



National Science Foundation  
University of Pennsylvania

Mentors:

Kristen W Lynch, Ph.D.

Professor and Chair, Department of Biochemistry and Biophysics

Yoseph Barash, Ph.D.

Associate Professor, Department of Genetics



***Alternative mRNA splicing redefines the landscape of commonly dysregulated genes across the acute myeloid leukemia patient population.***



**Osvaldo D. Rivera**

Cancer Biology Ph.D. Program  
Biomedical Graduate Studies

**Contact us:**

<http://www.kwlynchlab.org/>  
<https://www.biociphers.org/>

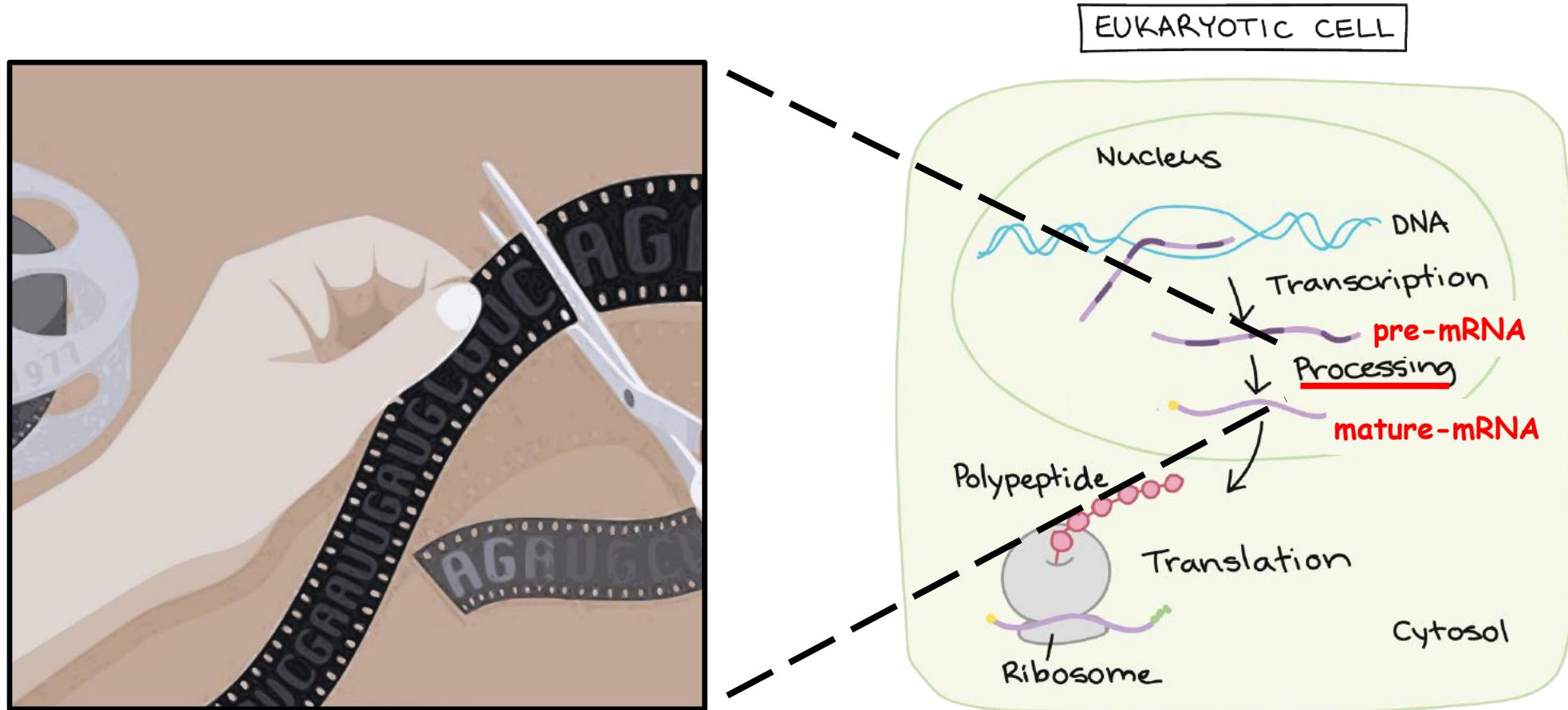
**BiōCiphers**

Send inquiries to: [orivera@pennmedicine.upenn.edu](mailto:orivera@pennmedicine.upenn.edu)

