

In silico prediction of retained intron-derived neoantigens in leukemia

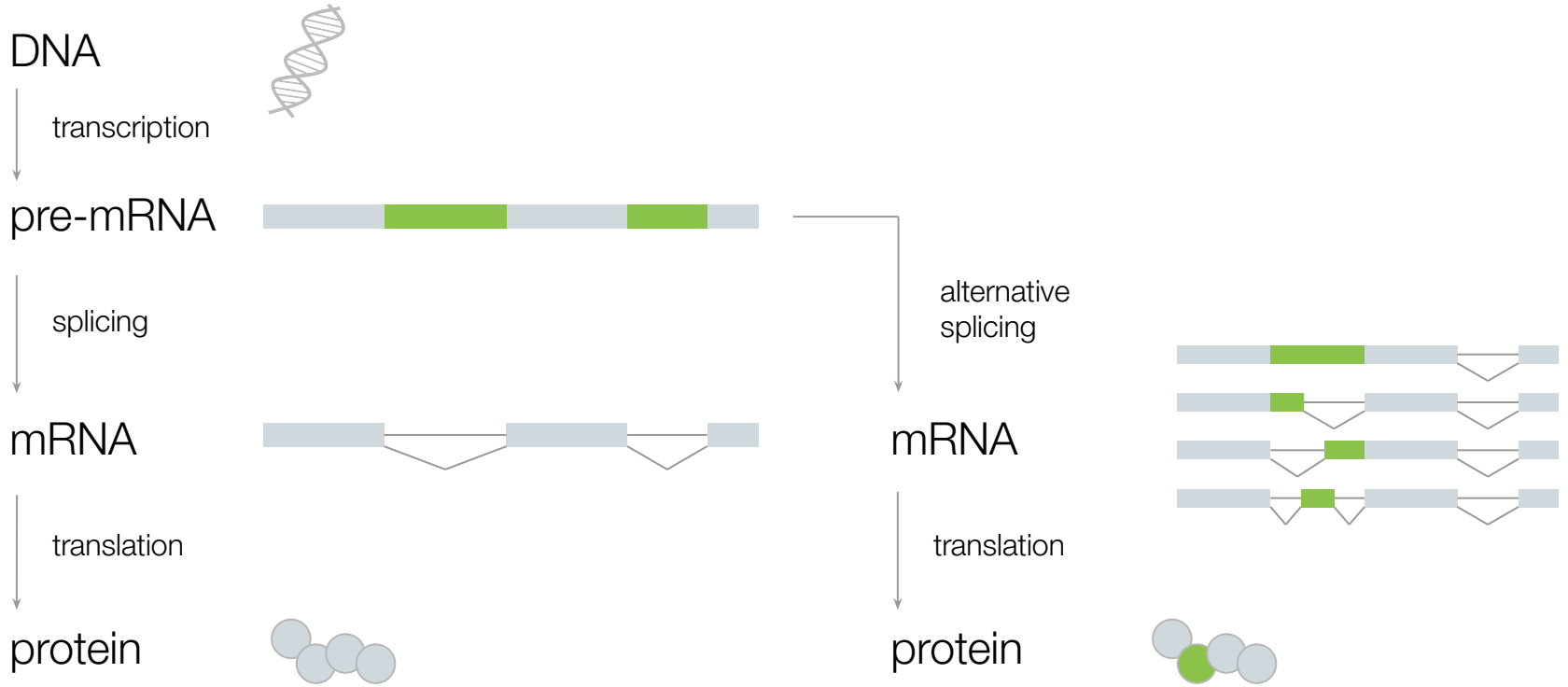
Sarah Chen

MIT PRIMES Computational Biology

Under the direction of Dr. Nicoletta Cieri and Kari Stromhaug

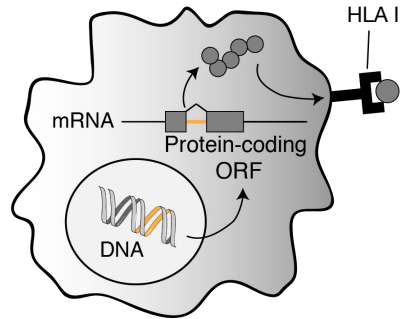
October 18, 2020

Alternative splicing diversifies the cancer transcriptome

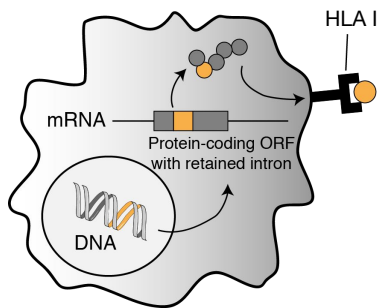


