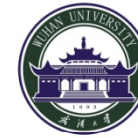




UNIVERSITÉ
PARIS-SUD 11



武汉大学
WUHAN UNIVERSITY



PTB's positional rule in regulating alternative splicing and RNA design under motif constraints

Yu Zhou

14-04-2009

Supervisors:

Prof. Alain Denise, Laboratoire de Recherche En Informatique

Prof. Yi Zhang, College of Life Science

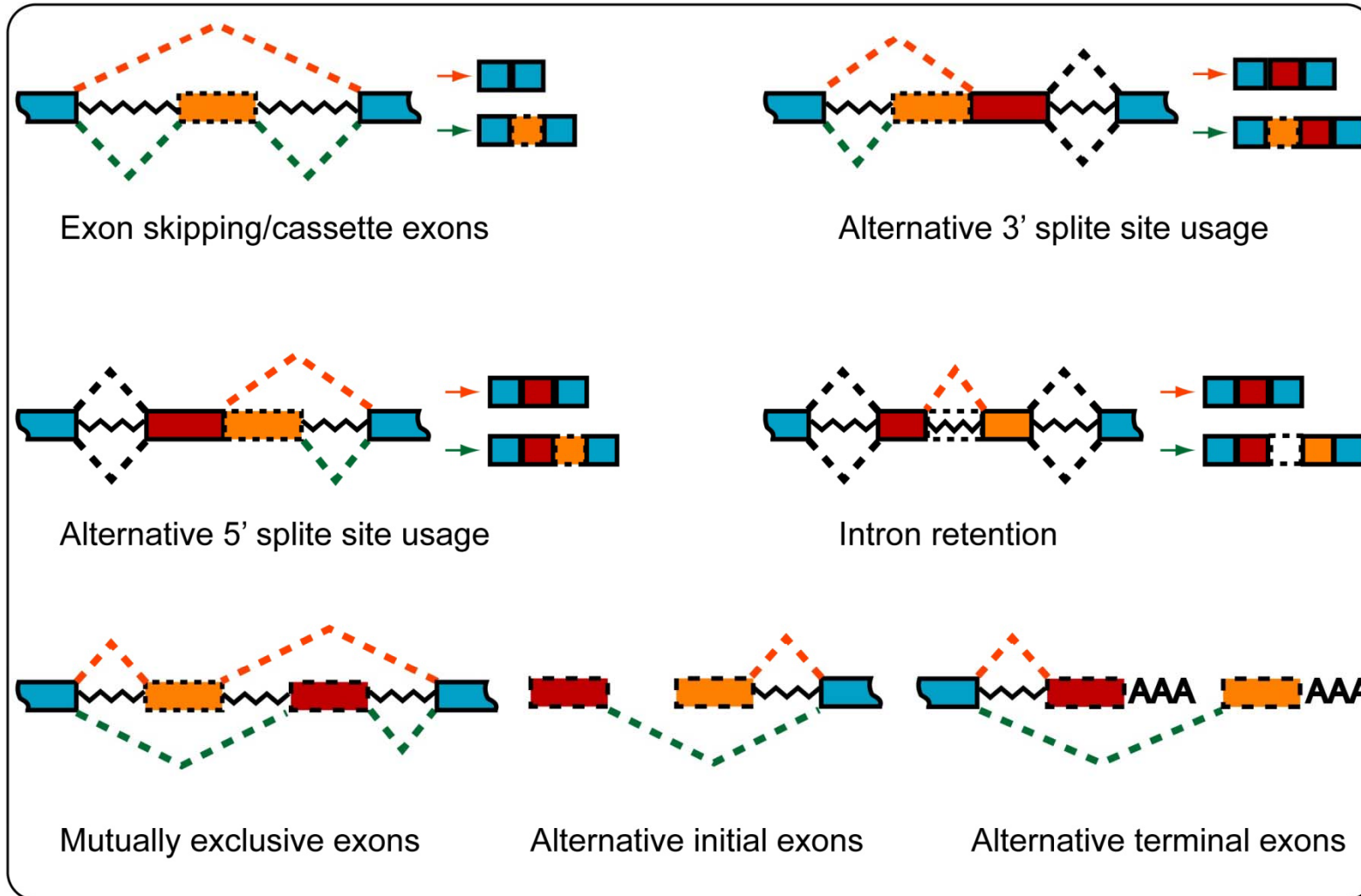


Genome-wide analysis of PTB binding by CLIPseq reveals a positional rule for PTB to positively or negatively regulate alternative splicing in mammalian cells

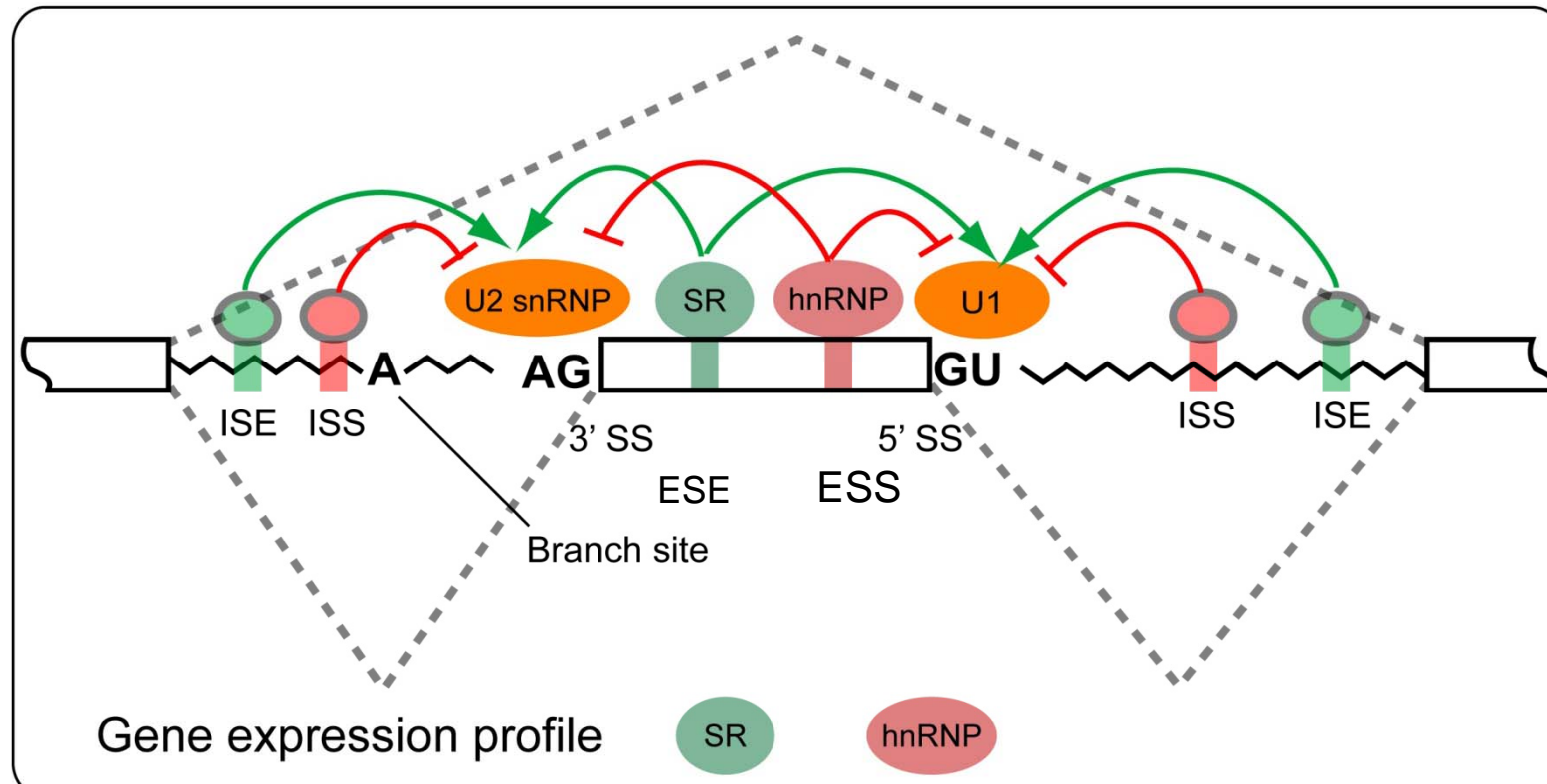
Yuanchao Xue, Yu Zhou, ..., Xiang-dong Fu*, Yi Zhang*

Alternative splicing (AS)

a mechanism to generate structurally and functionally distinct protein variants



Large number of cis-elements and trans-acting factors involve in AS regulation.



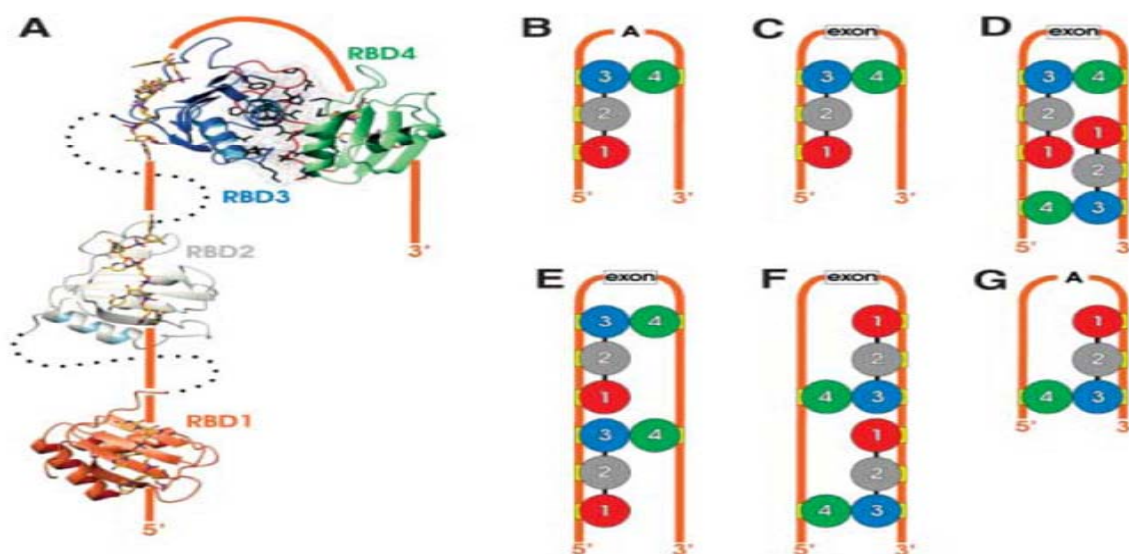
ESE: Exonic Splicing Enhancer



PTB

(Polypyrimidine Tract Binding protein, hnRNP I)

- PTB is a general splicing repressor.
- SELEX reveals it binds to CU-rich motifs.
- Monomer in solution (Monie et al., 2005) .
- Looping model to repress splicing.



