UNIVERSITY of IOWA
CARVER COLLEGE
OF MEDICINE
University of Iowa Health Care

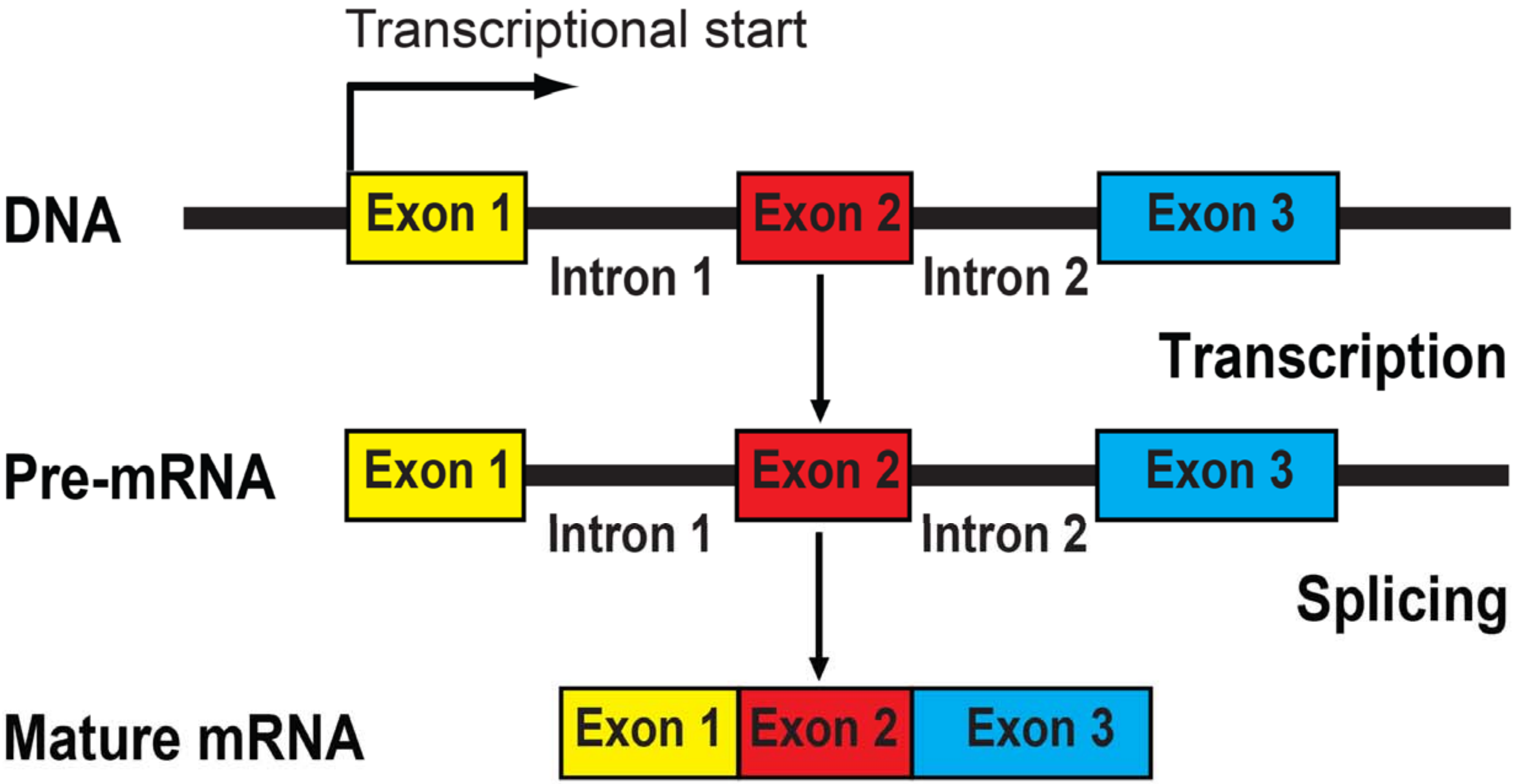
RNA Regulatory Networks in Health and Disease

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

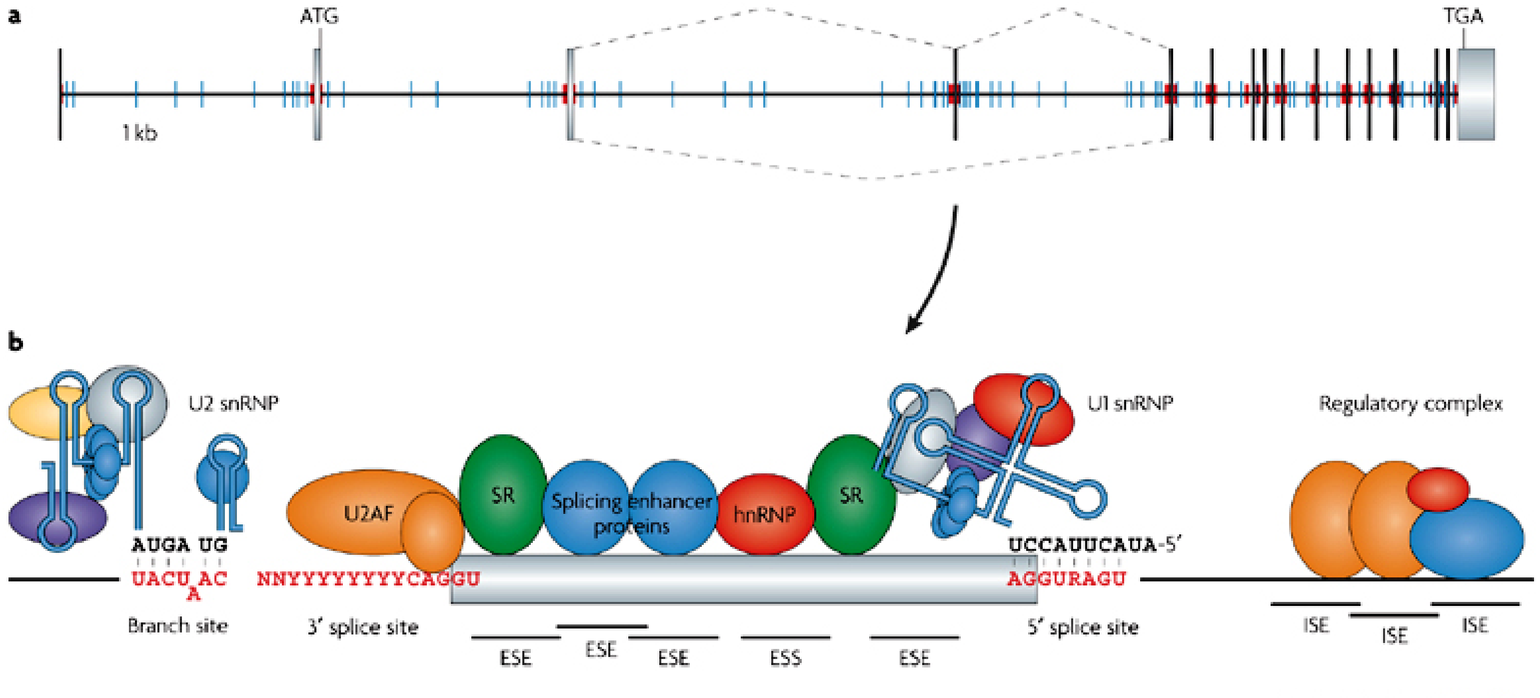
Outline

- Background on RNA splicing and alternative splicing.
- Alternative splicing network during the Epithelial Mesenchymal Transition.
- Evolution of new exons in primates.

RNA Splicing



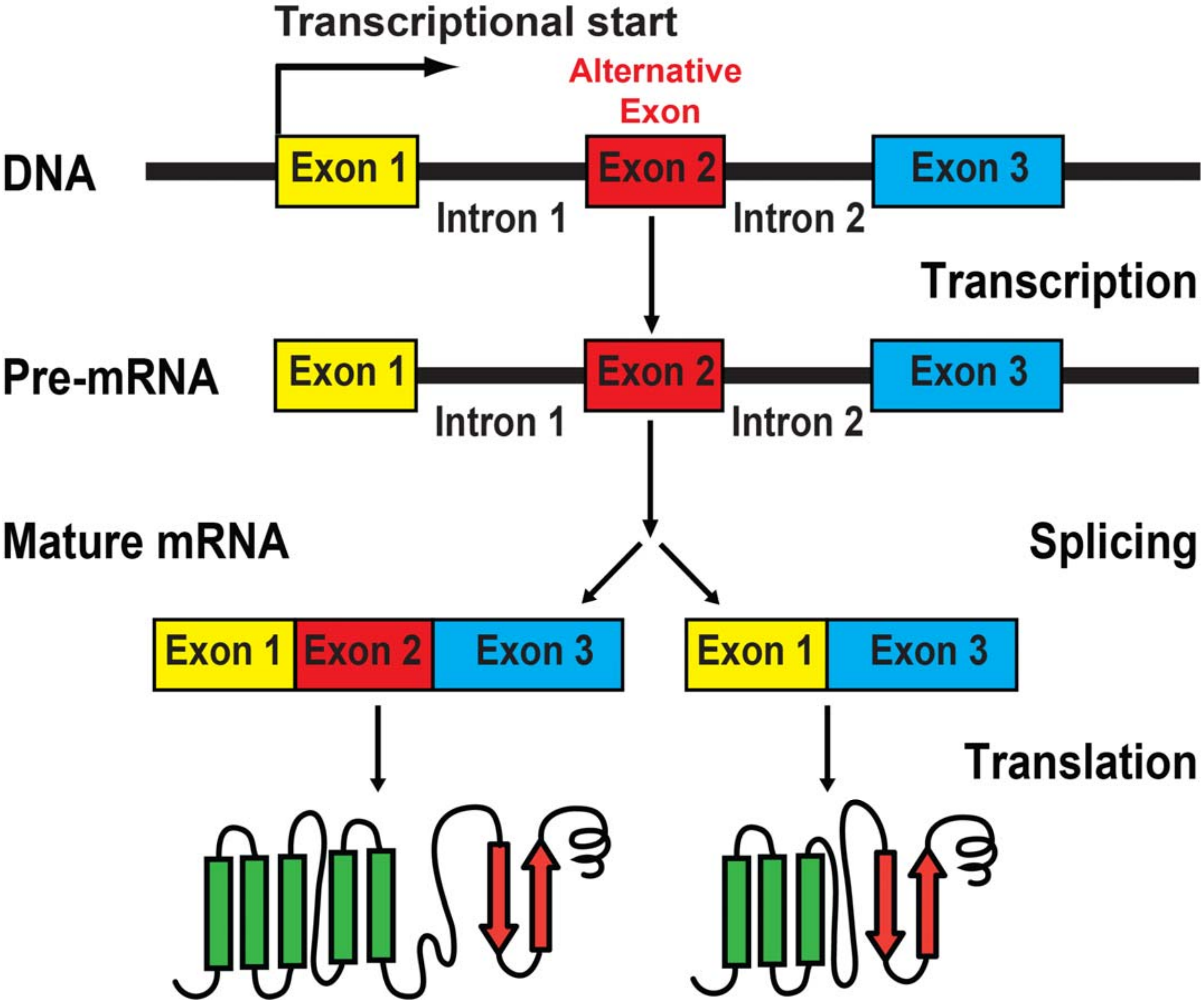
Regulation of pre-mRNA splicing



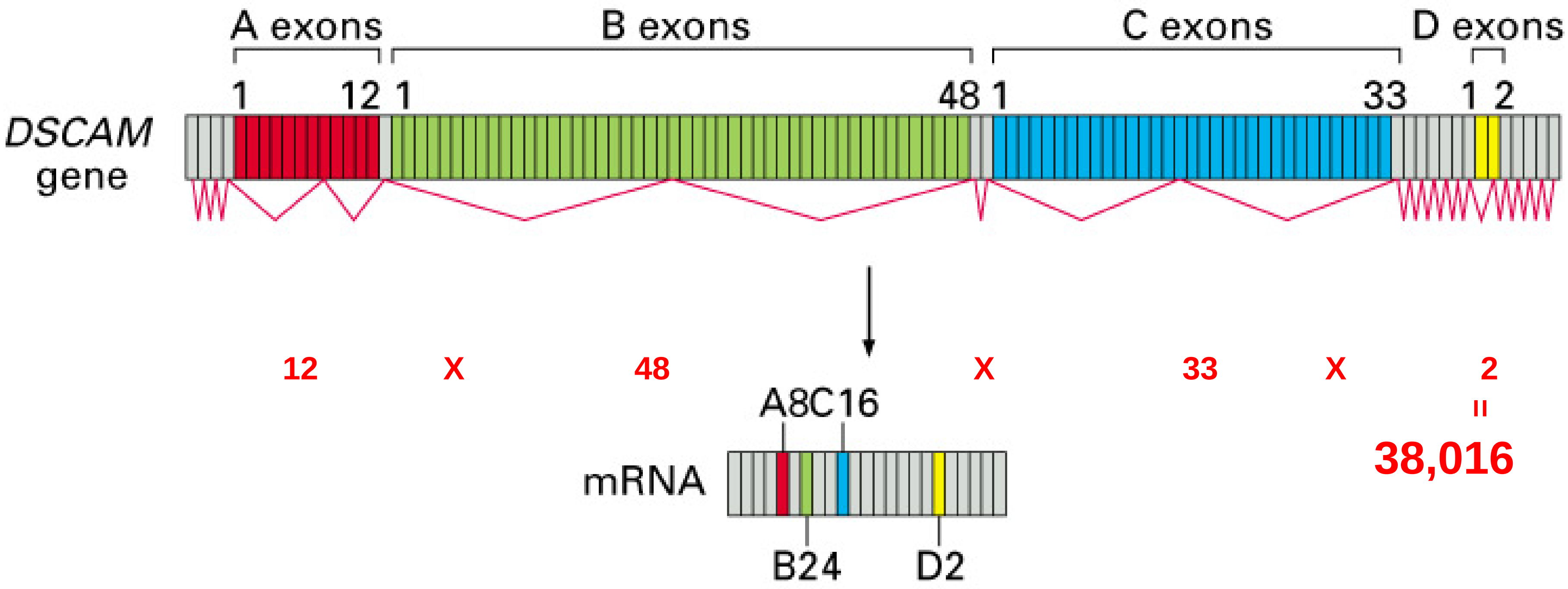
Nature Reviews | Genetics

Wang and Cooper, *Nature Reviews Genetics* 8, 749-761

Alternative Splicing



DSCAM alternative splicing



12 X 48 X 33 X 2

=

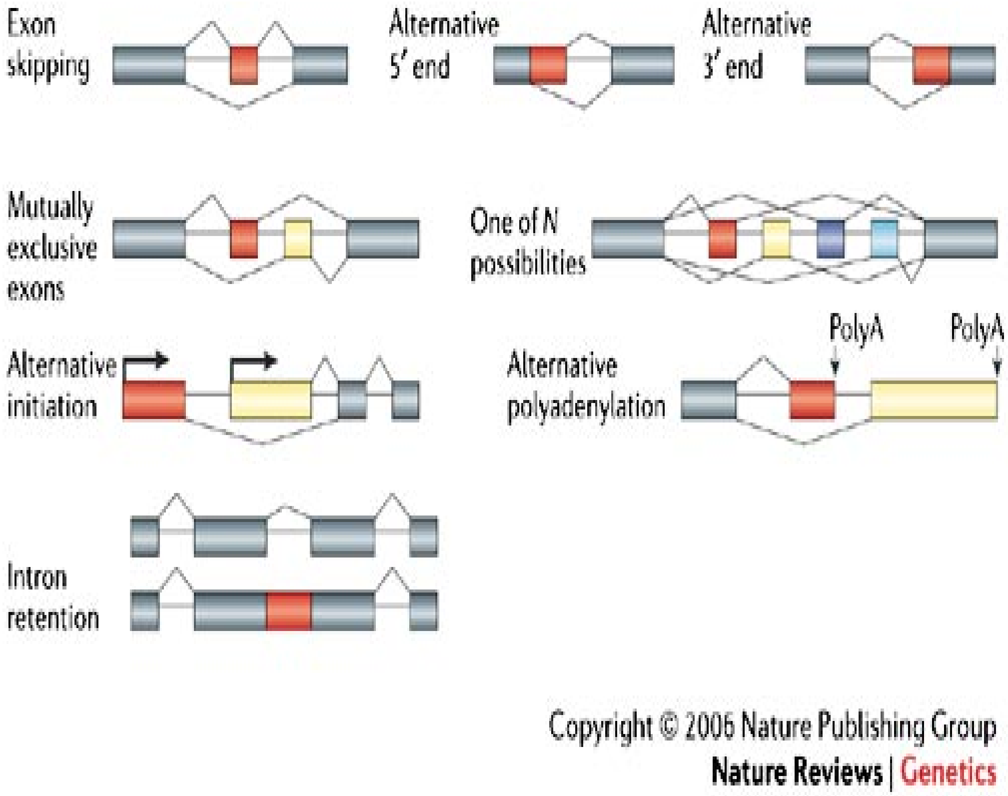
38,016

one out of 38,016 possible splicing patterns

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

Importance of Alternative Splicing

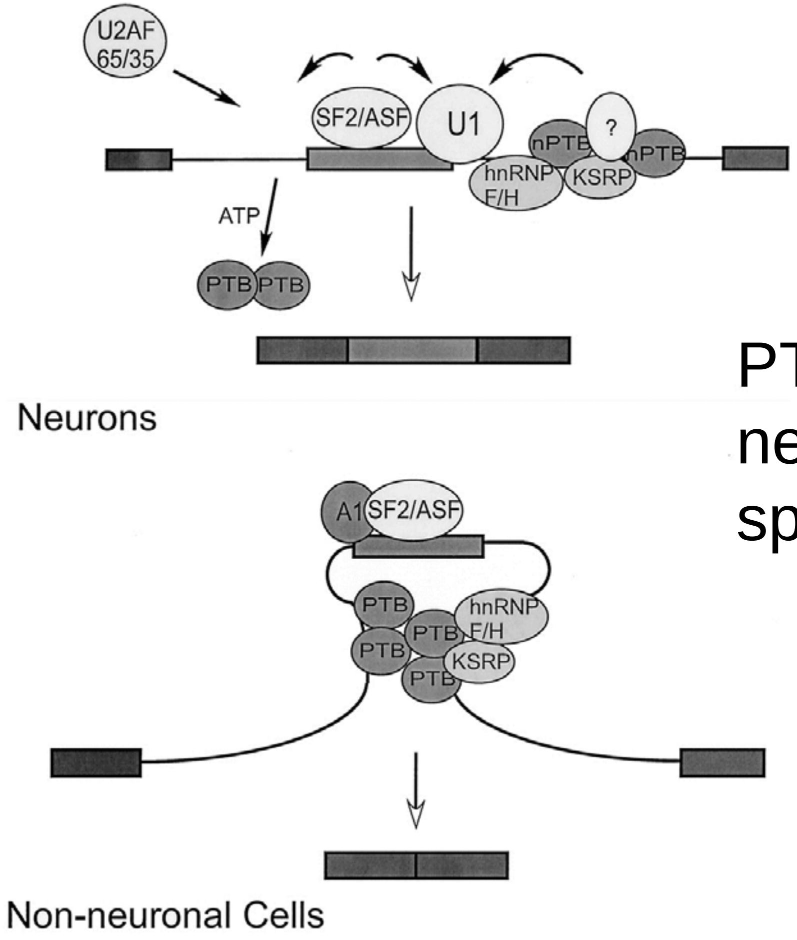
- >90% of human multi-exon genes undergo alternative splicing.
- Important in regulation of gene function.
- Aberrant splicing is a major cause of human diseases [1].
- An important mechanism for acquisition of evolutionary novelties [2-3].