Alternative splicing in cancer is a potential source of neoantigens

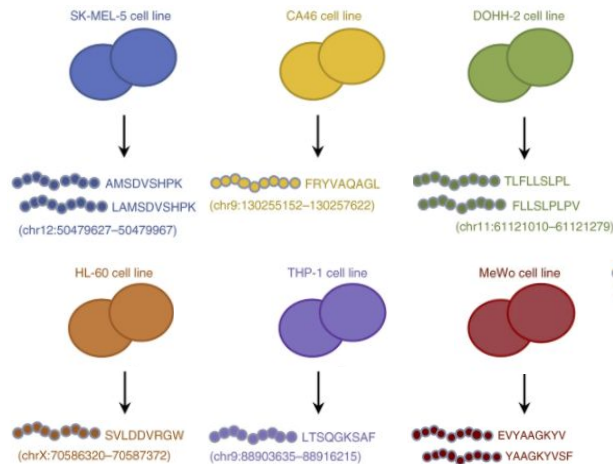
Canonical splicing



Alternative splicing (retained intron)

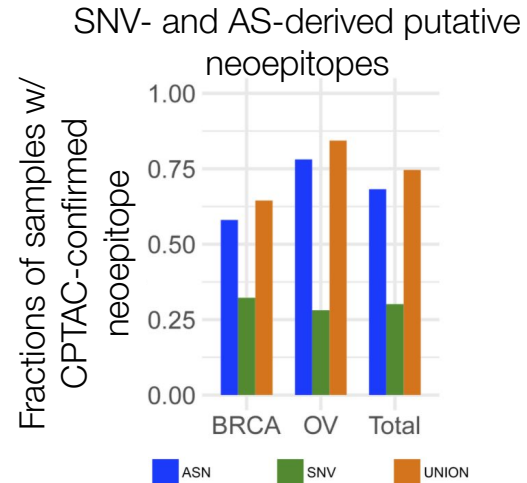


Intron retention:



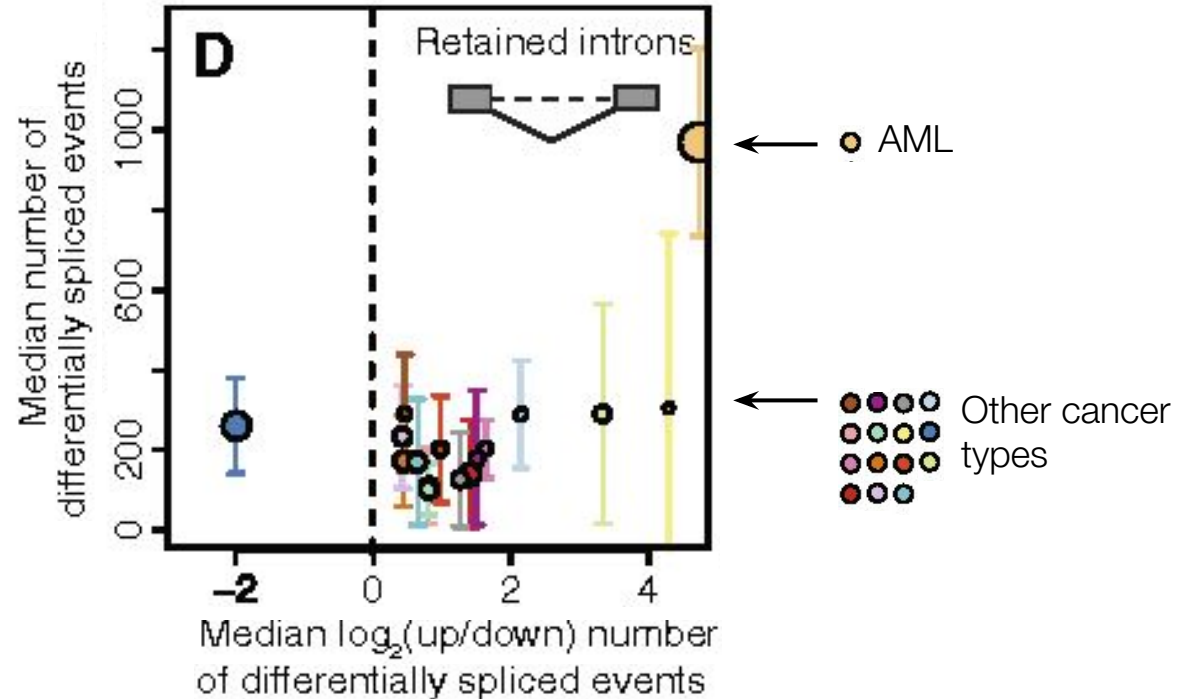
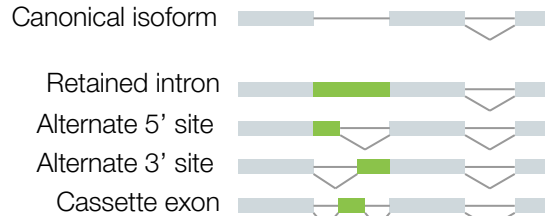
[Smart et al. Nature Biotechnology. 2018](#)