Model,
No experimental evidence

Oberstrass et al., 2005, Science

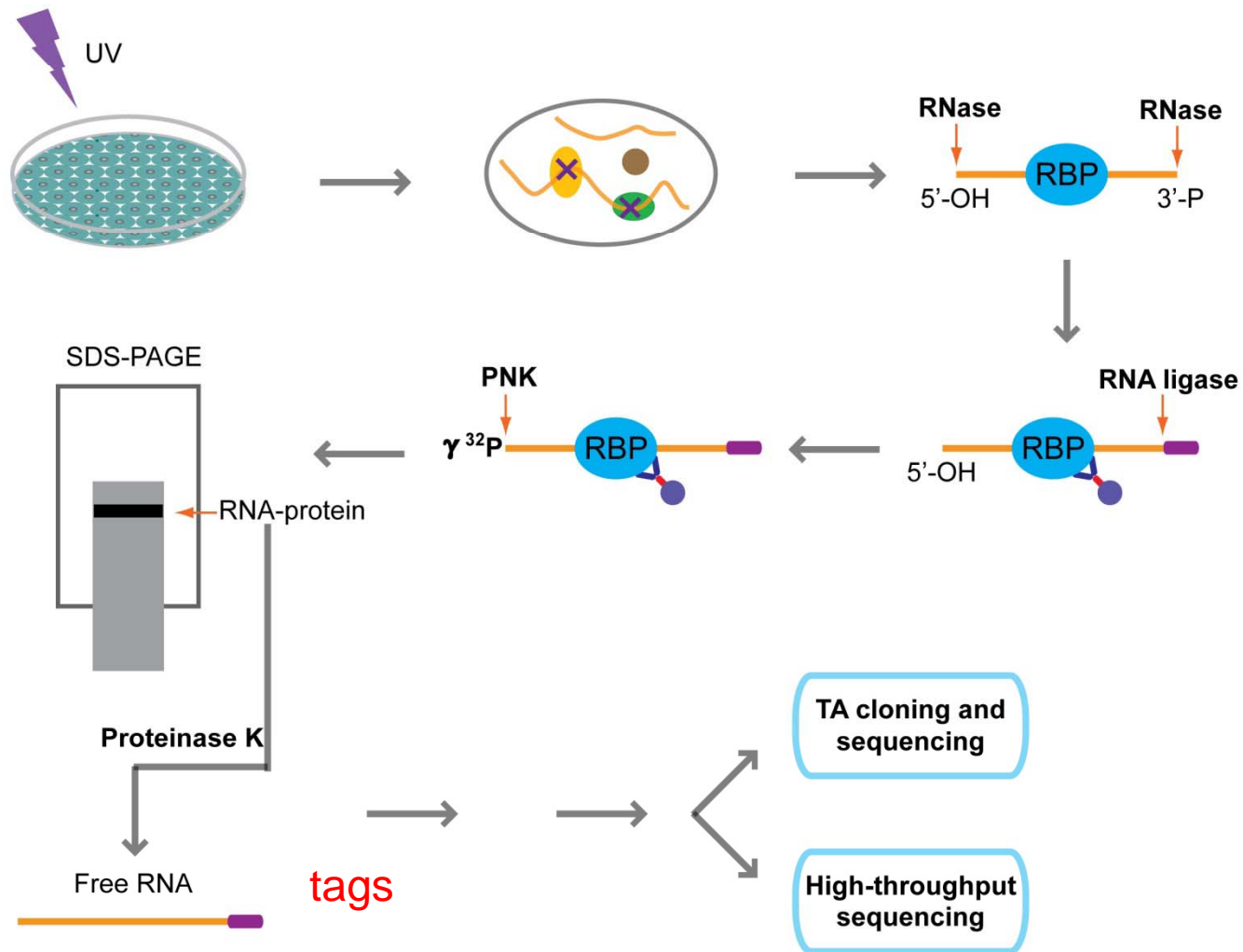


Remaining questions:

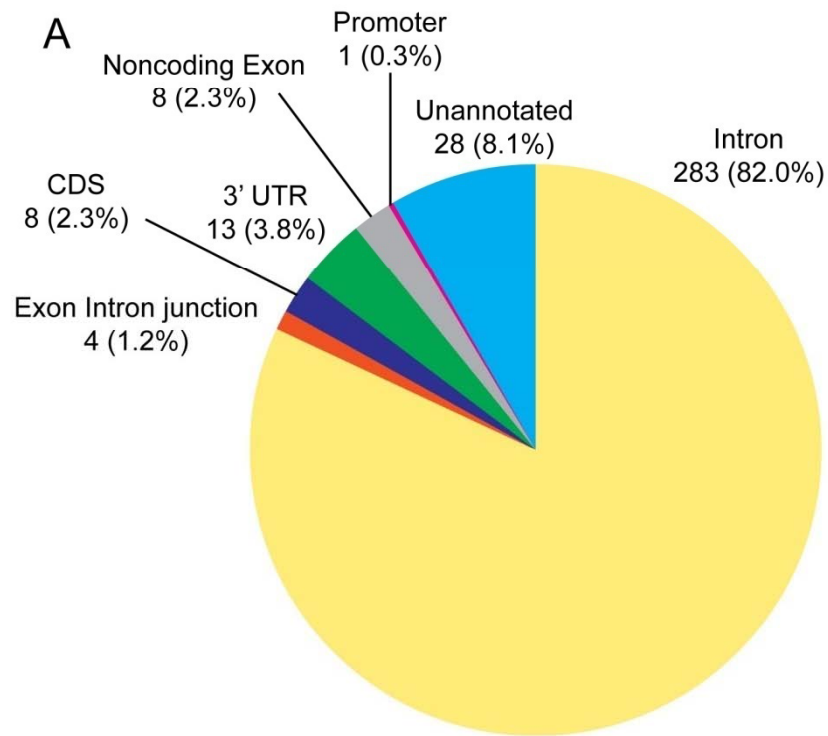
- **Monomer or dimer *in vivo*?**
- **Binding motifs *in vivo*?**
- ***In vivo* targets of PTB?**
- **The mechanism of PTB repression of splicing?**
- **New function?**