Nature Reviews | Genetics

1. Xing and Lee, *Nature Reviews Genetics*, 2006, 7: 499-510.
2. Xing and Lee, *PNAS*, 2005, 102(38): 13526 - 13531.
3. Calarco*, Xing*, Caceres*, et al, *Genes & Dev*, 2007, 21:2963-2975.

Control of Alternative Splicing by Tissue-specific Splicing Factors

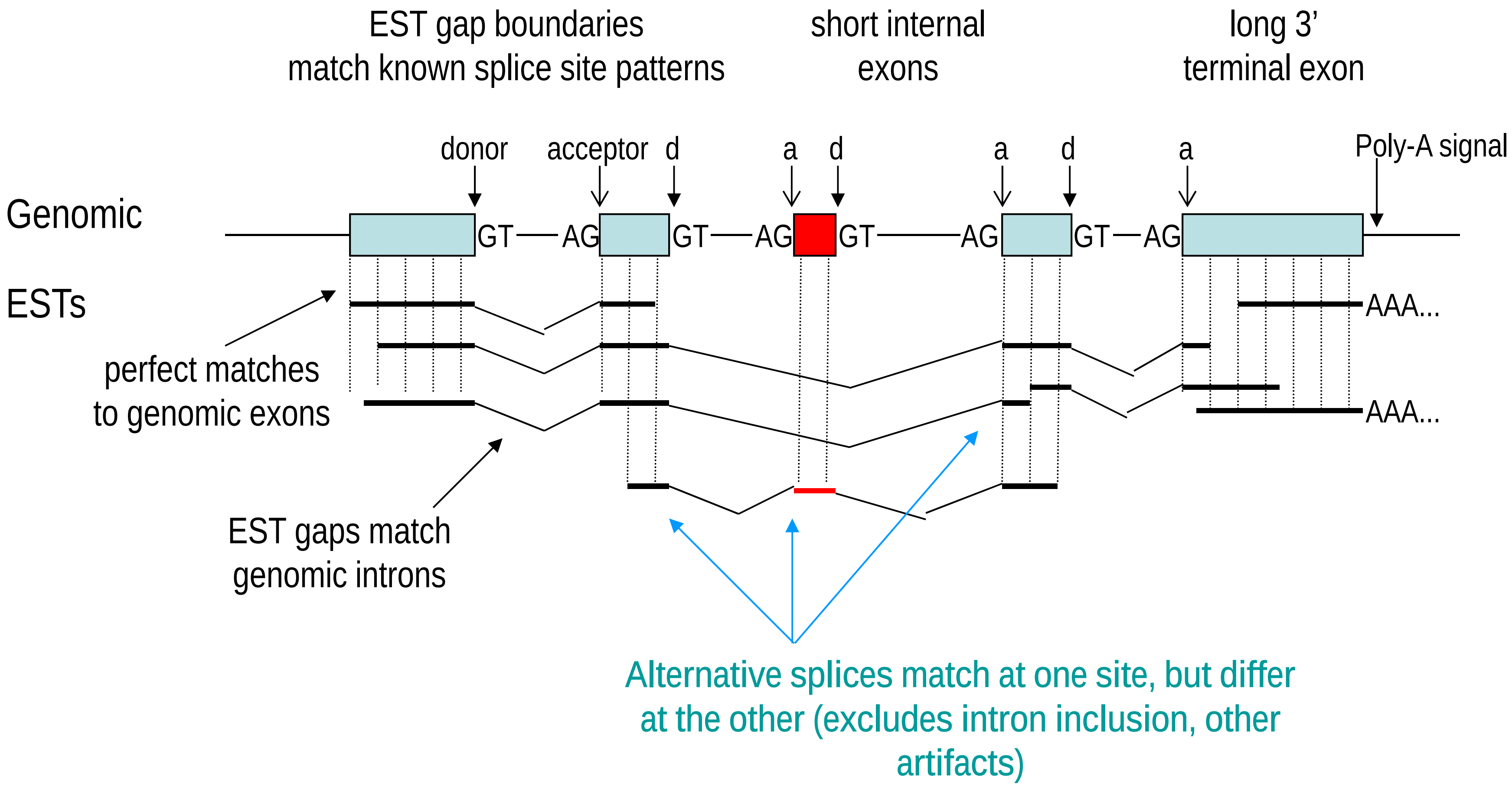


PTB: a switch for neuronal-specific splicing

Black D, Annu Rev Biochem. 2003;72:291-336.

Boutz P, et.al. Genes Dev. 2007, 21(13):1636-52.

EST analysis: first wave of alternative splicing discovery



EST gap boundaries match known splice site patterns

short internal exons

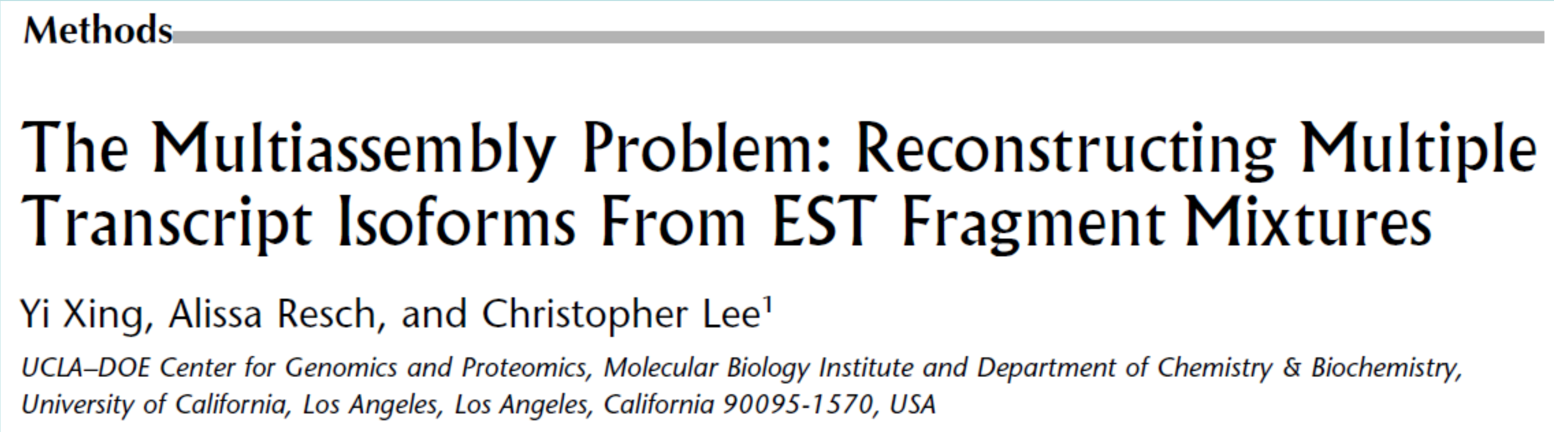
long 3’ terminal exon

perfect matches to genomic exons

EST gaps match genomic introns

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

Modrek & Lee, *Nature Genetics* 30:13-9 (2002).



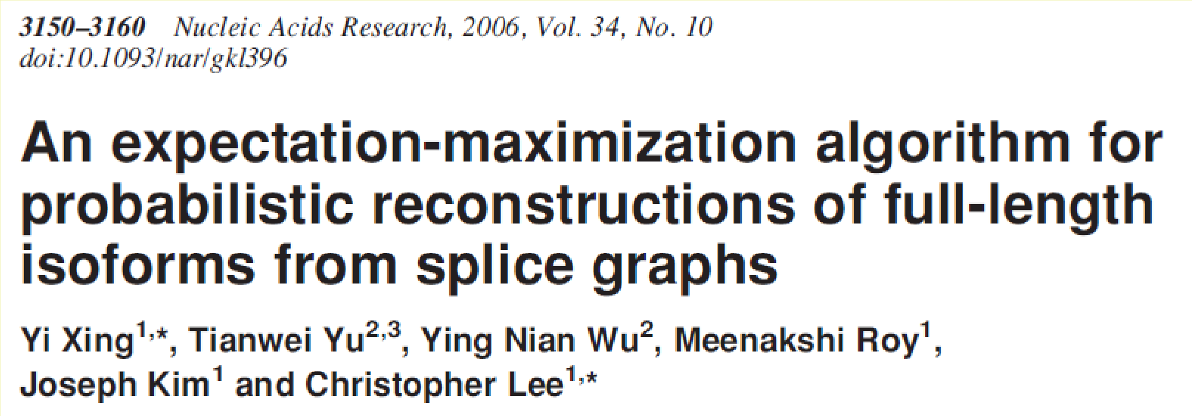
Methods

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

Yi Xing, Alissa Resch, and Christopher Lee[1]

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

Genome Research, 2004

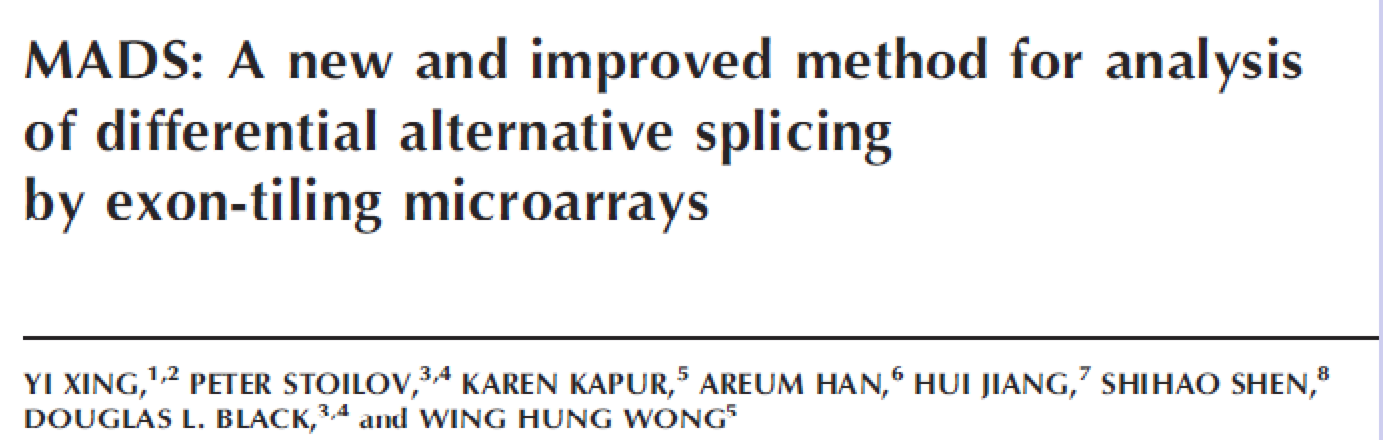


3150–3160 *Nucleic Acids Research, 2006, Vol. 34, No. 10*
doi:10.1093/nar/gkl396

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

Yi Xing[1,*], Tianwei Yu[2,3], Ying Nian Wu[2], Meenakshi Roy[1], Joseph Kim[1] and Christopher Lee[1,*]

Nucleic Acids Research, 2006



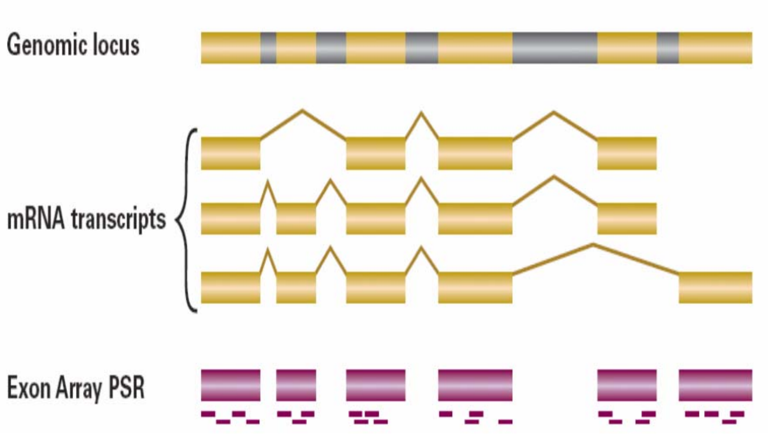
MADS: A new and improved method for analysis of differential alternative splicing by exon-tiling microarrays

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 WONG[5]

RNA, 2008

Genomic Approaches for Global Analysis of Alternative Splicing

High-density Exon Array



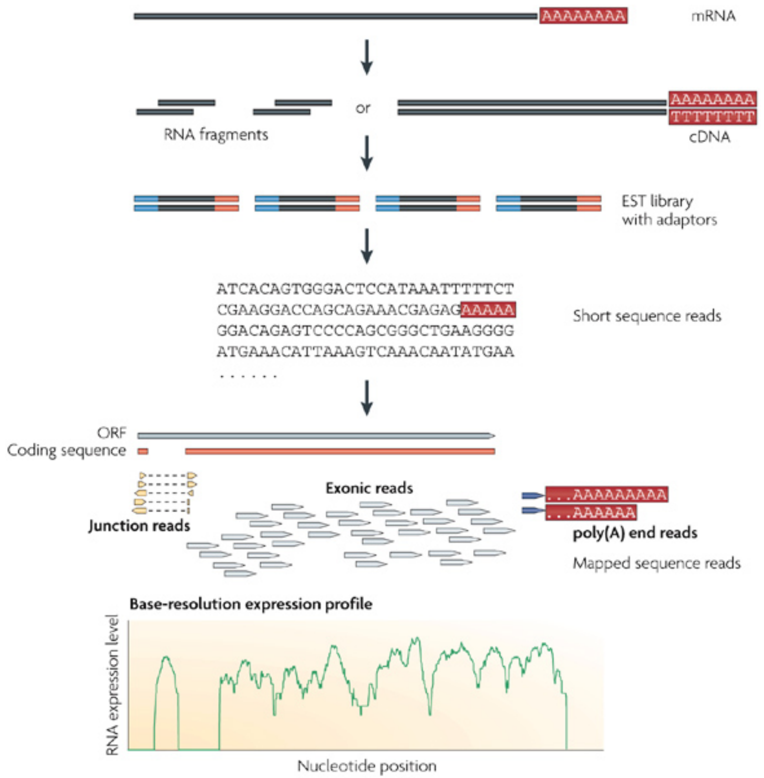
1 gene --- many probesets

Probes from each putative exon

1.4 Million probesets, >6 M probes

1. Kapur, Xing, Wong, *Genome Biology*, 8:R82, 2007
2. Xing, Kapur, Wong WH, *PLoS ONE*, 20;1:e88, 2006
3. Kapur, Jiang, Xing, Wong, *Bioinformatics*, 24:2887-2893, 2008
4. Xing et.al., *RNA*, 14(8): 1470-1479, 2008
5. Shen et.al., *Bioinformatics,* 26:268-269, 2010

Ultra-deep RNA Sequencing



Wang et al., *Nat Rev Genet*. 10(1):57-63.

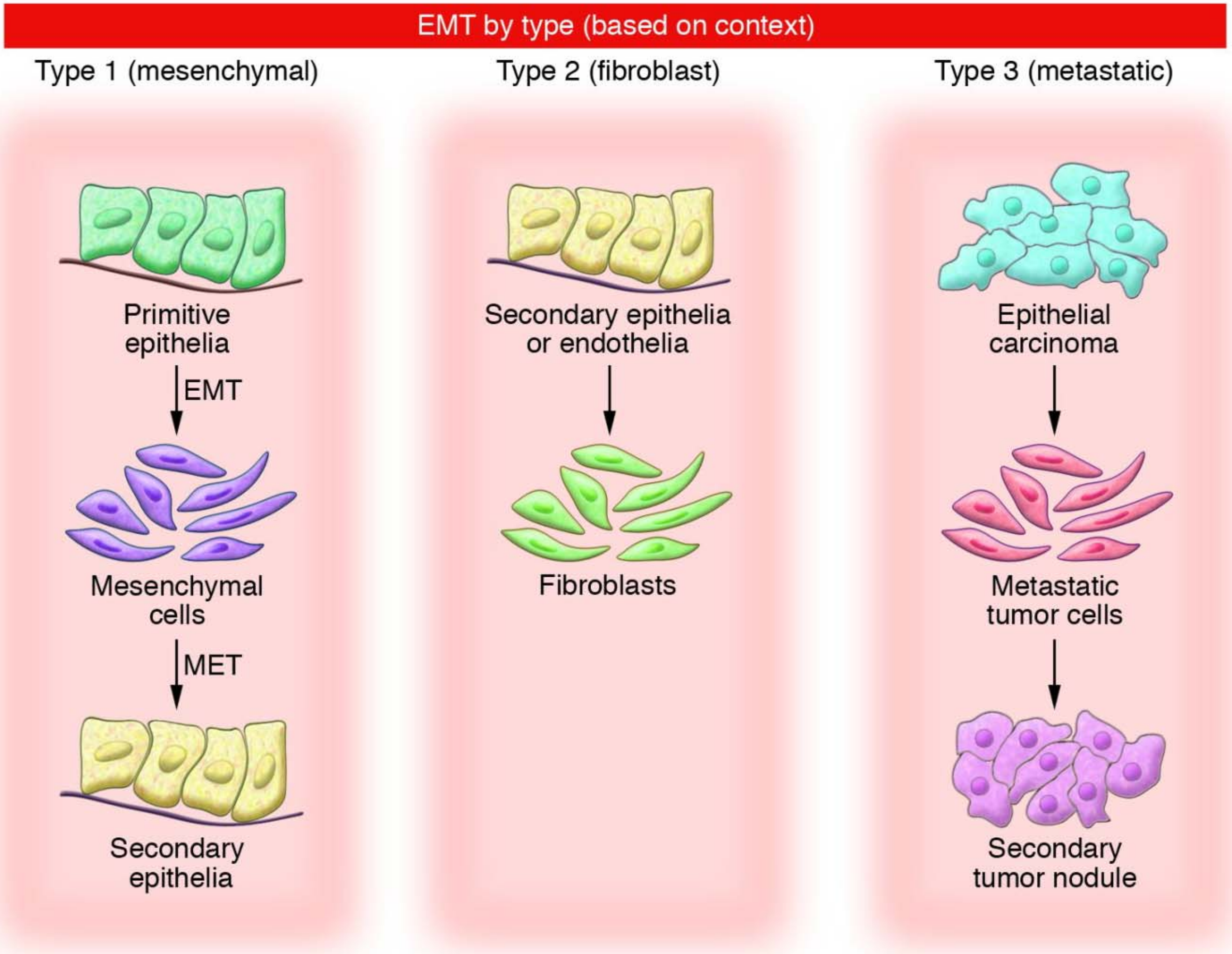
Nature Reviews | Genetics

1. Xing, Resch, Lee, *Genome Research*, 14:426-41, 2004
2. Xing et al., *Nucleic Acids Research*, 34:3150-60, 2006
3. Au et al., *Nucleic Acids Research*, 38:4570-8, 2010
4. Shen*, Lin* et al., *PNAS*, 108:2837-42, 2011
5. Shen et al., *Nucleic Acids Research*, in revision