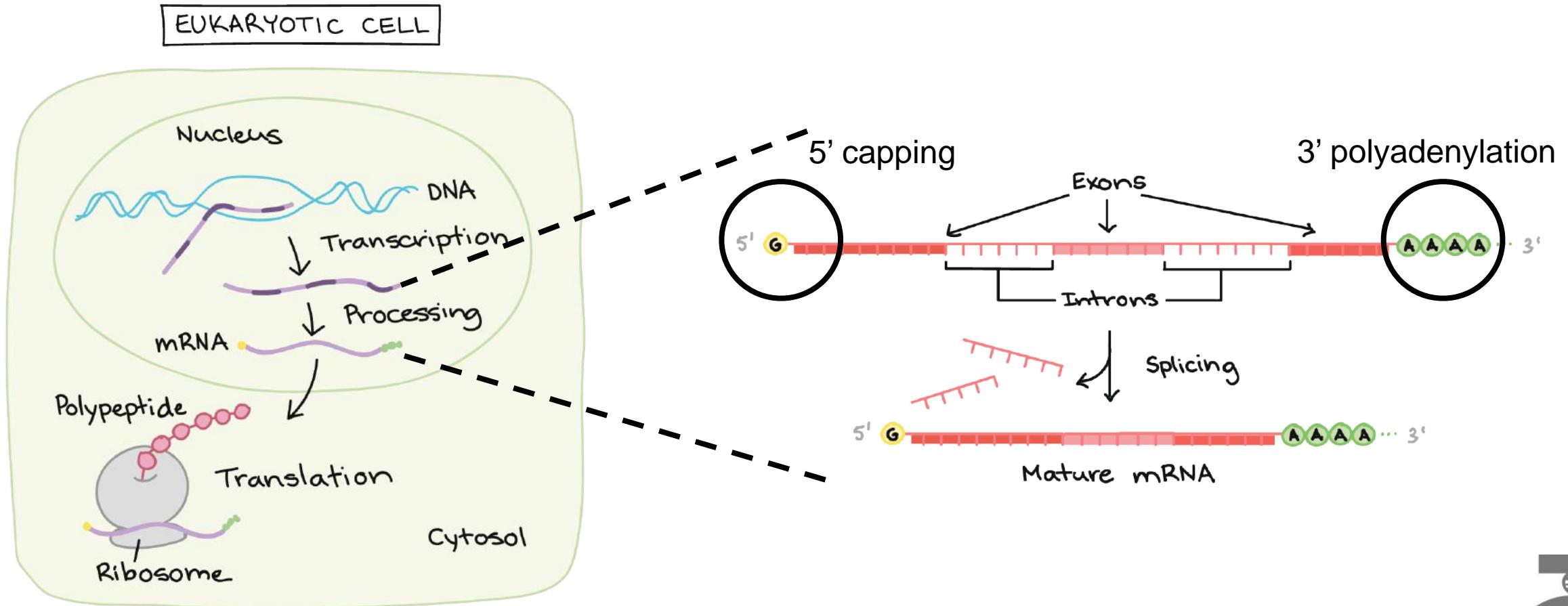
- **Introduction:** mRNA splicing quantification and leukemia biology
- **Results and Discussion:** Discovered splicing patterns in leukemia patients
- **Conclusions and Future Directions**
- **Closing Remarks**

