RNA Regulatory Networks in Health and Disease

Yi Xing
Departments of Internal Medicine, Biomedical Engineering
University of Iowa
Outline

• Background on RNA splicing and alternative splicing.

• Alternative splicing network during the Epithelial Mesenchymal Transition.

• Evolution of new exons in primates.
Regulation of pre-mRNA splicing

Wang and Cooper, *Nature Reviews Genetics* 8, 749-761
Alternative Splicing
DSCAM alternative splicing

Figure 7–89. Molecular Biology of the Cell, 4th Edition.
Importance of Alternative Splicing

• >90% of human multi-exon genes undergo alternative splicing.

• Important in regulation of gene function.

• Aberrant splicing is a major cause of human diseases [1].

• An important mechanism for acquisition of evolutionary novelties [2-3].

Control of Alternative Splicing by Tissue-specific Splicing Factors

PTB: a switch for neuronal-specific splicing

EST analysis: first wave of alternative splicing discovery

Alternative splices match at one site, but differ at the other (excludes intron inclusion, other artifacts)

The Multiassembly Problem: Reconstructing Multiple Transcript Isoforms From EST Fragment Mixtures

Yi Xing, Alissa Resch, and Christopher Lee¹

UCLA–DOE Center for Genomics and Proteomics, Molecular Biology Institute and Department of Chemistry & Biochemistry, University of California, Los Angeles, Los Angeles, California 90095-1570, USA

An expectation-maximization algorithm for probabilistic reconstructions of full-length isoforms from splice graphs

Yi Xing¹,*, Tianwei Yu²,³, Ying Nian Wu², Meenakshi Roy¹, Joseph Kim¹ and Christopher Lee¹,*

MADS: A new and improved method for analysis of differential alternative splicing by exon-tiling microarrays

Yi Xing,¹,² Peter Stoilov,³,⁴ Karen Kapur,⁵ Areum Han,⁶ Hui Jiang,⁷ ShiHao Shen,⁸ Douglas L. Black,³,⁴ and Wing Hung Wong⁵
Genomic Approaches for Global Analysis of Alternative Splicing

**High-density Exon Array**

1 gene --- many probesets
Probes from each putative exon
1.4 Million probesets, >6 M probes

**Ultra-deep RNA Sequencing**

5. Shen et al., *Bioinformatics*, 26:268-269, 2010

5. Shen et al., *Nucleic Acids Research*, in revision
Outline

• Background on RNA splicing and alternative splicing.

• Alternative splicing network during the Epithelial Mesenchymal Transition.

• Evolution of new exons in primates.
The Epithelial to Mesenchymal Transition (EMT): Roles in development, fibrosis and metastasis

Mutually Exclusive Alternative Splicing of Fibroblast Growth Factor Receptor 2 (FGFR2) Exons 8 and 9

Ligand Binding Specificity:
- **FGFR2-E8**: FGF-3, 7, 10, 22
- **FGFR2-E9**: FGF-2, 4, 5, 6, 8, 9, 17
ESRP – A master splicing switch of epithelial-mesenchymal transition

• ESRPs
  – Epithelial Splicing Regulatory Proteins
  – ESRPs promote the inclusion of FGFR2 exon 8 and repress the inclusion of exon 9

ESRP expression is restricted exclusively to epithelial cells

PNT2: Human Prostate Epithelial Cells

siRNA: GFP

siRNA: ESRP1 and ESRP2

RNA

RNA

MDA-MB-231: Human Breast Cancer Mesenchymal Cells

Retrovirus: EGFP

Retrovirus: mEsrp1

RNA

RNA

Affymetrix HJAY exon junction array

RNA-Seq (Illumina)

Genome-wide discovery of ESRP targets using RNA-Seq

PNT2: Human Prostate Epithelial Cells

siRNA: GFP

siRNA: ESRP1 and ESRP2

Total RNA

RNA-Seq Library Preparation

76bp Sequencing

59M reads

MDA-MB-231: Human Breast Cancer Mesenchymal Cells

Retrovirus: EGFP

Retrovirus: mEsrp1

Total RNA

RNA-Seq Library Preparation

76bp Sequencing

120M reads

76bp Sequencing

136M reads
Discovery of ESRP Targets by RNA-Seq

Shihao Shen, MATS: Multivariate Analysis of Transcript Splicing
Discovery of Novel ESRP Targets by RNA-Seq