Alternative splicing:

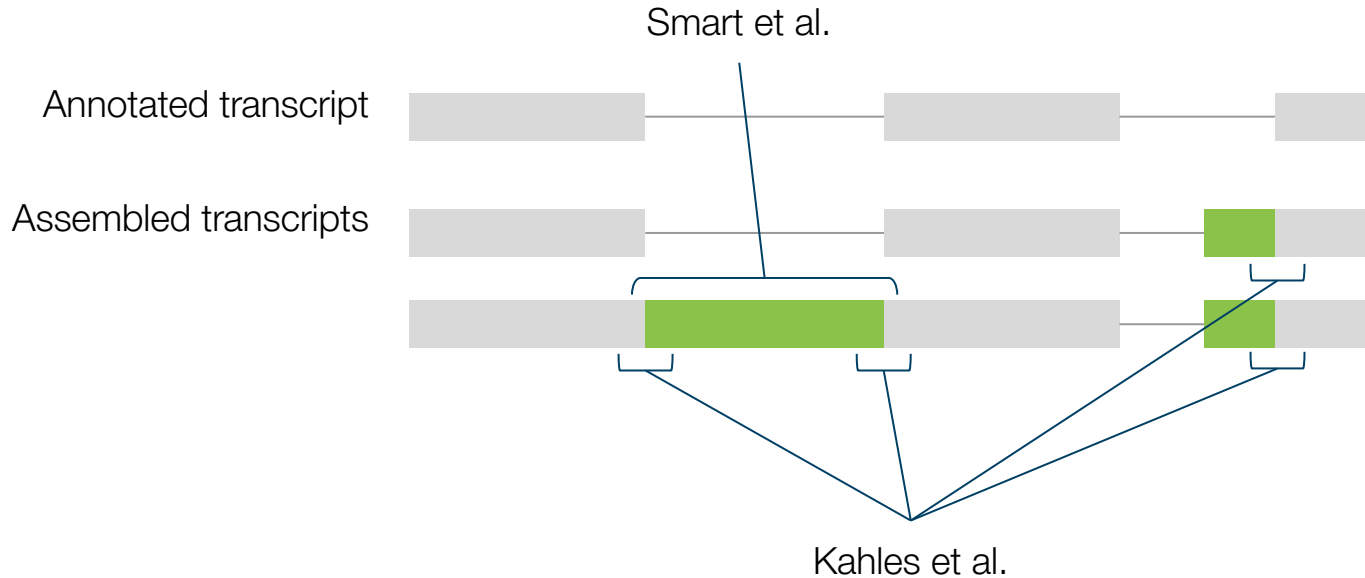


[Kahles et al. Cancer Cell. 2018](#)

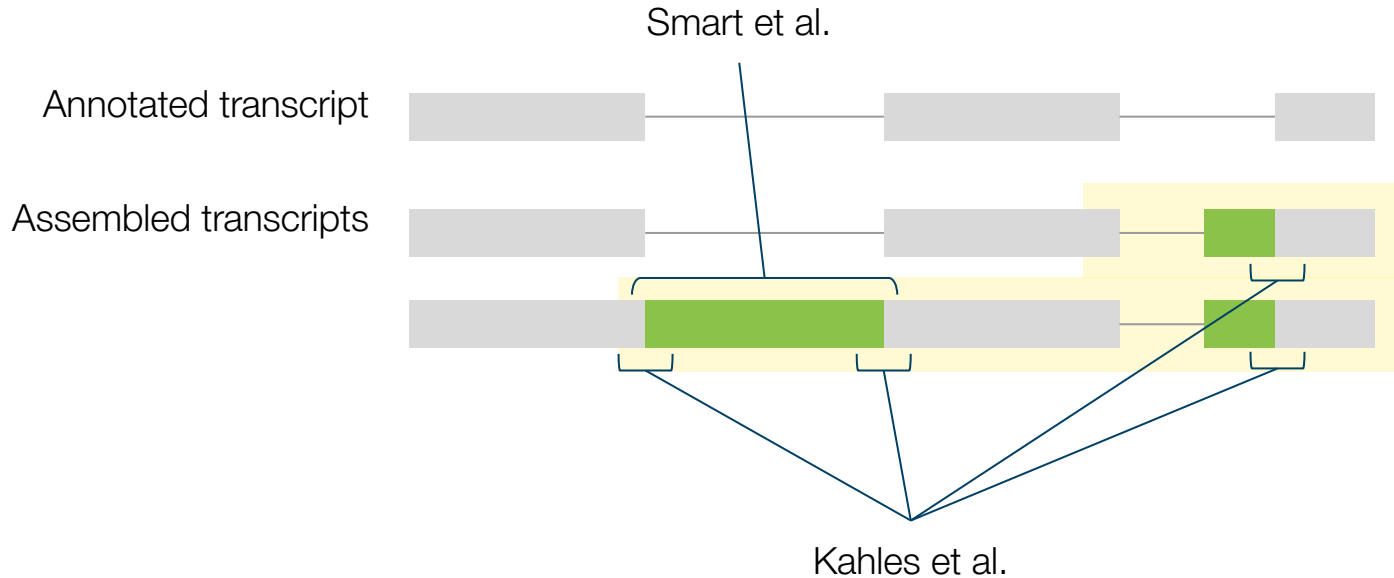
Upregulation of intron retention is widespread in cancer and in AML in particular



