

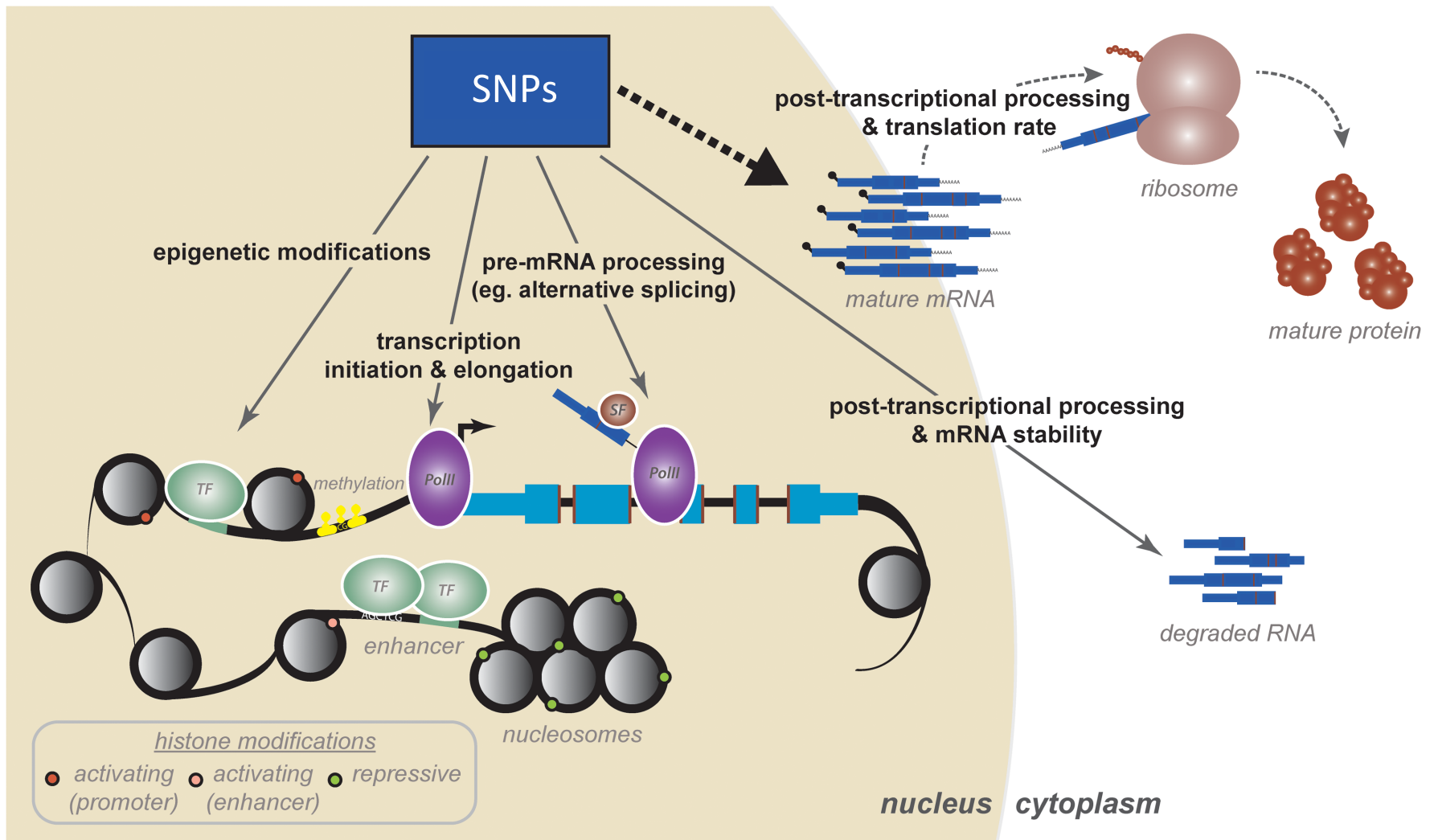
Genetic Variants and Associated Proteins in the Regulation of Pre- mRNA Splicing

Xinshu (Grace) Xiao, Ph.D.