Outline

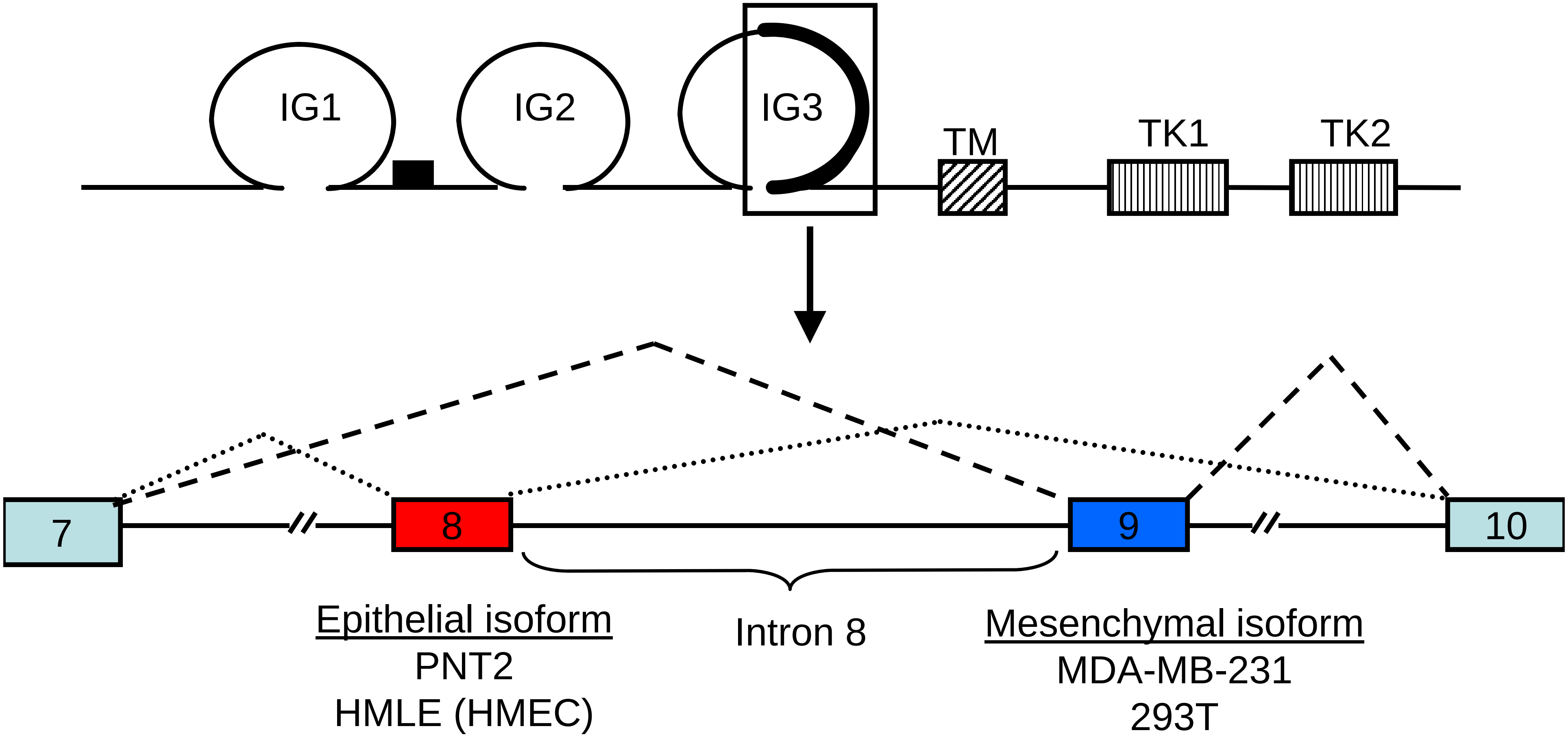
- Background on RNA splicing and alternative splicing.
- Alternative splicing network during the Epithelial Mesenchymal Transition.
- Evolution of new exons in primates.

The Epithelial to Mesenchymal Transition (EMT): Roles in development, fibrosis and metastasis



Zeisberg M., and Neilson, E.G. (2009) *JCI* 119:1429

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



Epithelial isoform
PNT2
HMLE (HMEC)

Mesenchymal isoform
MDA-MB-231
293T

Ligand Binding Specificity:

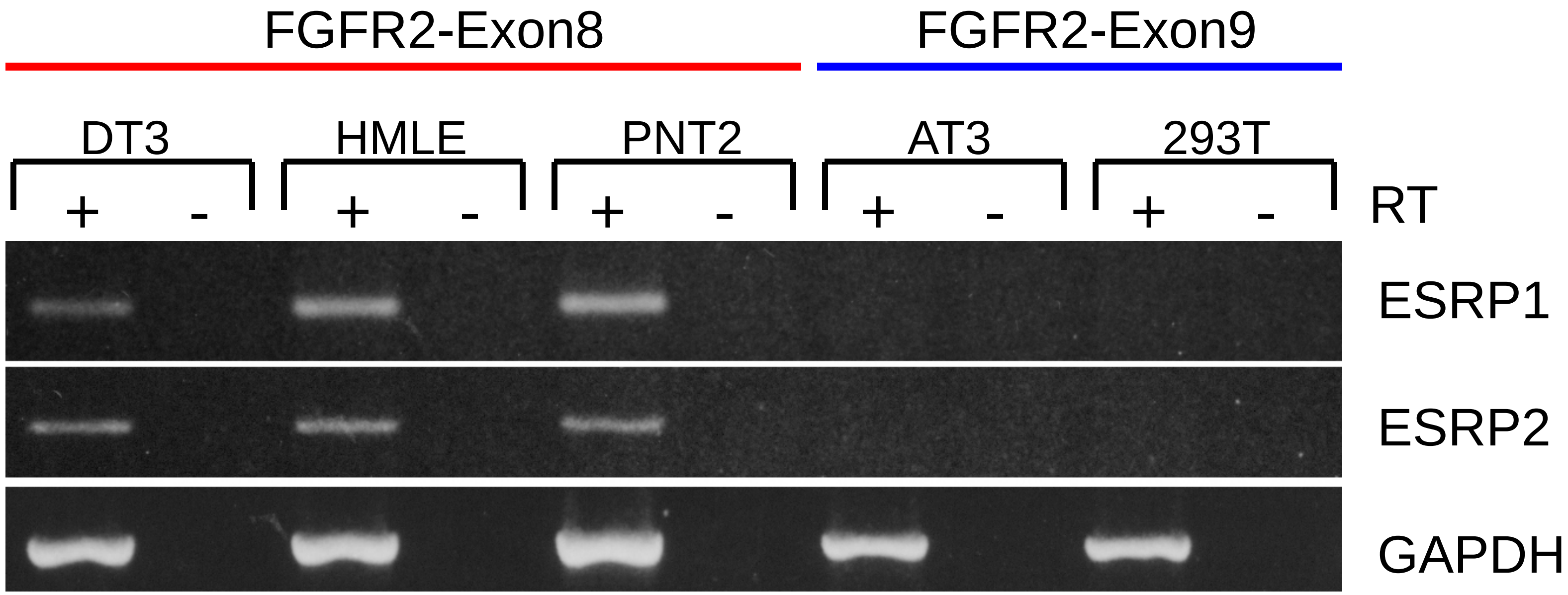
FGFR2-E8: FGF-3, 7, 10, 22

FGFR2-E9: FGF-2, 4, 5, 6, 8, 9, 17

ESRP – A master splicing switch of epithelial-mesenchymal transition

Russ
Carstens
(Penn)

- 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

Warzecha, Sato, Nabet, Hogenesch, and Carstens. *Molecular Cell*, 33(5): 591-601

PNT2: Human Prostate
Epithelial Cells

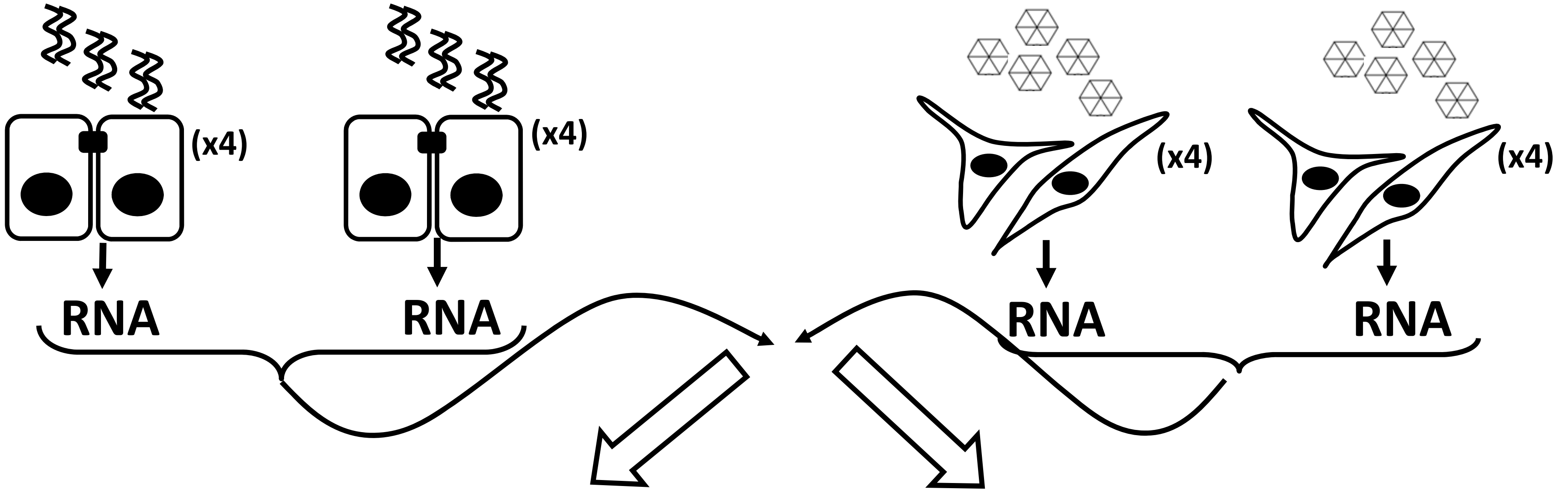
MDA-MB-231: Human Breast Cancer
Mesenchymal Cells

siRNA: GFP

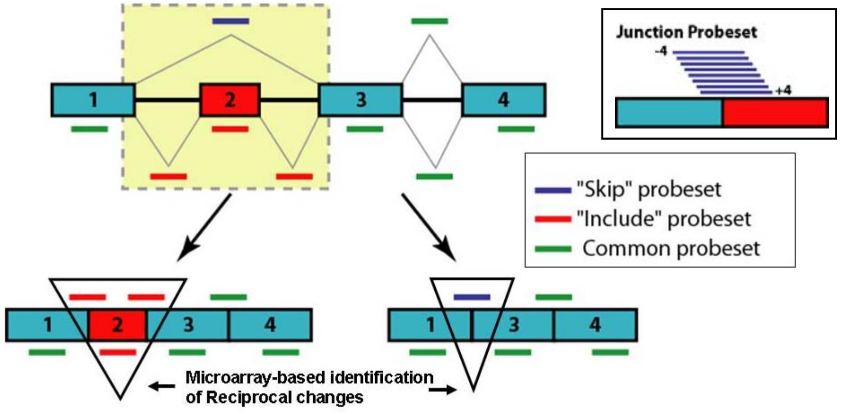
siRNA: ESRP1 and ESRP2

Retrovirus: EGFP

Retrovirus: mEsrp1



Affymetrix HJAY exon junction array



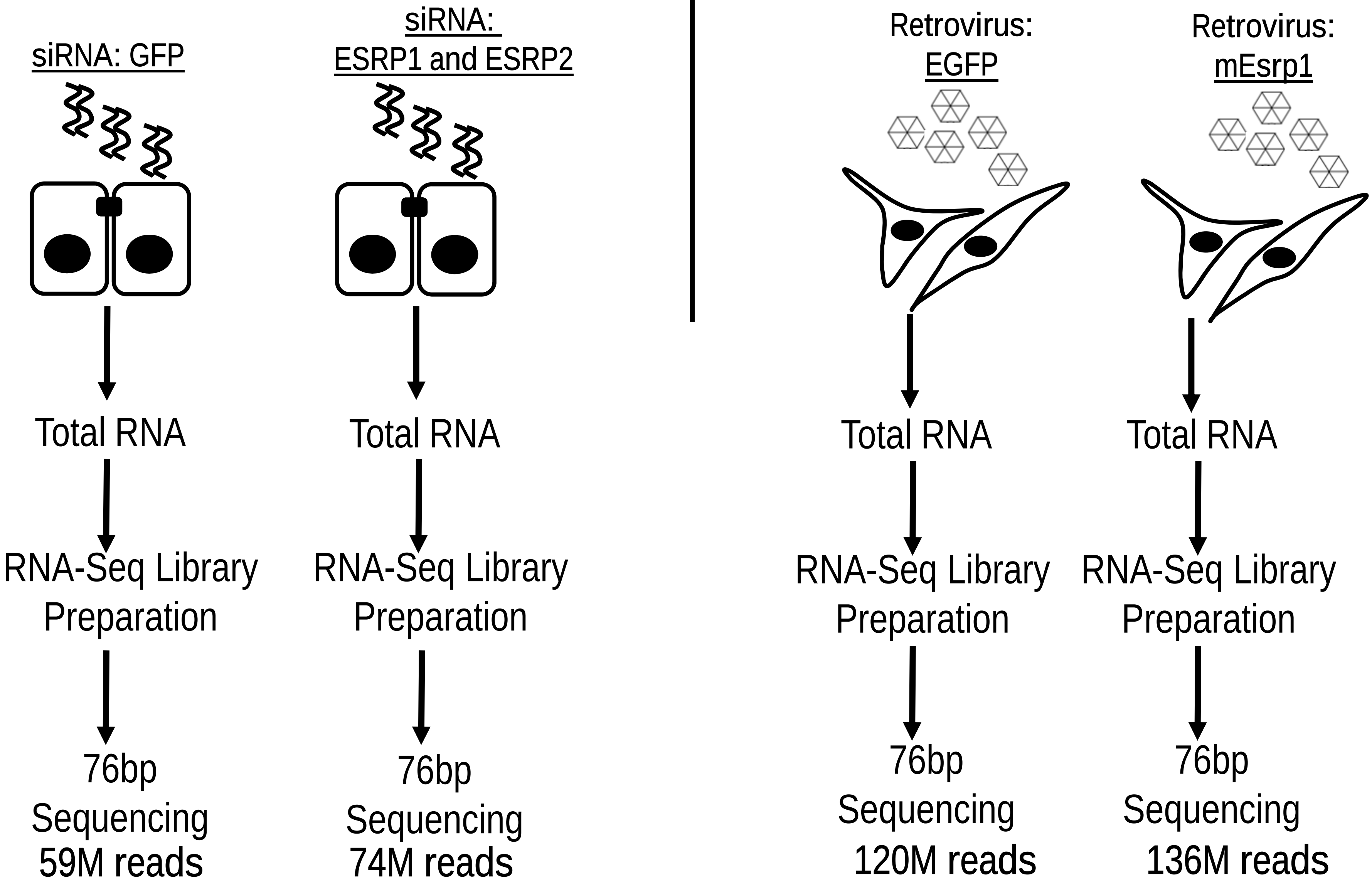
RNA-Seq (Illumina)

Warzecha et al., *EMBO J*, 2010.

