

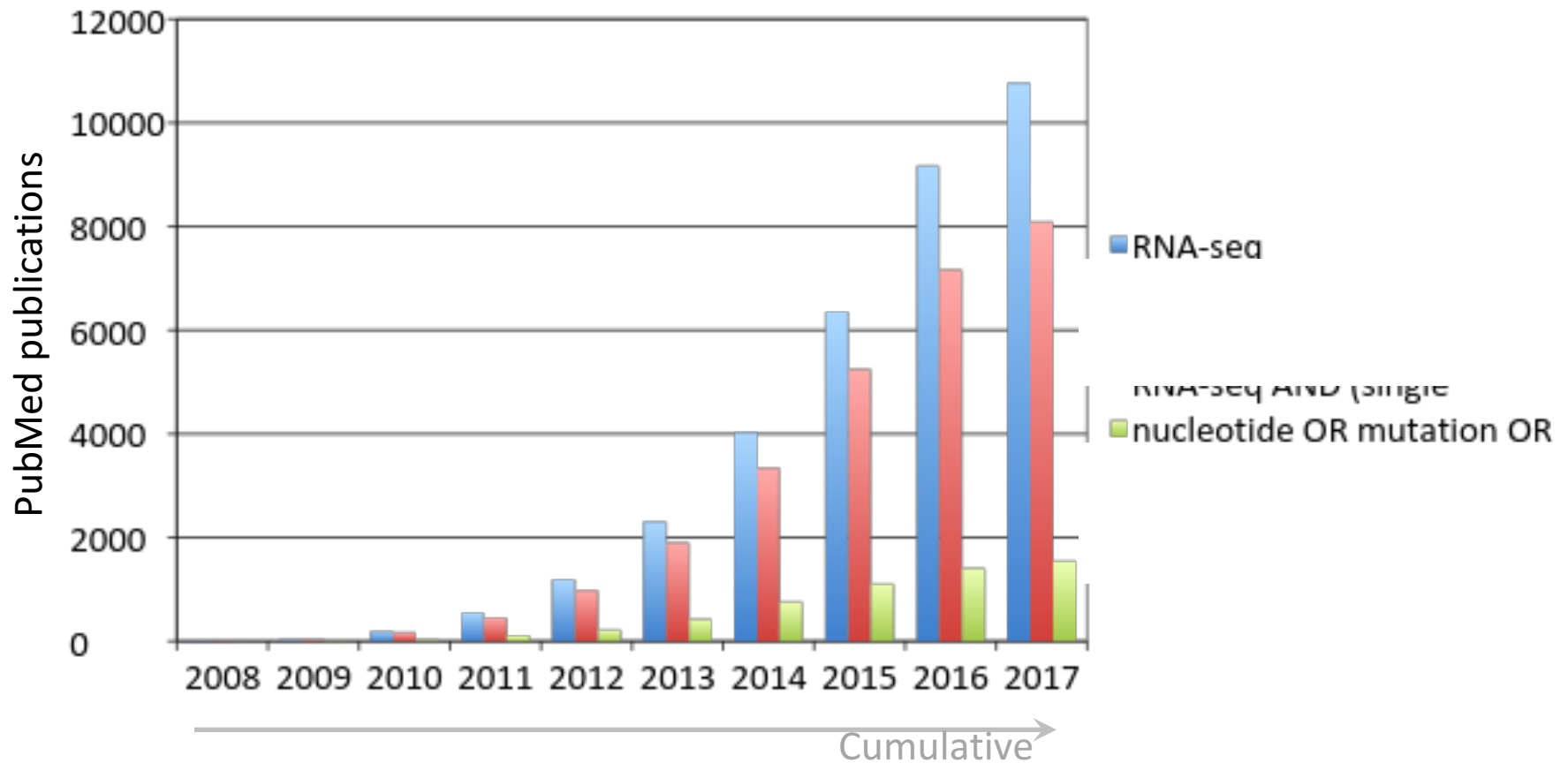
Informatic Tools for Single-Nucleotide Analysis of Cancer RNA-Seq

Xinshu (Grace) Xiao

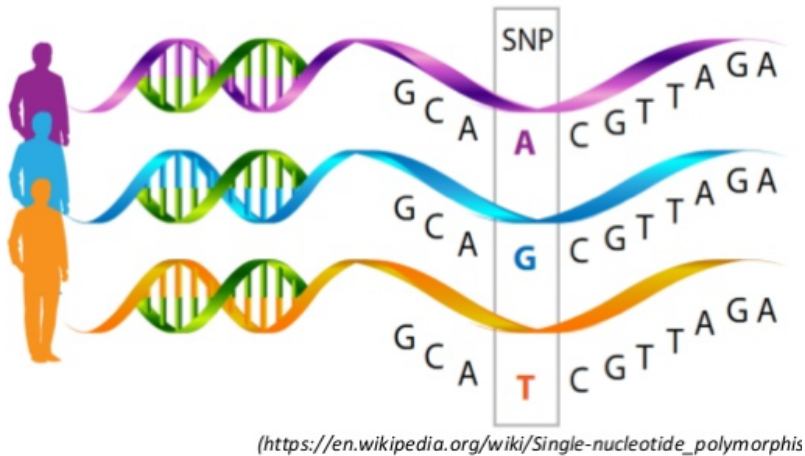
Professor

Integrative Biology and Physiology
University of California, Los Angeles

Single-Nucleotide Analysis of RNA-Seq Data



Why Single-Nucleotide Analysis?