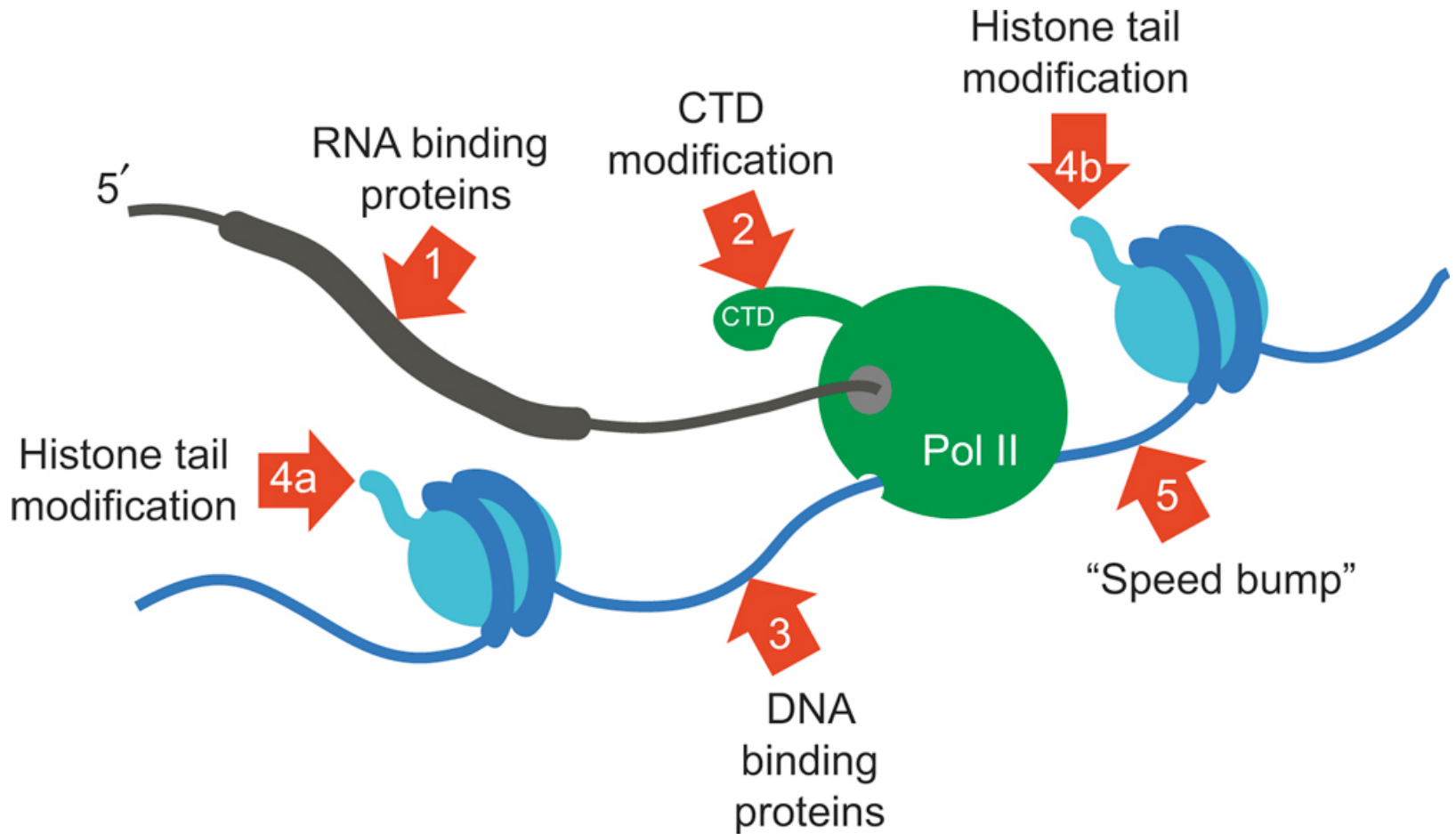
University of California, Los Angeles

2016 ENCODE Users Meeting

Impact of Genetic Variation on Gene Regulation

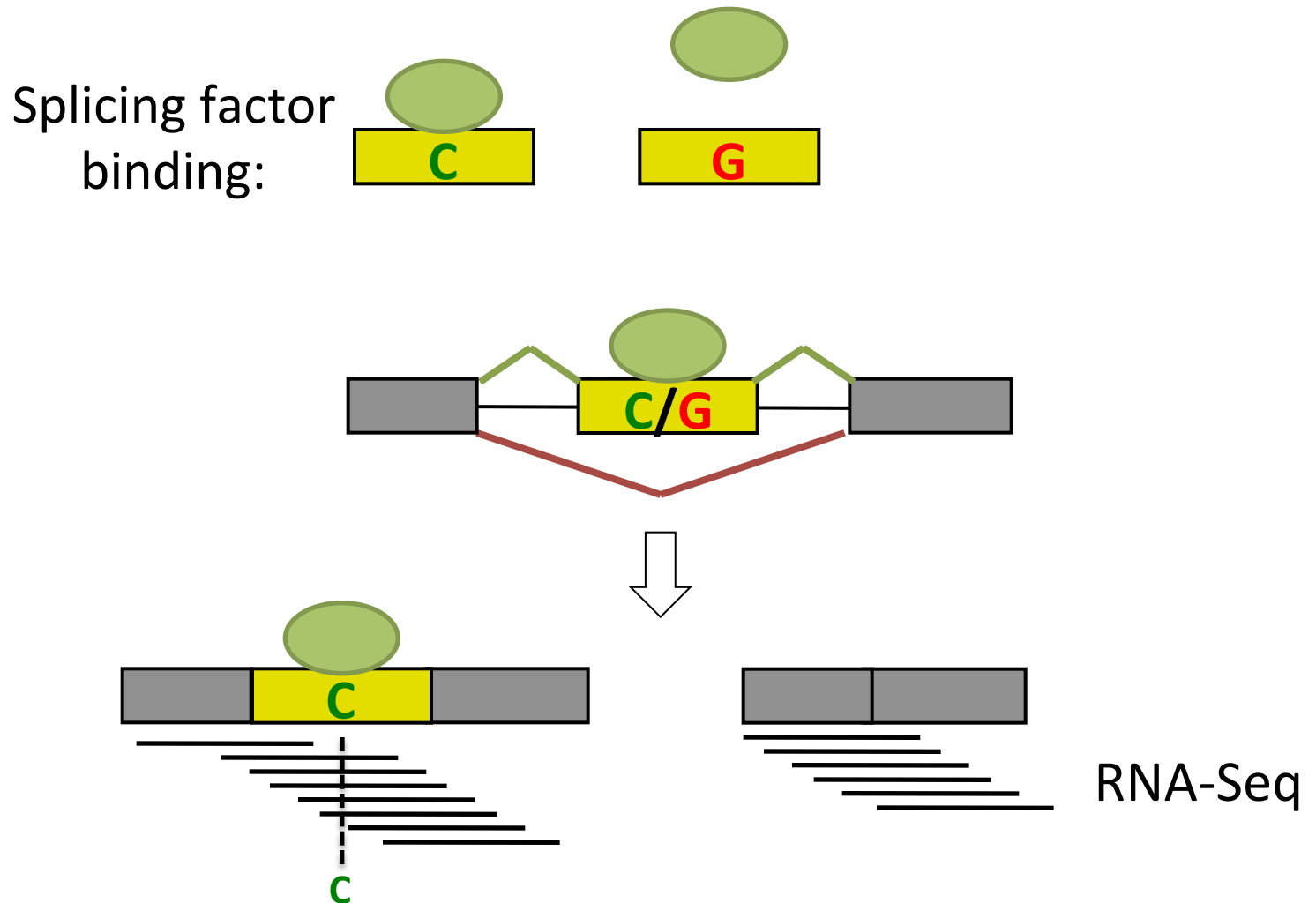


Regulation of Splicing



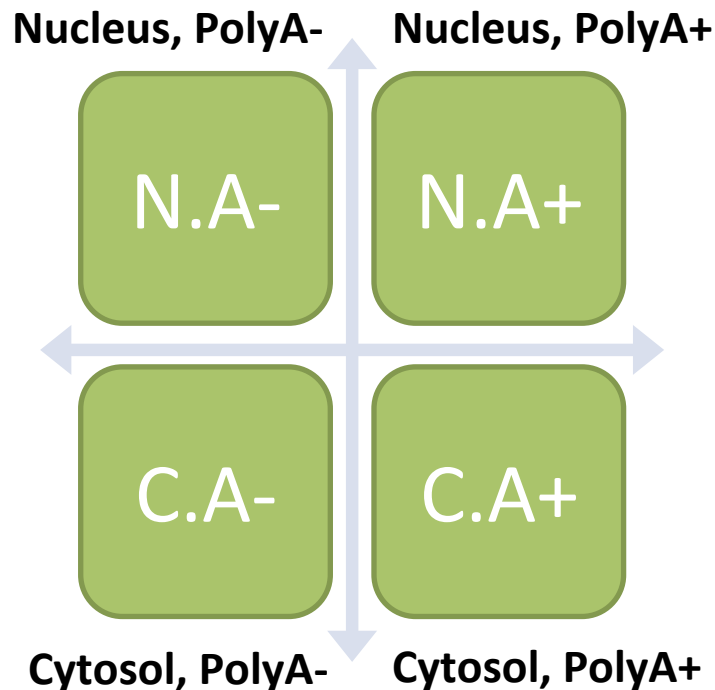
Allele-Specific RNA Splicing

Hypothetical model