Genome-wide discovery of ESRP targets using RNA-Seq

PNT2: Human Prostate Epithealial Cells

MDA-MB-231: Human Breast Cancer Mesenchymal Cells



<u>siRNA: GFP</u>

Total RNA

RNA-Seq Library Preparation

76bp Sequencing
59M reads

<u>siRNA: ESRP1 and ESRP2</u>

Total RNA

RNA-Seq Library Preparation

76bp Sequencing
74M reads

Retrovirus: <u>EGFP</u>

Total RNA

RNA-Seq Library Preparation

76bp Sequencing
120M reads

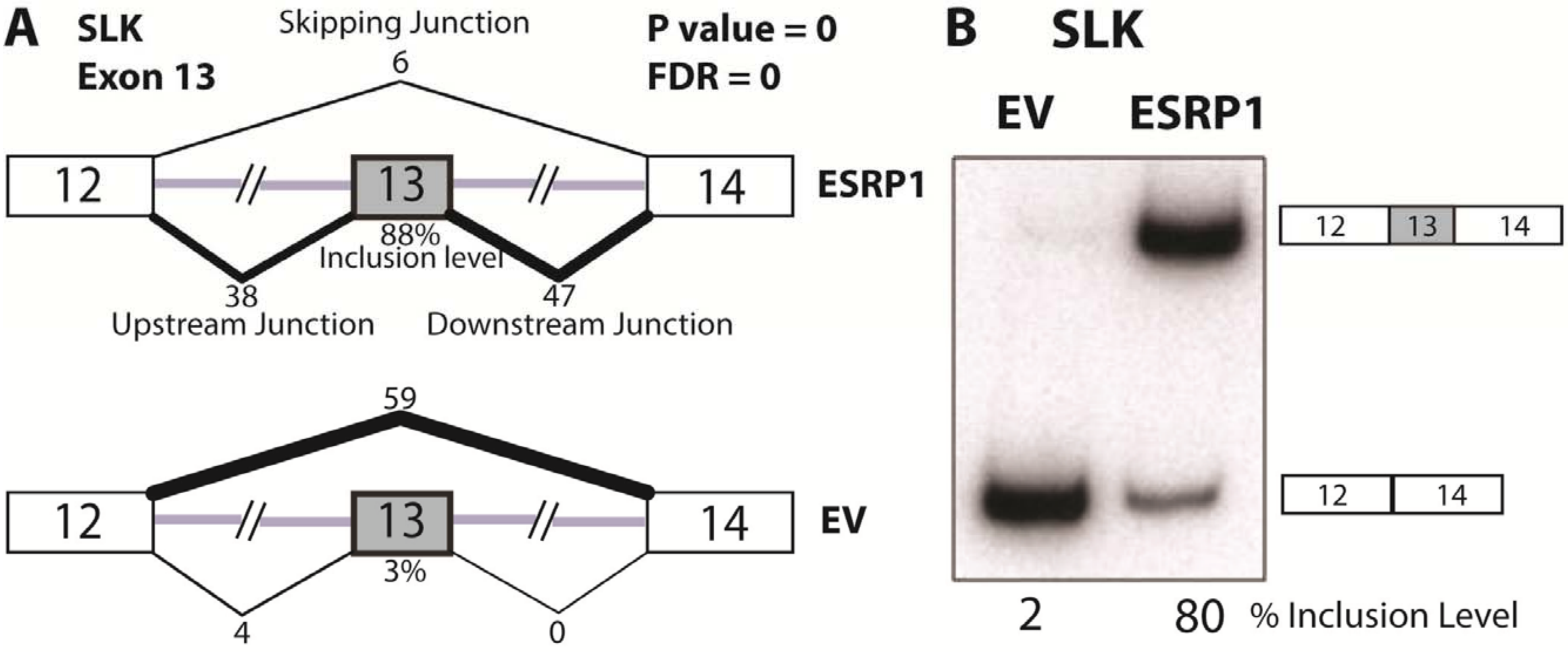
Retrovirus: <u>mEsrp1</u>

Total RNA

RNA-Seq Library Preparation

76bp Sequencing
136M reads

Discovery of ESRP Targets by RNA-Seq



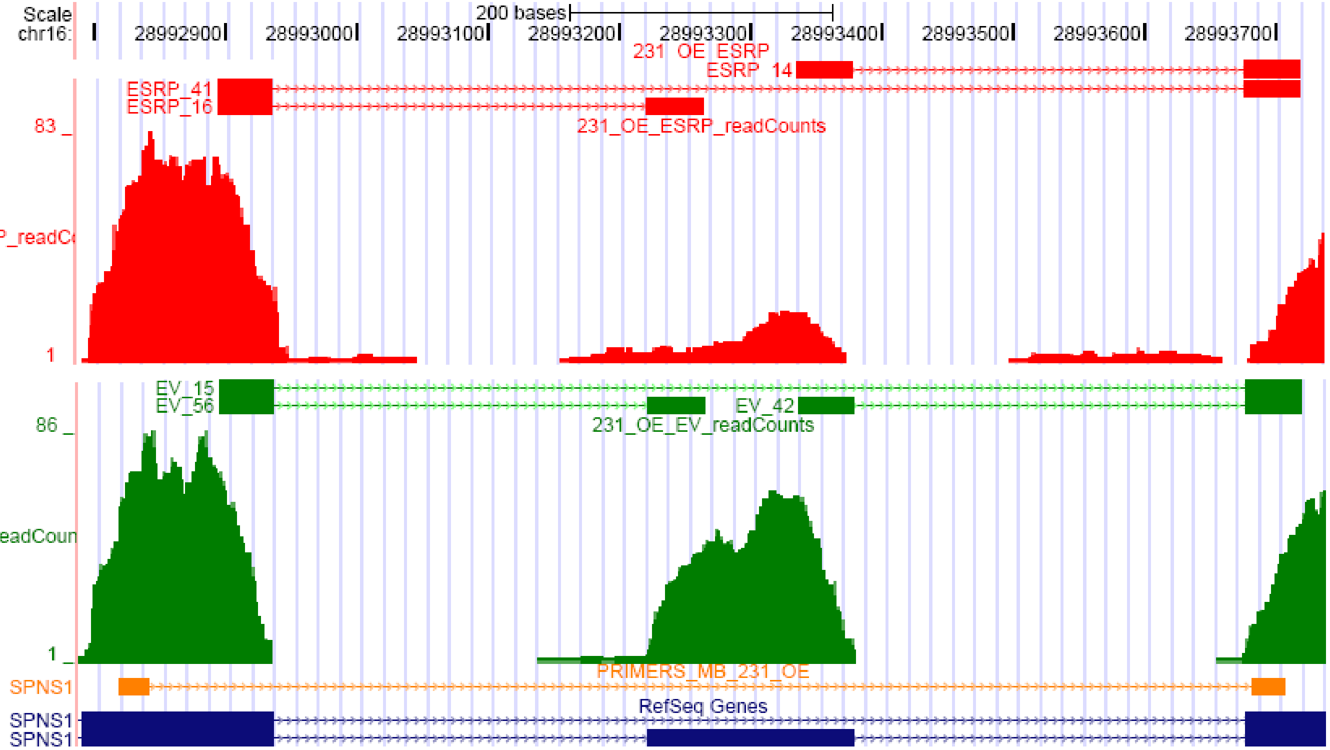
Shihao Shen, *MATS: Multivariate Analysis of Transcript Splicing*

Discovery of Novel ESRP Targets by RNA-Seq

SPNS1

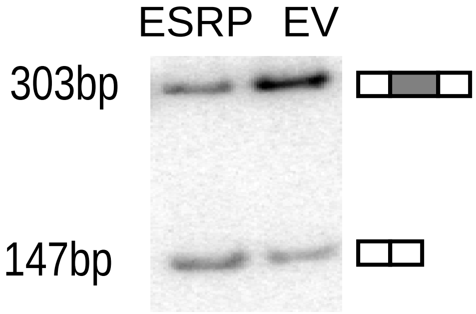
ESRP (136M)

EV (120M)

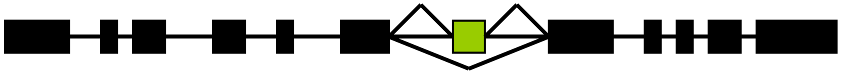


UJC: 16
DJC: 14
SJC: 41
incLvl: 27%

UJC: 56
DJC: 42
SJC: 15
incLvl: 76%



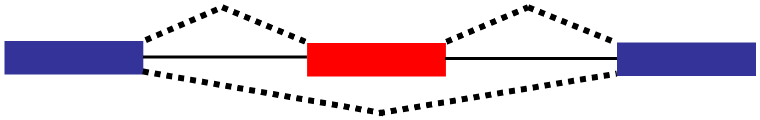
Exon Inclusion Level	ESRP	EV	ESRP- EV
RNA-Seq	0.27	0.76	- 0.49
RT-PCR	0.36	0.75	- 0.39



Overall validation rate: **86%** (115 out of 134)

RNA-Seq Validation Summary

Cassette



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)

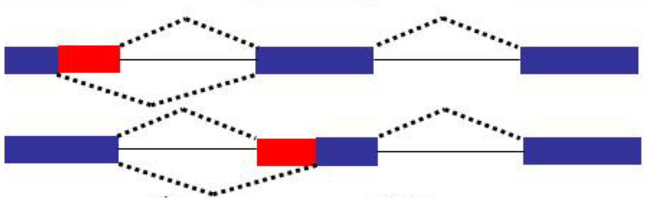
PNT2
(siRNAs vs. ESRP1/2)

35 predicted

RT-PCR Validations :

13/18 (72.2%) >5%
10/18 (55.6%) >10%
(+12 previously validated from HJAY)

Alt. 3’ or 5’ ss



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.

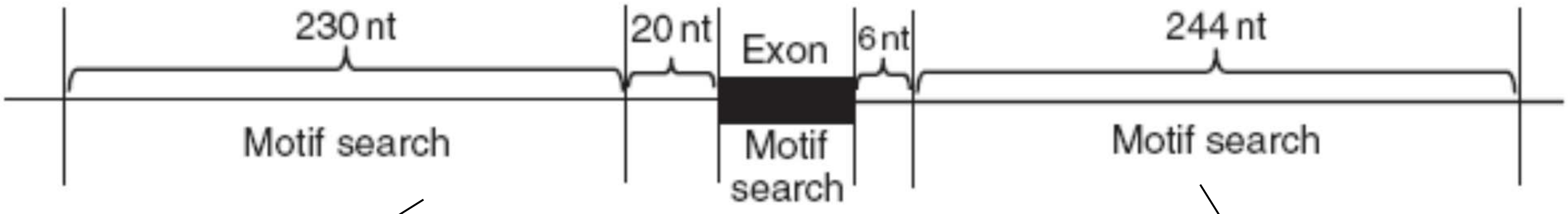
Enriched RNA Motifs Around ESRP-Regulated Exons

ESRP Enhanced

Motif	p-value
CCCTCC	4.36E-20
TGCCGA	7.54E-18
ATGACT	1.42E-16
TAATTC	3.97E-16
TGCATG	8.34E-16

Motif	p-value
ACTCCG	1.31E-21
CCGATG	1.35E-17
CGATGT	4.14E-17
TAACCC	4.01E-15
GTCGAC	1.50E-14

Motif	p-value
GTGGTG	7.05E-65
GGTGGT	7.84E-44
TGGTGG	1.64E-37
GCTGTG	7.53E-35
TGGTGC	4.20E-33



ESRP Silenced

Motif	p-value
GGGGCT	3.90E-40
TCCCCT	2.98E-27
GGGCTG	1.21E-25
GTCCCC	6.86E-24
CCTGGC	2.44E-23

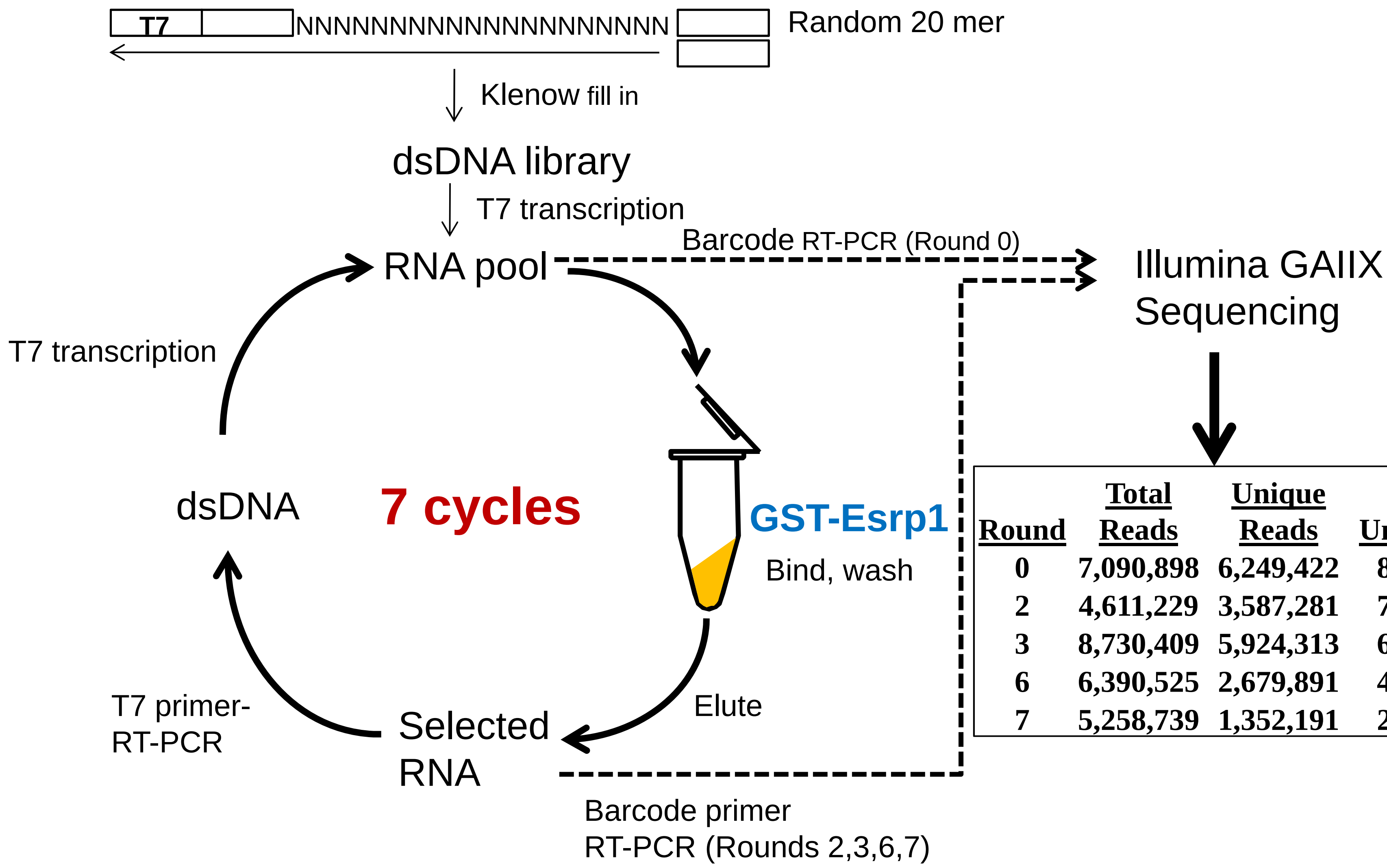
Motif	p-value
TGGTGG	2.14E-64
GGTGGT	2.38E-64
GGTGCT	1.04E-63
GTGGTG	6.79E-62
CTGGTG	9.42E-55

Motif	p-value
TGCATG	1.36E-46
GCAGCT	7.16E-30
GCATGG	2.72E-24
GCTTGG	1.20E-23
CCCTCA	7.14E-22

🟦 GT-rich motif

🟥 FOX-1/2 motif

Experimental determination/validation of a UGG-rich ESRP1 binding site by SELEX-Seq



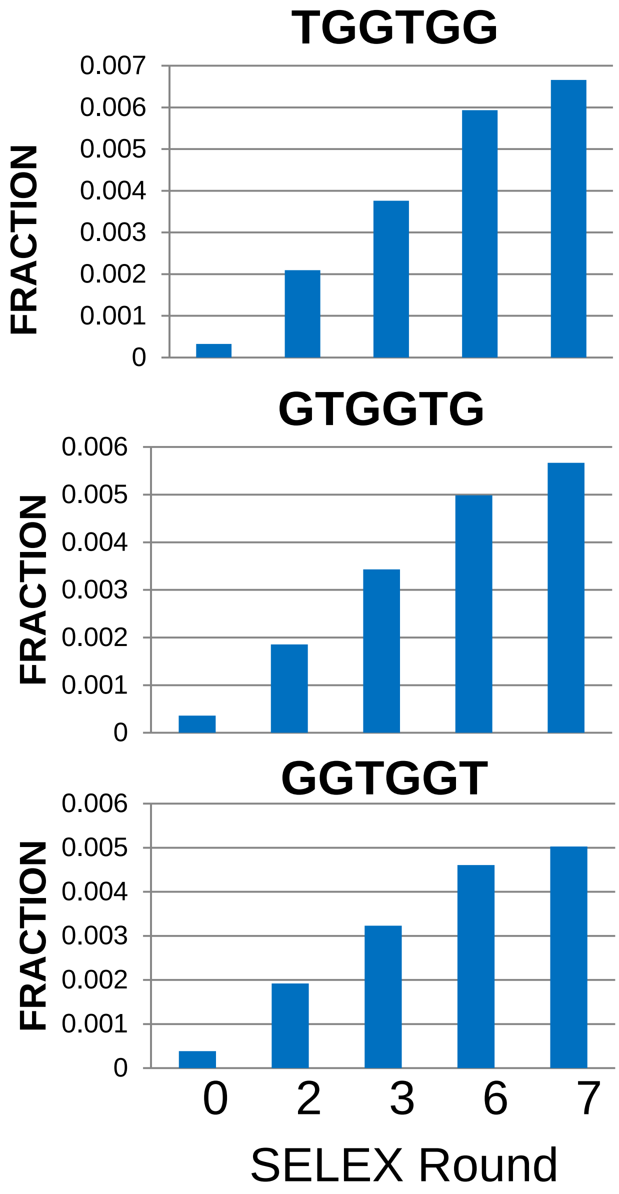
Round	Total Reads	Unique Reads	% Unique
0	7,090,898	6,249,422	88.1%
2	4,611,229	3,587,281	77.8%
3	8,730,409	5,924,313	67.9%
6	6,390,525	2,679,891	41.9%
7	5,258,739	1,352,191	25.7%

Systematic Evolution of Ligands by EXponential enrichment (SELEX)

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
GTGGGT
TGGGGT
GGGGGT
TGGGGG



Confirmed by gel mobility shift assay

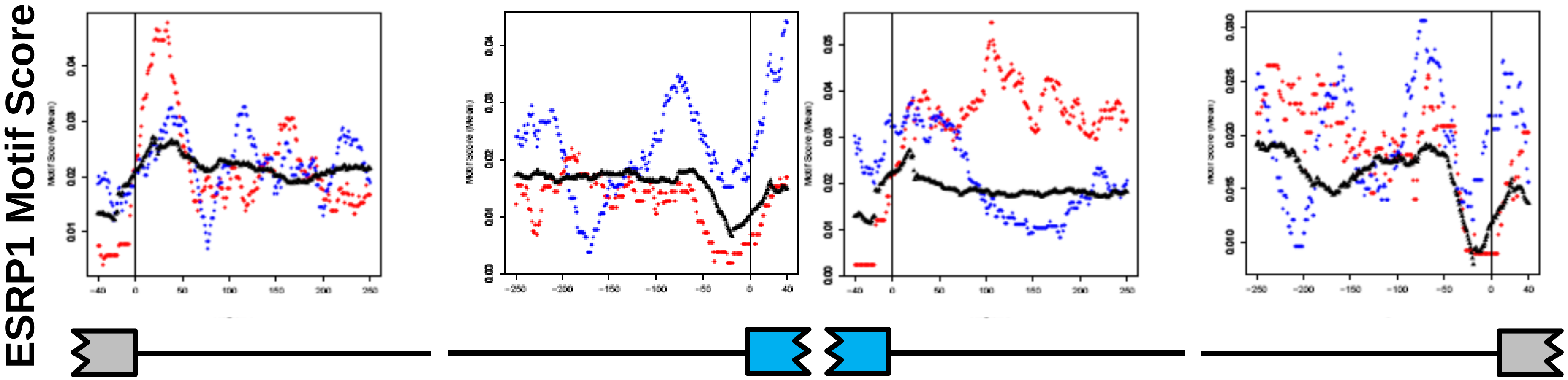
A SELEX-Seq motif score defines a position-dependent ESRP RNA map

● ESRP Enhanced exons (103)

◆ ESRP Silenced exons (173)