Expressed SNPs/mutations

Complementary

genotyping

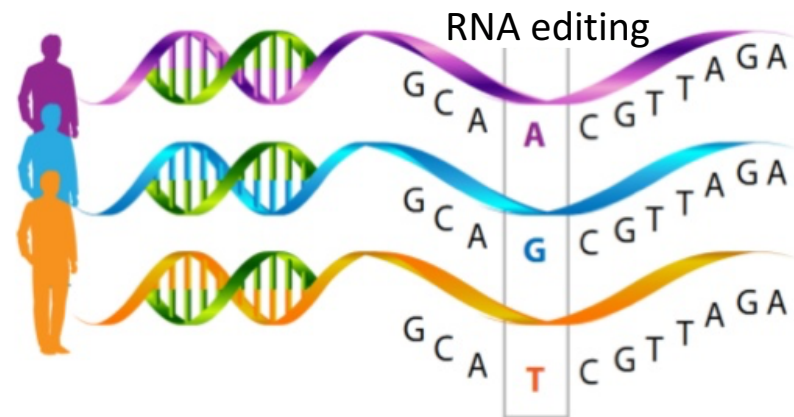
Prediction of mutation

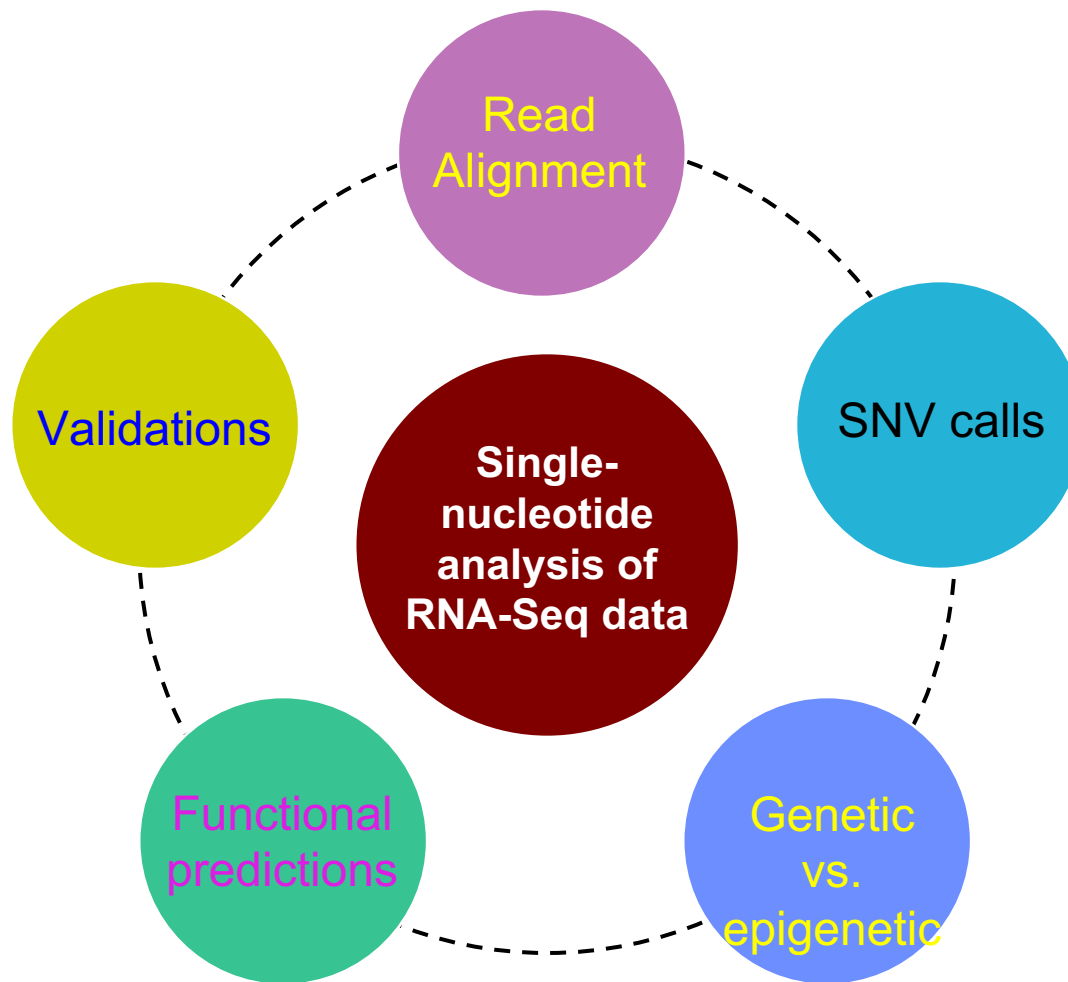
function

RNA editing

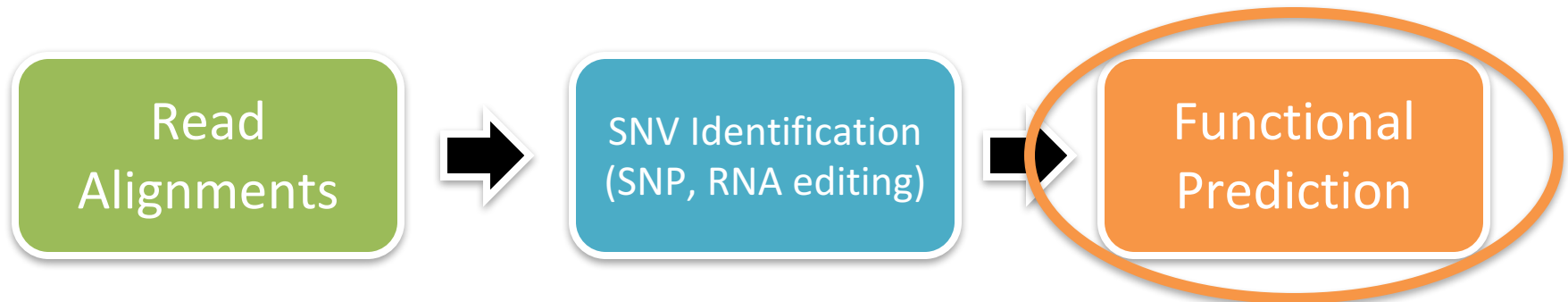
Disease association

Widespread alteration in
cancer





RNA-Seq Based SNV Analysis



RASER

Read Aligner for SNPs
and Editing sites in RNA
(*Bioinformatics*, 2015)