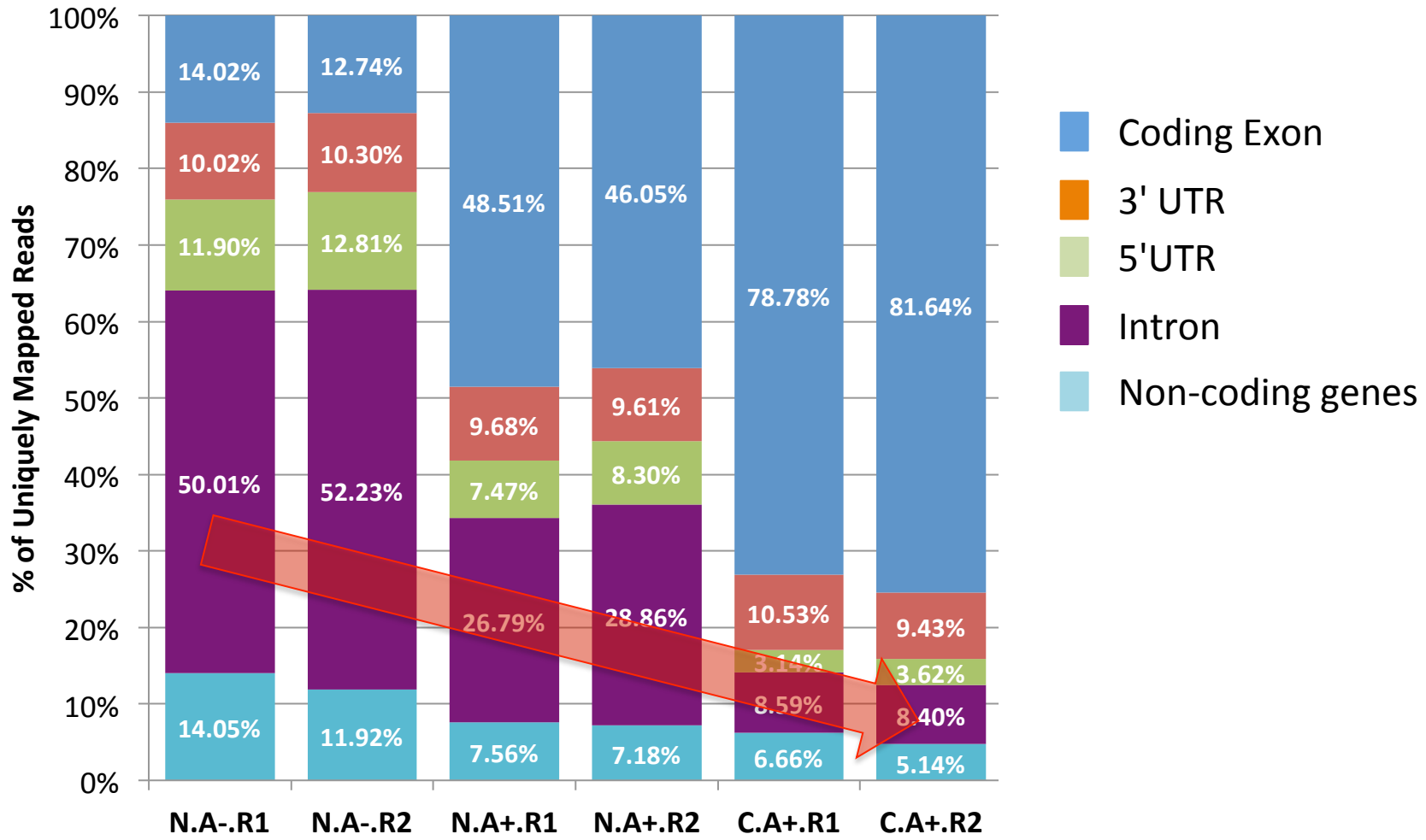


ENCODE RNA-Seq Data

- For a number of cell lines:

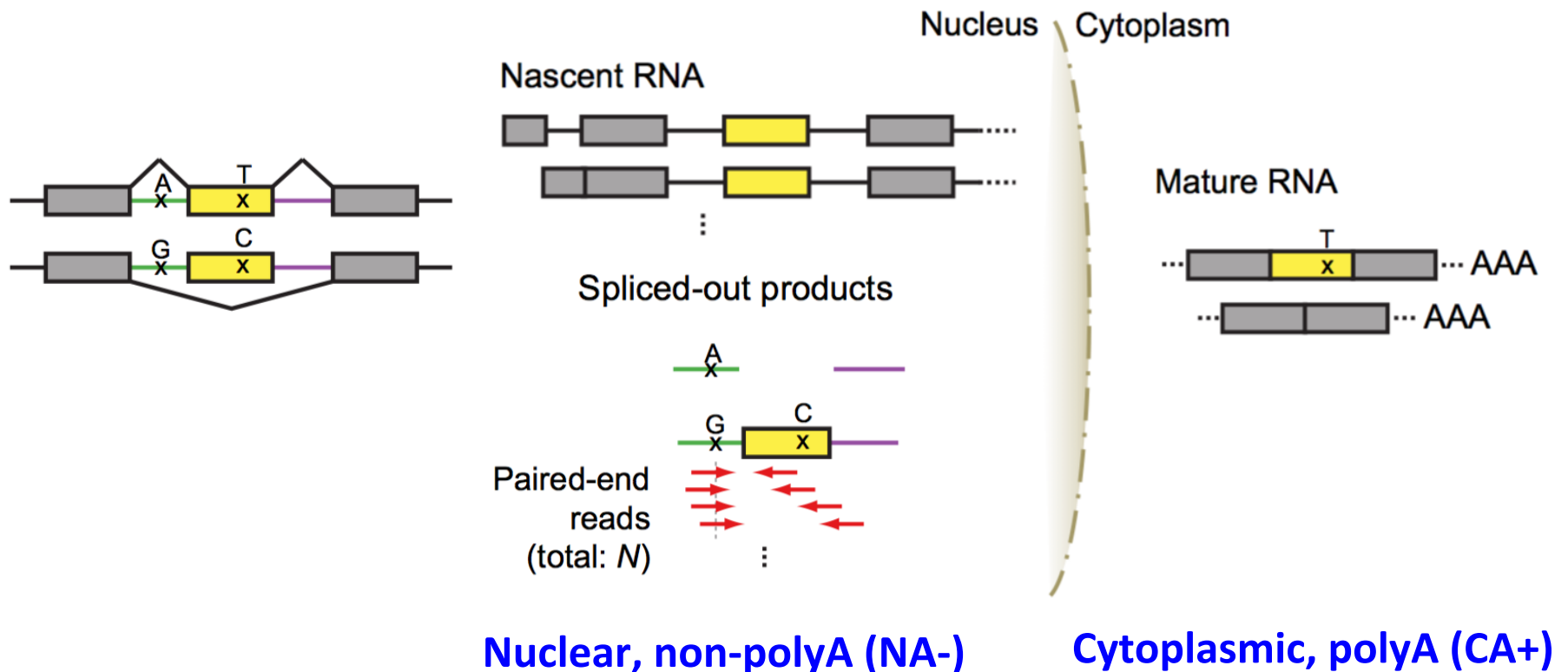


GM12878 Mapped Reads

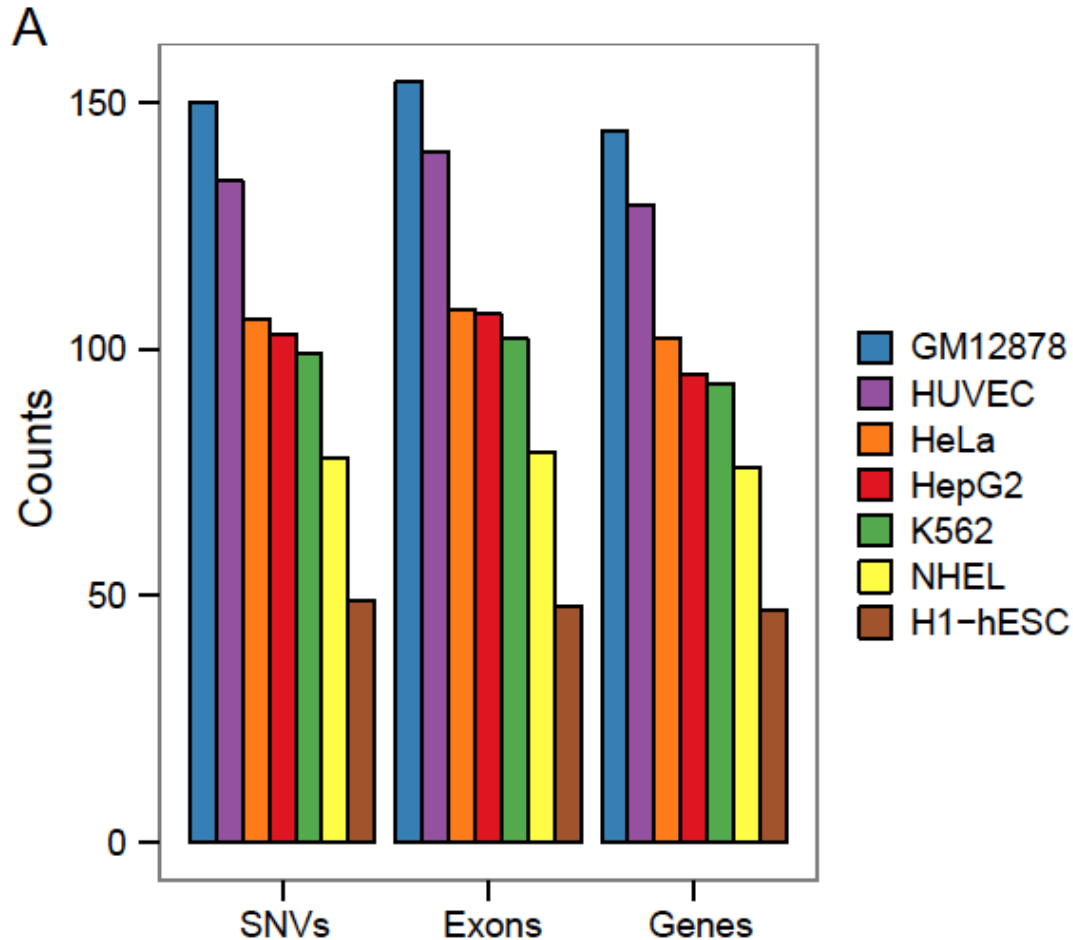


Amount of intronic reads as expected

Identification of Intronic tag SNVs for Genetically Modulated Alternative Splicing (iGMAS)