Past work failed to consider the full scope of potential derived neoantigens

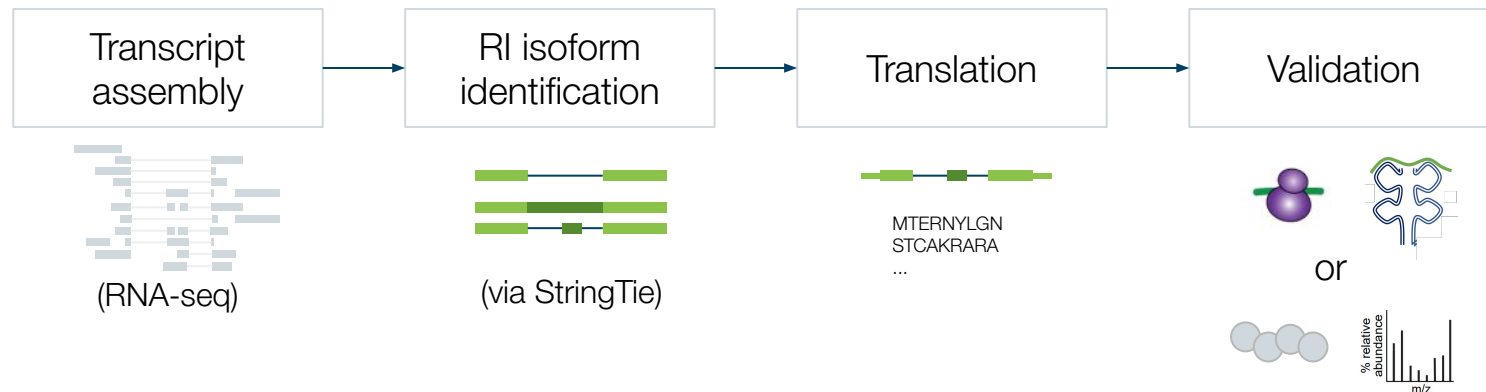
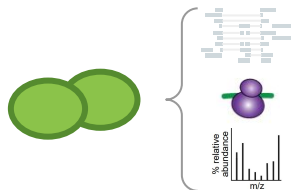


Past work failed to consider the full scope of potential derived neoantigens



Building a prediction pipeline for RI-derived neoantigens

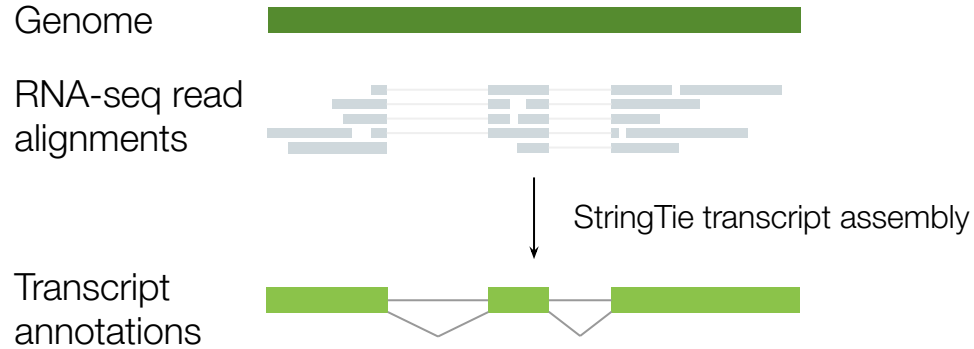
B721.221 cells (model system)