GIREMI

Genome-independent
Identification of RNA
Editing by Mutual
Information
(*Nature Methods*, 2015)

deSNP

De novo Expressed SNP
identification using
RNA-Seq alone
(*In prep*, 2017)

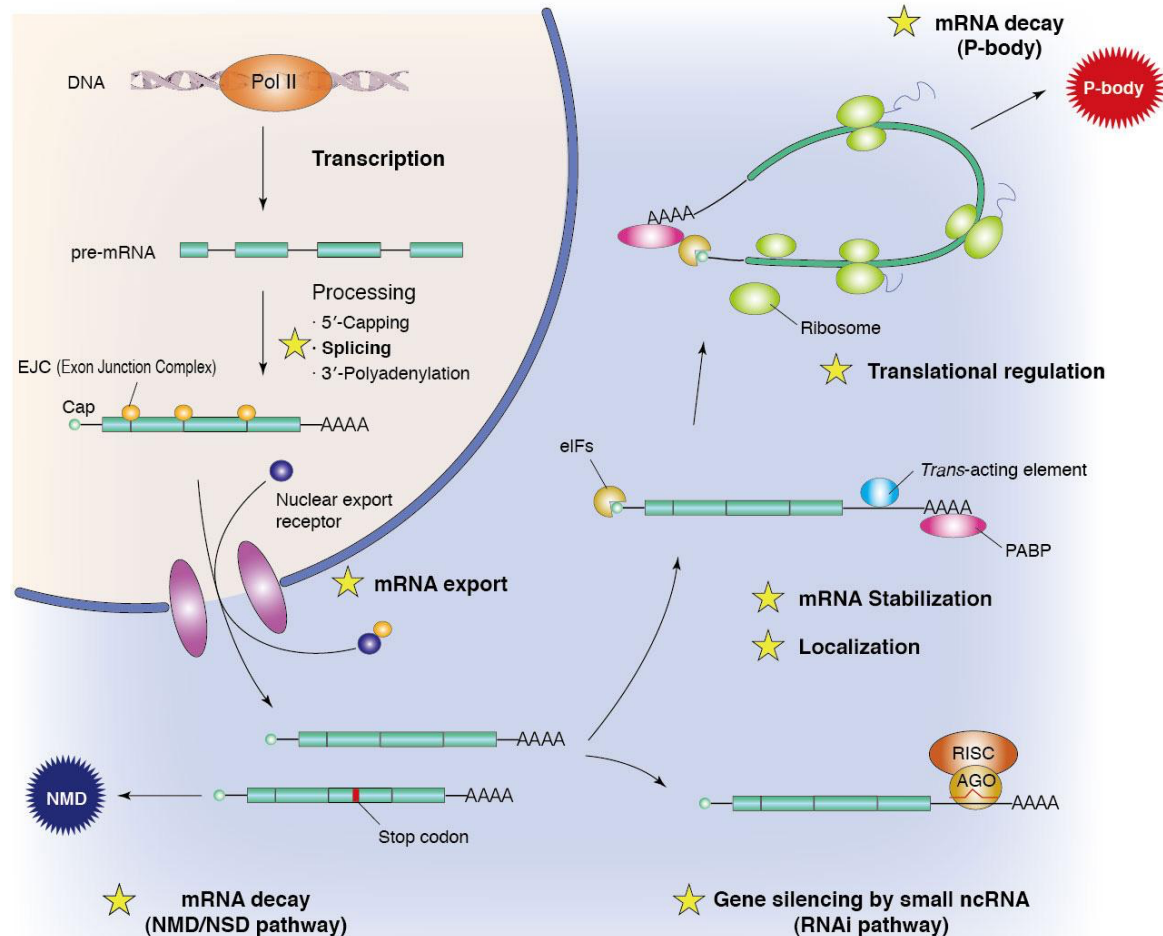
eGRAS, iGRAS

Predict functional SNPs in
mRNA processing
(splicing, alternative 3'
end formation)
(*Nucleic Acids Res*, 2012)
(*Genome Res*, 2016)

**Functional RNA editing
sites** in 3' UTRs or introns
(*under review*, 2017)

Post-Transcriptional Gene Regulation

Highlighting regulation by cis-elements & RNA binding proteins (RBPs)

