more likely active in same biological process

functional modules
contribution of AS seems minor (< 1%)

**BUT** 548 rewiring events in hematopoiesis are due to AS.

Rewiring events exclusively regulated by AS were enriched in GO terms related to:

- **post-elongation** processing of mRNA
- **cell cycle** (G2-M checkpoint and control of pre-replication complex)
- **transport of mRNA from the nucleus to the cytoplasm**, Hippo signaling, as well as Interleukin receptor SHC signaling
Identify **reduced set** of transcriptomic changes that
– explains all rewiring events (i.e. is very likely given the data) and
– is of small cardinality: **weighted set-cover problem**

\[ \rightarrow (\text{left}) \text{roteins in reduced set are hub proteins in the differential network,} \]
not in full reference network (right)

They tend to be **connectors** of **different functional modules**.

26. lecture SS 2018

Bioinformatics III
Rewiring HSC ➝ MPP

edge color:
direction of rewiring event
- emerging
- vanishing

edge thickness:
ocurrence of rewiring event
- 15/18
- 18/18

node color:
protein classification
- protein in minimal reasons set
- all other proteins

node size:
protein’s importance in rewiring
deepened with relevance for rewiring events

node shape:
regulatory mode
- differential expression
- alternative splicing

Bone marrow:
HSC ➝ MPP

Blood:
EB ➝ MEP ➝ GMP ➝ CLP ➝ CD4

26. lecture SS 2018

Bioinformatics III
involvement of epigenetics in alternative splicing?

V21 – V23 showed that chromatin state plays an essential role in regulating gene expression.

Although epigenetic signatures are mainly found to be enriched in promoters, it has become increasingly clear that they are also present in exon regions, indicating a potential link of epigenetic regulation to splicing.

Zhou et al. BMC Genomics 2012, 13:123
http://www.biomedcentral.com/1471-2164/13/123

RESEARCH ARTICLE

Epigenetic features are significantly associated with alternative splicing

Yuanpeng Zhou, Yulan Lu and Weidong Tian

26. lecture SS 2018 Bioinformatics III
Procedure to recognize AS events

Junction site annotation and alternative splicing recognition process.

Data sources:
- Annotated AS types from Ensembl
- DNA methylation data from Salk institute
- Chip-seq data from ENCODE and elsewhere


26. lecture SS 2018 Bioinformatics III
### Procedure to recognize AS events

<table>
<thead>
<tr>
<th>Sketch map</th>
<th>Recognition Code</th>
<th>AS Type</th>
<th>Number of Events</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Sketch" /></td>
<td>0 1 0</td>
<td>CNE</td>
<td>103,806</td>
</tr>
<tr>
<td><img src="image2" alt="Sketch" /></td>
<td>0 1 1</td>
<td>A3SS</td>
<td>2,521</td>
</tr>
<tr>
<td><img src="image3" alt="Sketch" /></td>
<td>0 0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image4" alt="Sketch" /></td>
<td>1 1 0</td>
<td>A5SS</td>
<td>1,884</td>
</tr>
<tr>
<td><img src="image5" alt="Sketch" /></td>
<td>1 0 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image6" alt="Sketch" /></td>
<td>1 1 1</td>
<td>IR</td>
<td>563</td>
</tr>
<tr>
<td><img src="image7" alt="Sketch" /></td>
<td>1 0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image8" alt="Sketch" /></td>
<td>0 1 0 1 0</td>
<td>ME</td>
<td>1,206</td>
</tr>
<tr>
<td><img src="image9" alt="Sketch" /></td>
<td>0 0 0 1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><img src="image10" alt="Sketch" /></td>
<td>0 1 0 1 0</td>
<td>ES</td>
<td>2,725</td>
</tr>
<tr>
<td><img src="image11" alt="Sketch" /></td>
<td>0 1 0 0 1 0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CNE: constitutively spliced exon (no AS)

The recognition code of AS events and the number of each type of splicing event

Coupling AS ↔ epigenetic modifications

The association of DNA methylation and nucleosome occupancy with AS.

(a) Distribution of genomic CpG levels around the splice sites of different types of AS events. 6% is the expected frequency of dinucleotides. However, CpG levels are lower on average (V21).

(b) Distribution of DNA methylation level (mCG) around the splice sites of different types of AS events. Both (a) and (b) use a sliding window of 147 bp.

CNE: constitutively spliced exon (no AS)
ES: exon skipping
ME: mutually exclusive exon
A5SS: alternative 5' splice site selection
A3SS: alternative 3' splice site selection
IR: intron retention.
Association of histone modification with AS

H3K36me3 (a) is the only histone PTM that is significantly associated with all types of AS events in all regions.

However the association patterns are different:
H3K36me3 levels are significantly lower in ME and ES and significantly higher in A3SS, A5SS and IR.

The levels of H3K4 methylation, including H3K4me1, H3K4me2, and H3K4me3 (b - d), are almost all significantly higher in A3SS and A5SS.

Association of histone modification with AS

For the other histone methylations,
(e) the level of H4K20me1 is significantly higher in A3SS, A5SS and IR;
(f) the level of H3K27me3 is significantly higher in the exonic region of ES;
(g) the level of H3K79me1 is significantly higher in A3SS and A5SS, and slightly higher in the intronic region of ES;
(h) the level of H3K79me2 (h) is significantly higher in ME, A3SS and A5SS, and most region of IR;
(i) the level of H3K9me1 is significantly higher in A3SS, A5SS and most regions of IR.
(j) However, H3K9me3 is not significantly associated with any type of ASE.
The binding levels of EGR1, GABP, SIN3A, SRF and RNA Pol II (a - e) are all significantly higher in A3SS, A5SS and IR, and significantly lower in ME and ES,

Their levels all steadily increase from ME, ES, CNE, A3SS, A5SS to IR. This is similar to the results for H3K36me3.
Clustering of associations

epigenetic modifications that are associated with AS

The epigenetic features are strongly associated with different types of AS.

The features showing higher level and lower level in AS events than in CNE are colored in red and green, respectively. The features inside the dashed black box are those common in both ESRP and ASSP; note their association patterns are very different in between ESRP and ASSP.
Coupling AS ↔ epigenetic modifications

Epigenetic features are strongly associated with AS.

This suggests that epigenetic regulation may be involved in AS.

Clustering yielded 4 tight clusters of epigenetic features that are associated with AS.

The AS events may be grouped into 2 classes on the basis of their association patterns with epigenetic features.
- the exon skipping related process (ESRP) (including ME and ES) and
- the alternative splice site selection process (ASSP) (including A3SS, A5SS and IR)

→ these 2 processes may involve different mechanisms of epigenetic regulation.