CLIP(-seq) to identify *in vivo* RNA targets of PTB

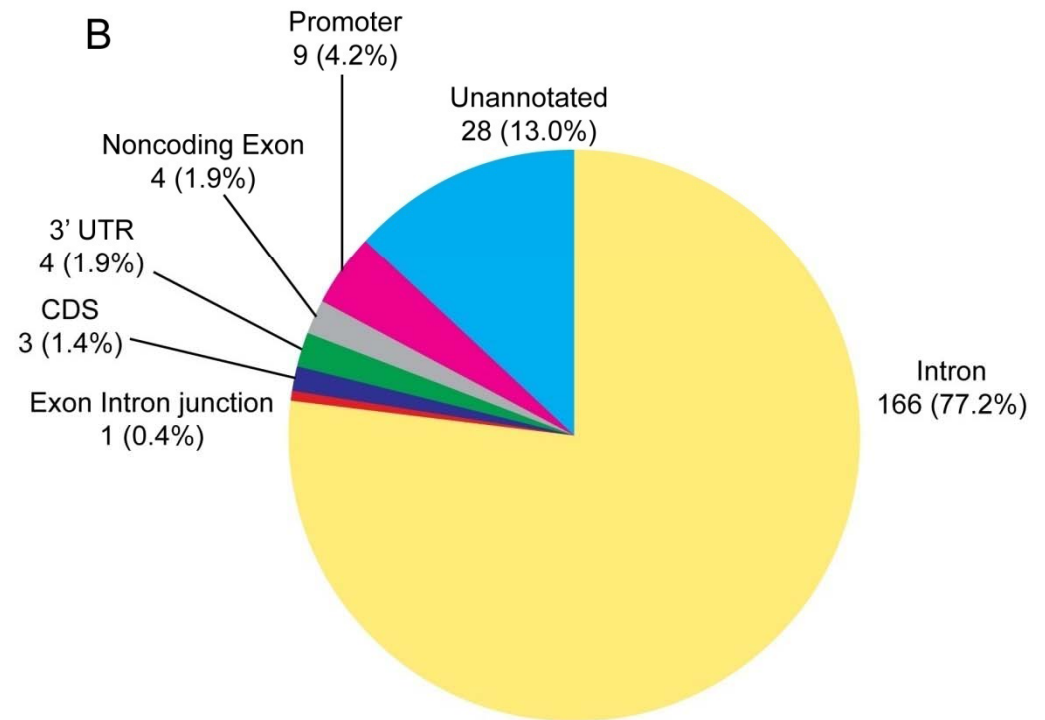
CrossLinking ImmunoPrecipitation



Cloning-sequenced tags are mostly in introns

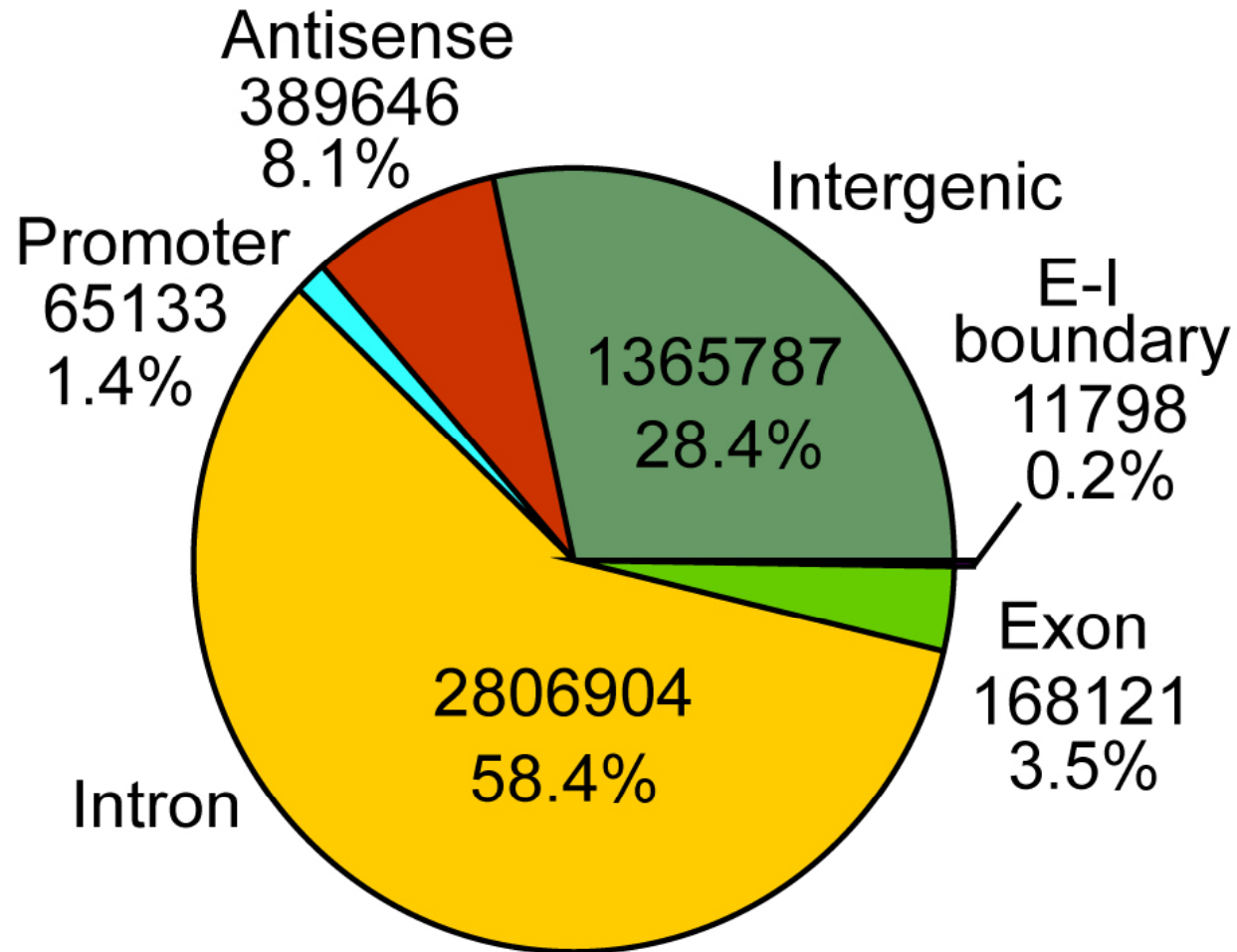


Monomeric tags



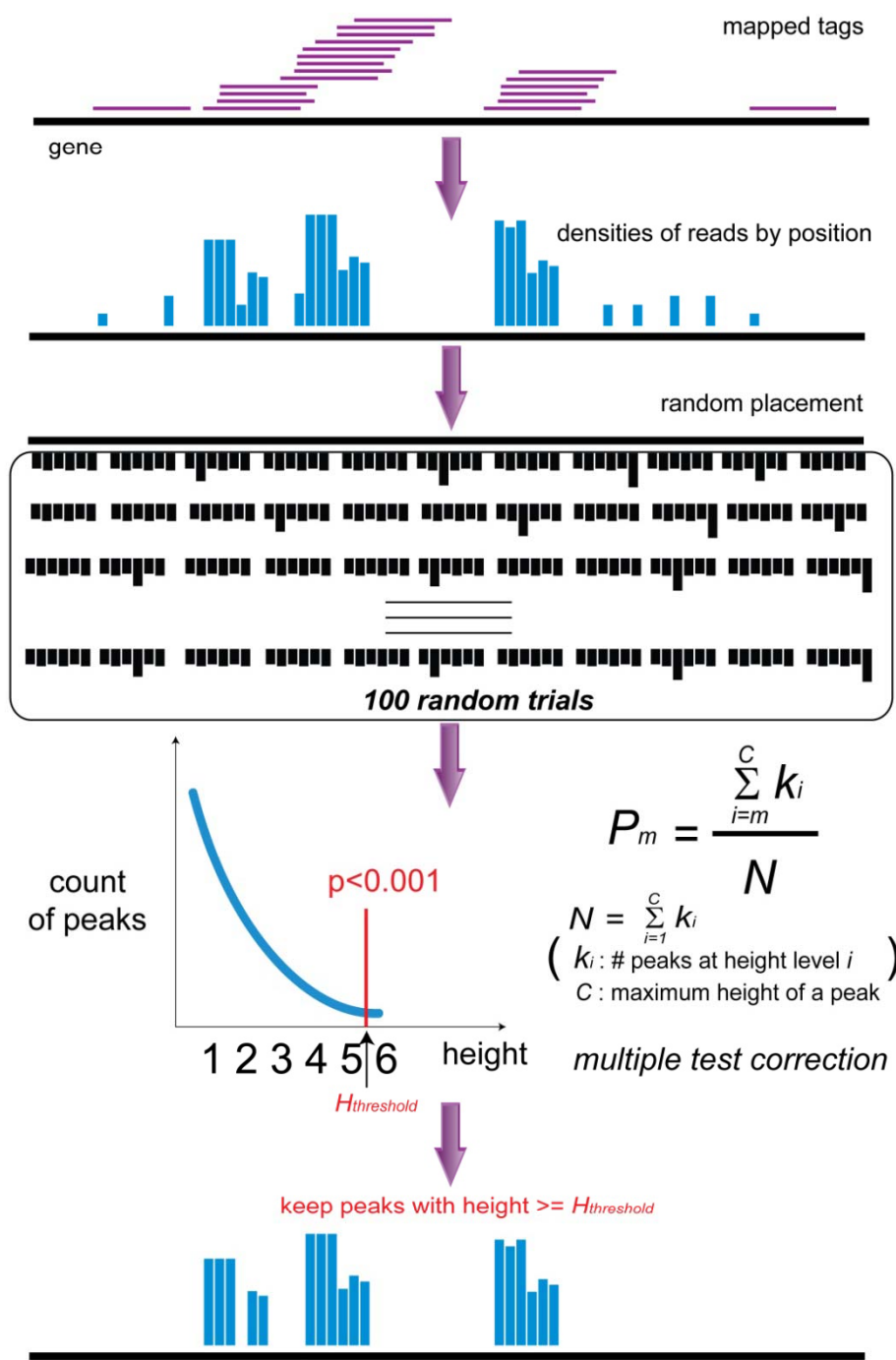
Dimeric tags

Tag distribution corresponding to knownGenes (UCSC)