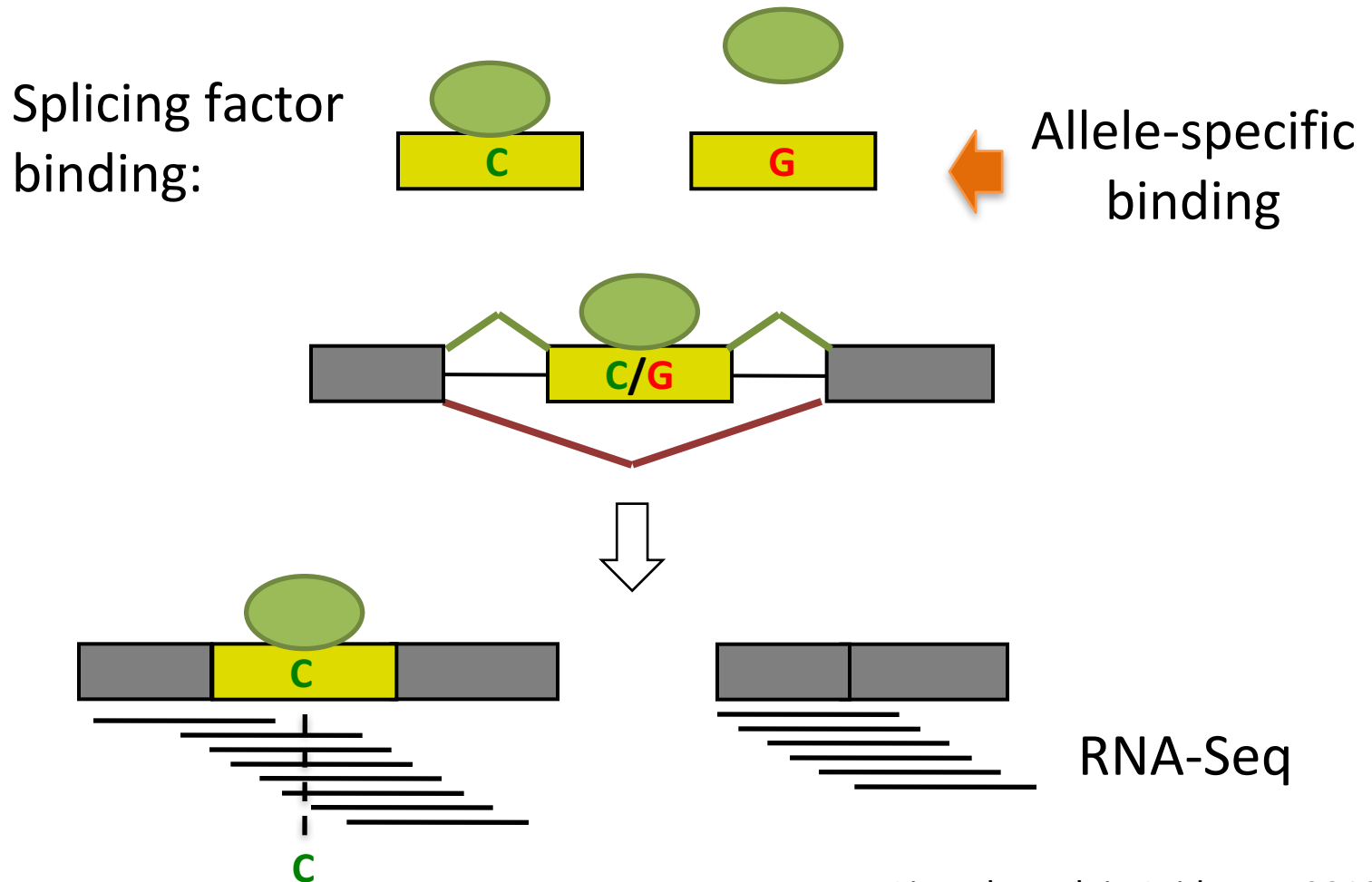


Other RBP functions

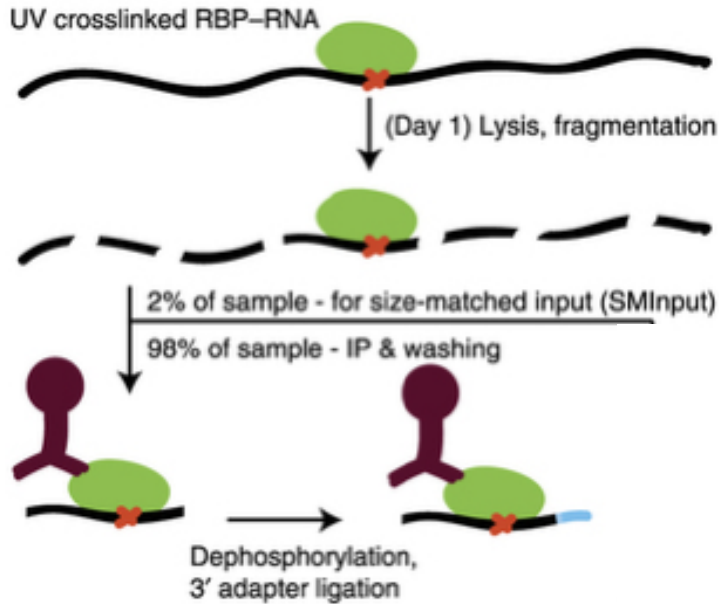
- ★ Transcriptional regulation
- ★ rRNA biosynthesis
- ★ tRNA biosynthesis
- ★ Mitochondrial biosynthesis
- ★ Virus replication
- ★ RNA modification