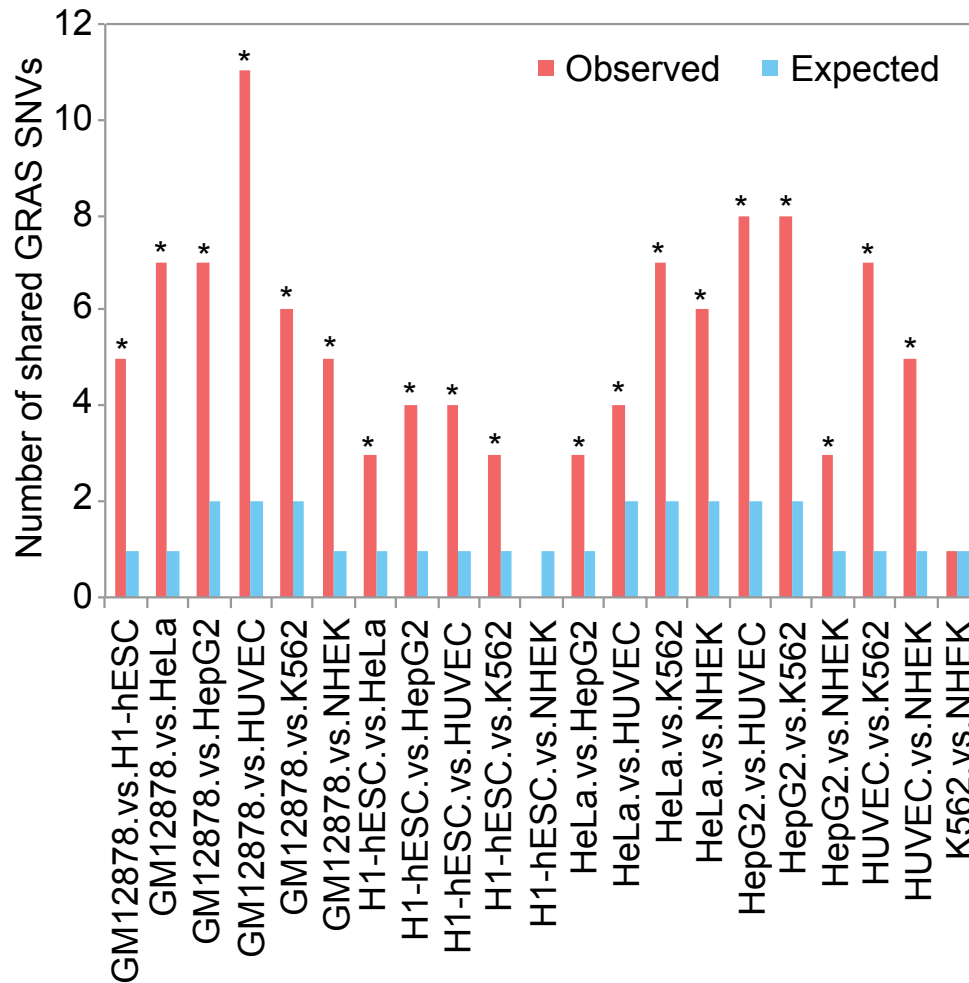
Total GMAS Events



- FDR by read count randomization: **3%**
- Experimental validation (splicing reporter assay, HeLa cells): **80% validated**
- **90% experimental validation rate if testing in multiple cells.**

GMAS events are highly cell-type independent

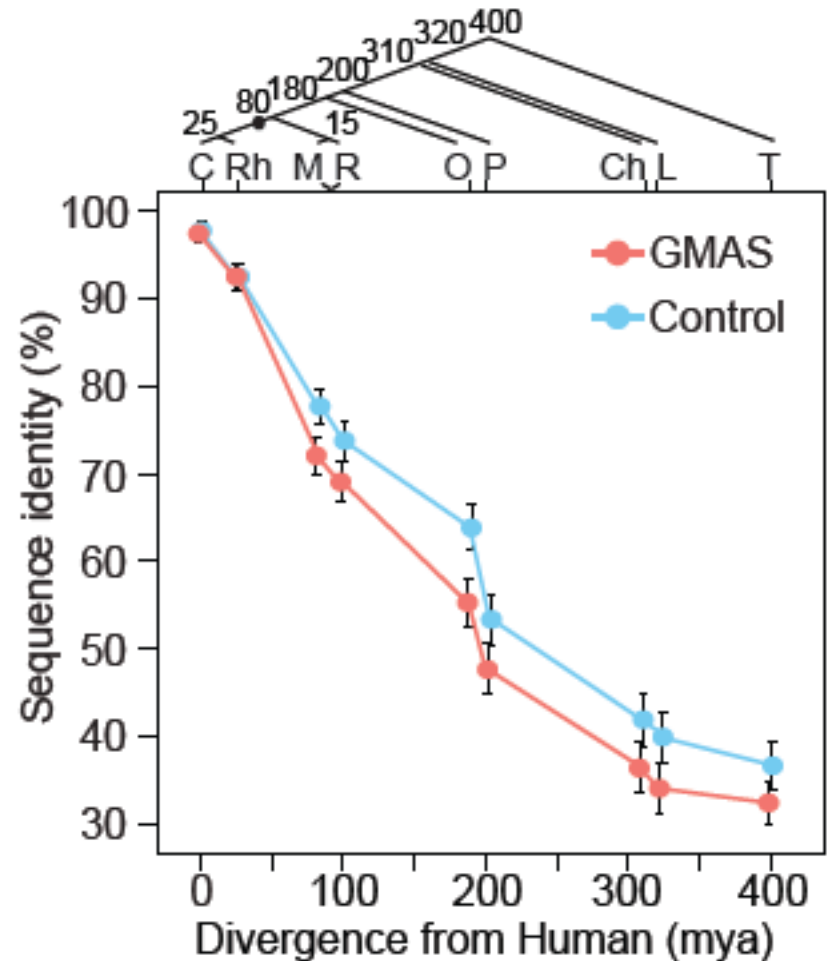
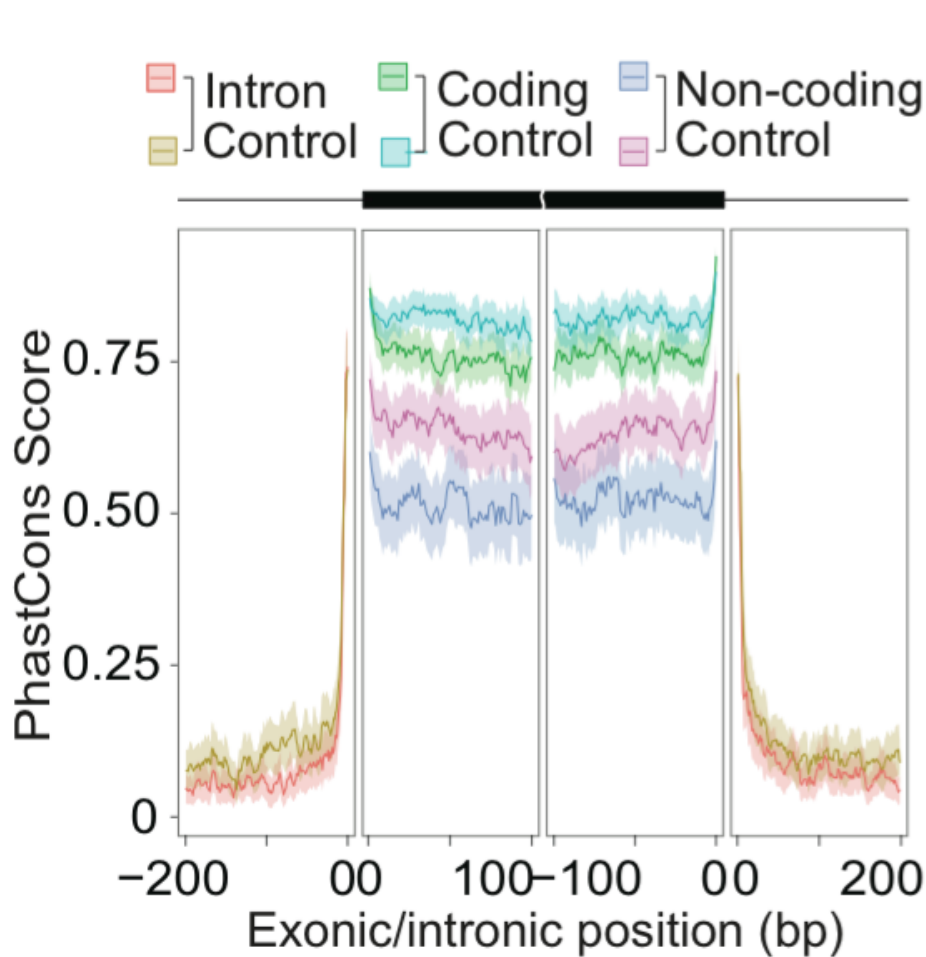
→ Indicating splicing altering variants could function ubiquitously.