<table>
<thead>
<tr>
<th>Exon Inclusion Level</th>
<th>ESRP</th>
<th>EV</th>
<th>ESRP- EV</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA-Seq</td>
<td>0.27</td>
<td>0.76</td>
<td>-0.49</td>
</tr>
<tr>
<td>RT-PCR</td>
<td>0.36</td>
<td>0.75</td>
<td>-0.39</td>
</tr>
</tbody>
</table>

Overall validation rate: **86%** (115 out of 134)
RNA-Seq Validation Summary

**MDA-MB-231**
(Ectopic Esrp1)
547 predicted

RT-PCR Validations:
- 115/134 (85.8%) >5%
- 104/134 (77.6%) >10%
(+55 previously validated from HJAY)

**PNT2**
(siRNAs vs. ESRP1/2)
35 predicted

RT-PCR Validations:
- 13/18 (72.2%) >5%
- 10/18 (55.6%) >10%
(+12 previously validated from HJAY)

**MDA-MB-231 only:**
33 predicted

RT-PCR Validations:
- 7/13 (53.8%) >5%
- 6/13 (46.2%) >10%
ESRP targets exhibit evidence of physiologically relevant co-regulated splicing

• In a number of cases the protein isoforms have been shown to have divergent functions consistent with differential morphologies of epithelial vs. mesenchymal cells (e.g. p120-catenin/CTNND1)

• Enriched in relevant protein interaction networks and canonical pathways including:
  • Tight Junction
  • Adherens Junction
  • Small GTPase regulator activity
  • Focal Adhesion
  • Integrin Signaling
  • ERK/MAPK Signaling
  • Protein localization and vesicle-mediated transport
  • Regulation of the actin cytoskeleton.
Enriched RNA Motifs Around ESRP-Regulated Exons

<table>
<thead>
<tr>
<th>Motif</th>
<th>p-value</th>
<th>Motif</th>
<th>p-value</th>
<th>Motif</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCCTCC</td>
<td>4.36E-20</td>
<td>ACTCCG</td>
<td>1.31E-21</td>
<td>GTGGTG</td>
<td>7.05E-65</td>
</tr>
<tr>
<td>TGCCGA</td>
<td>7.54E-18</td>
<td>CCGATG</td>
<td>1.35E-17</td>
<td>GGTGGT</td>
<td>7.84E-44</td>
</tr>
<tr>
<td>ATGACT</td>
<td>1.42E-16</td>
<td>CGATGT</td>
<td>4.14E-17</td>
<td>TGGTGG</td>
<td>1.64E-37</td>
</tr>
<tr>
<td>TAATTC</td>
<td>3.97E-16</td>
<td>TAACCC</td>
<td>4.01E-15</td>
<td>GCTGTC</td>
<td>7.53E-35</td>
</tr>
<tr>
<td>TGCAUG</td>
<td>8.34E-16</td>
<td>GTCGAC</td>
<td>1.50E-14</td>
<td>TGGTGC</td>
<td>4.20E-33</td>
</tr>
</tbody>
</table>

**ESRP Enhanced**

ESRP Silenced

---

GT-rich motif

FOX-1/2 motif
Experimental determination/validation of a UGG-rich ESRP1 binding site by SELEX-Seq

**Systematic Evolution of Ligands by EXponential enrichment (SELEX)**

**Table: Total Reads and Unique Reads**

<table>
<thead>
<tr>
<th>Round</th>
<th>Total Reads</th>
<th>Unique Reads</th>
<th>% Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>7,090,898</td>
<td>6,249,422</td>
<td>88.1%</td>
</tr>
<tr>
<td>2</td>
<td>4,611,229</td>
<td>3,587,281</td>
<td>77.8%</td>
</tr>
<tr>
<td>3</td>
<td>8,730,409</td>
<td>5,924,313</td>
<td>67.9%</td>
</tr>
<tr>
<td>6</td>
<td>6,390,525</td>
<td>2,679,891</td>
<td>41.9%</td>
</tr>
<tr>
<td>7</td>
<td>5,258,739</td>
<td>1,352,191</td>
<td>25.7%</td>
</tr>
</tbody>
</table>
SELEX defined ESRP-binding motifs validate previous bioinformatically predicted binding sites.

