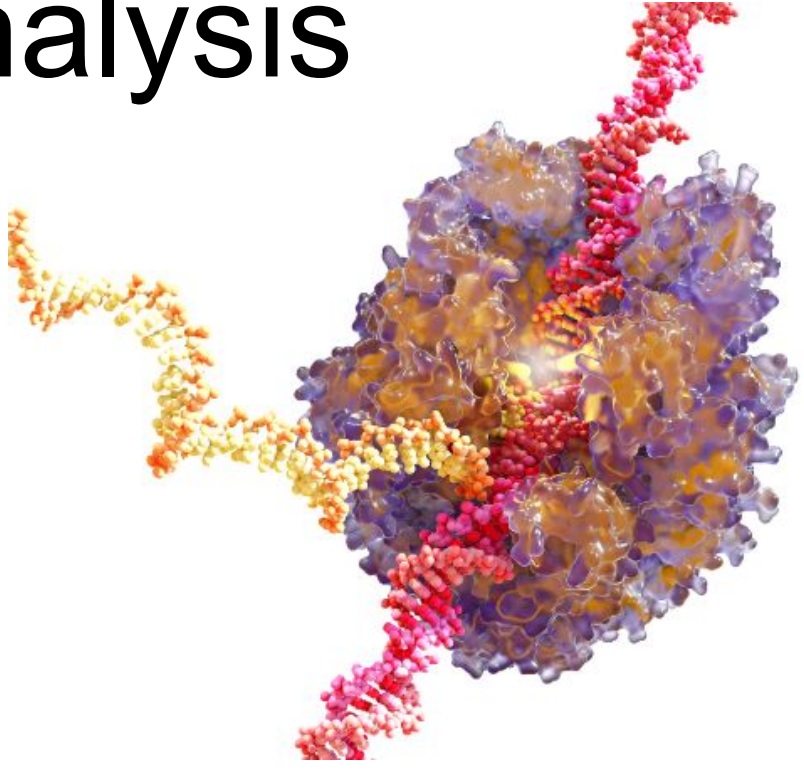
Differential hnRNPC Binding Analysis

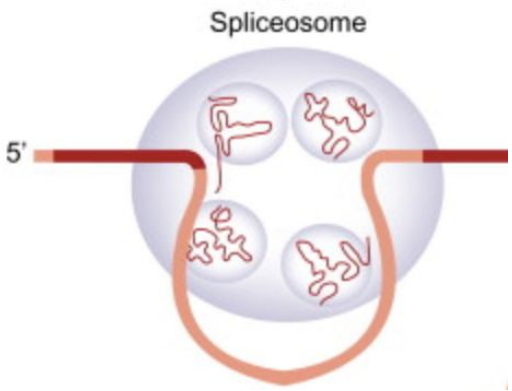


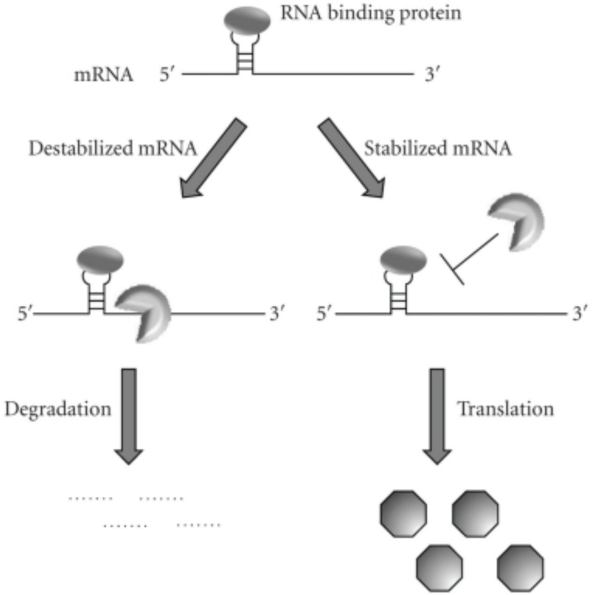
Davyd Sadovskyy
PATH-GDS Rotation 3
Dr. David Bentley Lab