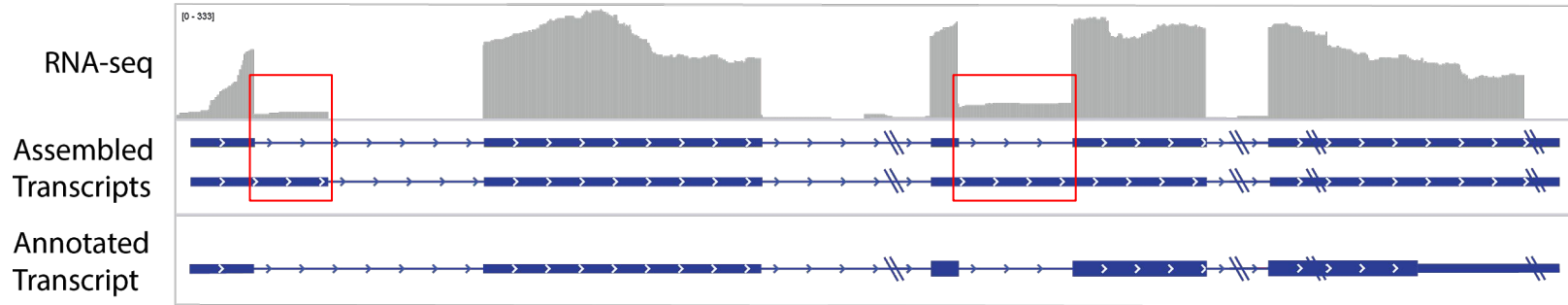
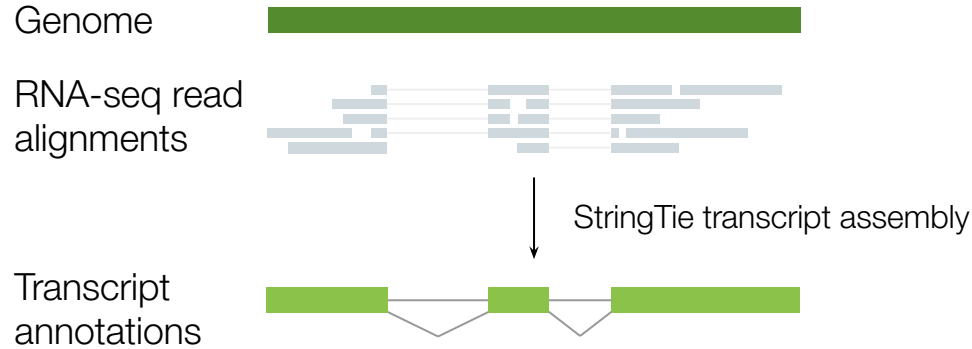
Overview