Top 12 6-mers after SELEX Round 7

- **SELEX Motif**
  - TGGTGG
  - GGTGGG
  - GTGGTG
  - GTGGGG
  - GTGTGG
  - GGTGTG
  - TGTGGG
  - GGTGGT
  - TGGGGT

Confirmed by gel mobility shift assay.
A SELEX-Seq motif score defines a position-dependent ESRP RNA map

- ESRP Enhanced exons (103)
- ESRP Silenced exons (173)
- HJAY array non-ESRP target background set (3508)

ESRP1 Motif Score

Scan window: 45nt and top 12 SELEX-Seq motif-based score
“A Splicing Mastermind for EMT”
An ESRP splicing signature that distinguishes epithelial cells from mesenchymal cells

Luminal Basal B

Epithelial cells Mesenchymal cells

A

B

Epithelial cell lines

Mesenchymal cell lines

% Exon inclusion

% Exon inclusion

0 20 40 60 80 100

0 20 40 60 80 100

ENAH FNIP1 ARFGAP2 RALGPS2 SLC37A2 ARHGEF11 MAGI1 SCRIB

ENAH FNIP1 ARFGAP2 RALGPS2 SLC37A2 ARHGEF11 MAGI1 SCRIB

% Exon inclusion

% Exon inclusion
The ESRPs regulate alternative polyadenylation (APA)

BCL2-associated athanogene (BAG1)

ESRP1
3’ DRS

ESRP1
RNA-Seq

Control
3’ DRS

Control
RNA-Seq

Extended UTR

Common region

MDA-MB-231 mesenchymal cells
Outline

• Background on RNA splicing and alternative splicing.

• Alternative splicing network during the Epithelial Mesenchymal Transition.

• Evolution of new exons in primates.
Alternative splicing and RNA selection pressure — evolutionary consequences for eukaryotic genomes

Yi Xing*15 and Christopher Lee*

Global analysis of alternative splicing differences between humans and chimpanzees

John A. Calarco,1,2,8 Yi Xing,3,4,8 Mario Cáceres,5,6,8 Joseph P. Calarco,1 Xinshu Xiao,7 Qun Pan,1 Christopher Lee,3 Todd M. Preuss,5,10 and Benjamin J. Blencowe1,2,9

Evolution of alternative splicing in primate brain transcriptomes

Lan Lin1,4,5, Shihao Shen2,4,5, Peng Jiang1, Seiko Sato1, Beverly L. Davidson1,3,4 and Yi Xing1,2,5,*
Some Exons Are Unique to Humans

Selenoprotein N, 1 (SEPN1)
Birth of New Exons

New exons are constantly added to existing functional genes via a variety of mechanisms:

- Insertion and exonization of transposable elements
- De novo exonization from intronic regions
- Exon duplication
Alu retrotransposons

• **Short interspersed nuclear elements (SINE) family**

• Primate-specific transposable elements

• Inserted in the genome of an ancestor of supraprimates at 60-65MYA

• The most abundant mobile elements in human genome
  – >1 million copies in human genome
  – 10% of the human genomic DNA
Alu exonization

A) Retrotransposition

B) Mutations

C) Alternative splicing

Rotem Sorek, RNA, 2007
Alu exonization

A) Retrotransposition

B) Mutations

C) Alternative splicing

Rotem Sorek, RNA, 2007
Alu exonization

• EST analysis revealed that nearly all exonized Alu elements are alternatively spliced; the vast majority are spliced into the transcript at low frequencies.

• It was thought that Alu exons are too young to acquire strong splicing activities; constitutive activation of Alu exons are almost exclusively associated with genetic disorders.

• How can we identify Alu exons with likely functional and regulatory roles, for example exons with tissue-specific splicing in human tissues?
Exon Array Analysis of Alu Exons

Exon array dataset
Public Affymetrix human exon 1.0 array dataset on 11 human tissues (three replicates per tissue)
- Breast, cerebellum, heart, kidney, liver, muscle, pancreas, prostate, spleen, testes, thyroid

Exon array analysis of Alu exons
Internal spliced exons in the UCSC Genome Browser database
Covered by Alu elements for at least 50% of the exon length
Final list: 330 Alu-derived exons, each with at least 3 reliable probes

Xing et al., *RNA*, 14: 1470-1479, 2008
Shen et al., *Bioinformatics*, 26:268-269, 2010
Detection of Alu exons “correlated” with gene expression

“Correlated” exon: at least 3 of the 4 probes show at least 0.6 Pearson correlation coefficient with gene expression level.