* Hypergeometric test, $P < 0.05$

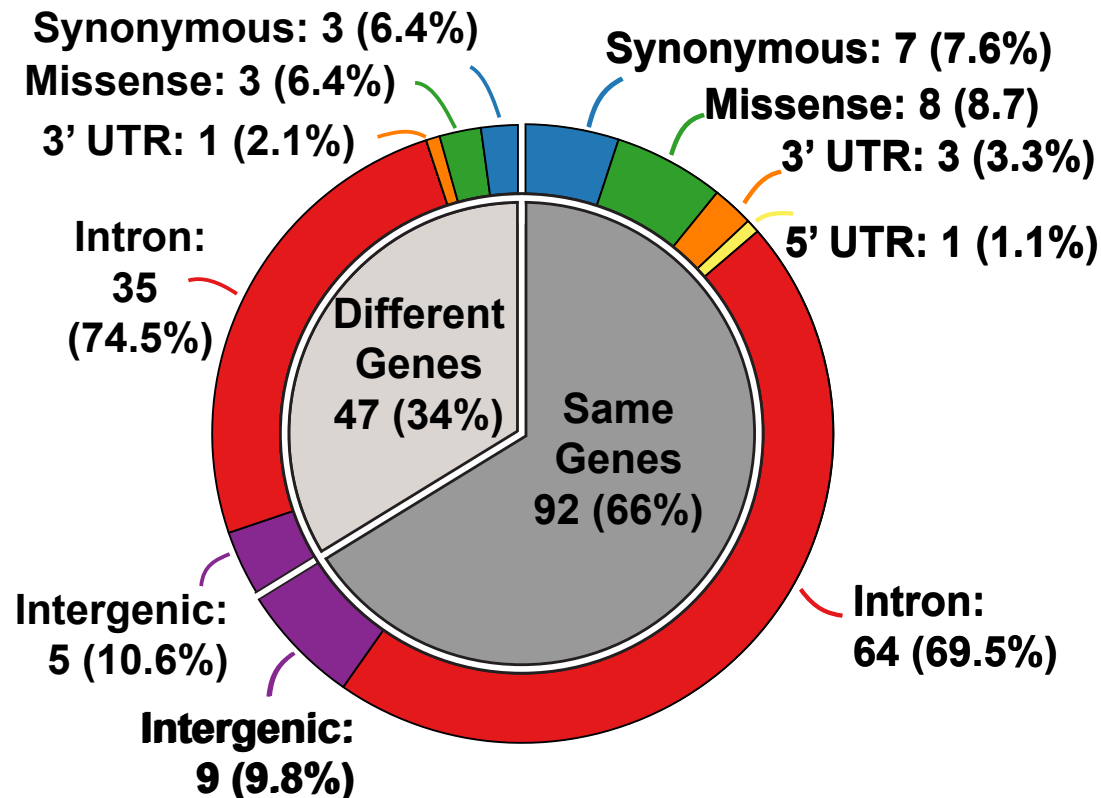
GMAS events demonstrate accelerated evolution

- GMAS regions are less conserved than controls.
- GMAS exons have accelerated sequence evolution in primate lineages.



More than 100 GMAS SNVs are in LD with GWAS SNPs

- **18% (N = 116)** GMAS SNVs in LD (and within 200kb) with GWAS SNPs (associated with cognitive performance, metabolic traits, and diseases).
- About 71% of GMAS-associated GWAS SNPs are in **introns**.



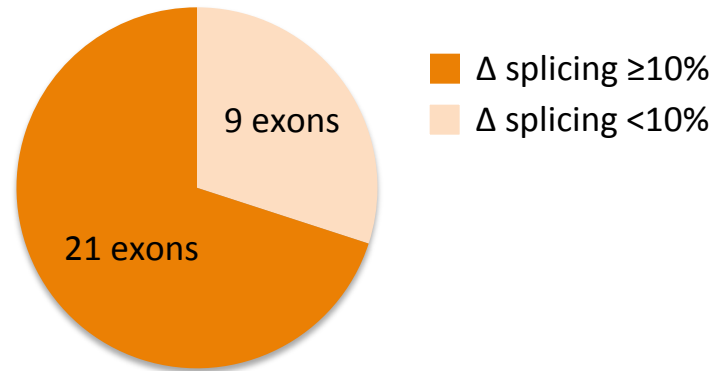
Mechanisms underlying GMAS?

Predicted Splicing Factors Regulating GMAS

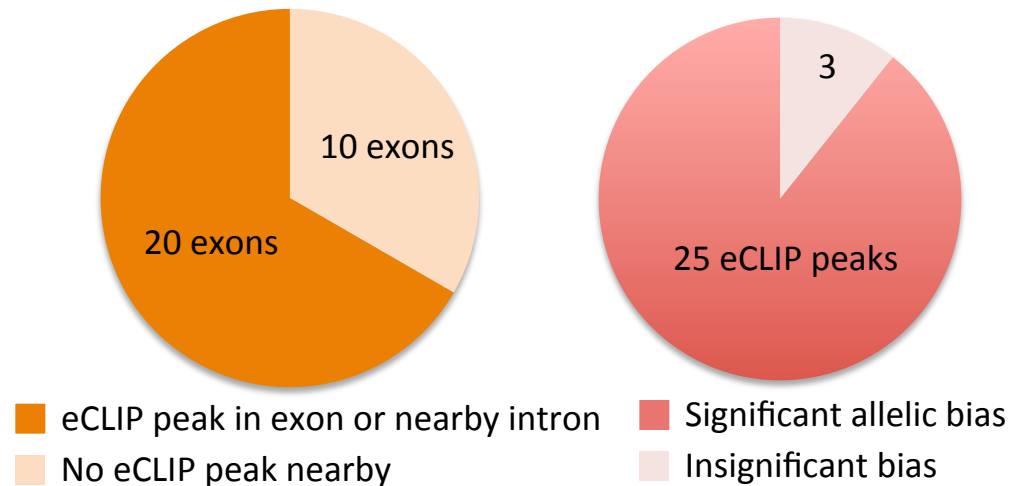
SF	# GMAS SNV
SRSF1	31
SNRPA	15
SRSF2	14
SNRPB2	14
ELAVL2	12
PCBP4	9
PCBP3	9
RBM45	8
RBM4	6
ELAVL3	6
RBM4B	6
QKI	6
CELF3	5
HNRNPA3	5
HNRNPA1	5