Identifying RI isoforms via StringTie transcript assembly



Identifying RI isoforms via StringTie transcript assembly

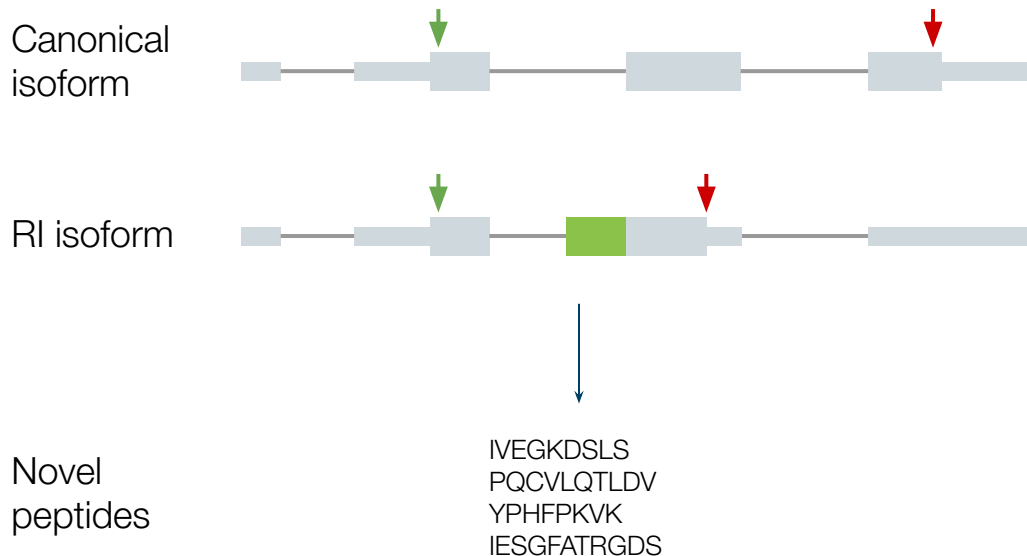


RI identification



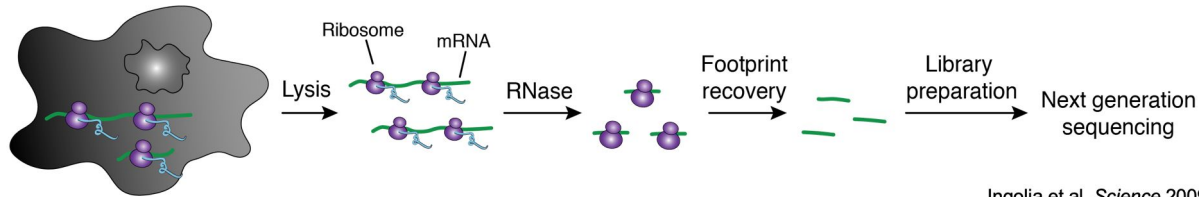
Translating identified RI isoforms

RI isoforms are translated from canonical start codons to the first downstream in-frame stop codon

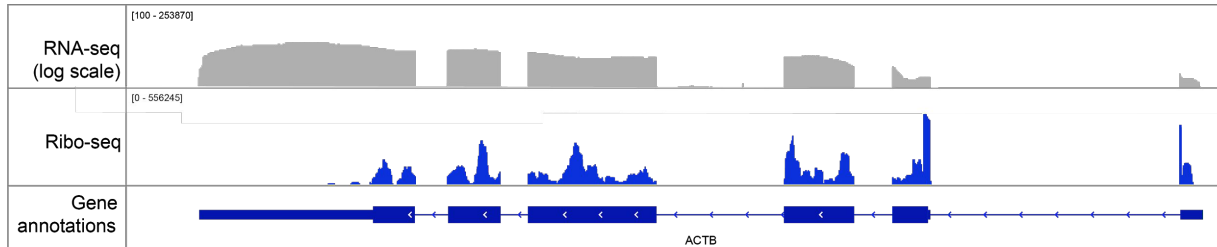


Validating the translation of RIs through Ribo-seq

Ribo-seq provides evidence that a sequence is **translated**



Ingolia et al, *Science* 2009

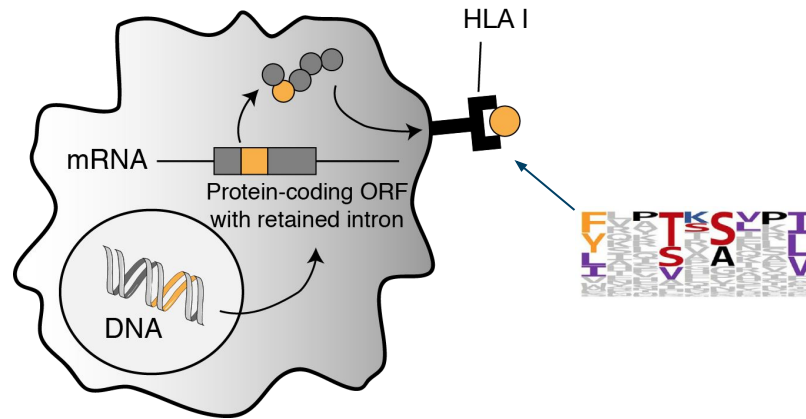


Validation



Estimating the likelihood of HLA presentation for predicted RIs

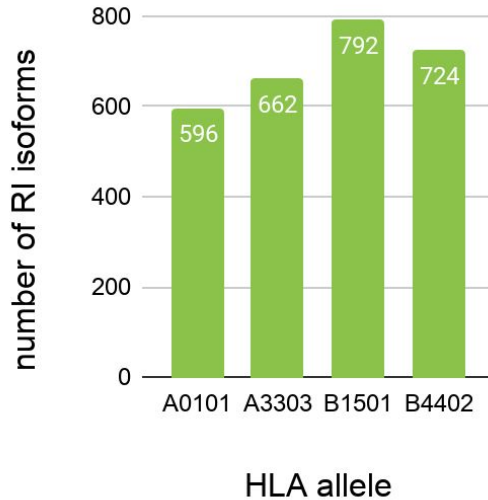
HLAthena predicts whether peptides are **strong HLA binders** using CNNs.