University of Colorado
Anschutz Medical Campus

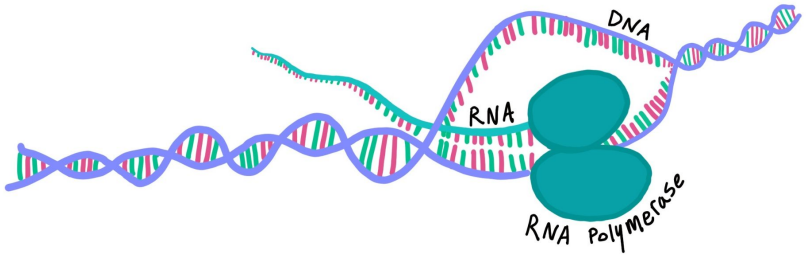
What is hnRNP?

- hnRNP (heterogeneous nuclear ribonucleoprotein) is a an RNA-binding protein (RBP) that binds to newly transcribed RNA
- Functions in RNA stabilization, splicing, and sorting transcripts (mRNA Export)
- **hnRNPC** (my project) recognizes unstructured single-stranded RNA regions (~200-300 bp long)



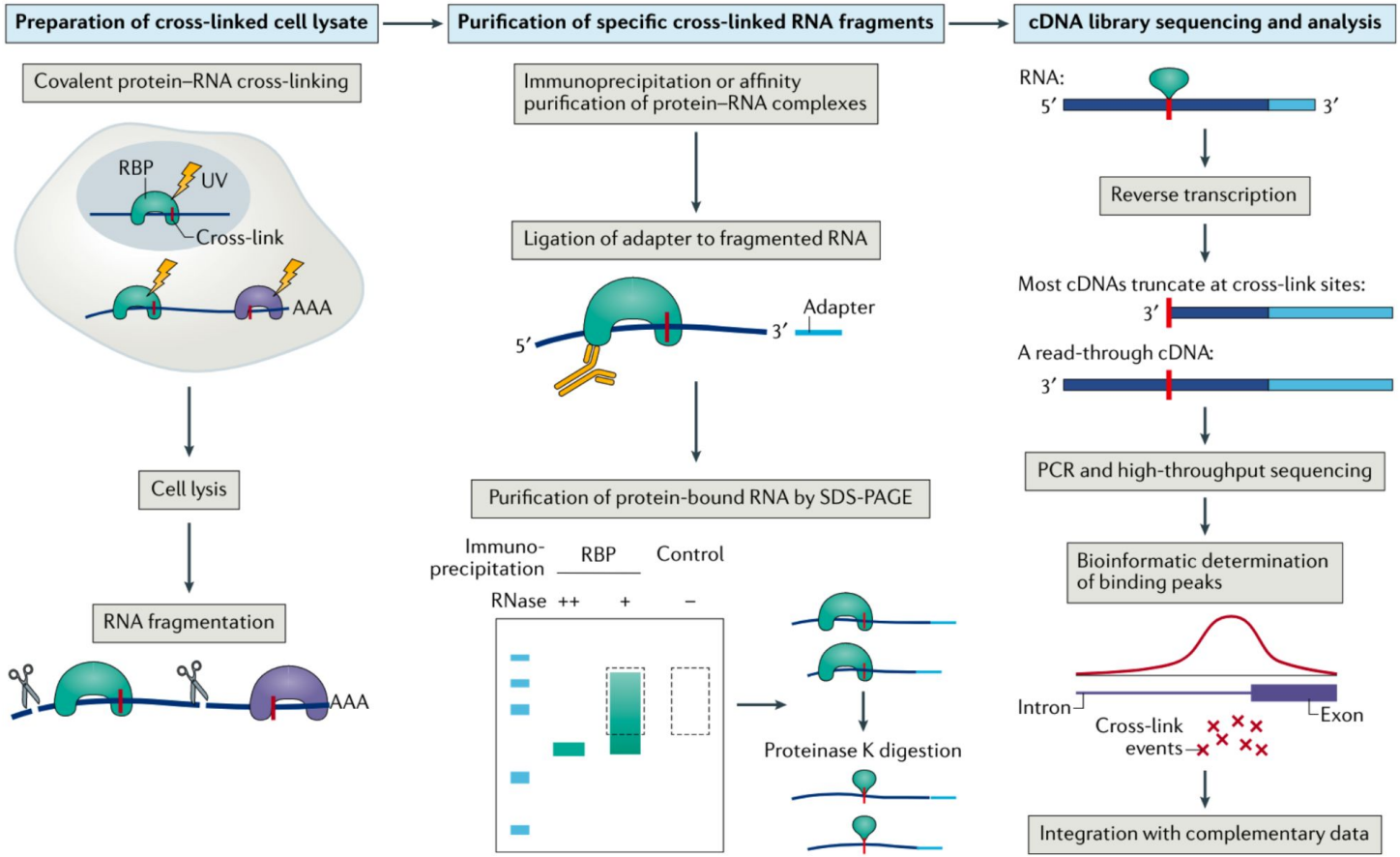


Fast and Slow RNA polymerase



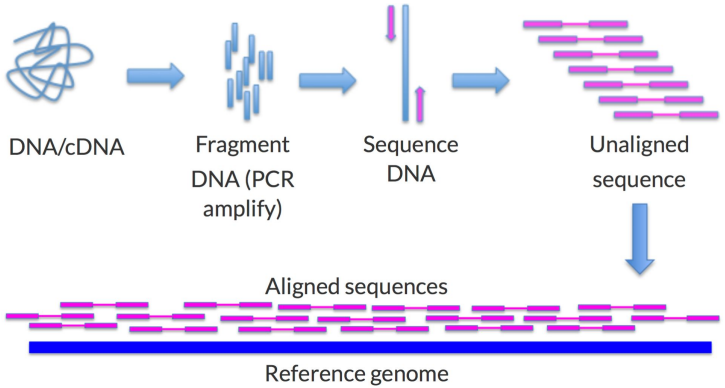
- Fast Pol II (Wild-Type)
  - ~3-4 kb per minute
  - Pre-mRNA folds before proteins can bind
- Slow Pol II (R749H Mutation / C4)
  - ~2 kb per minute
  - More unstructured RNA available for hnRNPC binding
- **Research Question**
  1. Is there differential binding of hnRNPC in fast vs slow pol?