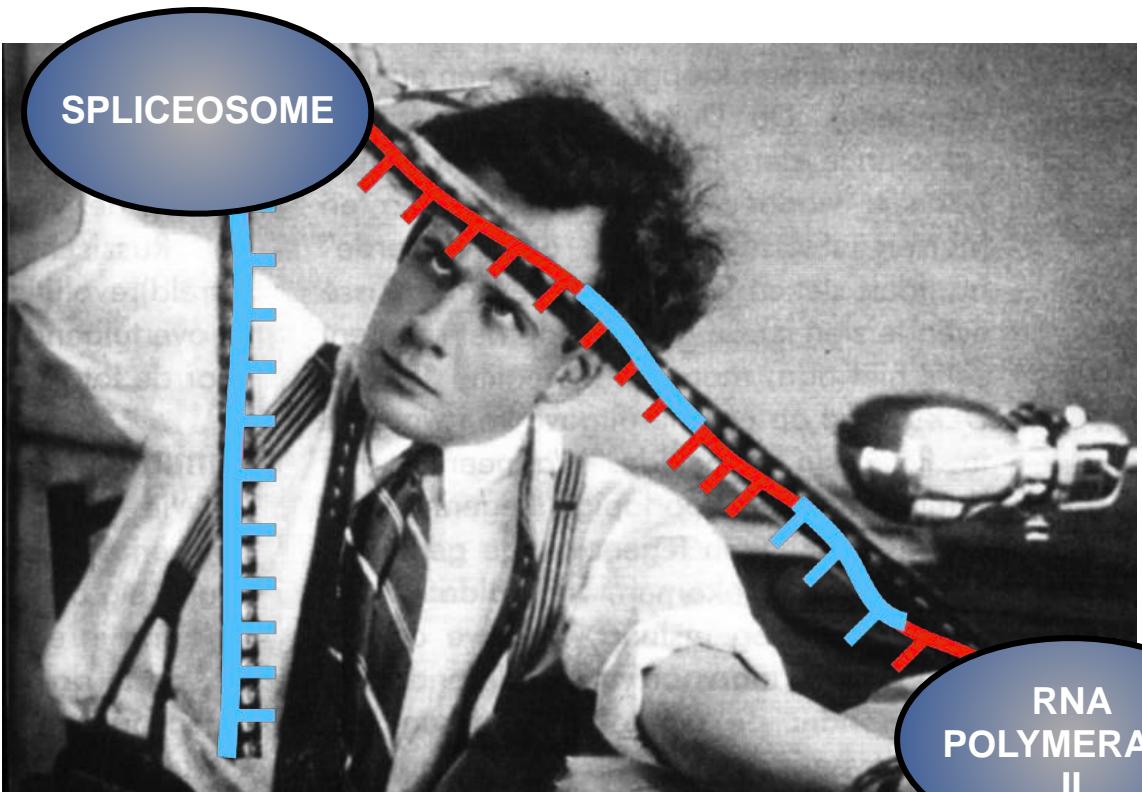


“DNA” is the *blueprint* ► “**pre-mRNA**” is the *raw film* ► “**mature-mRNA**” is the *processed film*