We considered peptides with a binding score the top 0.1% and 0.5% of HLAthena's background decoys

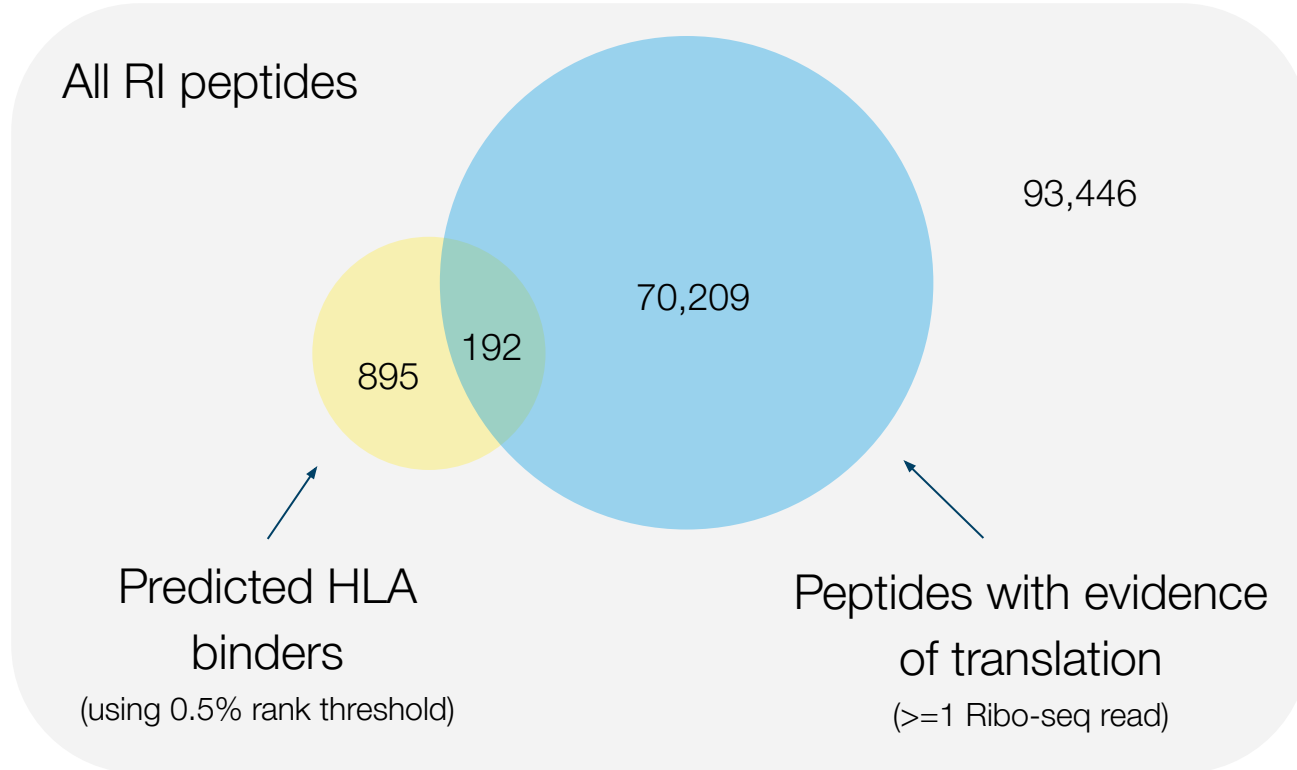
RIs identified in the B721.221 Model System

Number of RI isoforms



The union of RIs identified across 4 replicates yielded **164,742 8-11-mer potential peptides** not inside the canonical proteome