  2. Where in genome are binding site occupancy differences?

iCLIP Method



iCLIP Raw Data

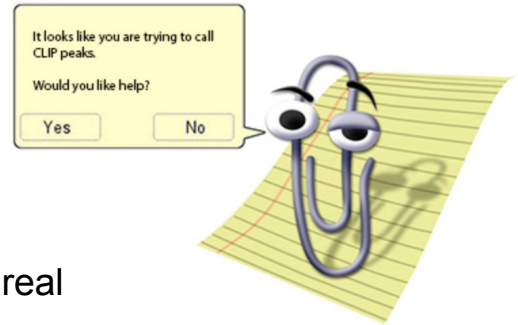
- The sequenced reads (cDNAs) are aligned to a reference genome using a tool like Bowtie or STAR.



- **BAM File**
  - Records every mapped cDNA read
  - Chromosome
  - BP position of alignment.
  - Strand information (which direction it maps to).
  - Sequence quality scores.

Peak Calling - Clippy

.bam file



.bed file

- not all reads correspond to real cross-link evens (random fragmentation, PCR bias, or sequencing errors)
- Different RNA regions have different background levels

- Peak calling identifies clusters of cross-linked reads, ranks them by intensity, and assigns binding confidence scores.

- start and end positions of peaks.
- Strand (+ or -).
- Peak intensity (more cross-links = higher score).