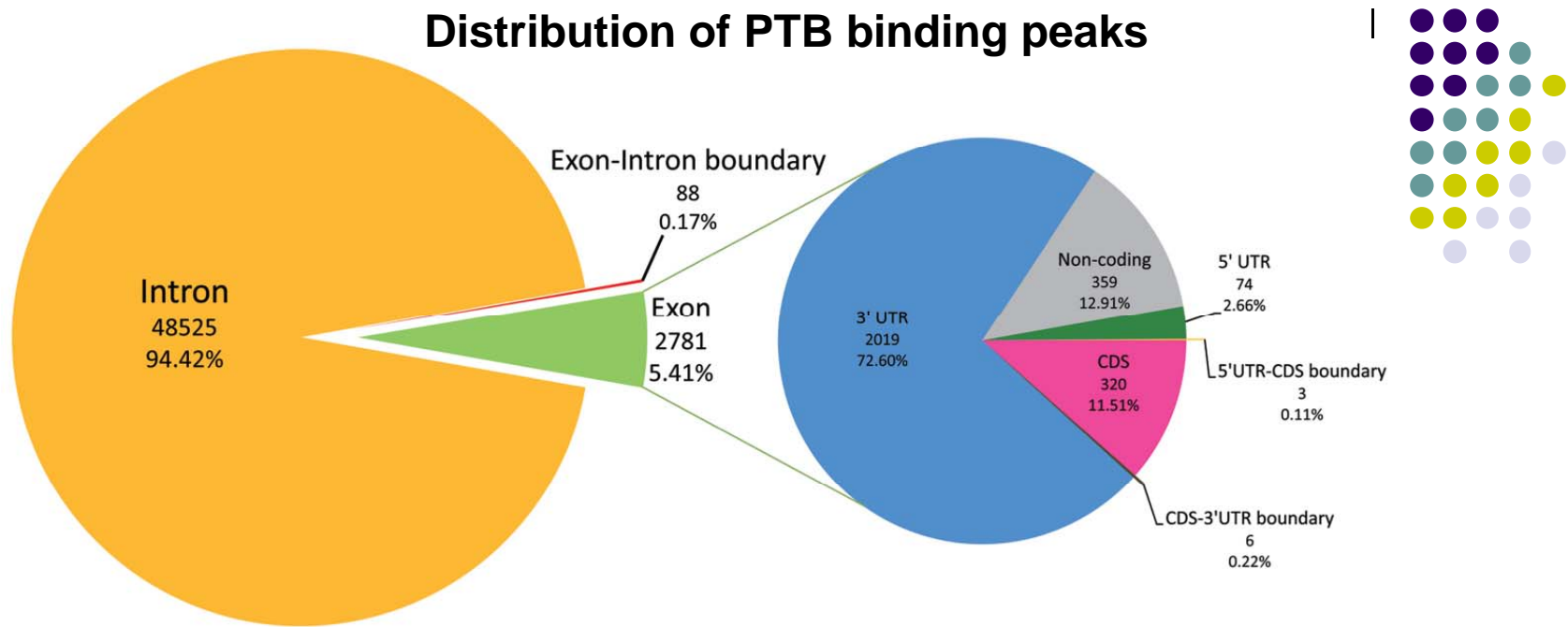
Flowchart of peak detection



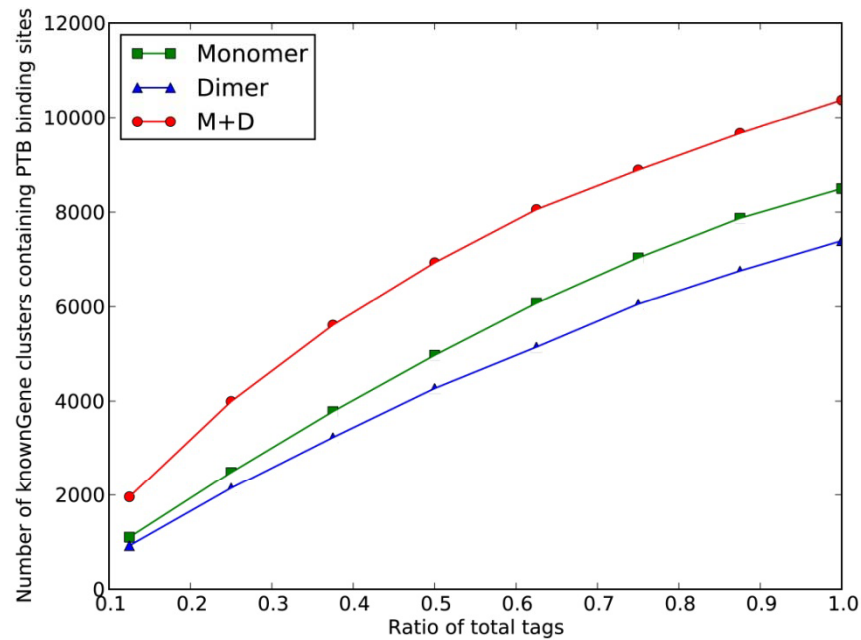
- Get PTB binding locus enriched of tags
- Determine the threshold in gene dependent manner

Strategy similar with *NSMB* paper (Gene Yeo et.al 2009)

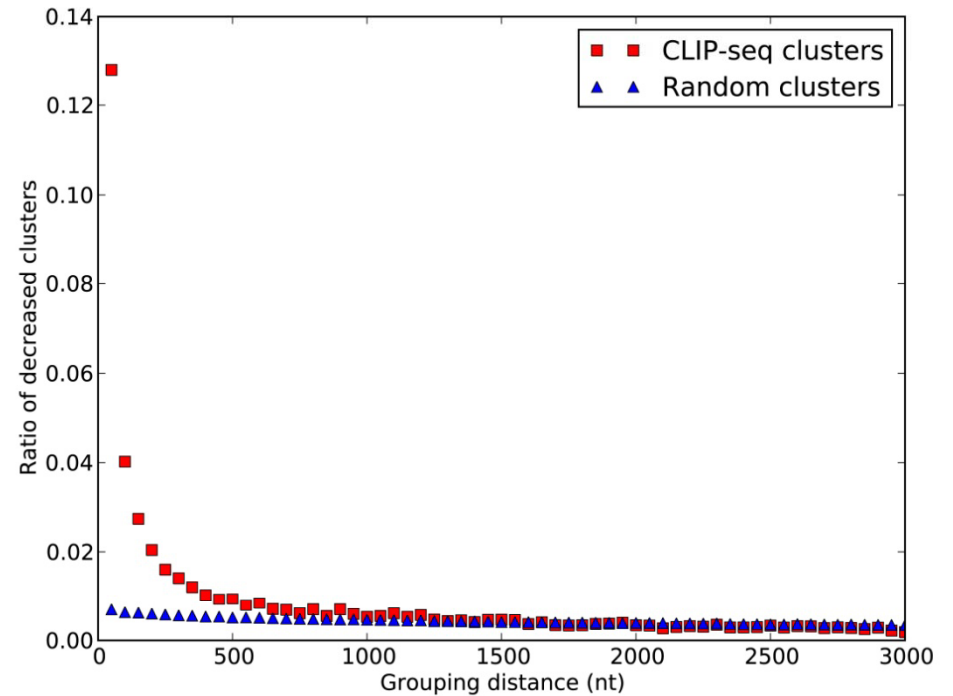
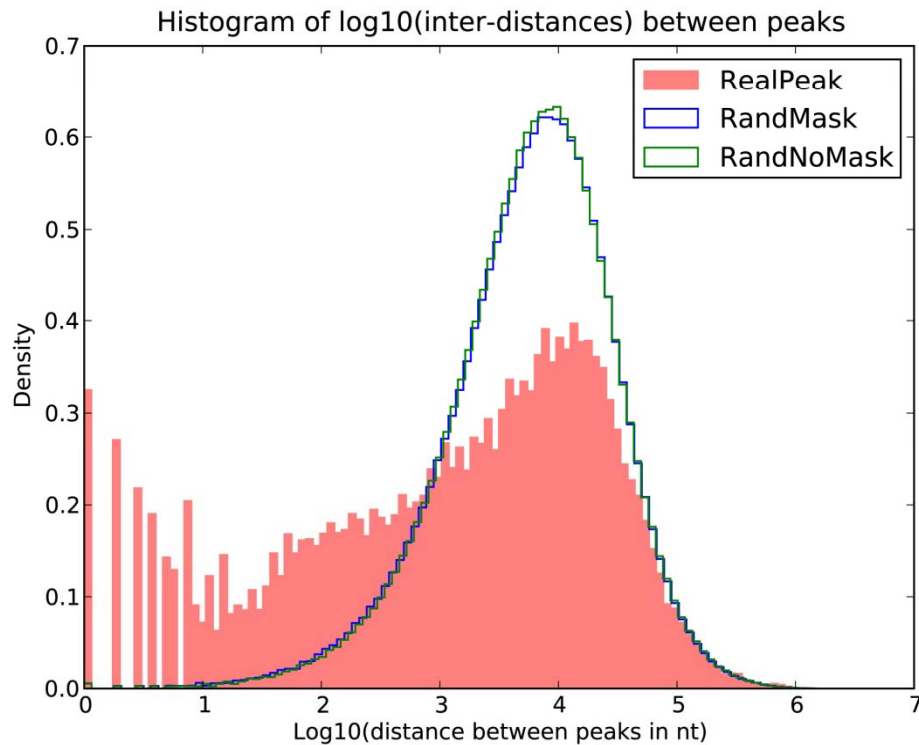
Distribution of PTB binding peaks



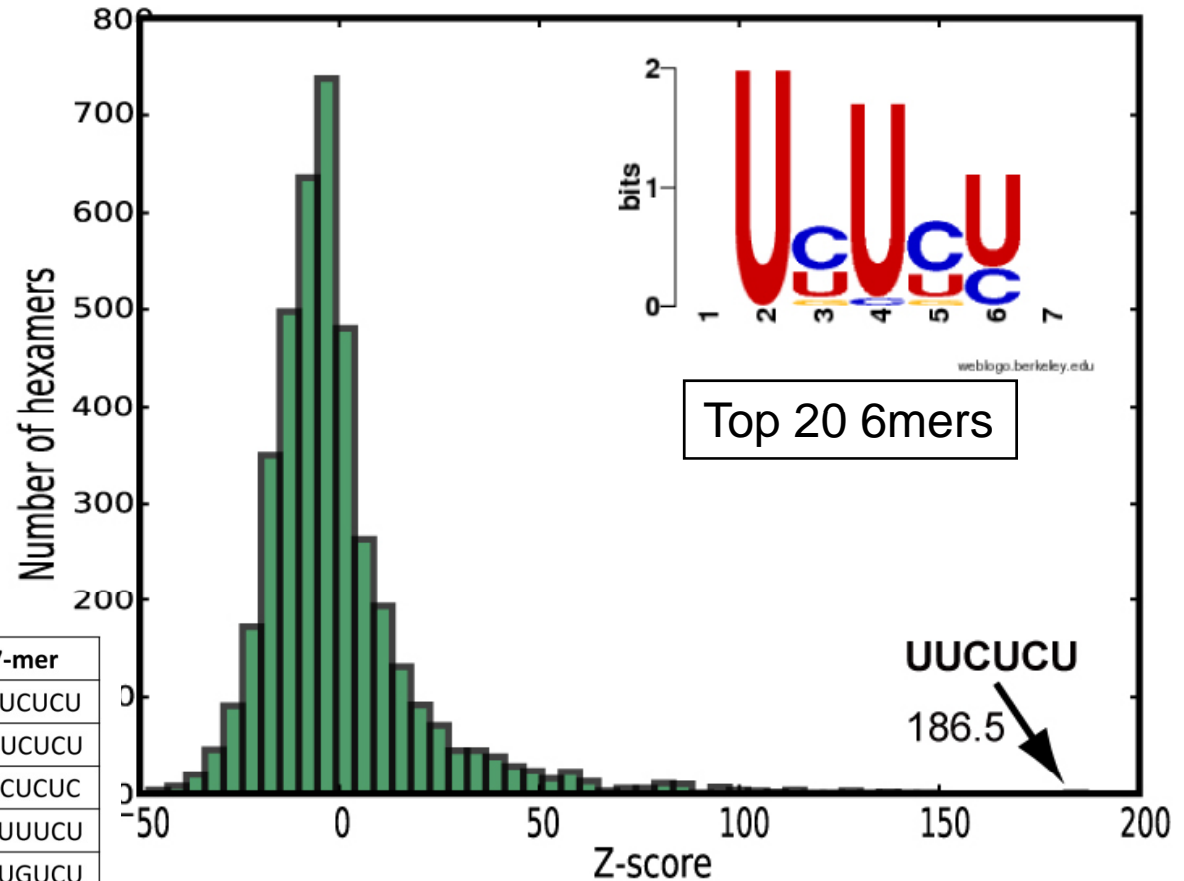
PTB may target more than 42.55% genes, as sequencing not saturated.



PTB binding sites tend to cluster together.



PTB binding clusters significantly enrich known PTB binding motifs.