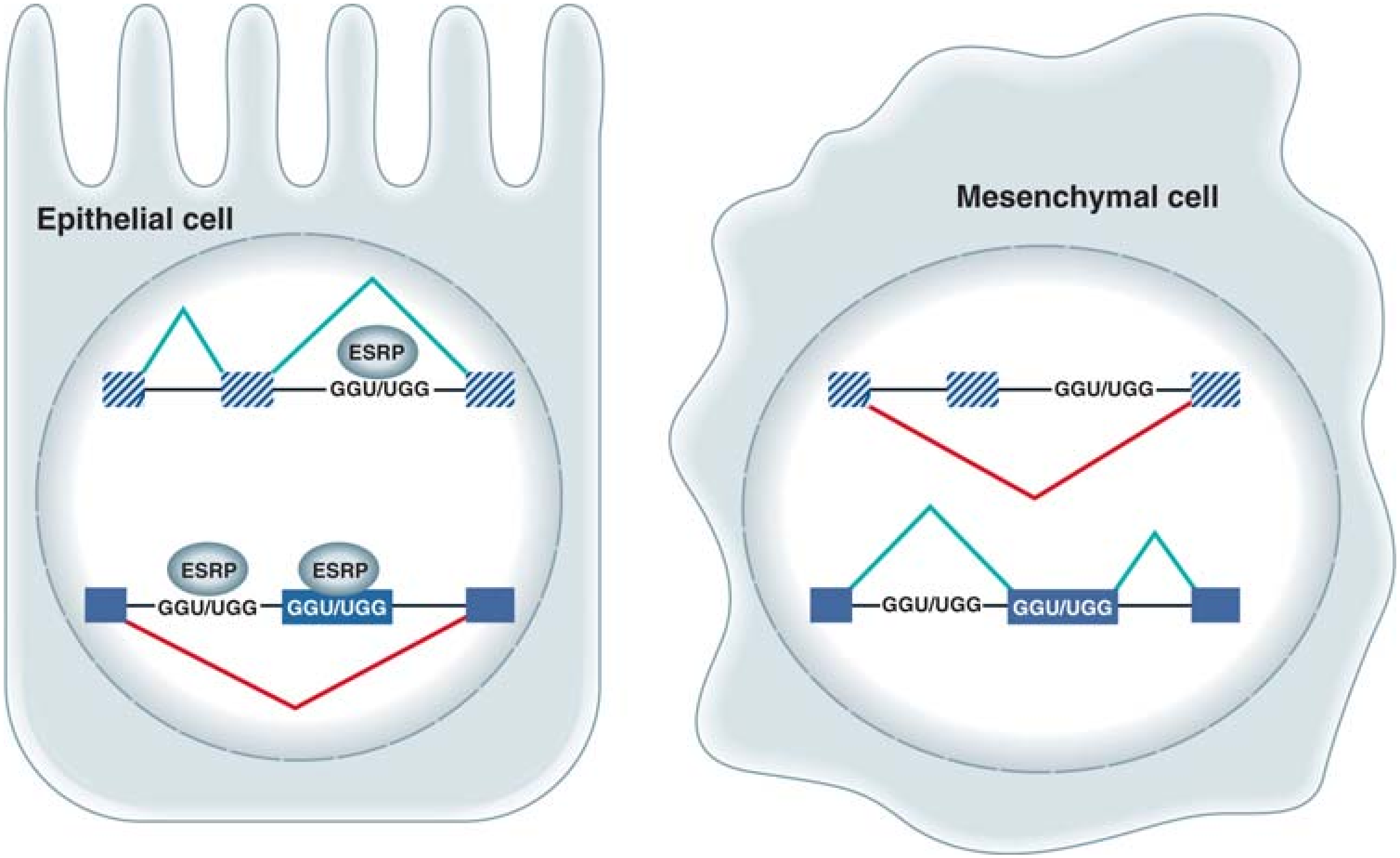
} RT-PCR Validated with >10% change

▲ HJAY array non-ESRP target background set (3508)



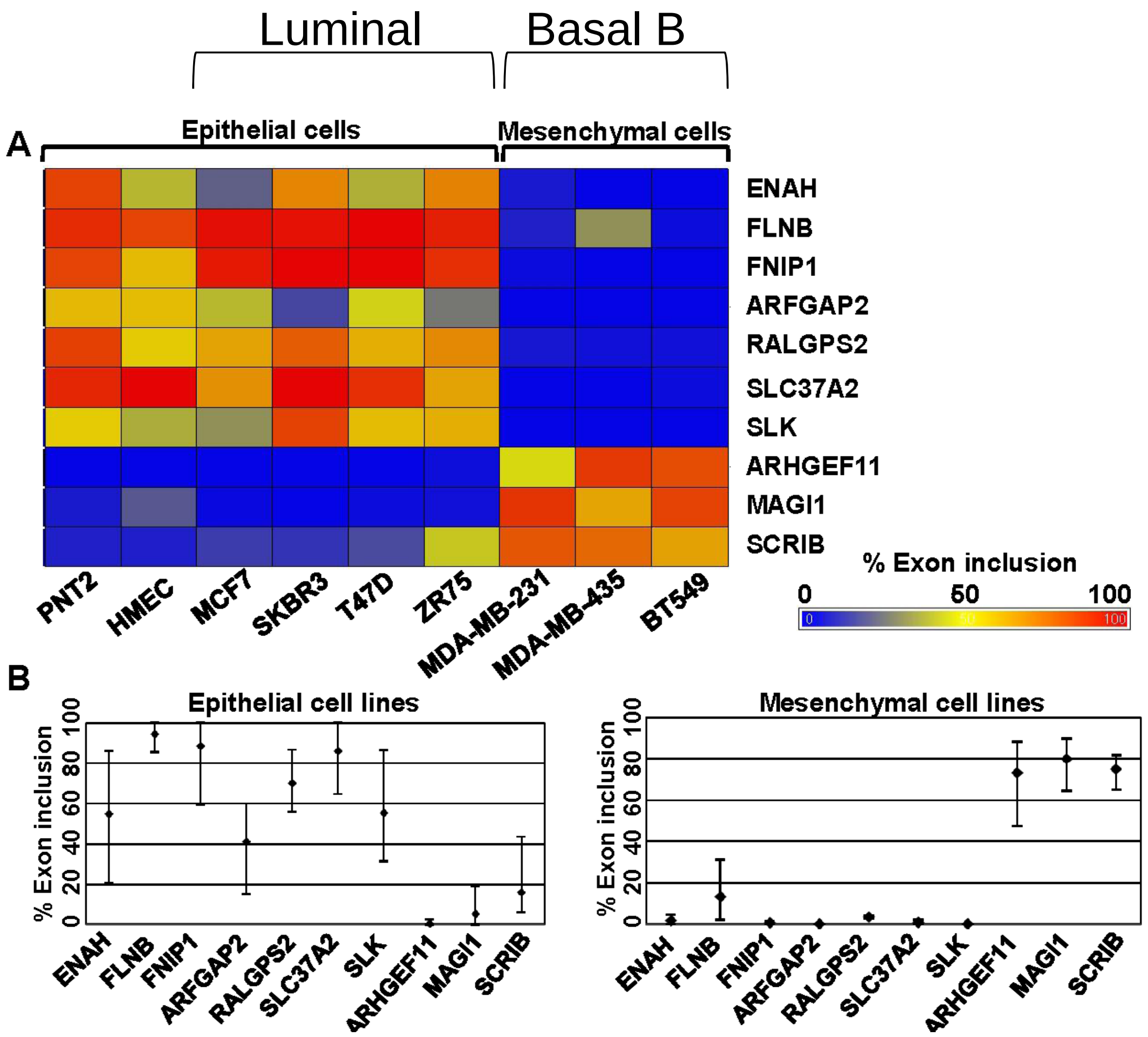
Scan window: 45nt and **<u>top 12</u>** SELEX-Seq motif-based score

“A Splicing Mastermind for EMT”



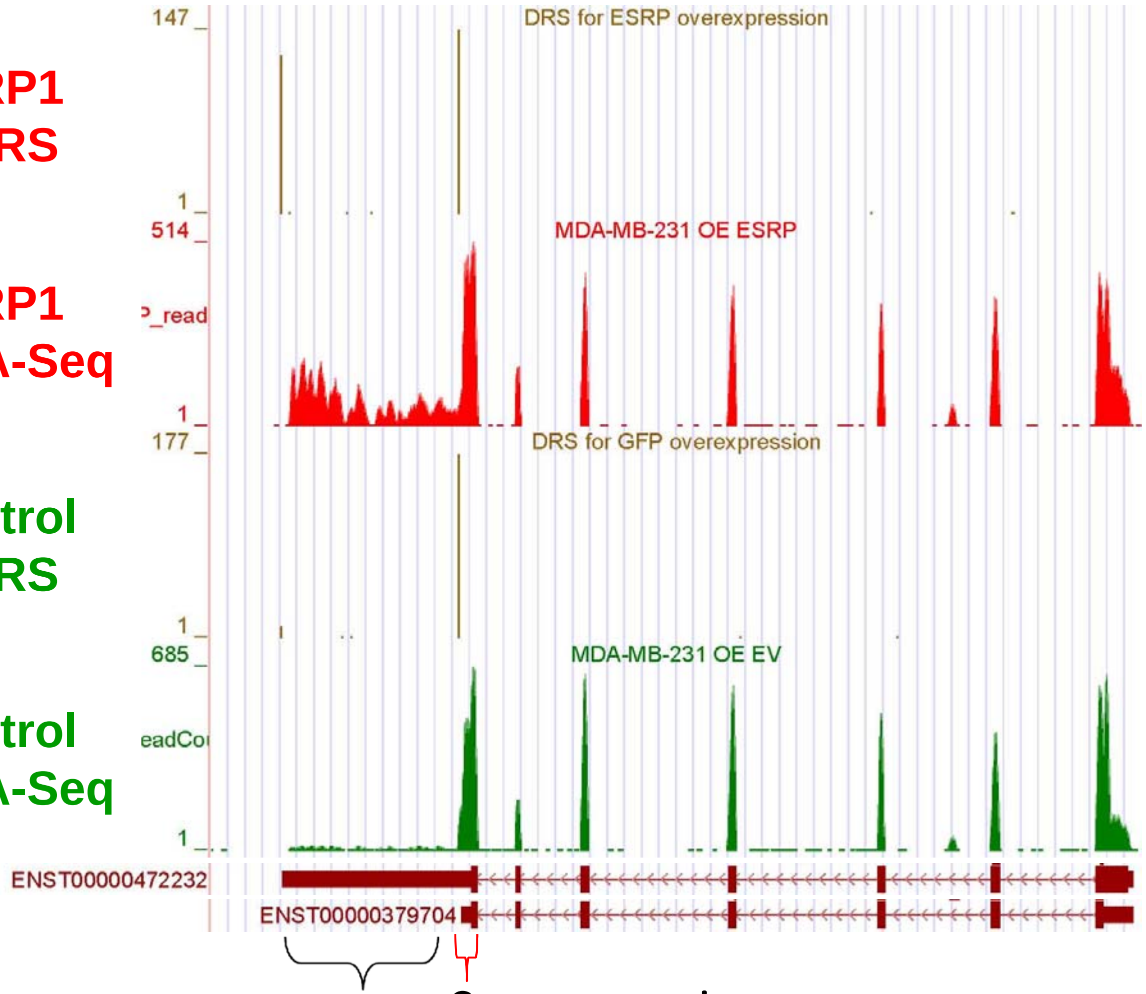
EMBO J, Tavanez JP, Valcárcel J. (2010) 29, 3217 - 3218

An ESRP splicing signature that distinguishes epithelial cells from mesenchymal cells



The ESRPs regulate alternative polyadenylation (APA)

BCL2-associated athanogene (BAG1)



ESRP1
3’ DRS

ESRP1
RNA-Seq

Control
3’ DRS

Control
RNA-Seq

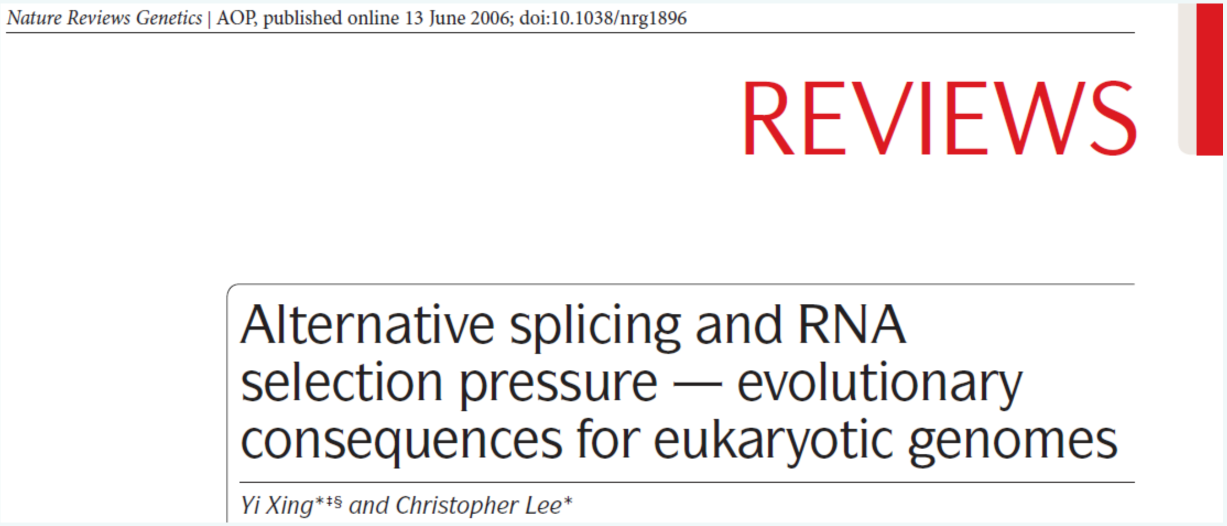
Extended UTR

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.



Nature Reviews Genetics | AOP, published online 13 June 2006; doi:10.1038/nrg1896

REVIEWS

Alternative splicing and RNA selection pressure — evolutionary consequences for eukaryotic genomes

Yi Xing[‡§] and Christopher Lee[*]*

Nature Reviews Genetics, 2006

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. Blencowe[1,2,9]

Genes and Development, 2007

Human Molecular Genetics, 2010, Vol. 19, No. 15 ***2958–2973***
doi:10.1093/hmg/ddq201
Advance Access published on May 11, 2010

Evolution of alternative splicing in primate brain transcriptomes

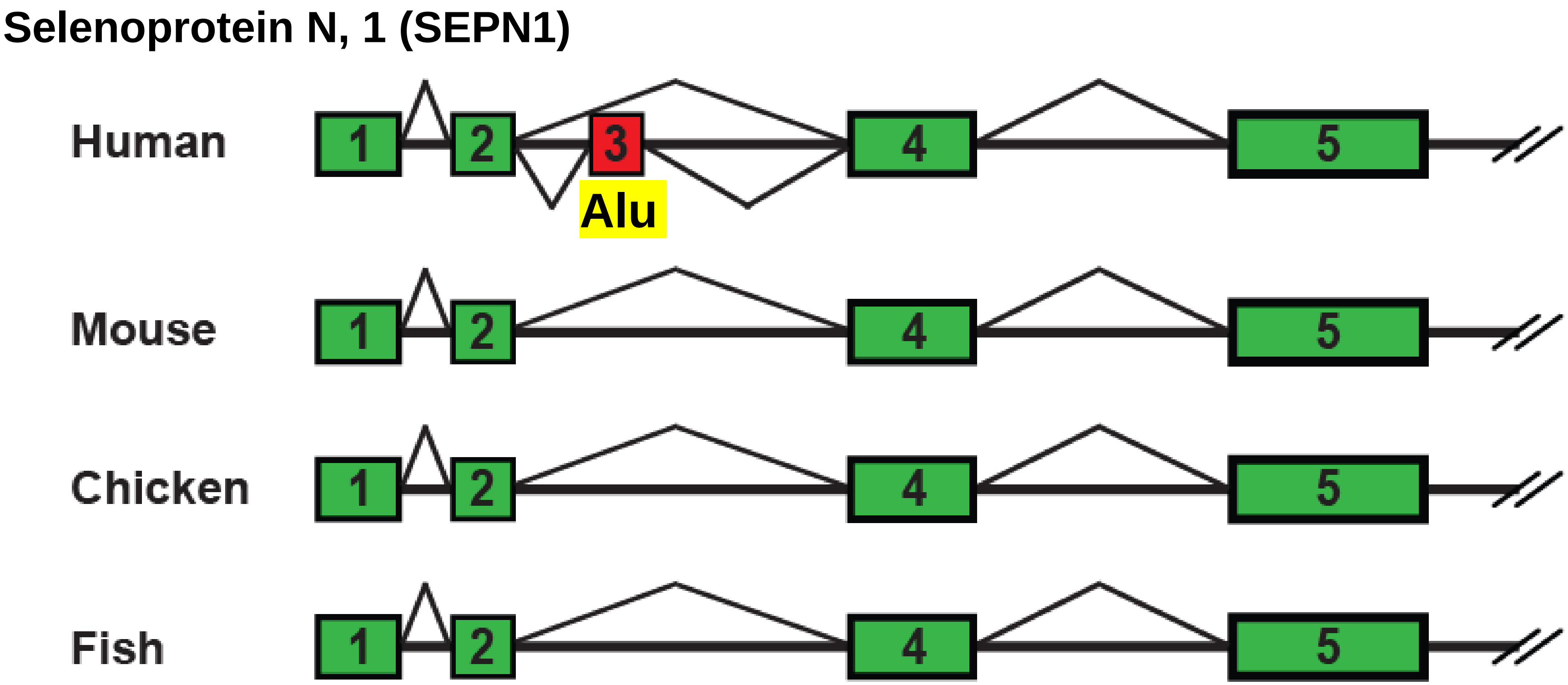
Lan Lin[1,†,‡], Shihao Shen[2,†,‡], Peng Jiang[1], Seiko Sato[1], Beverly L. Davidson[1,3,4] and Yi Xing[1,2,5,*]

[1]Department of Internal Medicine, [2]Department of Biostatistics, [3]Department of Molecular Physiology and Biophysics, [4]Department of Neurology and [5]Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242, USA

Received February 27, 2010; Revised April 16, 2010; Accepted May 7, 2010

Human Molecular Genetics, 2010

Some Exons Are Unique to Humans



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

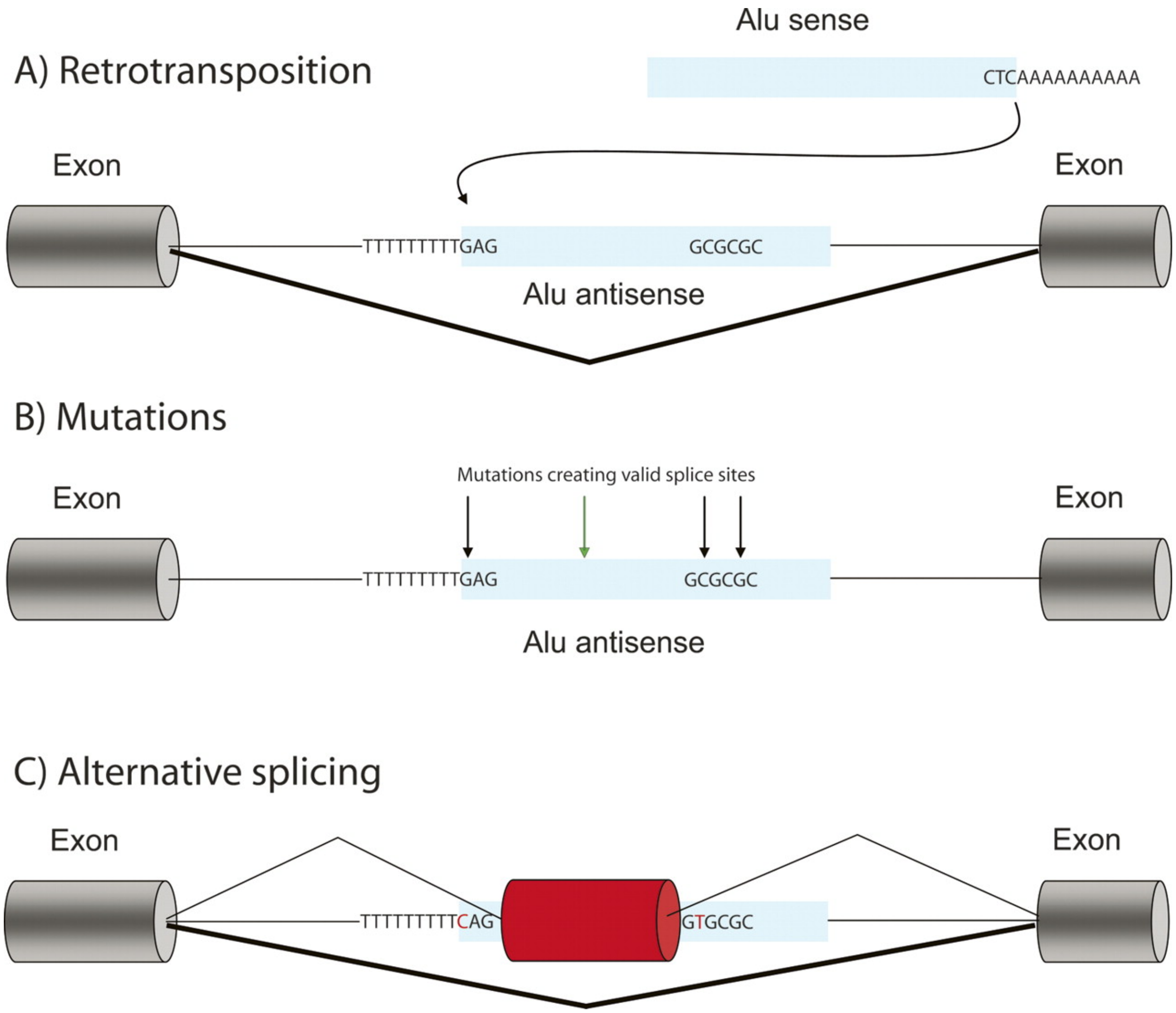
- Short interspersed nuclear elements (SINE) family