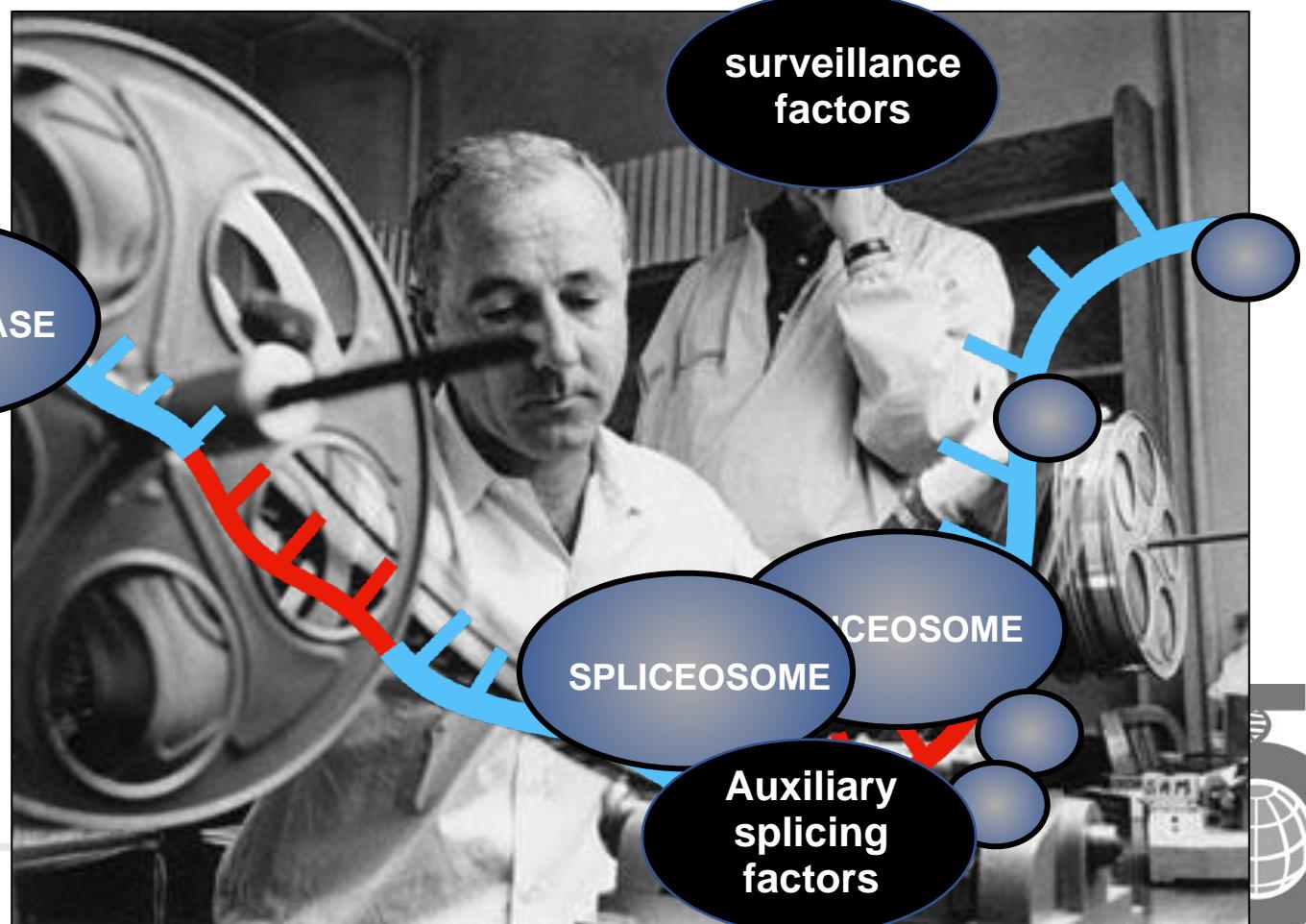


# mRNA splicing removes non-coding sequence introns to allow for the sequences of exons to sequentially code a functional protein.



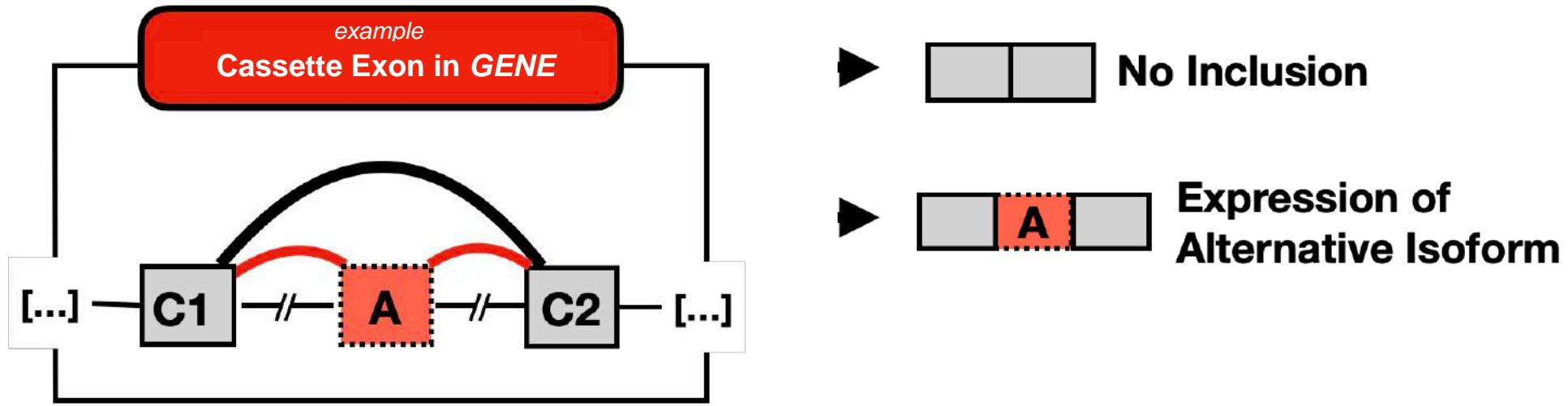


SPLICEROSOME



We know now that splicing is a very intricate process requiring of complex cellular machinery