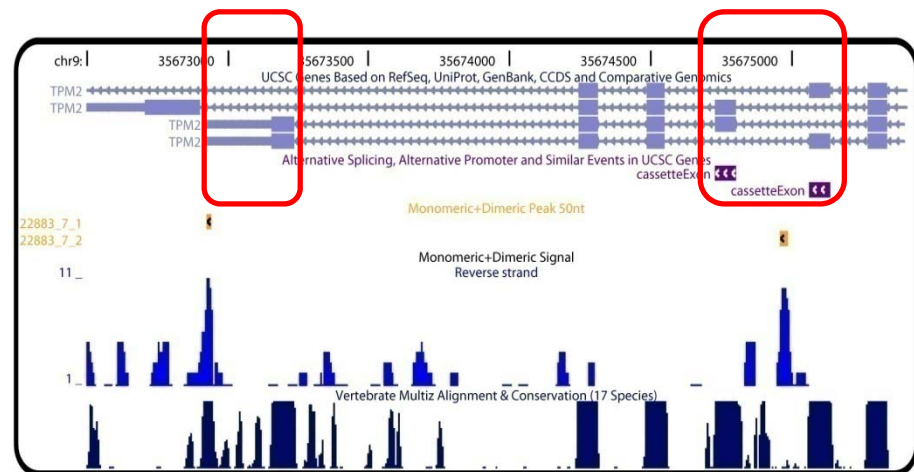
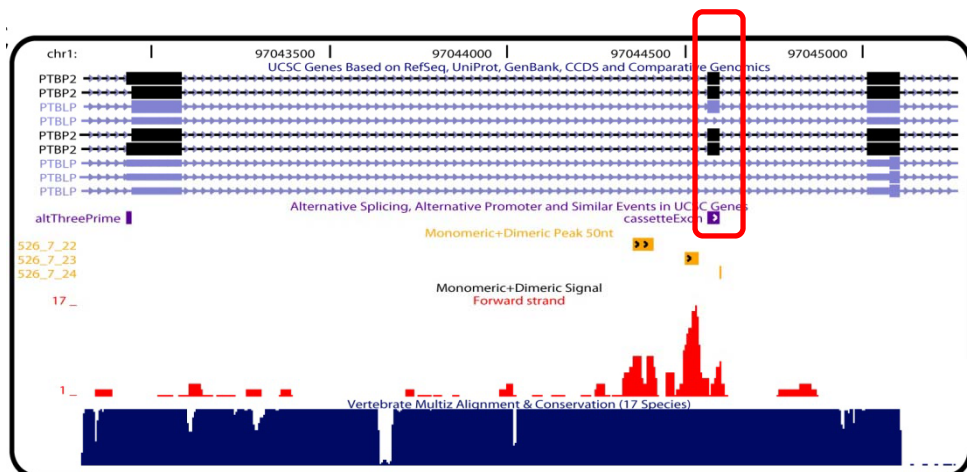


rank	2-mer	3-mer	4-mer	5-mer	6-mer	7-mer
1	CU	UCU	UUCU	UCUCU	UUCUCU	UCUCUCU
2	UC	CUU	CUCU	UUUCU	CUCUCU	UUUCUCU
3	UU	UUC	UCUU	UCUUU	UCUCUC	CUCUCUC
4	UG	CUC	UCUC	UCUGU	CUUUCU	UCUUUCU
5	CC	CUG	UUUC	UUCUU	UCUCUG	UCUGUCU
6		CCU	UCUG	CUUCU	UUUCUU	UCUCUGU
7		UUU	CUUU	UUCUC	UCUUCU	CUUCUCU
8		UCC	UCCU	CUUUC	UUUUCU	UUUCUUU
9		UGU	CUUC	UCUUC	UCUCUU	UUCUCUC
10		UUG	CUGU	UUCUG	UCUUUC	CUUUUCU

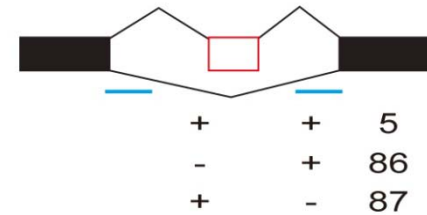
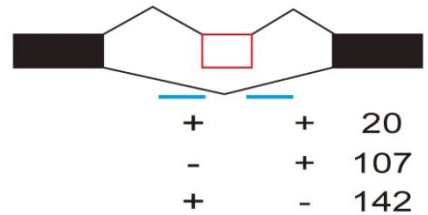
PTB binding locations are significantly associated with alternative splicing event.



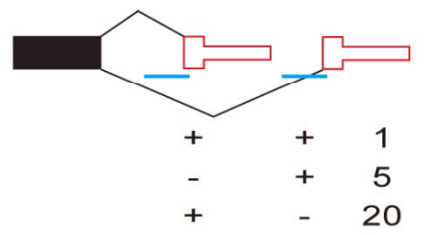
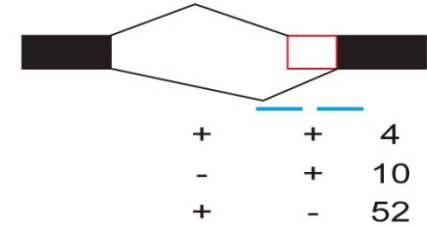
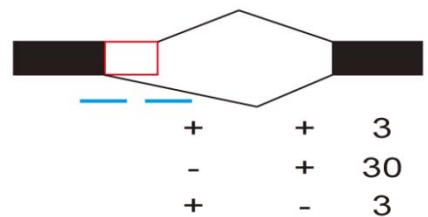
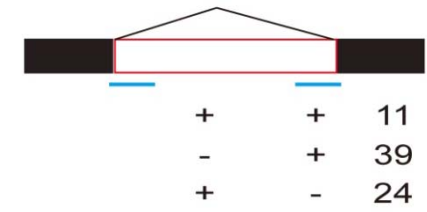
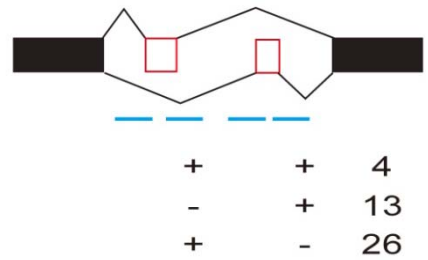
Alt event	# Total events	# PTB cluster associated		Z-score (100 random trials)
		Observed	expected	
Skipped exon	7449	5824	5053	14.11
Alt Terminal	909	815	661	8.56
Retained intron	1446	147	96	6.59
Mulx exon	522	662	581	4.04
Alt5Prime	1970	582	524	3.24
Alt3Prime	3207	805	748	2.70