- Primate-specific transposable elements

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

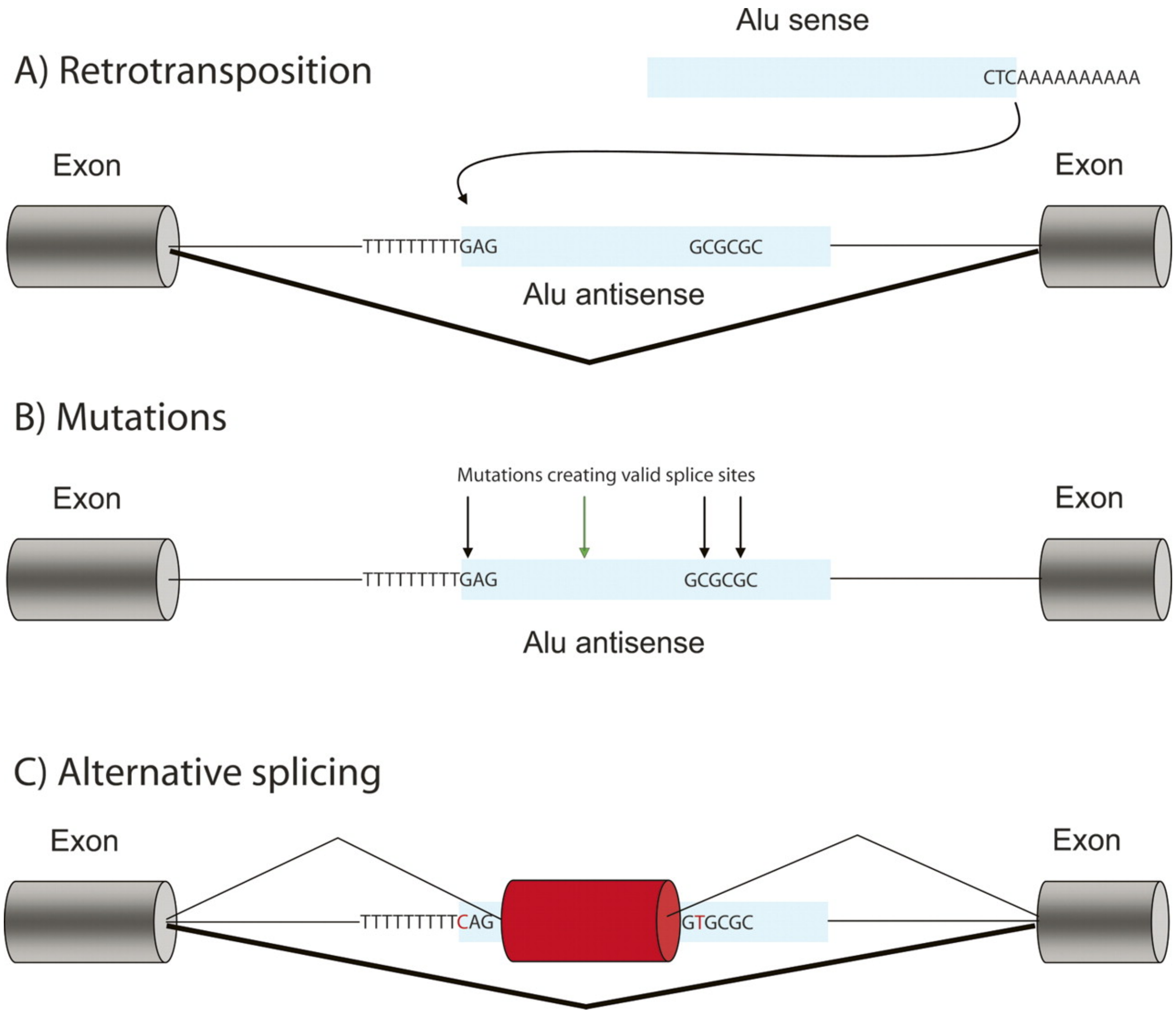
- The most abundant mobile elements in human genome
  - >1 million copies in human genome
  - 10% of the human genomic DNA

Alu exonization



Rotem Sorek, RNA, 2007

Alu exonization



Rotem Sorek, RNA, 2007

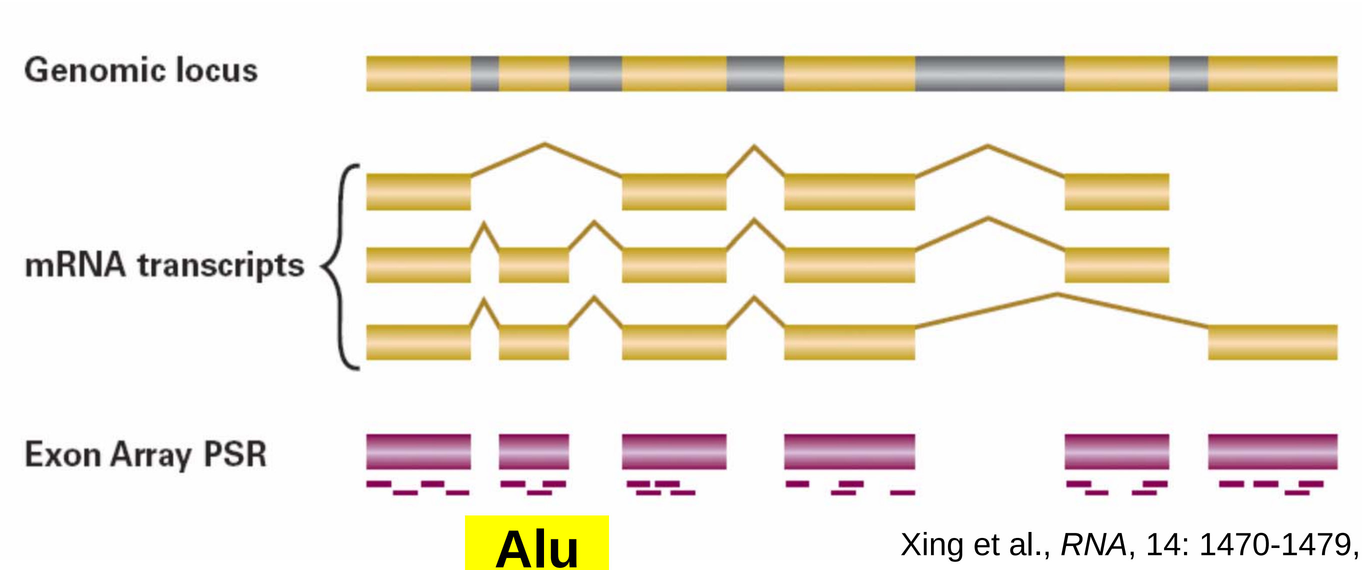
Alu exonization

- 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



Xing et al., *RNA*, 14: 1470-1479, 2008
Shen et al., *Bioinformatics,* 26:268-269, 2010

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, pancrease, prostate, spleen, testes, thyroid

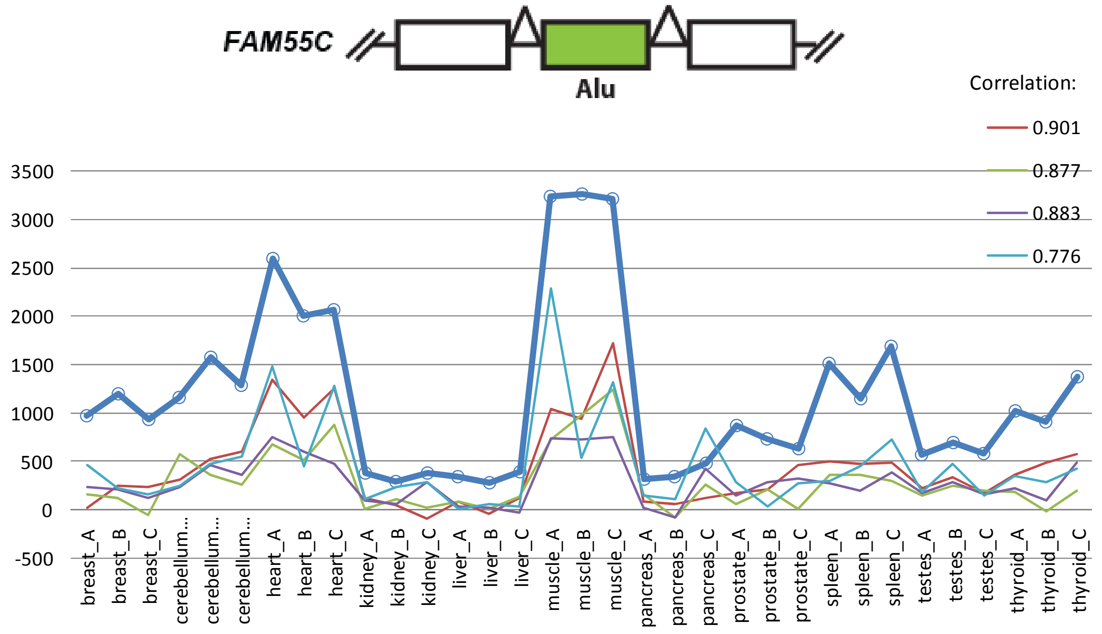
Exon array analysis of Alu exons

Internal spliced exons in the UCSC Genome Browser database

Covered by Alu elements for at least 50% of the exon length

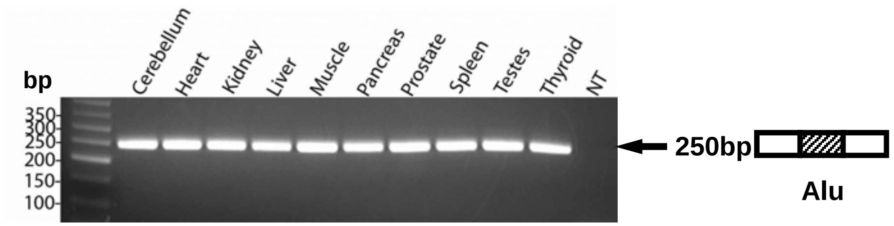
Final list: 330 Alu-derived exons, each with at least 3 reliable probes

Detection of Alu exons “correlated” with gene expression

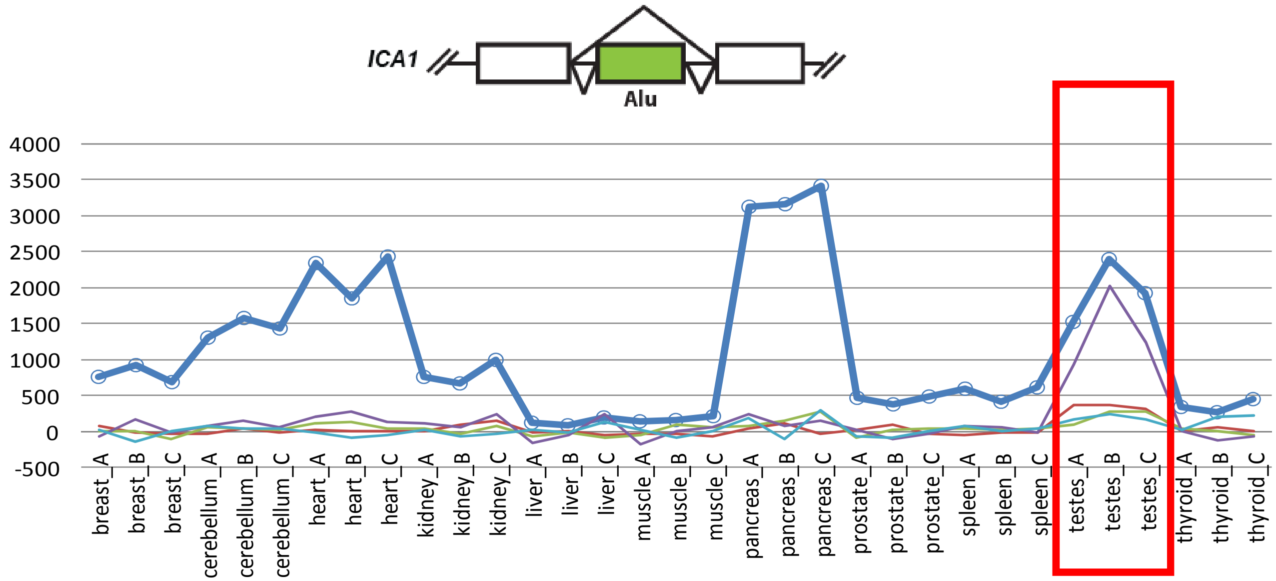


“Correlated” exon: at least 3 of the 4 probes show at least 0.6 Pearson correlation coefficient with gene expression level.

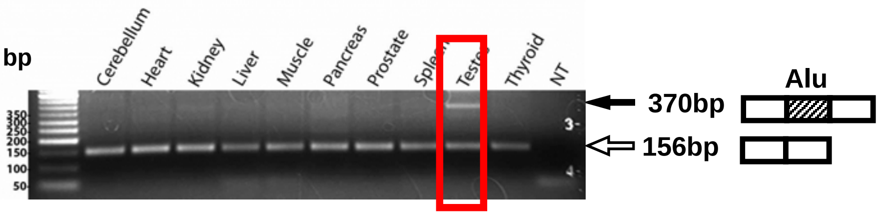
FAM55C (Inclusion 250bp, skipping 130bp)



Examples of tissue-specific Alu-derived exons



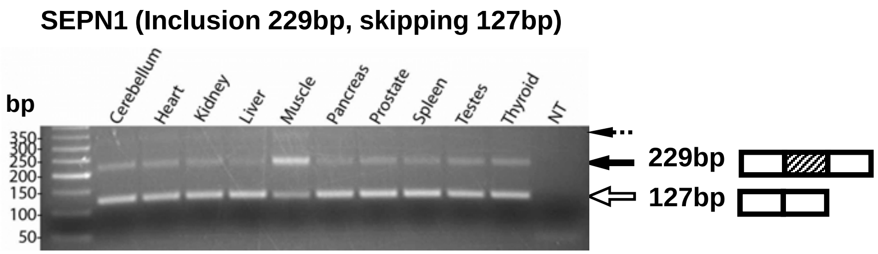
ICA1 (Inclusion 370bp, skipping 156bp)



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

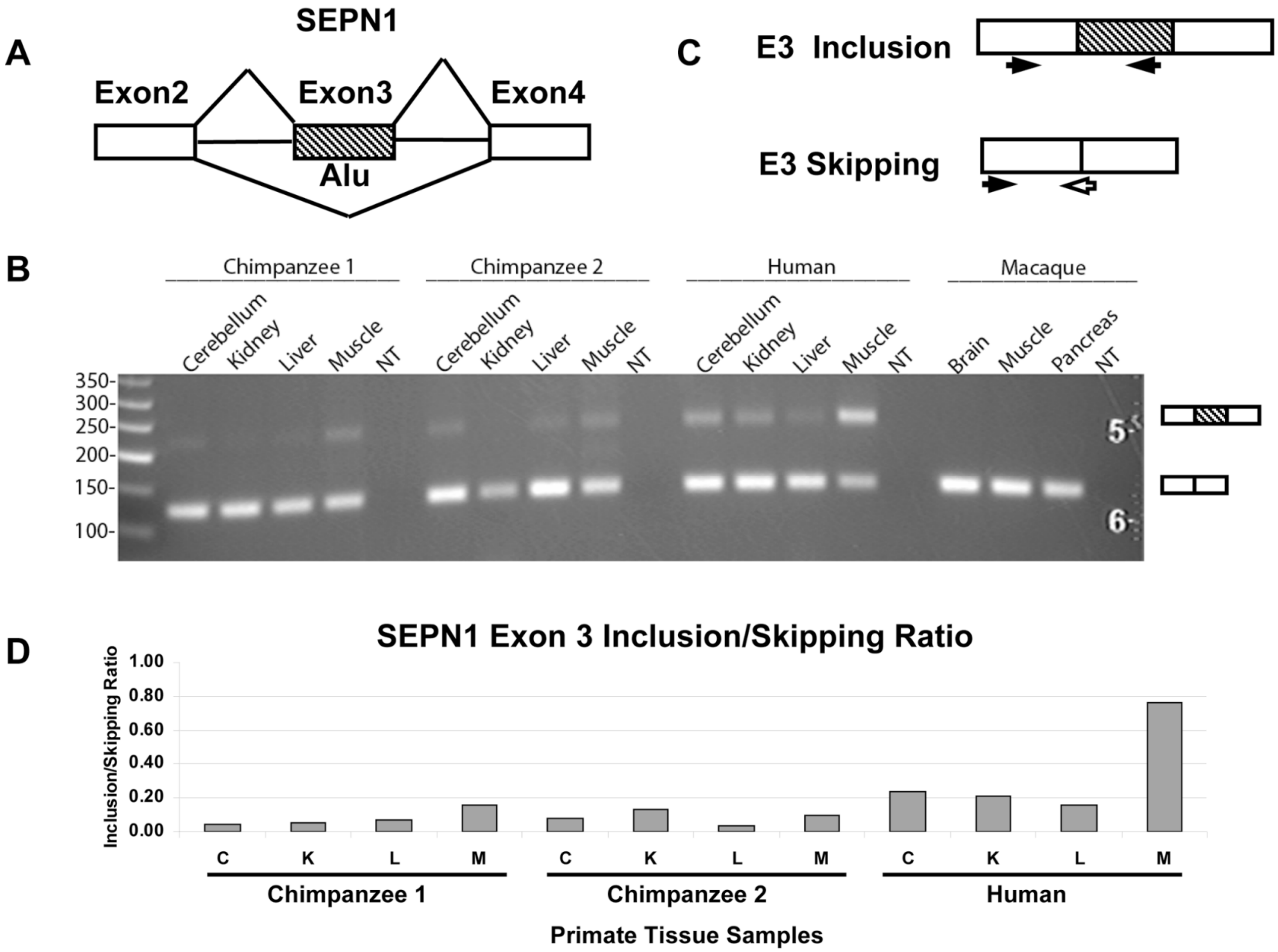

SEPN1 (Inclusion 229bp, skipping 127bp)