FAM55C (Inclusion 250bp, skipping 130bp)
Examples of tissue-specific Alu-derived exons

ICA1 (Inclusion 370bp, skipping 156bp)

Testes specific inclusion
Muscle specific alternative splicing of Selenoprotein N, 1 (SEPN1)

- Expressed in skeletal muscle
- Protection against oxidant damage
- Mutations were linked to one form of congenital muscular dystrophy.
- Two alternative spliced isoforms
  - Full-length isoform contains an Alu-derived exon
  - Predicted to be the minor isoform based on EST data

### Sequence Comparison

<table>
<thead>
<tr>
<th></th>
<th>Homo sapiens</th>
<th>Mus musculus</th>
<th>Gallus gallus</th>
<th>Danio rerio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length</td>
<td>131</td>
<td>102</td>
<td>96</td>
<td>104</td>
</tr>
<tr>
<td>Sequence</td>
<td>LSLLRSTPAASCEEEELPPDPSEETLTIEARFQPLLPETMTKSK</td>
<td>SVPVANYEEELHDPSEETLTIEARFQPLLMETMTKSK</td>
<td>VTPVSDFEED--APDPNGETLSIVAFQPLVMETMTKSK</td>
<td>VAPPPEYEEE-IPHDPNGETLTLHAKMQPLLESMTKSK</td>
</tr>
</tbody>
</table>

Alu-derived exon 3
Evolution of SEPN1 Alu-exon Splicing

The next challenge will be to pin down how these new exons affect the function of the genes in which they reside.

RNA Sequencing (RNA-Seq)

RNA-Seq Analysis of Alu Exons

A

ZNF445

Skipping Junction

1

2

79% Inclusion level

14

Upstream Junction

2

Downstream Junction

B

ZNF445

C

RT-PCR exon inclusion estimates vs.
RNA-Seq exon inclusion levels
Spearman’s $r = 0.60$ (p = 1.1e-5)

123 million reads for the human cerebellum

Shen*, Lin* et al. (2011) PNAS, 108:2837-2842
Alu Exons are Enriched in Zinc Finger Transcription Factors

A) Genes with highly included Alu exons
- 8 ZNF genes
- 27 non-ZNF genes
- 487 ZNF genes
- 16008 non-ZNF genes

p = 7.1e-6

B) Inclusion level
- p = 0.004

C) Gene Expression
- p = 0.007
Alu Exons are Enriched in the 5’-UTR

A

- Internal cassette exons: 2205
- Internal Alu exons: 290
- Alu exons ≥5 reads: 64
- Alu exons ≥50% exon inclusion: 14

B

Exon inclusion level

- CDS
- 5’-UTR

p = 0.05
Alu Exons in 5’-UTR
Regulate Protein Translation
Structural organization of eukaryotic mRNA

5'-UTR of an mRNA:
- Length
- Thermal stability
- GC content
- Secondary structures
- uORFs (upstream ORFs)
- IRES
- Binding sites for proteins

Chatterjee, S et al., Biol Cell. 2009
Alu exons repress translation by creating or elongating uORFs

A  NOSIP 5’-UTR

Exon 2
AUG
UUG
AAG
AUC

B  ZNF81 5’-UTR

Exon 2
UCA
UAA
UAG
UGA

Translational efficiency (%)

Skipping
Inclusion
Inclusion a65t
Inclusion t66a
Inclusion g67c

Translational efficiency (%)

Skipping
Inclusion
Inclusion c131a
Inclusion c131a-a132g
Inclusion c131g
Alu exonization: Regulating the regulators

Pol III → Alu Retrotransposon → Ancestral Element → Mutations

Exonization → Translation → Master transcriptional regulator → Gene activation or repression
Acknowledgements

Lab Members

Peng Jiang
Juw Won Park
Jinkai Wang
Keyan Zhao
Seth Brown
Shihao Shen
Ji Wan
Collin Tokheim

Lan Lin
Zhixiang Lu
Elizabeth Kenkel
Mallory Stroik
Sara Miller
Jennifer Dozier
Jingzhu Xu

Collaborators

Russ Carstens
(U. Penn)

Qing Zhou
(UCLA)

Jeff Murray
(U. Iowa)

James Cai
(TAMU)

Funding

NIH (NHGRI, NIGMS, NHLBI, NIDDK, CTSA-KL2)
Burroughs Wellcome Fund
March of Dimes Foundation
Edward Mallinckrodt Jr. Foundation
Hereditary Disease Foundation
FSH Society