Patterns of PTB binding competing with splice sites



No predominant looping

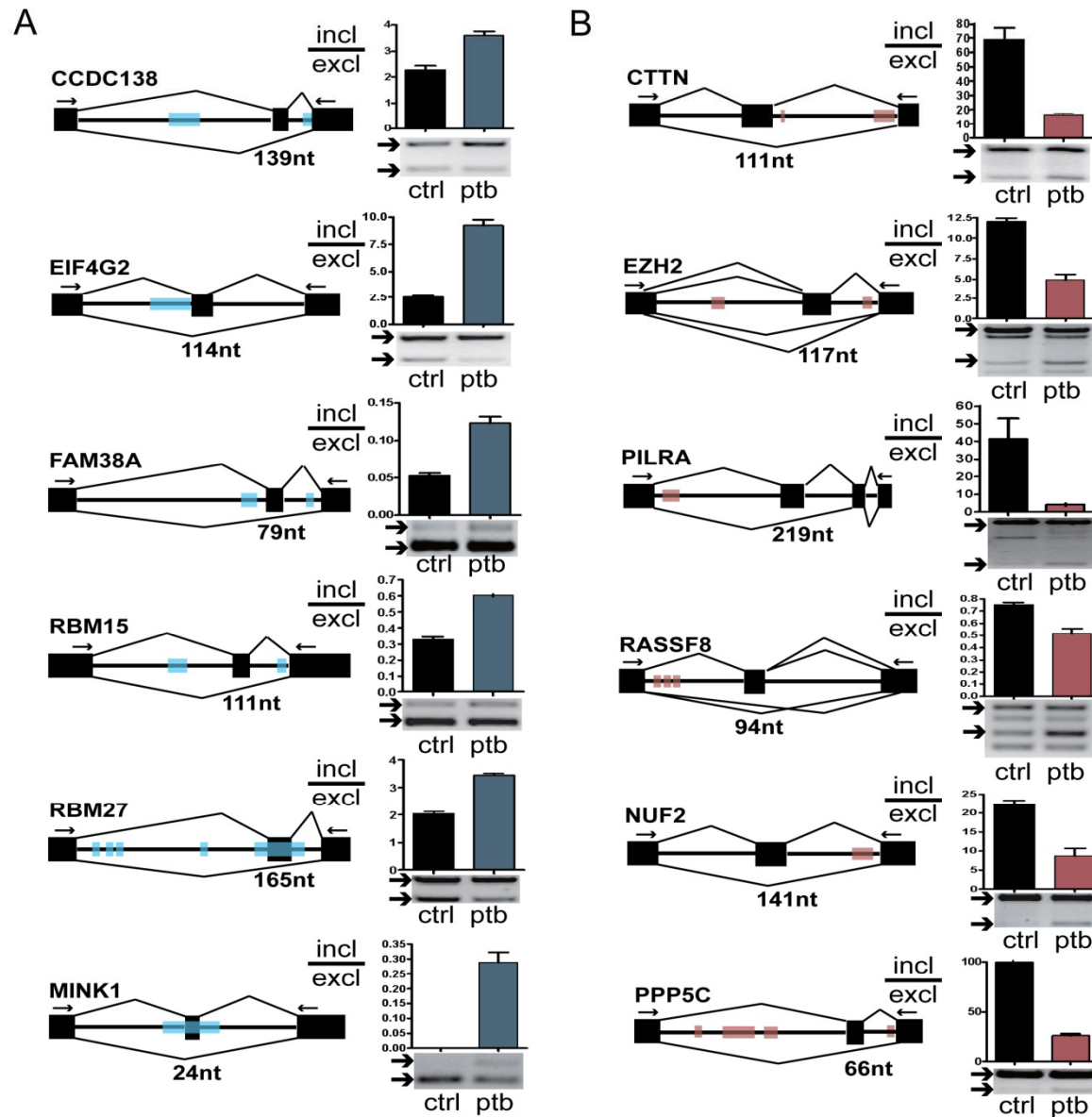


PTB-dependent repression or enhancement of alternative splicing *in vivo*



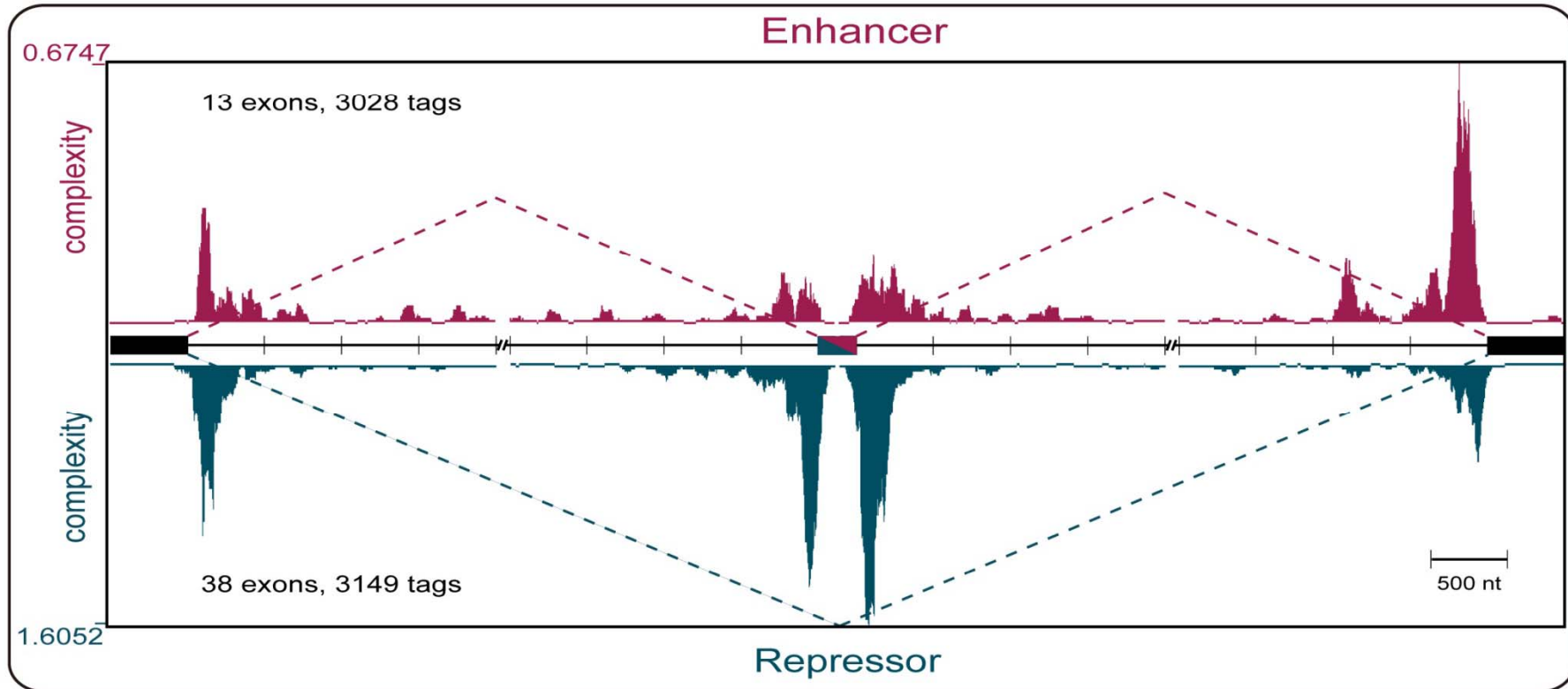
Repressor

Enhancer

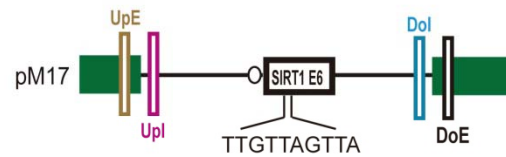


PTB composite RNA functional map

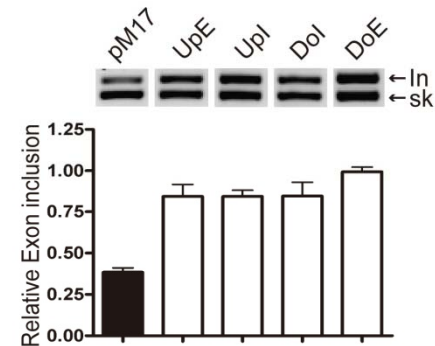
Balance of Dynamic Competition



Minigene test



UpE, Upl, Dol Insertion: CTCTCTCTTCTTCTT
 DoE insertion: TCTTCTTCTTCTTCT



Summary



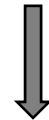
- PTB exists *in vivo* as dimer.
- PTB associates with different splicing events.
- PTB binding profile provides mechanistic insights to understand the undefined splicing regulation of many genes.
- PTB dominant binding near constitutive splice sites induces exon inclusion, whereas dominant binding close to alternative sites causes exon skipping (dynamic competition mechanism).
- This is a new positional rule for general splicing suppressors PTB to positively or negatively regulate alternative splicing in mammalian cells.

Further works



1. An integrated package for analyzing CLIP-seq data for individual RNA binding protein.

- mapping tags to genome,
- peak identification,
- saturation analysis,
- landscape comparison,
- motif analysis,
- annotation analysis,
- association with AS,
- patterns with different splicing events
- ...



One splicing code

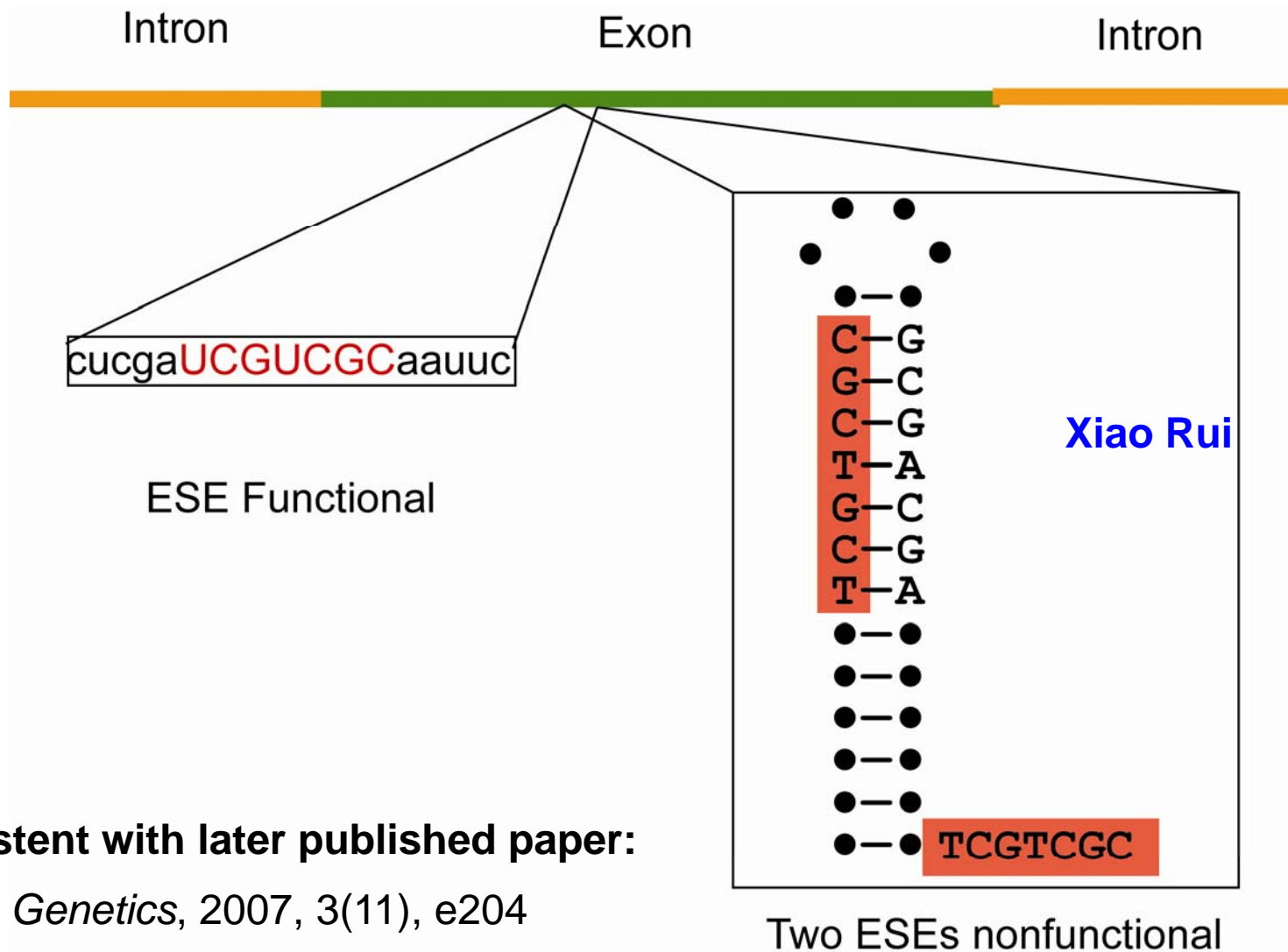
2. Combine different splicing codes from different splicing factors.