Wilcoxon Rank-Sum Test

- **Does slow pol have different peak intensity compared to fast pol?**
- non parametric test - no assumption about distribution

Steps for the test:

1. Rank all values together
2. Sum ranks for each group
3. Compute test statistic. R1 = sum of ranks for the smaller group.
   n1 = size of the smaller group

$$U = R_1 - \frac{n_1(n_1+1)}{2}$$

4. compare U to all possible rank sums that could occur if the groups were randomly shuffled. The p-value is the proportion of times the observed U value is as extreme (or more extreme) than expected under the null.

Results

Total hnRNPC Binding Peaks in Each Condition

Condition	Total_Peaks
WT_Rep3	40855
C4_Rep3	115584
WT_Rep4	9116
C4_Rep4	29475

Shared hnRNPC Binding Peaks Between Conditions

Condition	Shared_Peaks
WT_Rep3	29363
C4_Rep3	29363
WT_Rep4	4556
C4_Rep4	4556

Unique hnRNPC Binding Peaks in Each Condition

Condition	Unique_Peaks
WT_Rep3	16651
C4_Rep3	91274
WT_Rep4	5435
C4_Rep4	25777

Average hnRNPC Binding Peak Score in Each Condition

Condition	Avg_Peak_Score
WT_Rep3	14.3153
C4_Rep3	16.1859
WT_Rep4	18.6286
C4_Rep4	15.0892

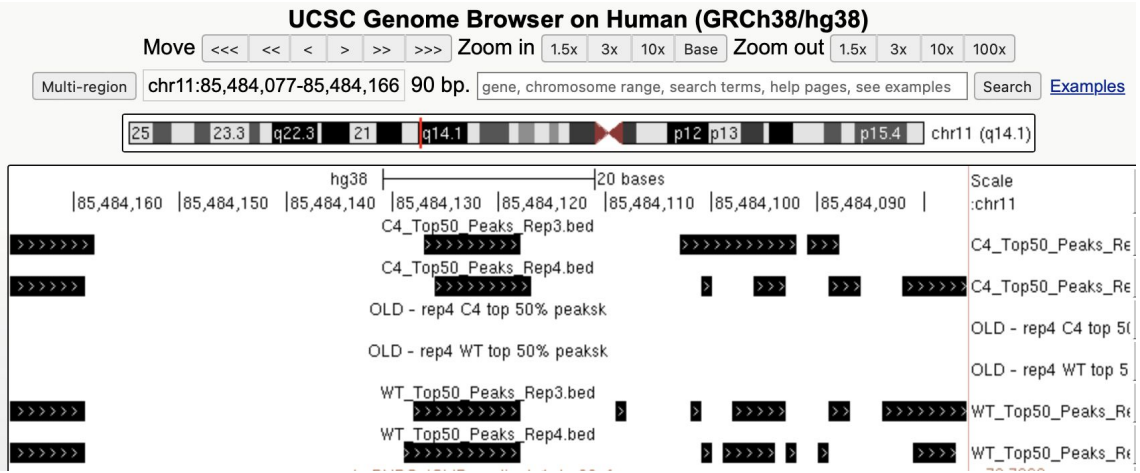
Wilcoxon Rank-Sum Test for Binding Strength Differences

Comparison	Wilcoxon_Statistic	Wilcoxon_p_value
WT_Rep3 vs C4_Rep3	-19.91	0.0e+00
WT_Rep4 vs C4_Rep4	-4.74	2.1e-06

- Rep3: p-value = 3.6e-88 (VERY significant). C4 had higher peak scores (stronger hnRNPC binding with slow Pol II).
- Rep4: p-value = 2.1e-6 (still significant, but less extreme). C4 had lower peak scores than in Rep3 but still different from WT

Next Steps

- Do further statistical analysis identify exact genes or genomic regions with different hnRNPC binding



- Validate differences in alternative splicing transcripts with more experiments.
  - RT-PCR or qPCR – Check if alternative exons are included or skipped differently in WT vs. C4 cells.
  - RNA-seq Analysis – Perform differential splicing analysis to identify transcripts with altered exon inclusion.

Acknowledgments

Thank you to David Bentley and Ben Erickson for your help and guidance in this project