Functional Genetic Variants in Splicing

Hypothetical model

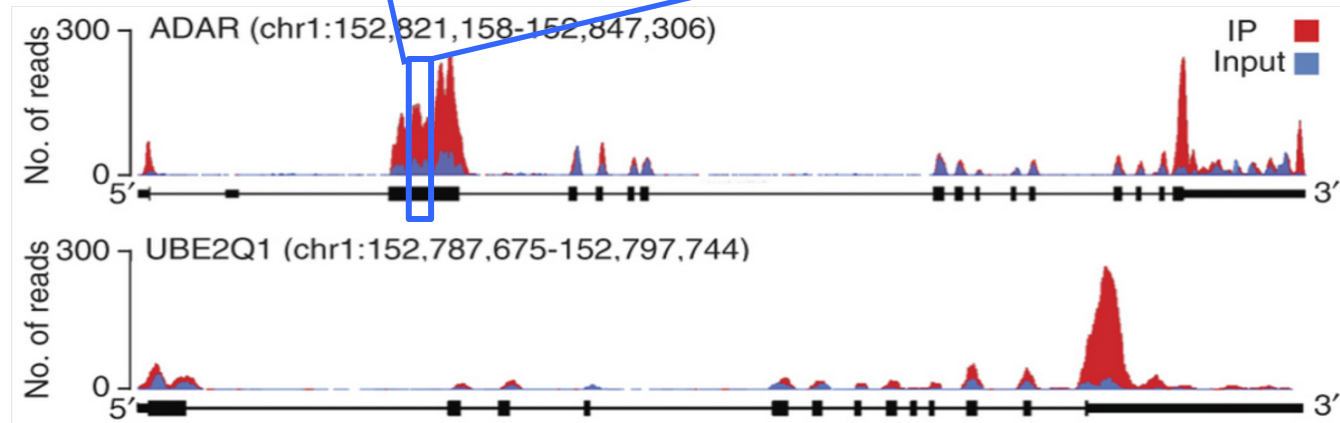


Allele-Specific Binding (ASB) of RBPs in CLIP-Seq data

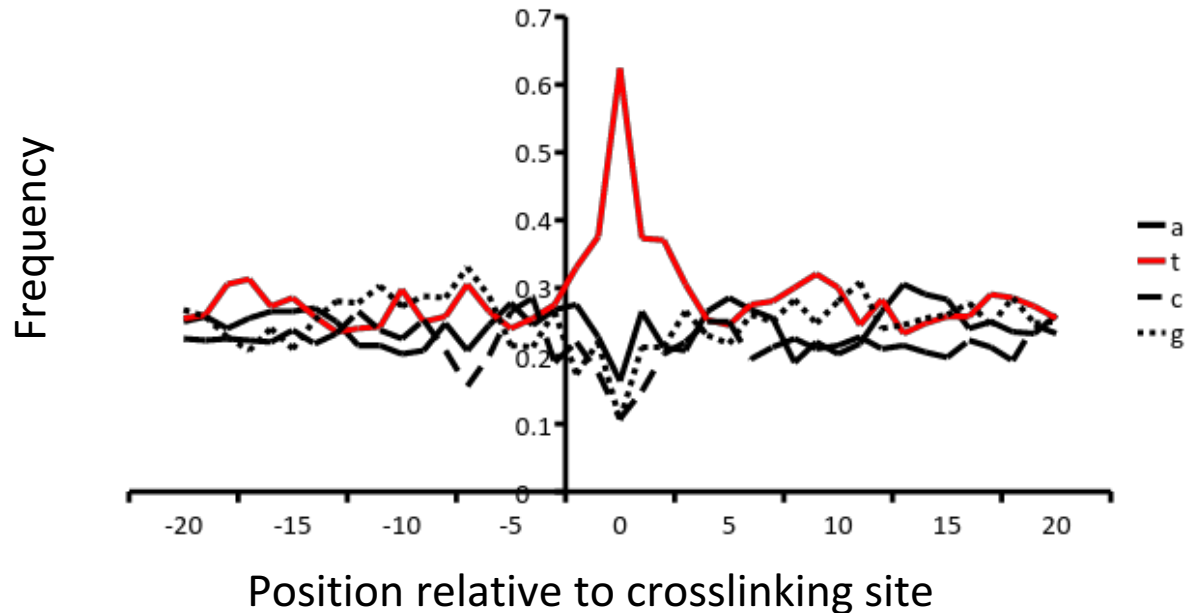


```

AGAGGAGGATGTTGCCGTTGTCTCCCGCCGCTTGGCTCTGAC
AGAGGAGGATGTTGCCGTTGTCTCCCGCCGCTT
GAGGAGGATGTTGCCGTTGTCTCCCGCCGCTTG
GGAGGGTGTTGCCCTTGTCTCCCGCCGCTTGGC
GAGGATGTTGCCGTTGTCTCTCGCCGCTTGGCT
GAGGATGTTGCCCTTGTCTCCCGCCGCTTGGCT
GGGTGTTGCCGTTGTCTCCCGCCGCTTGGCTCT
TGTTGCCCTTGTCTCCCGCCGCTTGGCTCTGAC
    
```