Alu-derived exon 3

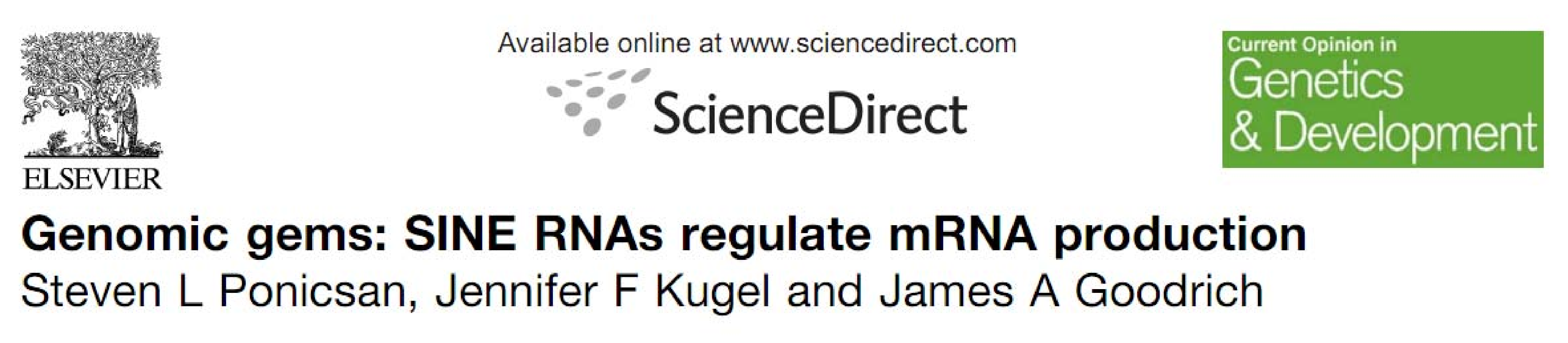
Homo sapiens	87	`ISPEEFKPIAEKLTGSCSVTQTGVQWCSHSSLQPQLPWLNUSSC`	130
Mus musculus	87	`ISPEEFKPIAEKLTG-----------------------------`	101
Gallus gallus	81	`LSPEEFKPIAEKLTG-----------------------------`	95
Danio rerio	89	`LSPEEFKPIAEKLTG-----------------------------`	103

Homo sapiens	131	`LSLLRSTPAASCEEEELPPDPSEETLTIEARFQPLLPETMTKSK`	174
Mus musculus	102	`-----SVPVANYEEEELPHDPSEETLTIEARFQPLLMETMTKSK`	140
Gallus gallus	96	`-----VTPVSDFEED--APDPNGETLSIVAKFQPLVMETMTKSK`	132
Danio rerio	104	`-----VAPPPEYEEE-IPHDPNGETLTLHAKMQPLLLESMTKSK`	141

Evolution of SEPN1 Alu-exon Splicing



Lin L et al. *PLoS Genetics*, 2008, 4(10): e1000225.



ELSEVIER

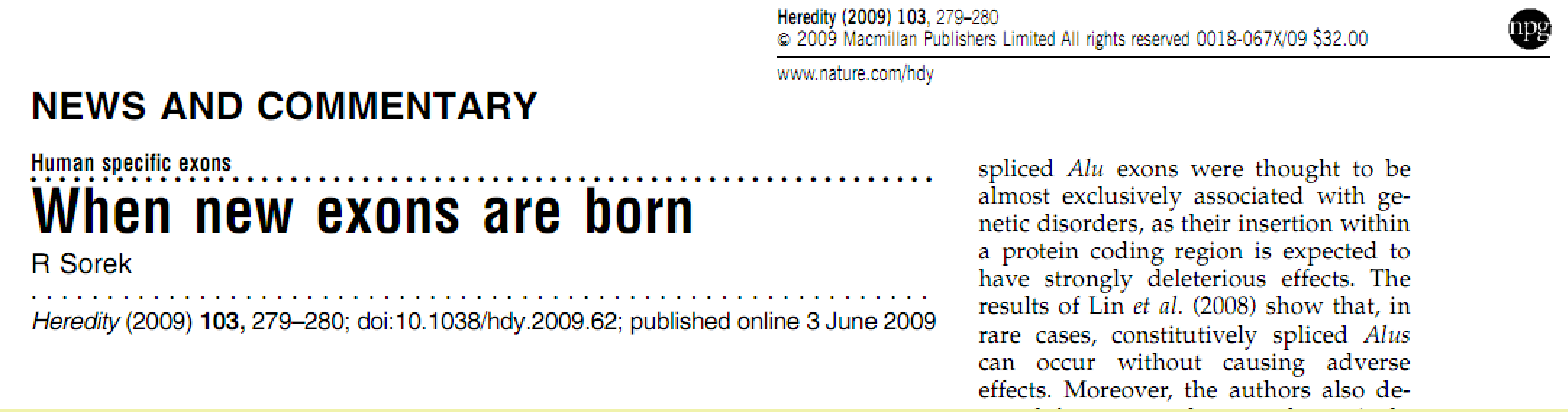
Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in
Genetics
& Development

Genomic gems: SINE RNAs regulate mRNA production

Steven L Ponicsan, Jennifer F Kugel and James A Goodrich



Heredity (2009) 103, 279–280
© 2009 Macmillan Publishers Limited All rights reserved 0018-067X/09 $32.00
www.nature.com/hdy

npg

NEWS AND COMMENTARY

Human specific exons

When new exons are born

R Sorek

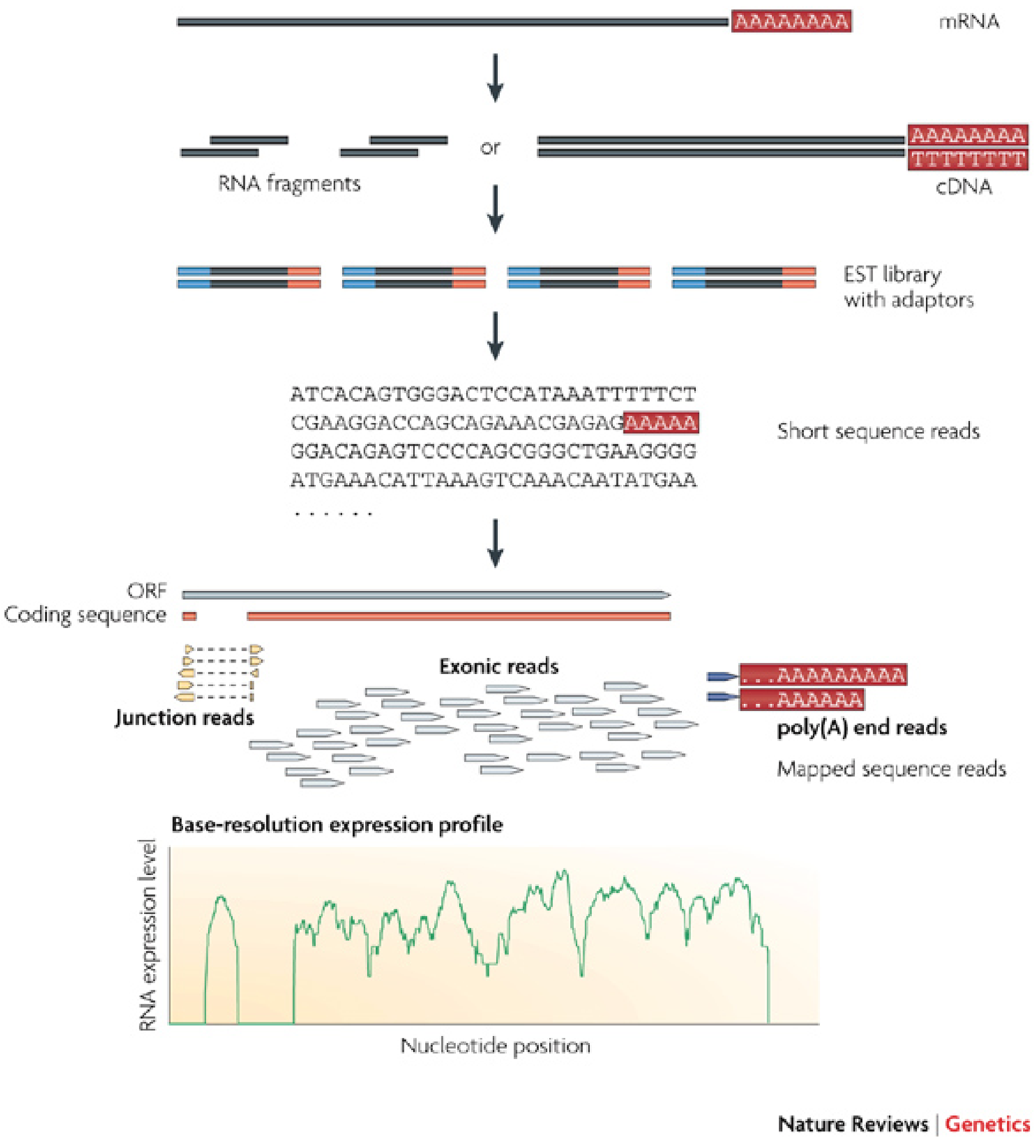
Heredity (2009) **103,** 279–280; doi:10.1038/hdy.2009.62; published online 3 June 2009

spliced *Alu* exons were thought to be almost exclusively associated with genetic disorders, as their insertion within a protein coding region is expected to have strongly deleterious effects. The results of Lin *et al*. (2008) show that, in rare cases, constitutively spliced *Alus* can occur without causing adverse effects. Moreover, the authors also de-

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

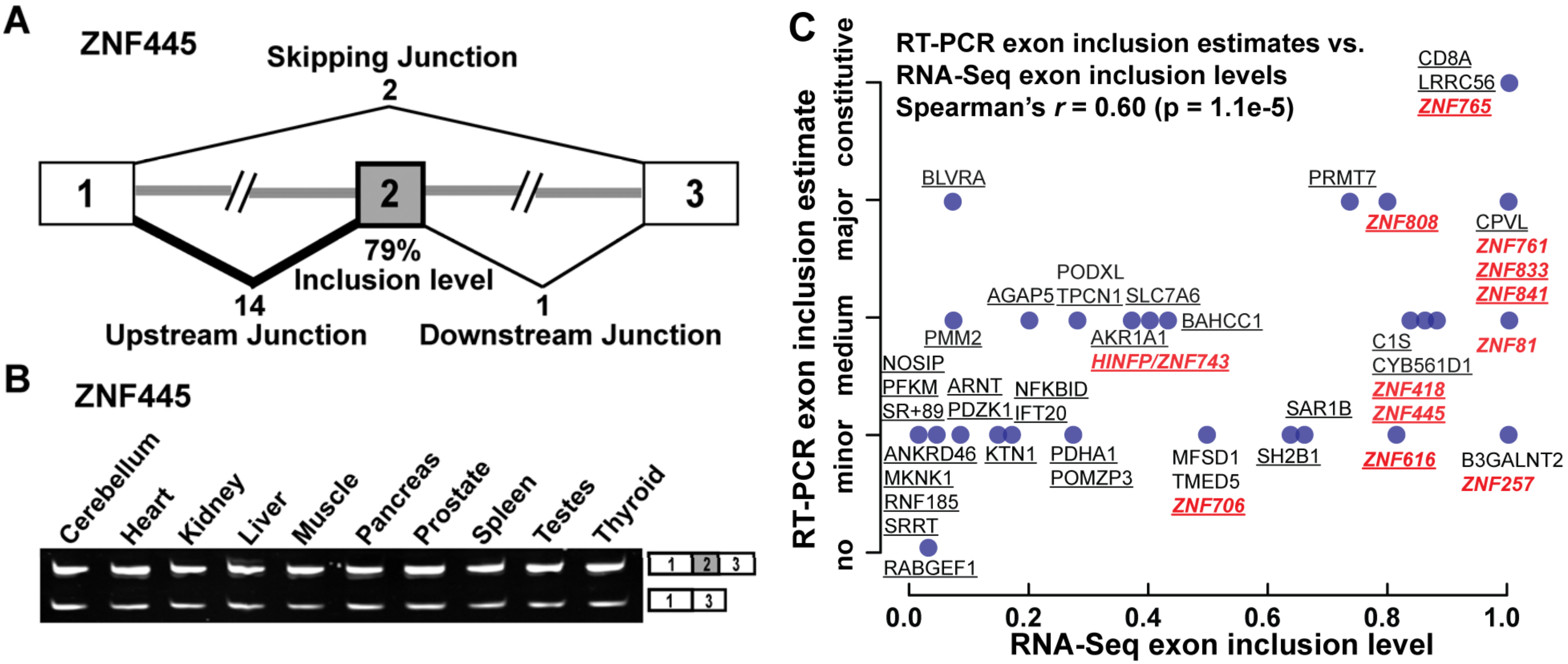
Sorek R, *Heredity* (2009) 103, 279–280

RNA Sequencing (RNA-Seq)



Wang et al., *Nat Rev Genet*. 2009, 10(1):57-63.

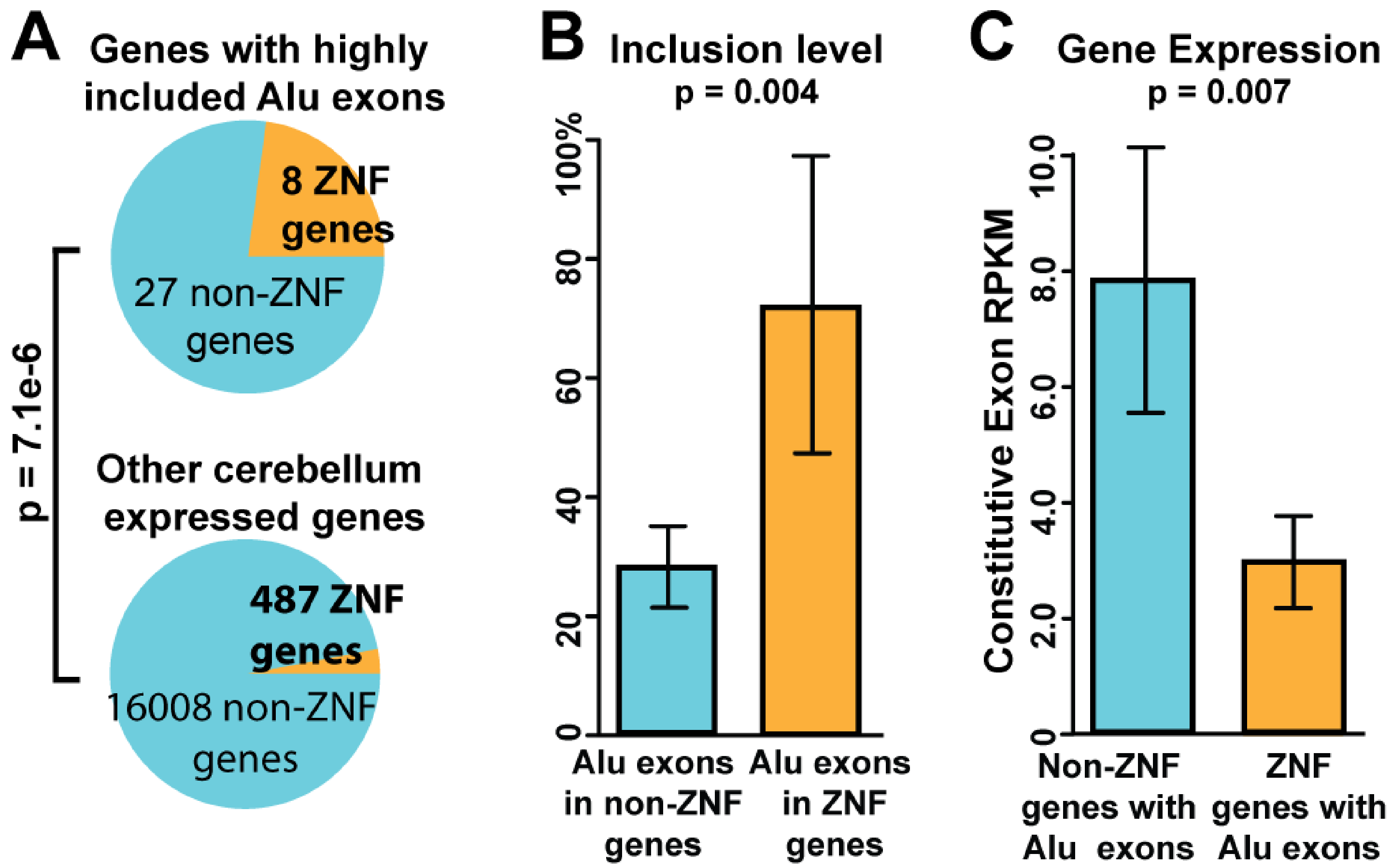
RNA-Seq Analysis of Alu Exons



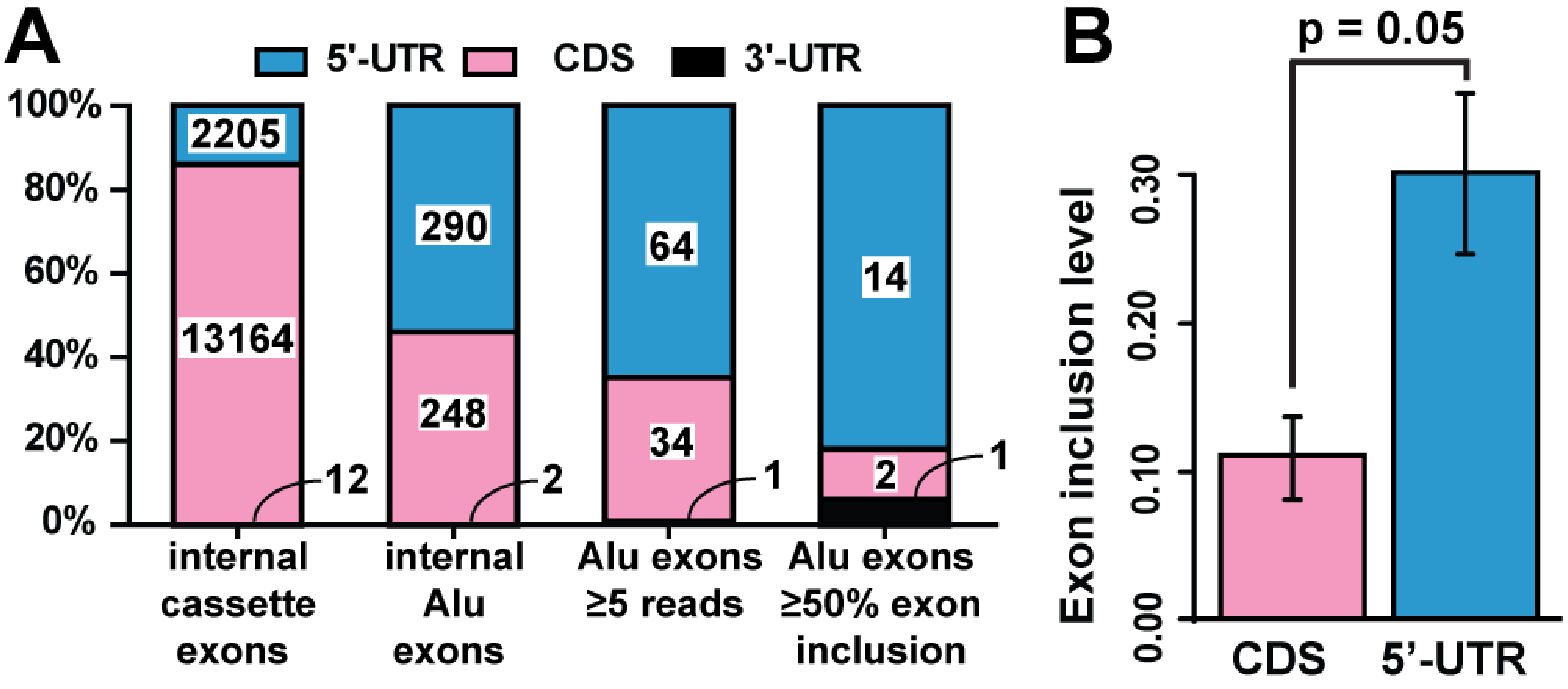
123 million reads for the human cerebellum