RNA design under motif constraints

Yu Zhou, Stephane Vialette, Yi Zhang, Alain Denise*

Strong secondary structure abolishes function of ESE within or nearby

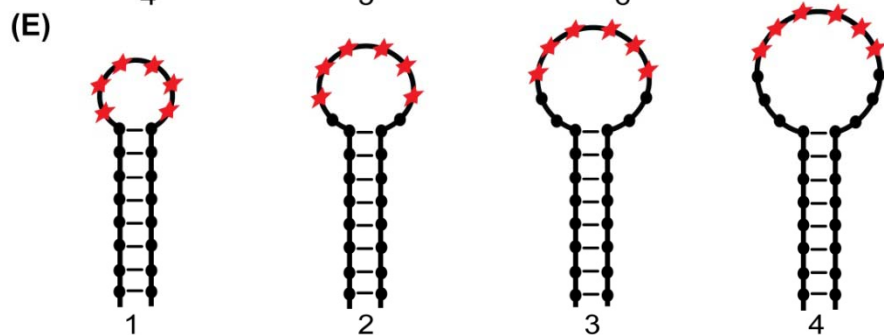
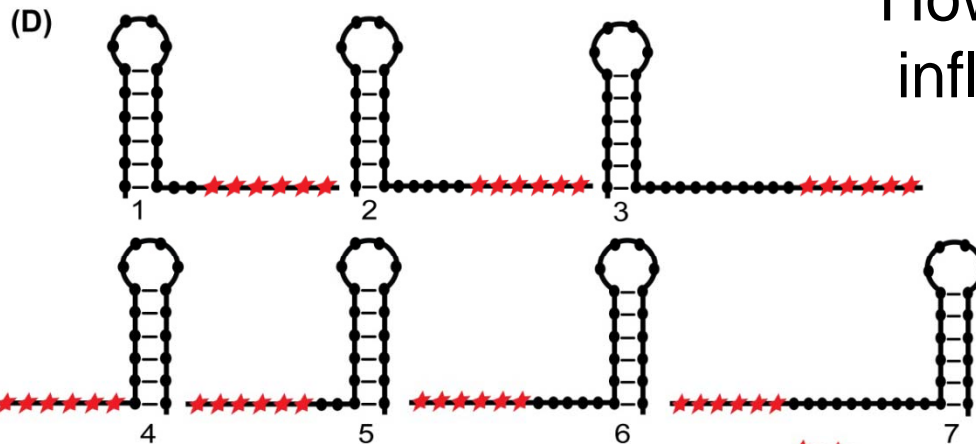
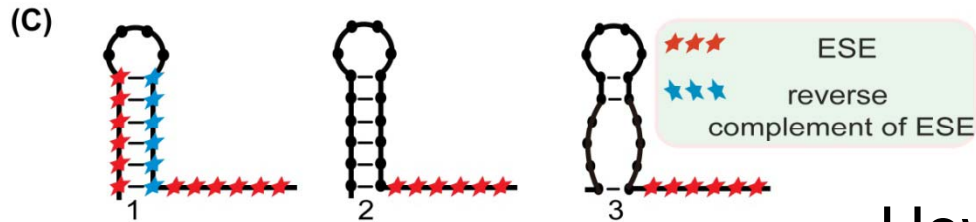
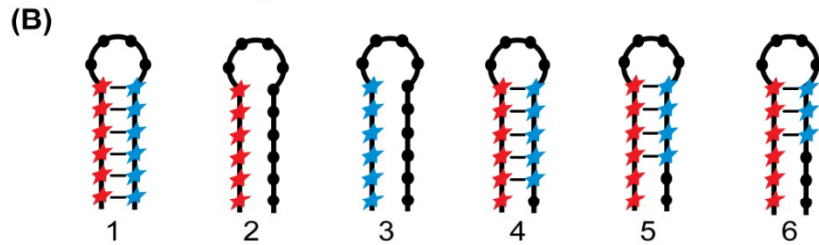


Consistent with later published paper:

Plos Genetics, 2007, 3(11), e204



Designed various mutated structures



Biological problem

How different secondary structures influence cis-element's function?

Liu Wei, Xiang-dong Fu



Computational Problem:

Design sequences folded to given structure under constraints of **mandatory motifs** and **forbidden motifs**.

Mandatory motif: one specific ESE, like UCGUCG.

Forbidden motifs: other ESEs except the one chosen to test.

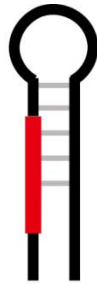


Current programs in RNA design

- RNAinverse
- RNA-SSD
- INFO-RNA

None of them can solve the constraint of forbidden motifs!

Strategy 1: (not efficient)

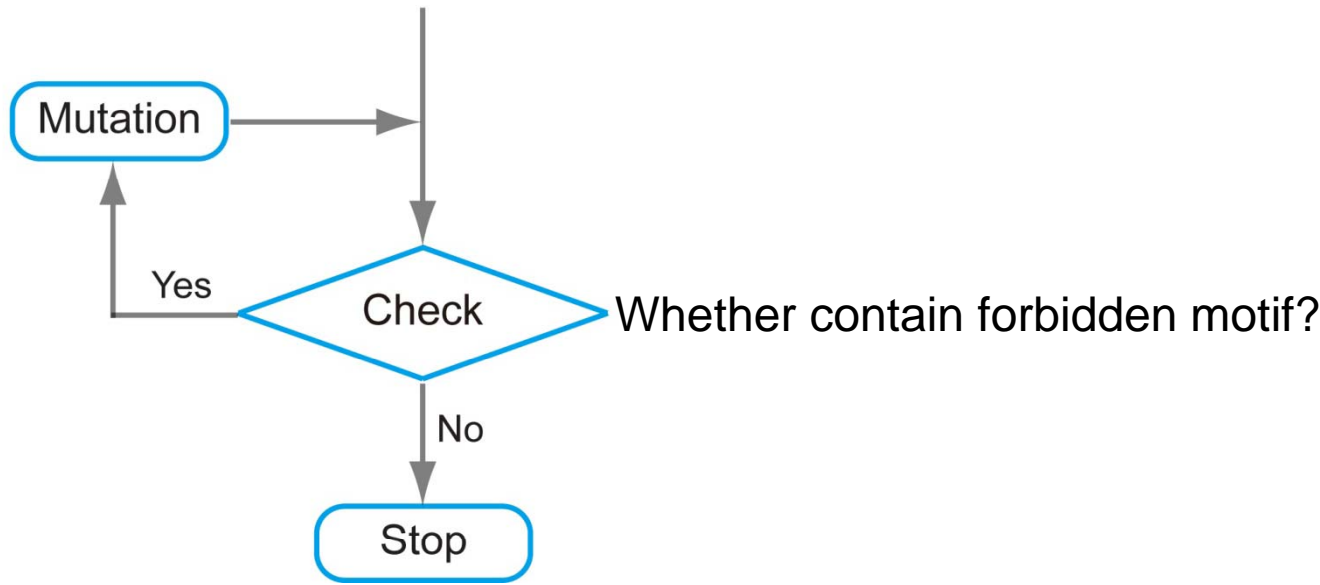


Forbidden motif set

AACCUUG
CCCCUUG
.....

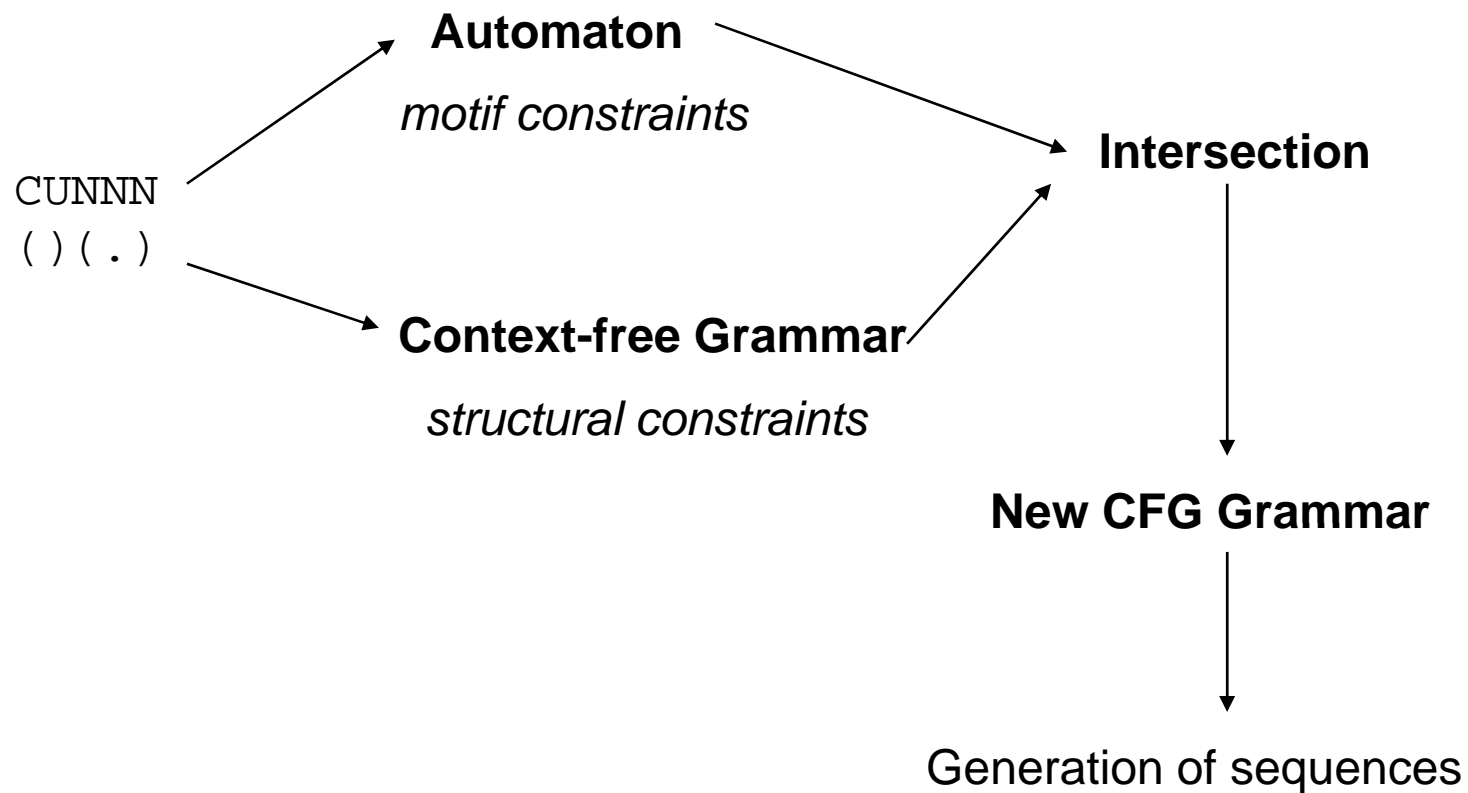


Give an initial solution satisfying mandatory motif constraint by random assignment of bases





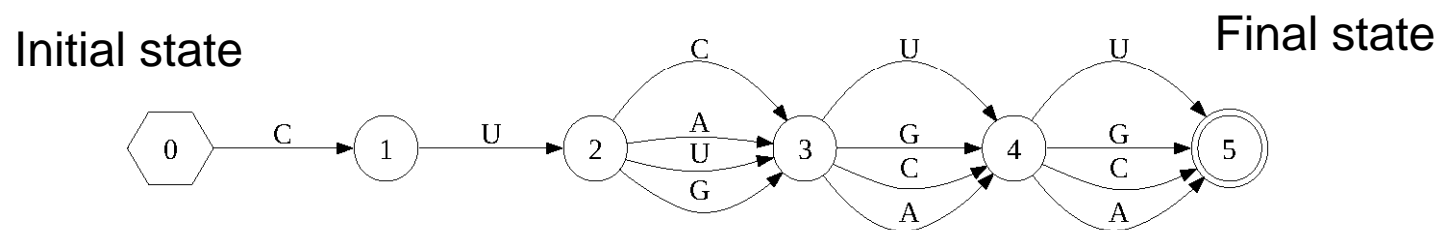
Strategy 2 (new):





A simple example: Automaton

CUNNN



C U C U U
C U A U U
C U G U U
.....