SRSF1 Target exons: Validation using ENCODE data

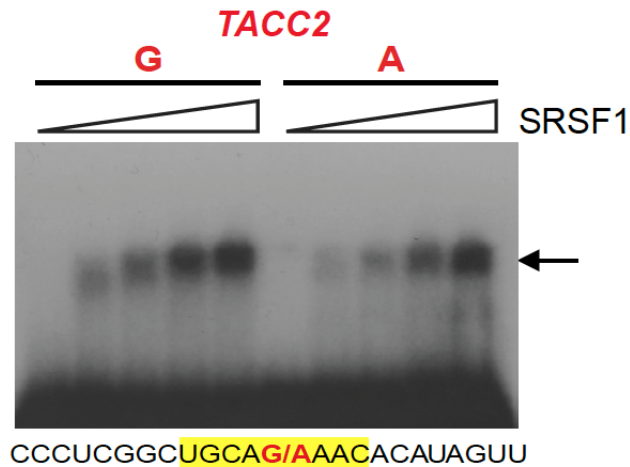
ENCODE RNA-Seq (SRSF1 knockdown)



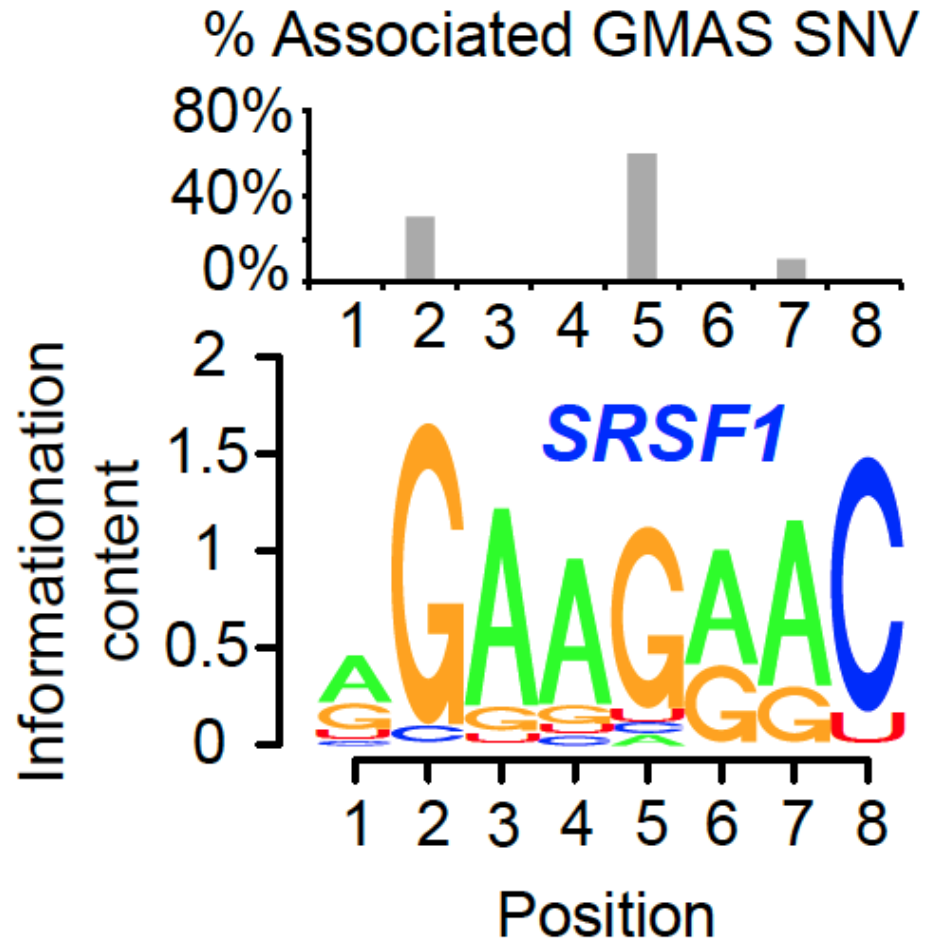
ENCODE eCLIP-Seq (SRSF1)



SRSF1: A Potent Regulator of GMAS



Gel shift assays with recombinant SRSF1

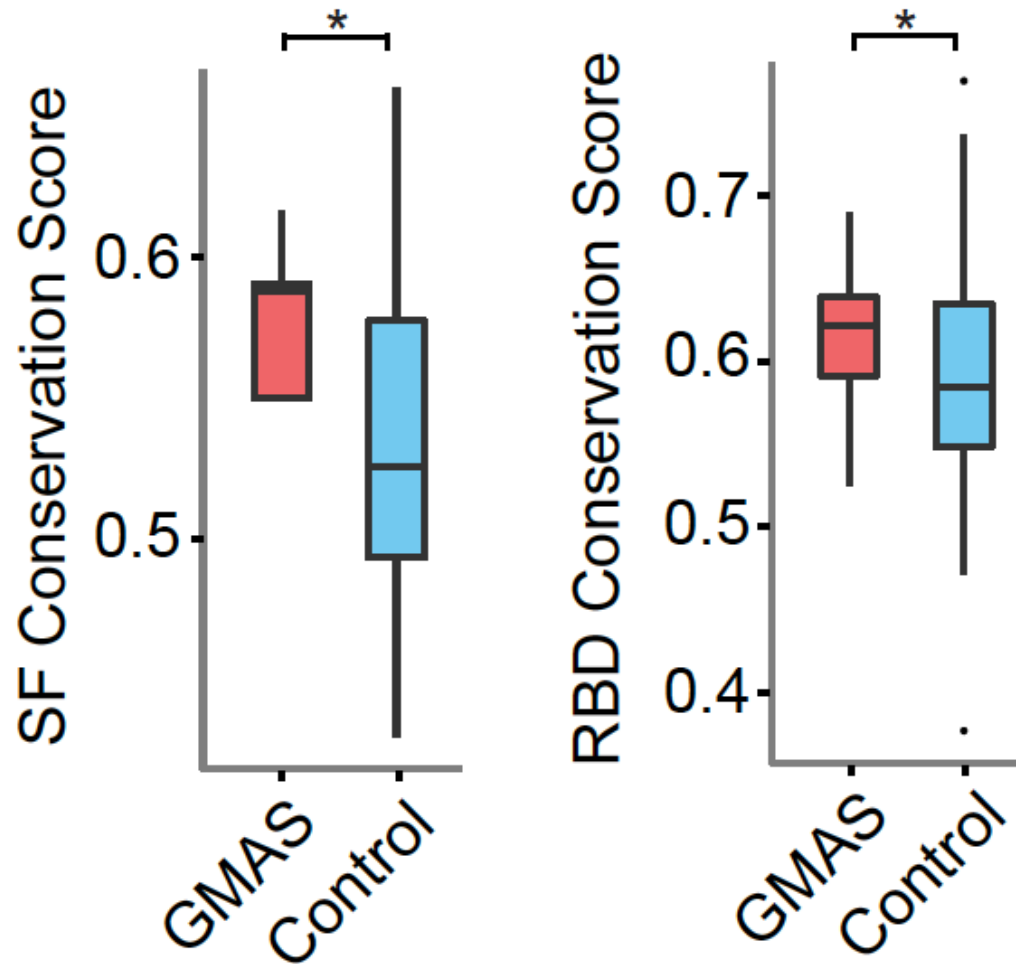


Predicted Splicing Factors Regulating GMAS

SF	# GMAS SNV
SRSF1	31
SNRPA	15
SRSF2	14
SNRPB2	14
ELAVL2	12
PCBP4	9
PCBP3	9
RBM45	8
RBM4	6
ELAVL3	6
RBM4B	6
QKI	6
CELF3	5
HNRNPA3	5
HNRNPA1	5

Ongoing:
ENCODE RBP KD,
eCLIP data used to
validate allele-specific
splicing, and to discover
novel events