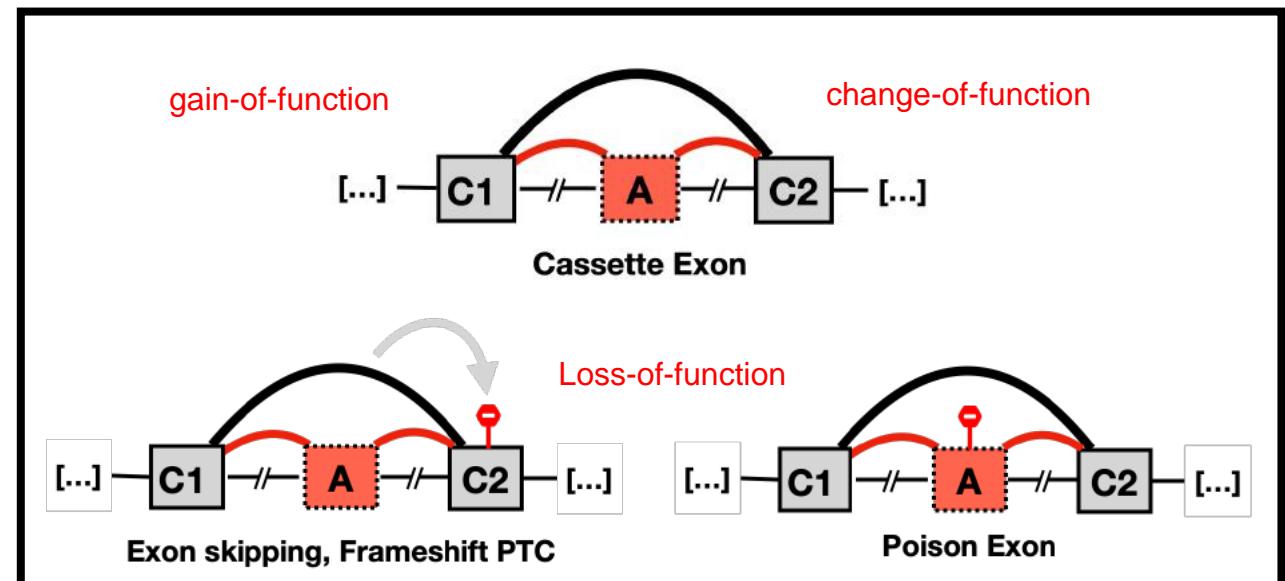
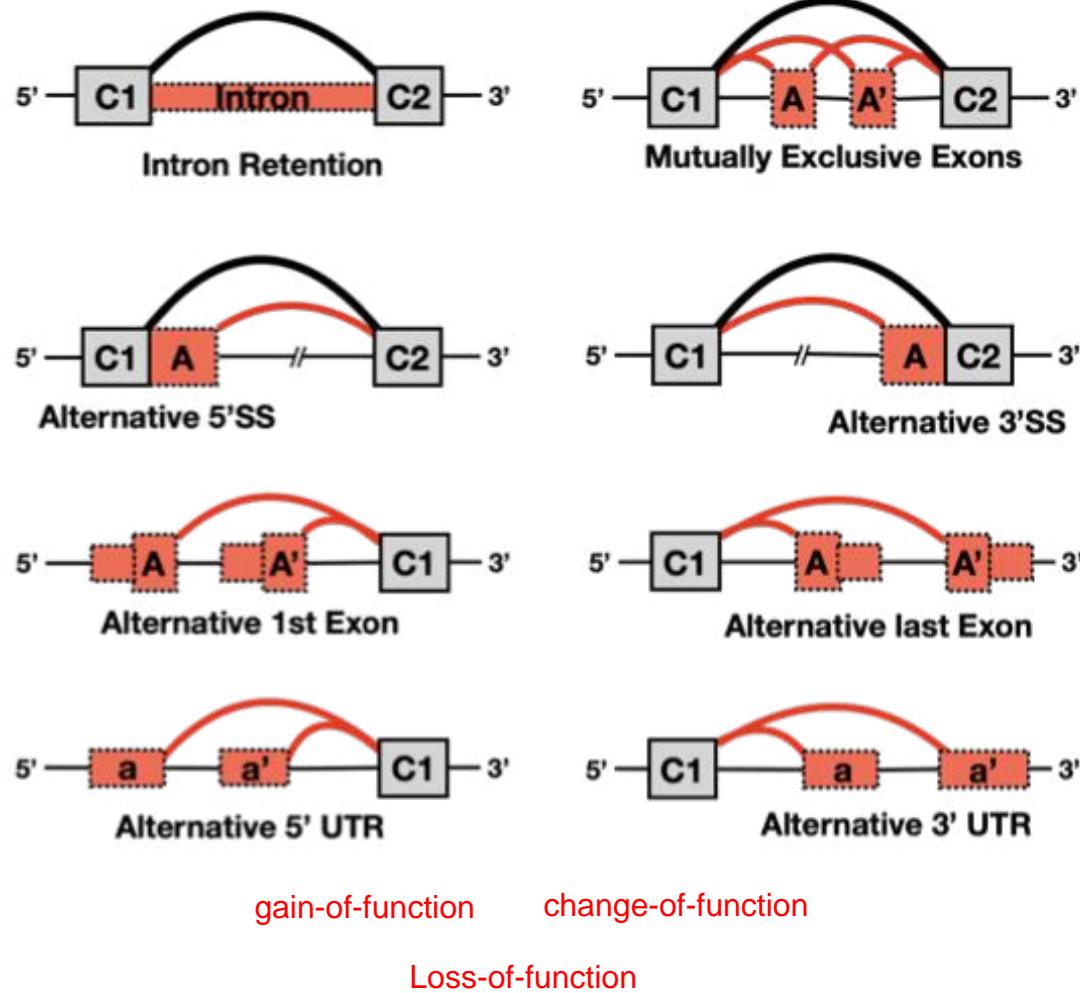
*In alternative splicing events, recognition and joining of a splice site pair by the spliceosome are **in competition with at least another splice site pair**.*