A simple example: CFG

structure: (.)()

Grammar with 17 productions (axiom state = S)

$V1 \rightarrow g \mid a \mid c \mid u$

$V2 \rightarrow gc \mid cg \mid au \mid ua \mid gu \mid ug$

$V3 \rightarrow gV1cV2 \mid cV1gV2 \mid aV1uV2 \mid uV1aV2 \mid gV1uV2 \mid uV1gV2$

$S \rightarrow V3$

$S \rightarrow V3 \rightarrow g V1 c V2 \rightarrow g g c V2 \rightarrow g g c g c$
 $(\ . \) (\)$

\swarrow

$a V1 u V2$



Intersected CFG

Grammar with 10 productions (axiom state = S)

S -> a0_V4_a5

a0_V4_a5 -> a0_V3_a5

a0_V3_a5 -> c a1_V1_a2 g a3_V2_a5

a1_V1_a2 -> u

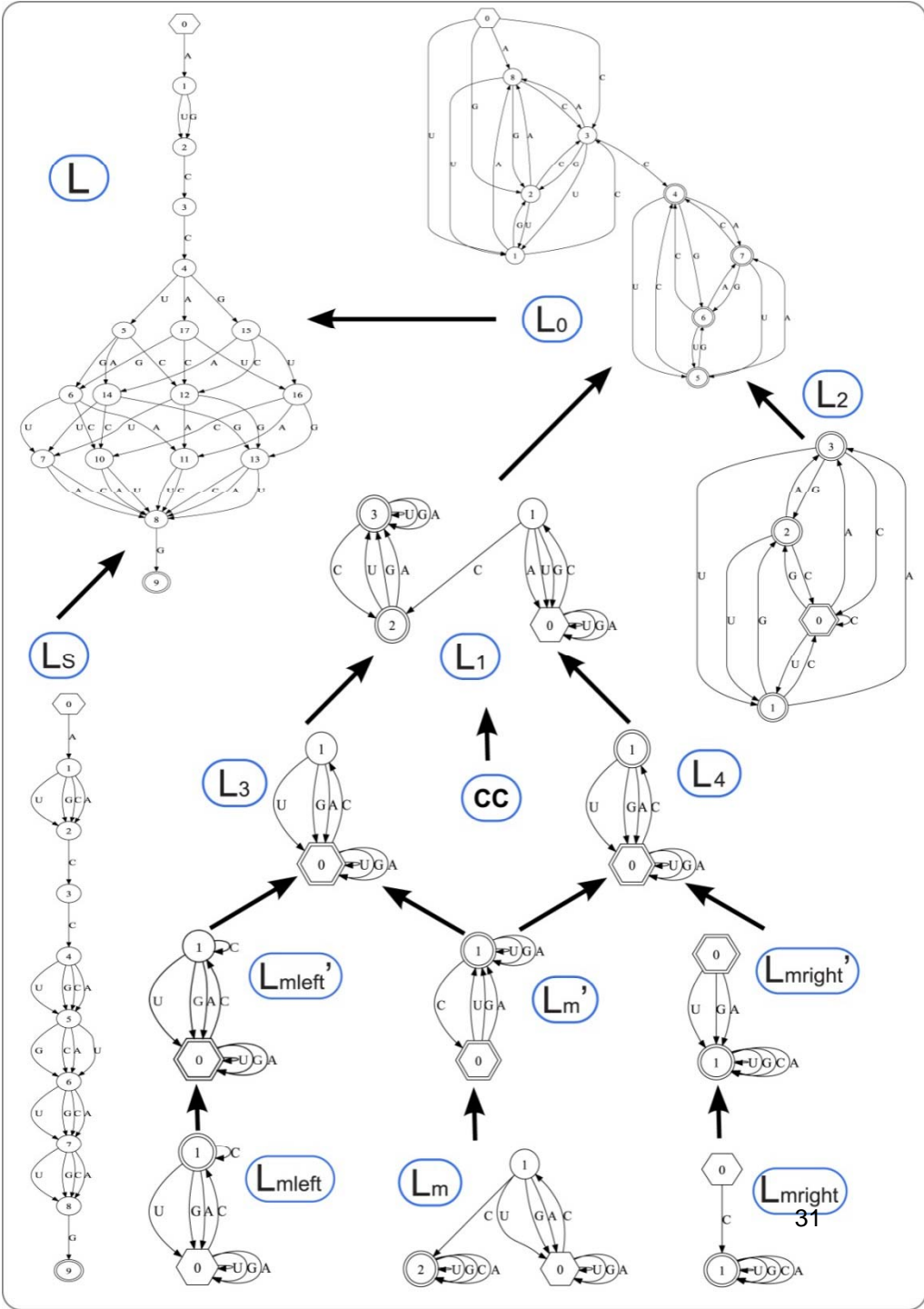
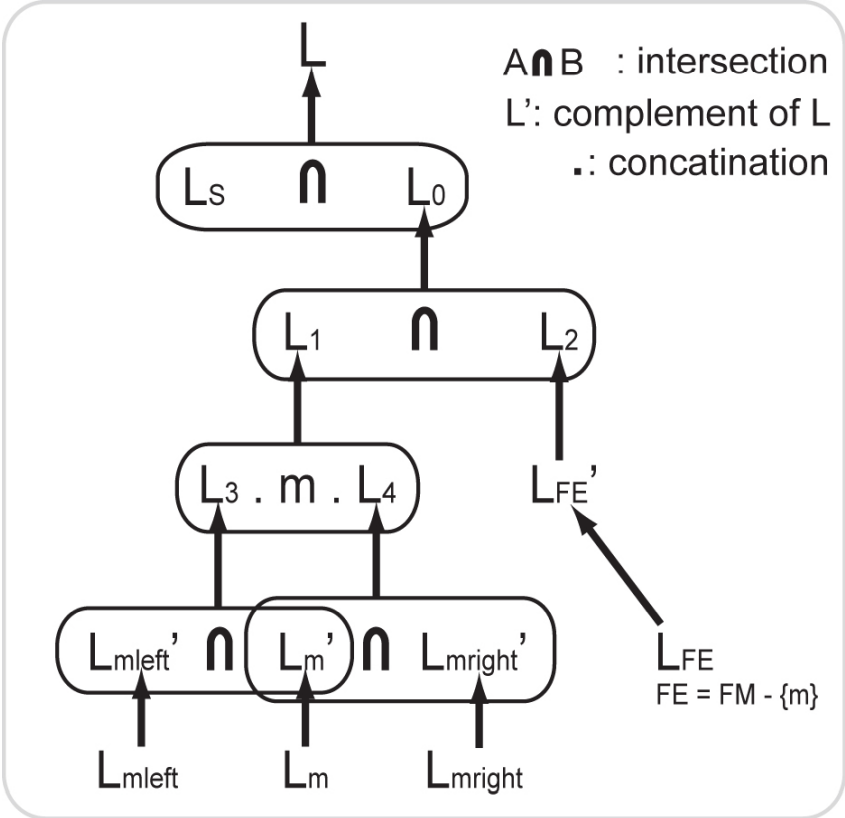
a3_V2_a5 -> gc | cg | au | ua | gu | ug

(.) ()
C U G G U
C U G C G
C U G G U
C U G G C
C U G G C
C U G U A
C U G G U

.....



A pipeline to build automaton for mandatory motif and forbidden motifs





The current program works for smaller sequences

T	Sequence	M	FSA			Grammar			#seq AG	Time(min)	
			#F	#state	#trans	GI	GA	GB		TA	TG
1	CUCGAACGCAANNNNNNNNNNNAAUUC (((((((.....)))))).....	ACGCAA	1	902	3062	14142, 17522	2040, 6405	630, 803	346	34.7m	55.3s
2	CUCGAACGCAANNNNNNNNNNNAAUUC	ACGCAA	1	902	3062	906, 3064	903, 3064	903, 3064	457647	34.3m	4.7s
3	CUCGAACGCAANNNNNNNNNNNAAUUC (((((((.....)))))).....	ACGCAA	1	902	3062	24220, 29483	3123, 10438	636, 820	1038	35.7m	1.7m
4	CUCGAACGCAANNNNNNNNNNNAAUUC (((((((.....)))))).....	ACGCAA	1	902	3062	38507, 48783	5434, 18910	645, 856	4844	34.2m	2.7m
5	CUCGAACGCAANNNNNNNNNNNAAUUC (((((((.....)))))).....	ACGCAA	1	902	3062	48087, 56849	5826, 20167	663, 914	7958	34.2m	3.3m.
6	CUCGANNNNNNUACAGANNNNNNNAAUUC (((((((.....)))))).....	UACAGA	1	222	712	16643, 47520	14946,45655	1998, 5005	13948	28.1m	1.2m
7	CUCGANNNNNNNNNNNNNNNNUACAGAAAUUC (((((((.....)))))).....	UACAGA	0	0	0	-	-	-	0	16.9m	-
8	NNNNNNNNNNUCGUCG (((.....))).....	UCGUCG	1	1117	3889	716553, 1710596	219601, 885384	18223, 34324	42234	49.4m	92m
9	UCGUCGNNNNNNNNNN (((.....)))	UCGUCG	1	939	3209	115152, 252205	32620, 122590	11469, 20532	35209	22.6m	6.3m



Further works

1. To improve the algorithm:

Do the intersection more efficiently

2. Selecting better candidates sequences according:

- UP (Unpaired Probability)
- Self-containment, (*PLOS comp. biology*, 2008)



Thanks to:

Prof. Yi Zhang and Alain Denise

Prof. Xiang-dong Fu (UCSD)

Yuanchao Xue

Stephane Vialette

Lab members in WHU and LRI