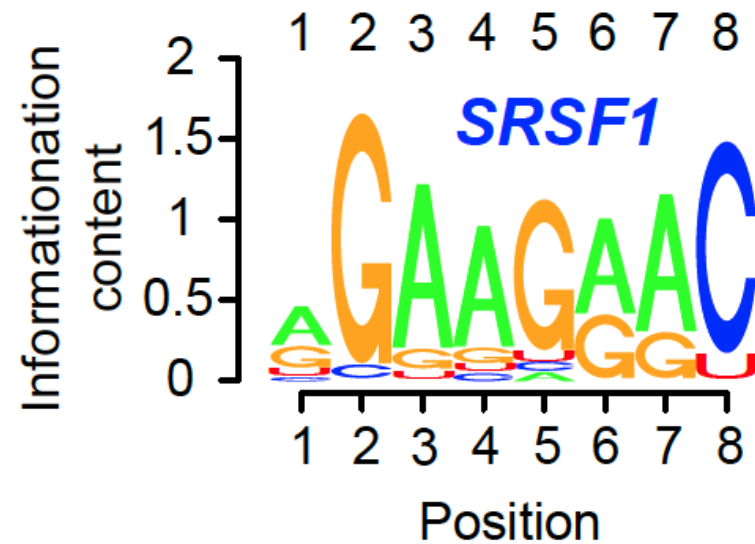
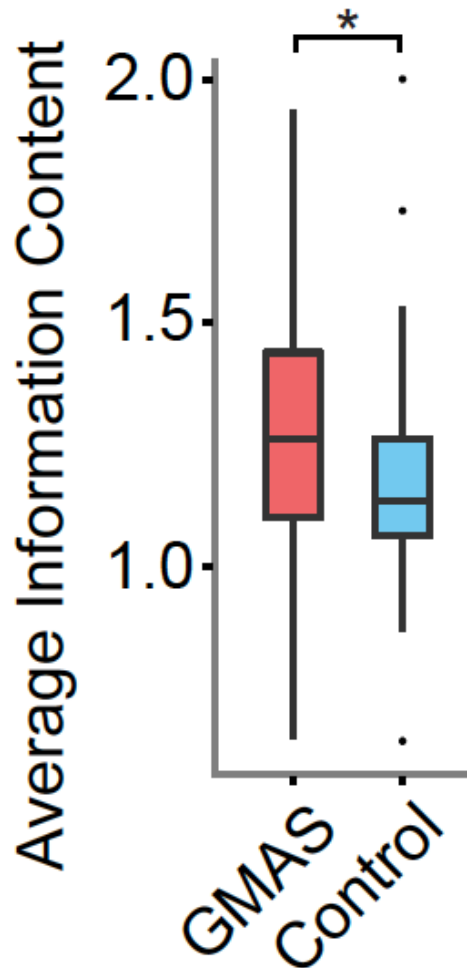
GMAS-Related Splicing Factors



SF: splicing factor
RBD: RNA binding domain

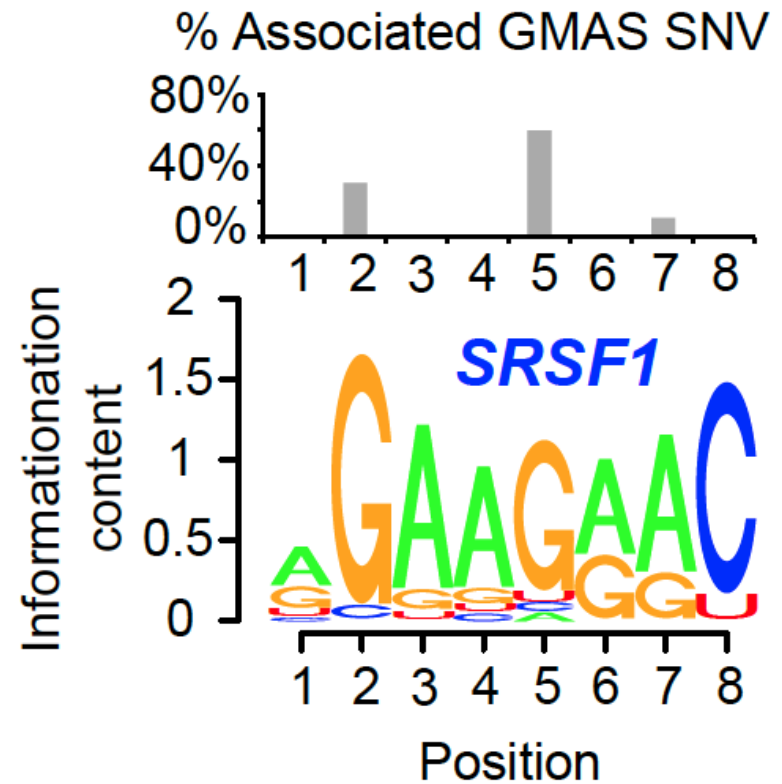
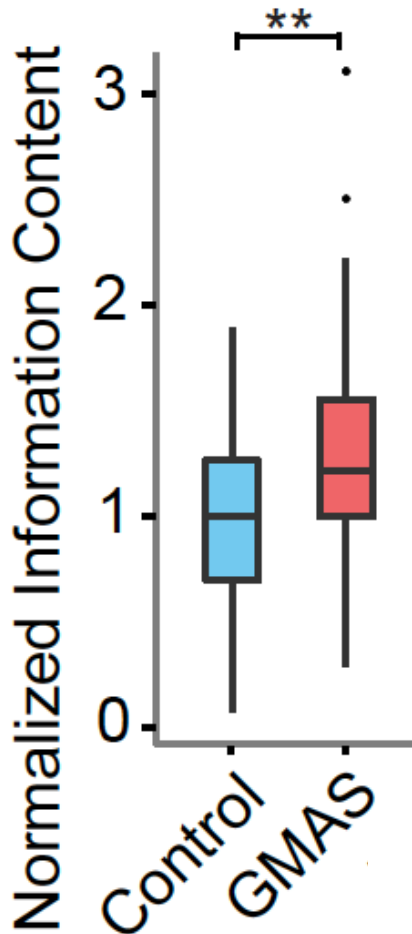
* $p < 0.05$, Wilcoxon Rank Sum test

Consensus Motif of Splicing Factors



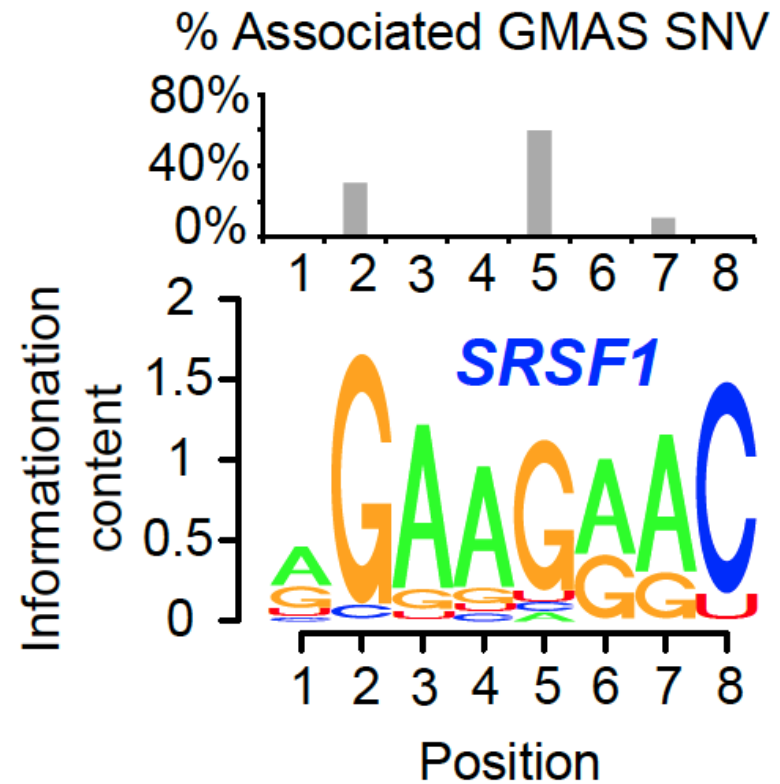
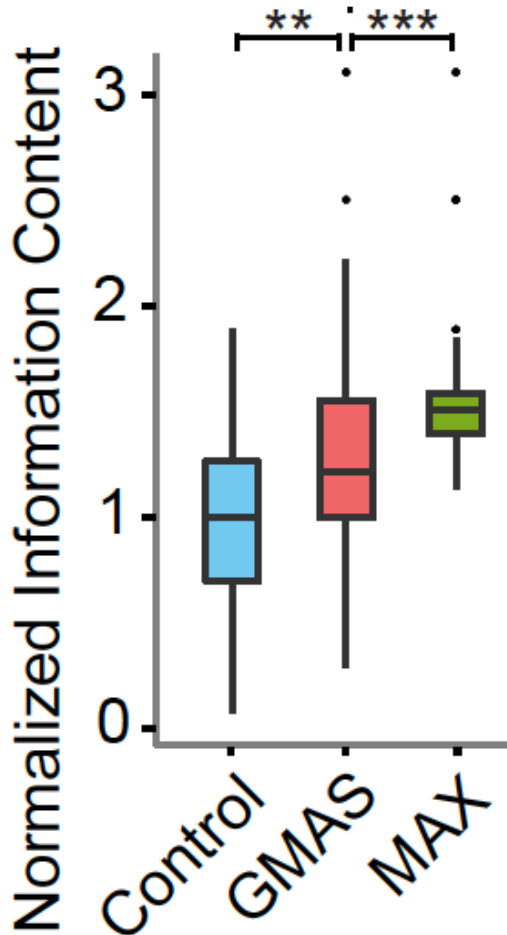
* $p < 0.01$, ** $p < 1e-5$, *** $p < 1e-10$
Wilcoxon Rank Sum test