192 RI peptides identified in B721.221 were translated and predicted to bind HLA

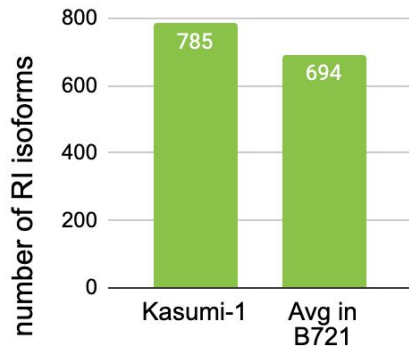


Applying the RI prediction pipeline to AML cell lines

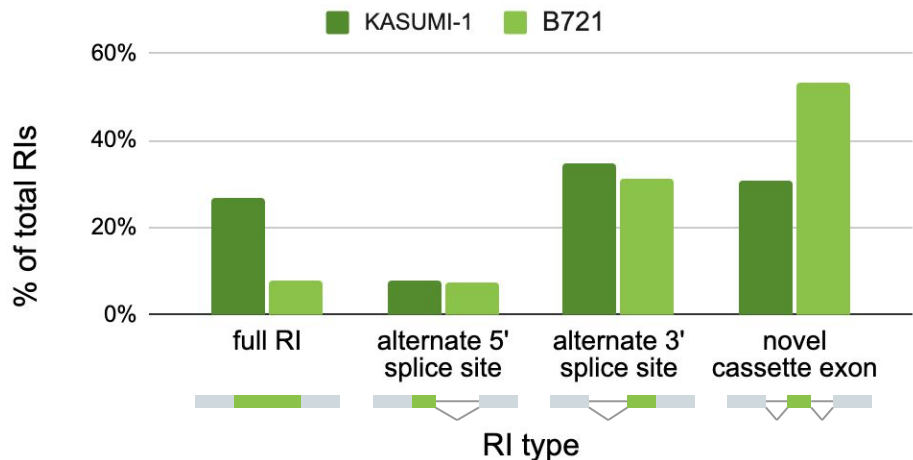
Potential neoepitopes include **72,636** 8-11-mers not found in the normal proteome

AML demonstrates a similar number of RIs but a distinct distribution of RI types

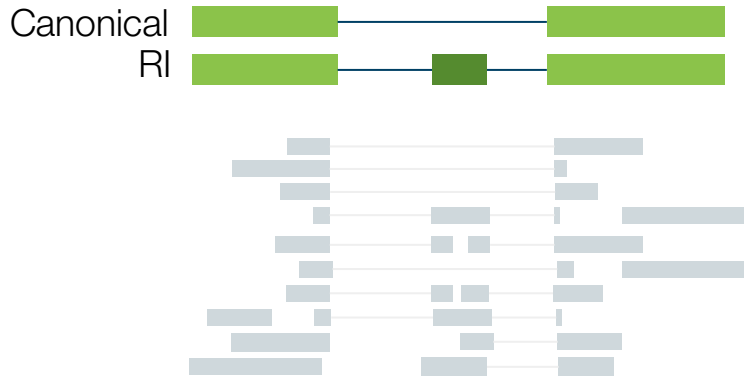
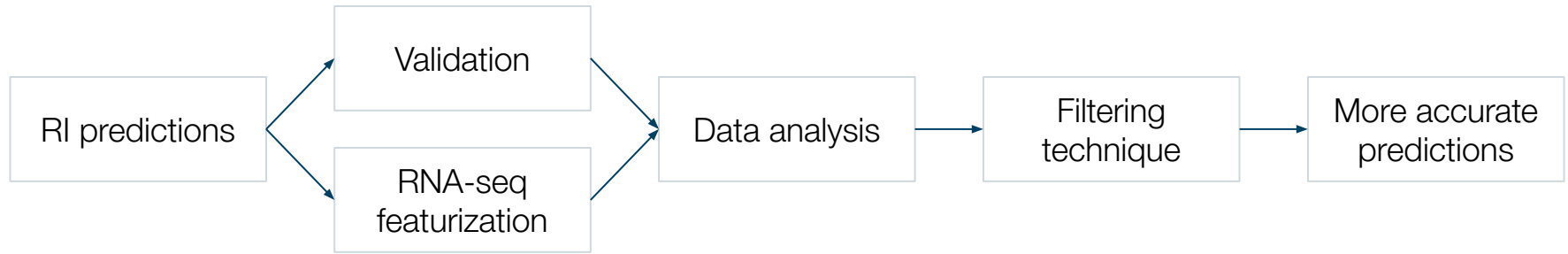
Number of RI isoforms in Kasumi-1 vs. B721.221



Distribution of RI types in B721 vs. Kasumi-1



RI features may improve pipeline accuracy when Ribo-seq/MS data are unavailable



Features include:

- Isoform TPM
- Read support of introns vs. adjacent exonic regions
- Proportion of multi-mapping and/or indel reads in introns vs. adjacent exonic regions

Summary & future directions

Summary

- Developed a pipeline to predict RI neoepitopes from RNA-seq data
- Validated **192 RI peptides** as translated and likely to be HLA presented in B721.221 cells, our model system
- Applied pipeline to **AML** cell line data

Future directions

- Leverage data generated in B721.221 to develop a filtering methodology to increase pipeline accuracy in settings without Ribo-seq and MS
- Apply to further AML cell lines and primary samples to detect cancer-specific splicing events and neoantigens shared across them

Acknowledgements



Cathy Wu



Aviv Regev



Slava Gerovitch



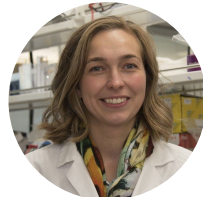
Brian Haas



Karl Clauser



Nicoletta Cieri



Tamara
Ouspenskaia



Kari Stromhaug



Travis Law