mRNA splicing patterns naturally increase the proportion of altered functional expression of proteins encoded by a particular genes.



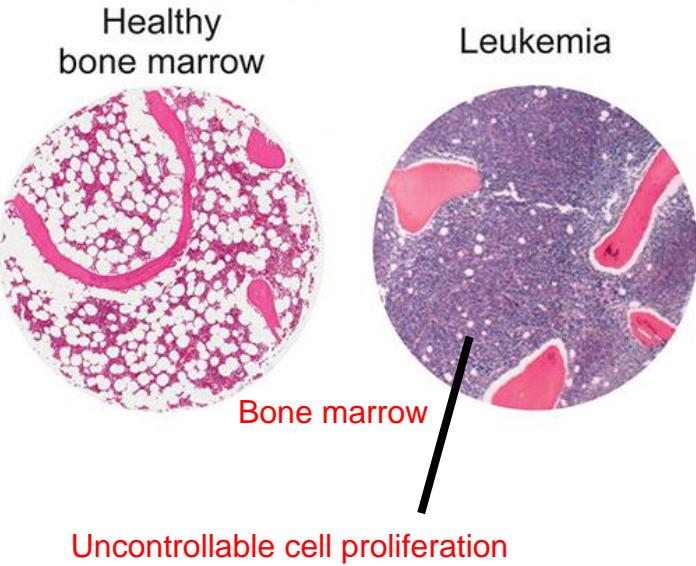
# Alternative splicing events can manifest complex levels of biological regulation



The fidelity of the splicing reaction has been observed commonly altered in heme malignancies such as **myeloid leukemias**.