Consensus Motif of Splicing Factors



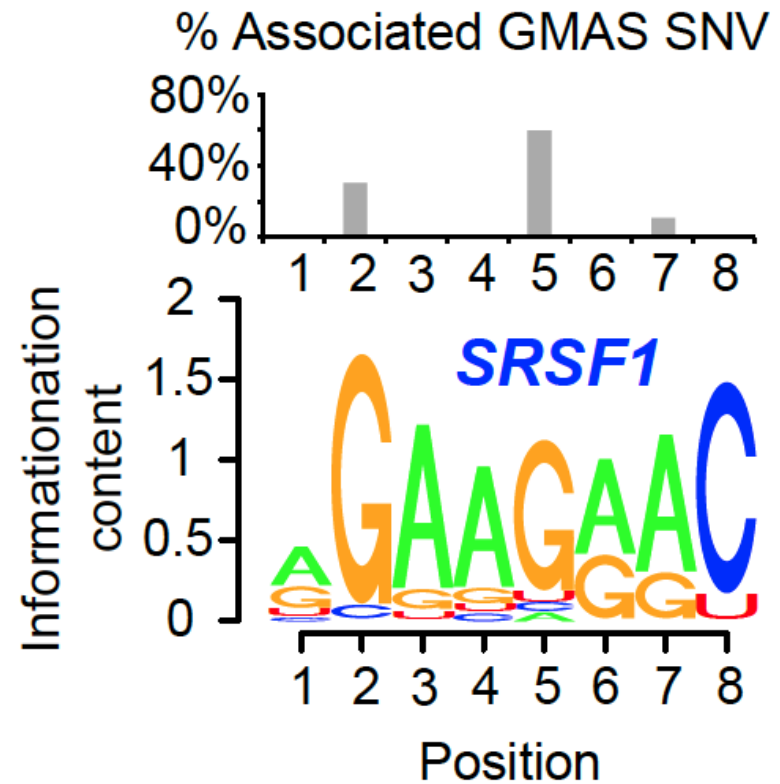
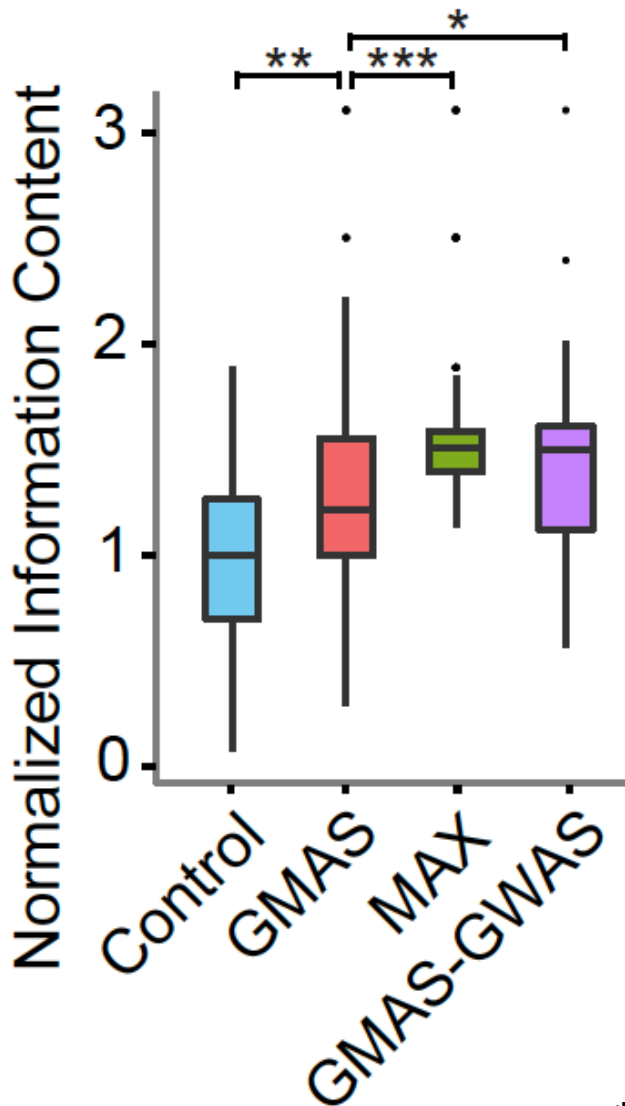
. <0.01, ** $p < 1e-5$, *** $p < 1e-10$
Wilcoxon Rank Sum test

Consensus Motif of Splicing Factors



* $p < 0.01$, ** $p < 1e-5$, *** $p < 1e-10$
Wilcoxon Rank Sum test

Consensus Motif of Splicing Factors

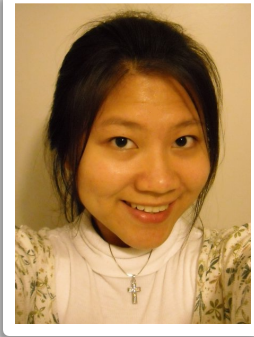


* $p < 0.01$, ** $p < 1e-5$, *** $p < 1e-10$
Wilcoxon Rank Sum test

Summary

- Using ENCODE RNA-Seq data, we identified a total of 622 GMAS exons.
- GMAS events: often regulated by splicing factors via allele-specific binding.
- GMAS exons demonstrated accelerated evolution, regulated by highly conserved proteins.
- ENCODE RBP knockdown and eCLIP data essential for inferring regulatory roles of proteins in GMAS.

Acknowledgements



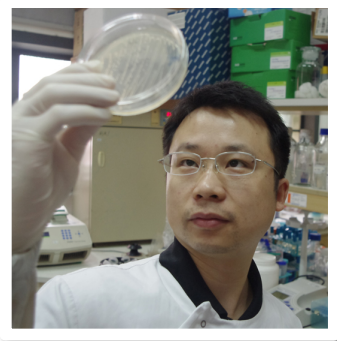
Esther Hsiao



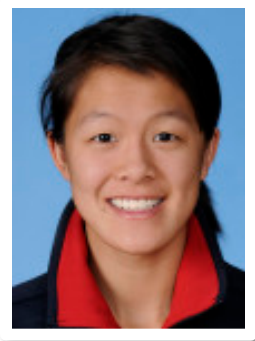
Dr. Cyrus Chan



Dr. Jae Bahn



Dr. Xianzhi Lin



Rena Wang

ENCODE Data Production

Brent Graveley group

Gene Yeo group

Tom Gingeras group

Funding:

NIH/NHGRI

Alfred P. Sloan Foundation

March of Dimes Foundation



National Human
Genome Research
Institute