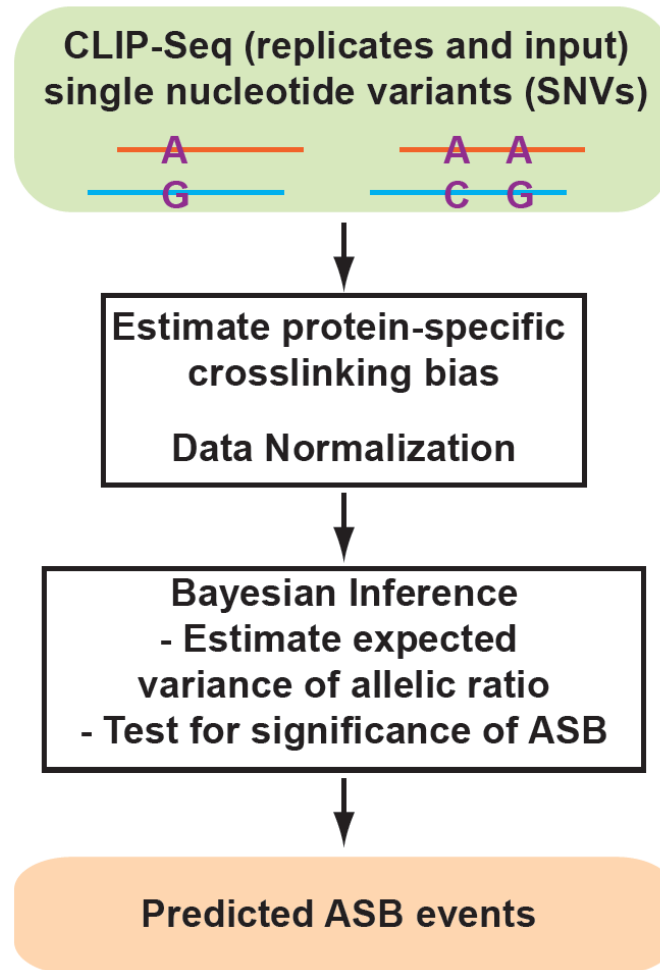
Nucleotide Bias Due to Crosslinking



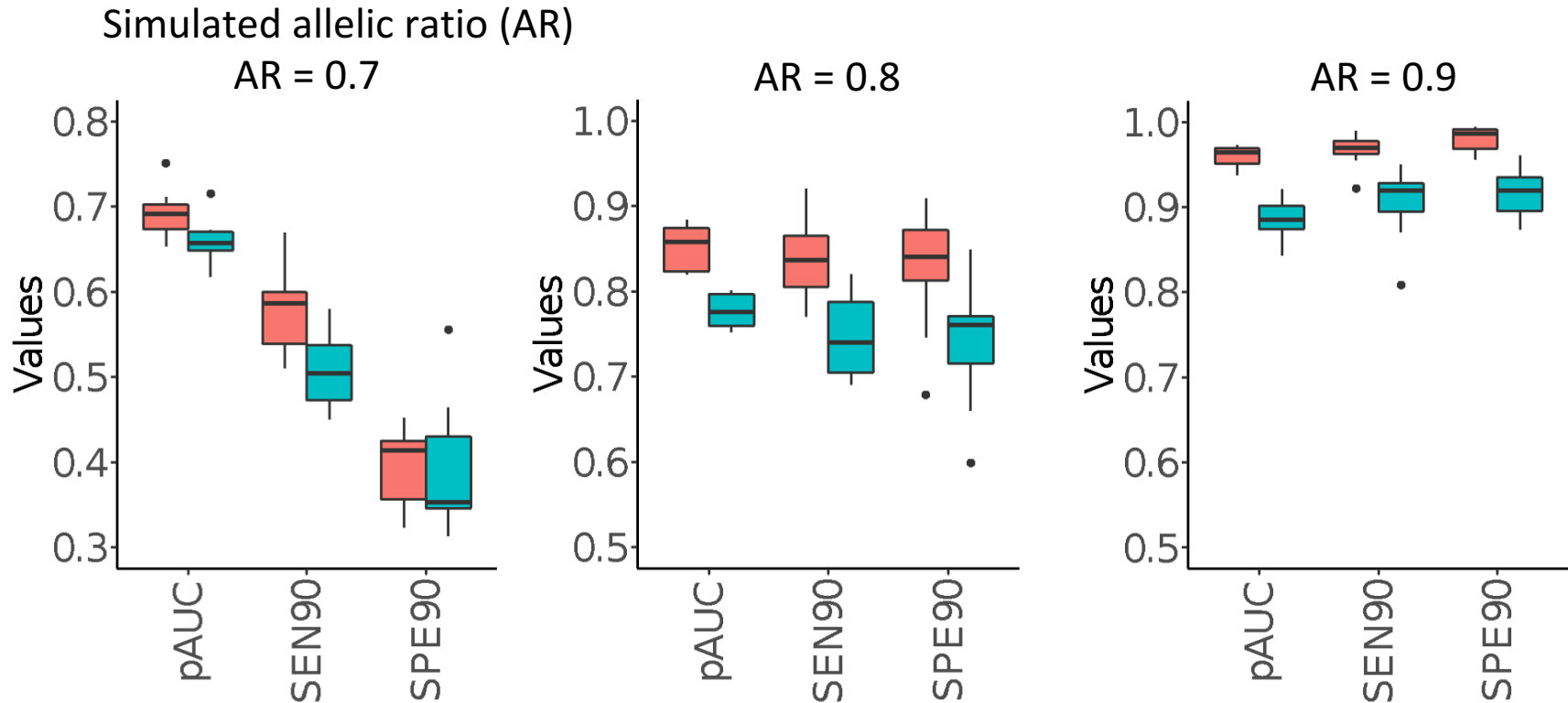
Statistical model to account for:

- Systematic bias due to crosslinking
- Variation between replicated experiments

Method to Identify ASB Events



Method Performance (Simulated Data)



pAUC: partial AUC (AUC at specificity range 90-100%)