Shen*, Lin* et al. (2011) *PNAS*, 108:2837-2842

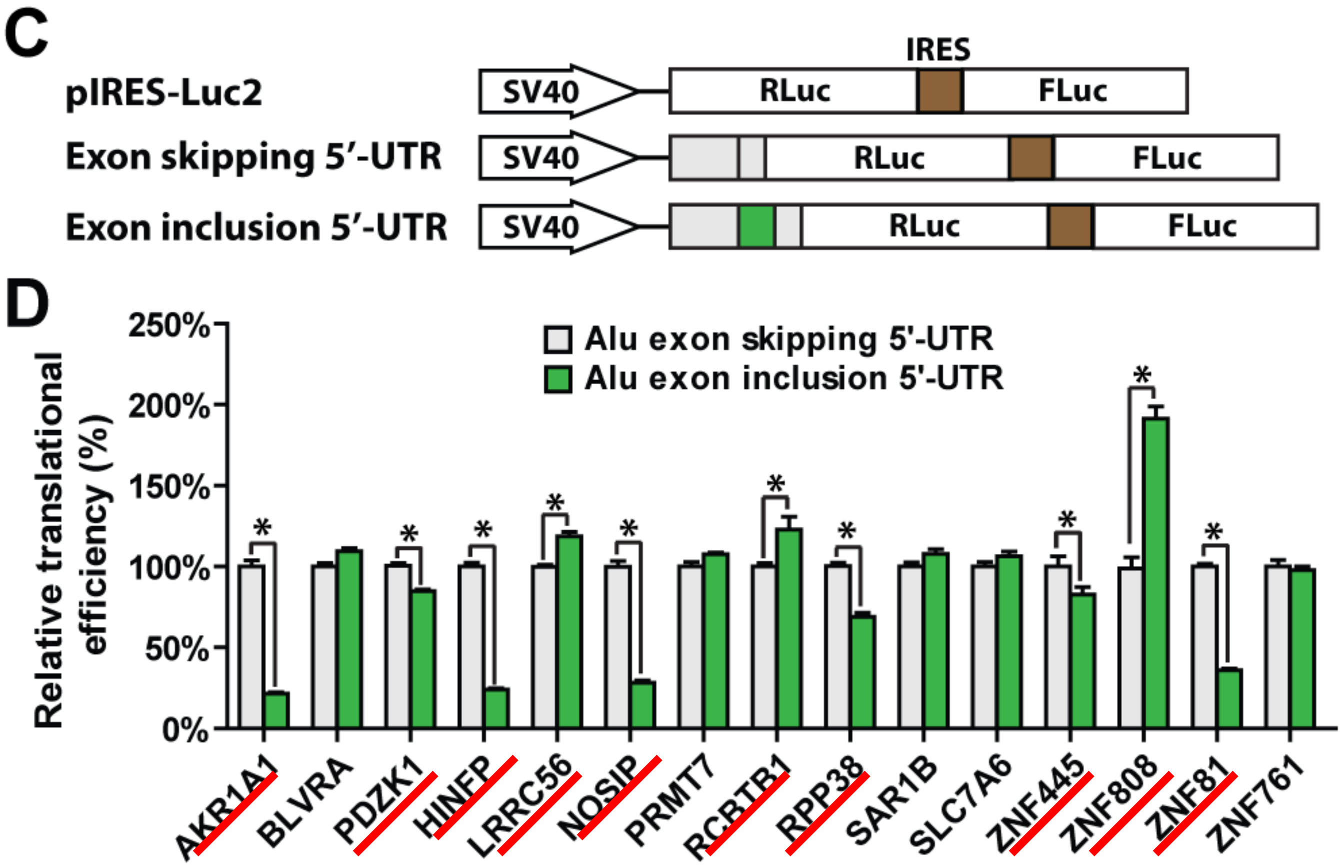
Alu Exons are Enriched in Zinc Finger Transcription Factors



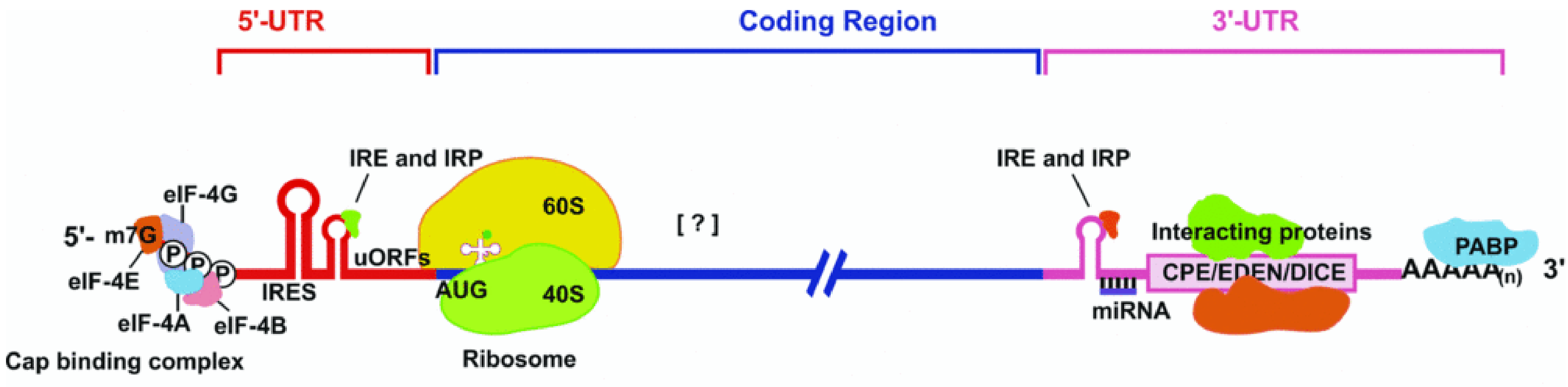
Alu Exons are Enriched in the 5’-UTR



Alu Exons in 5’-UTR Regulate Protein Translation



Structural organization of eukaryotic mRNA



Chatterjee, S et al., Biol Cell. 2009

5’-UTR of an mRNA:

Length

Thermal stability

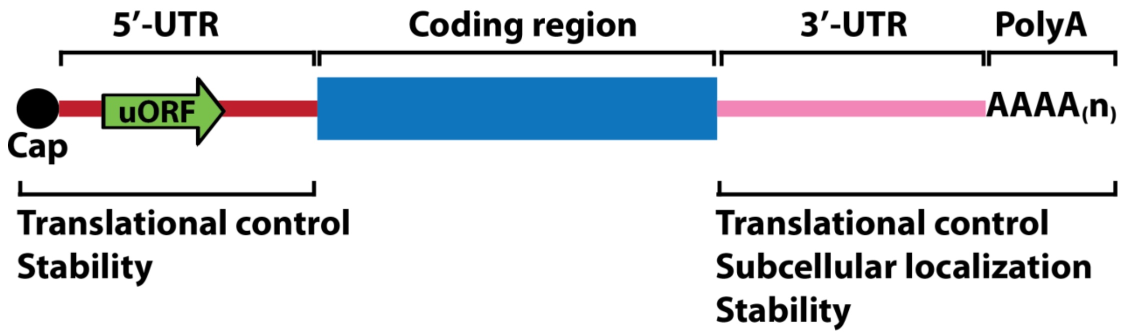
GC content

Secondary structures

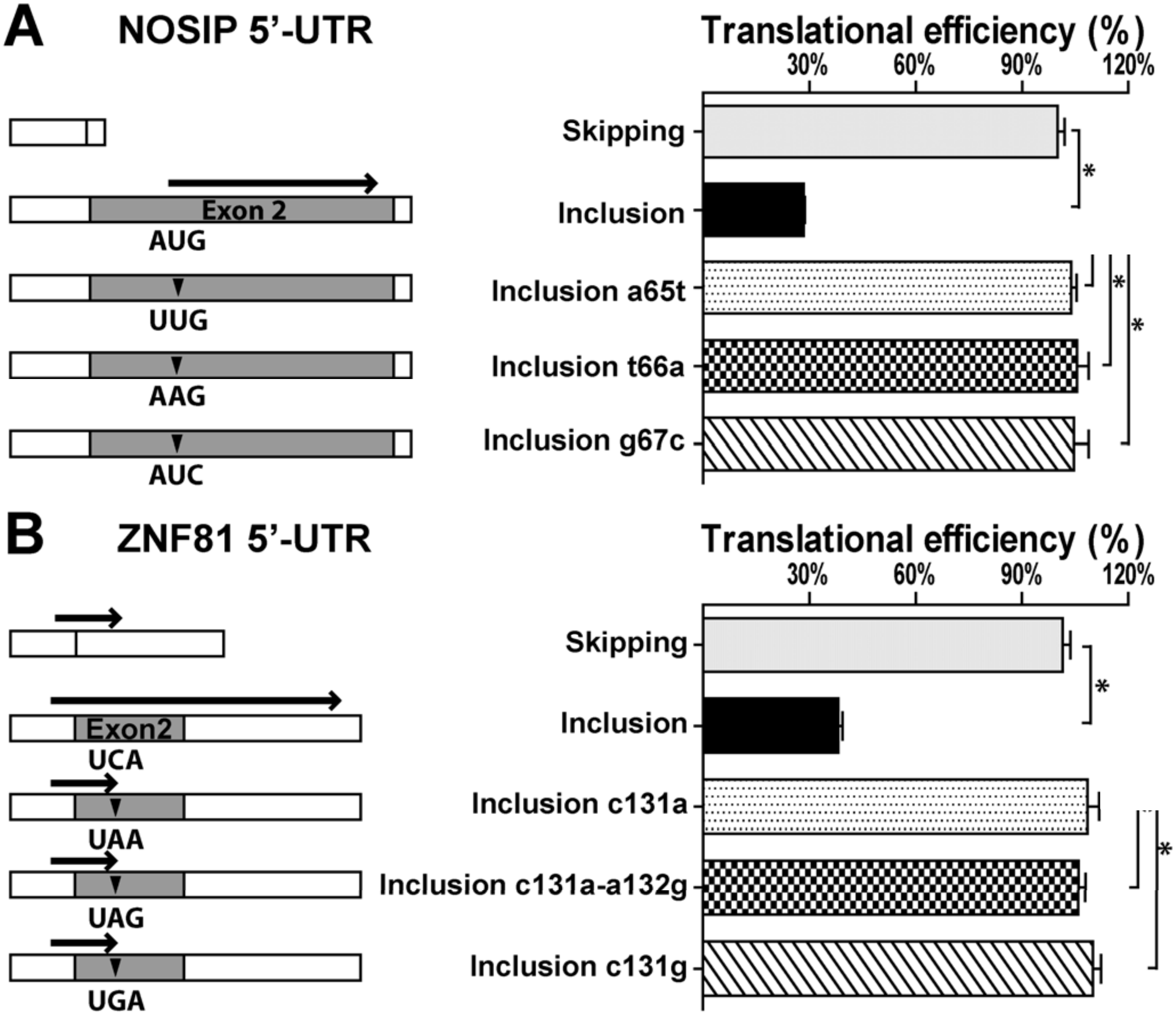
uORFs (upstream ORFs)

IRES

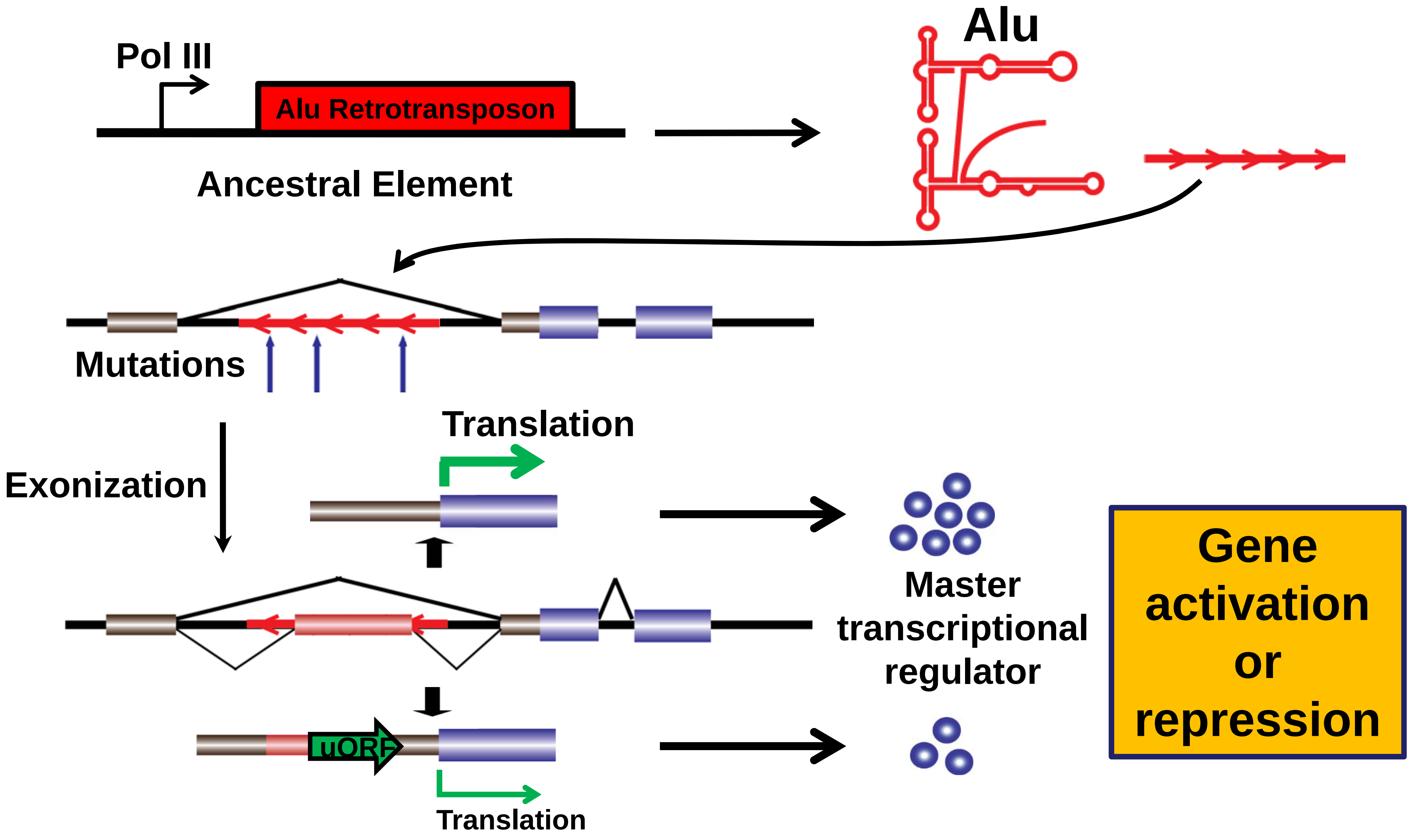
Binding sites for proteins



Alu exons repress translation by creating or elongating uORFs



Alu exonization: Regulating the regulators



Pol III

Alu Retrotransposon

Ancestral Element

Alu

Mutations

Exonization

Translation

uORF

Translation

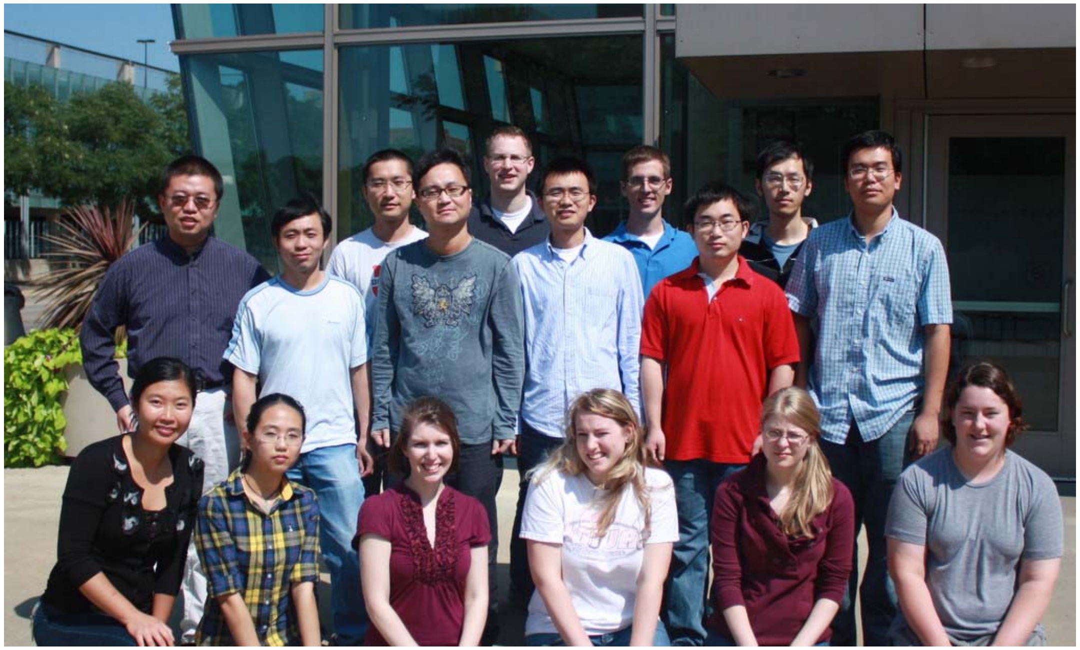
Master transcriptional regulator

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

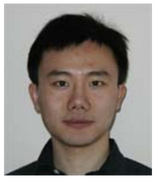


Collaborators

Russ Carstens
(U. Penn)

Jeff Murray
(U. Iowa)

Qing Zhou
(UCLA)



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