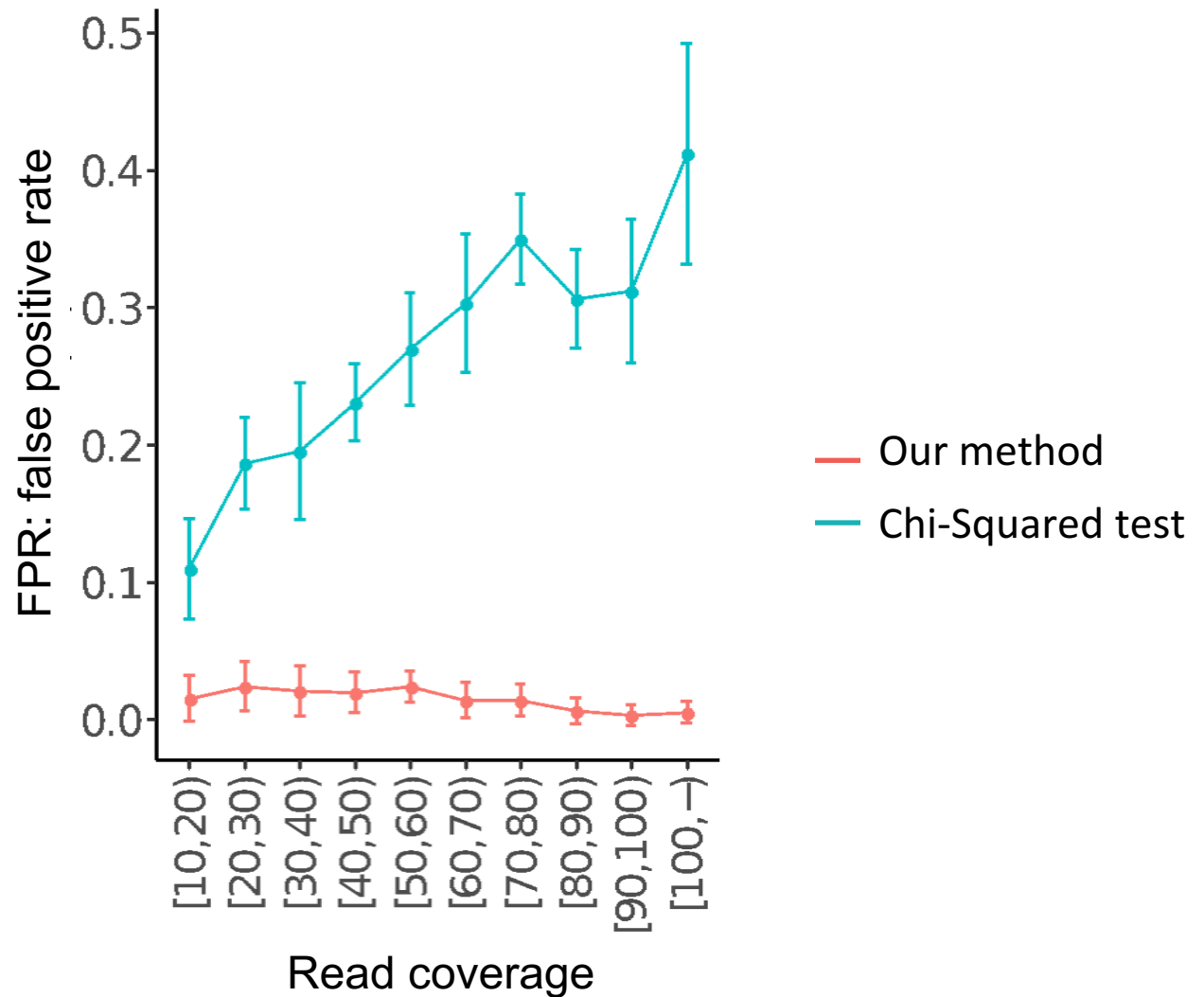
SEN90: sensitivity if specificity = 90%

SPE90: specificity if sensitivity = 90%

Our method

Chi-Squared test

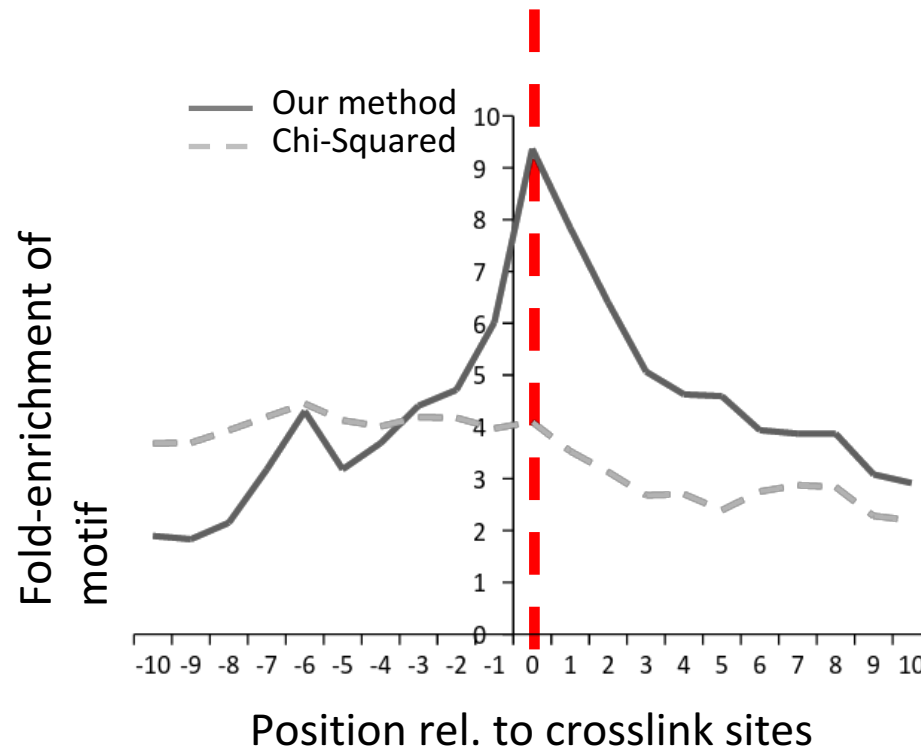
Over-Dispersion Issue of Chi-Squared Test



RBFox2 eCLIP Data Analysis

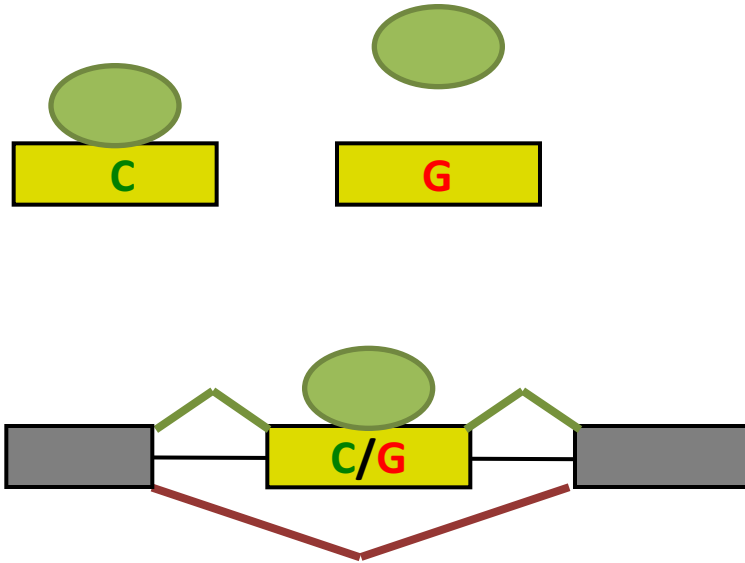


Motif enrichment around predicted ASB sites



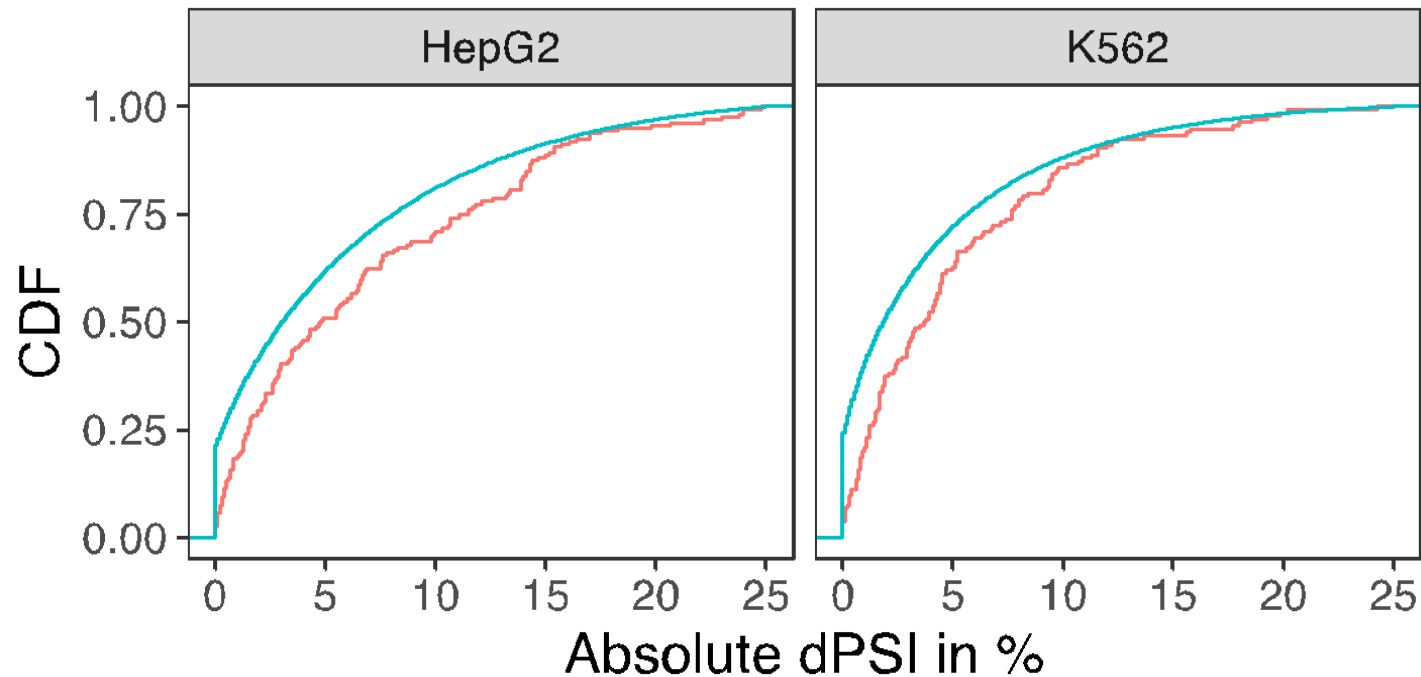
Does ASB Affect Splicing?

Splicing factor
binding:



Analyze splicing changes in ENCODE RNA-Seq data
Splicing factor knockdown vs. control shRNA

ASB Associated with Enhanced Splicing Changes



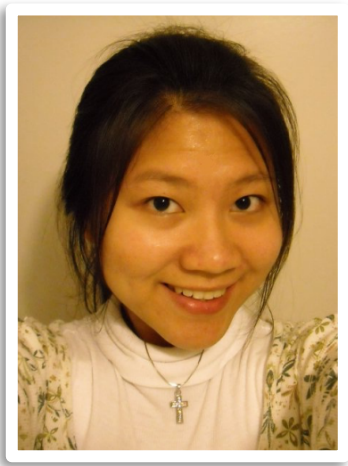
Splicing factor KD vs. control

- Exons with ASB events nearby
- All other exons with adequate read coverage

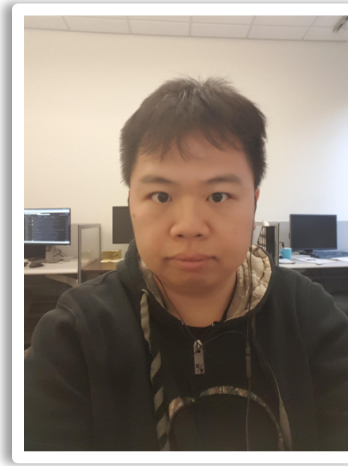
Summary

- We are developing a suite of tools to carry out different types of SNV analysis using sequencing data.
- The newly developed method to detect allele-specific binding of RNA-binding proteins has desirable performance.
- Ongoing: finalize software packages

Acknowledgements



Esther Hsiao



Ei-Wen Yang



Jae Bahn



NATIONAL CANCER INSTITUTE
Informatics Technology for
Cancer Research