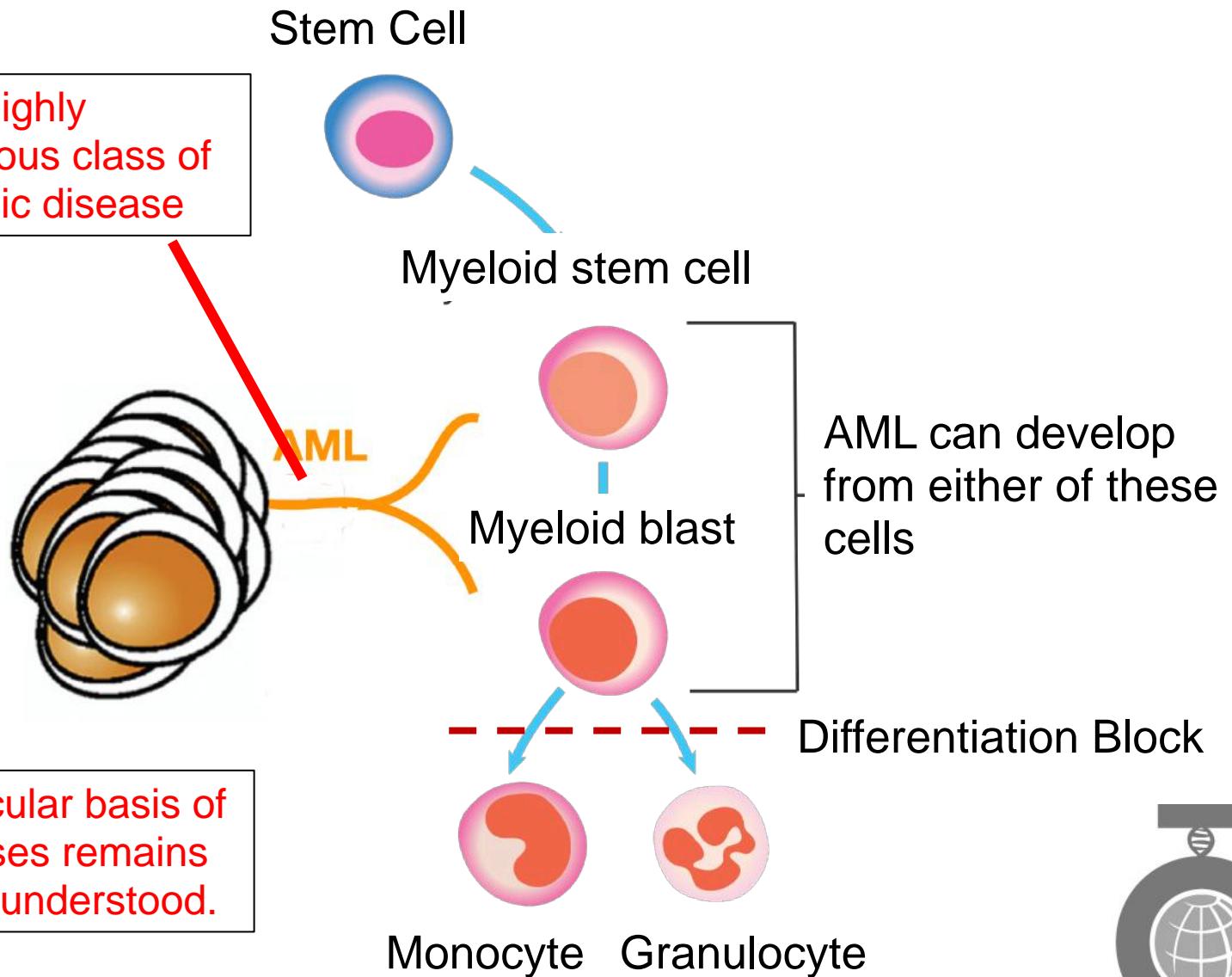
# Acute Myeloid Leukemia (AML) is an aggressive type of cancer characterized by the clonal expansion of poorly differentiated myeloid cell population



## Complications:

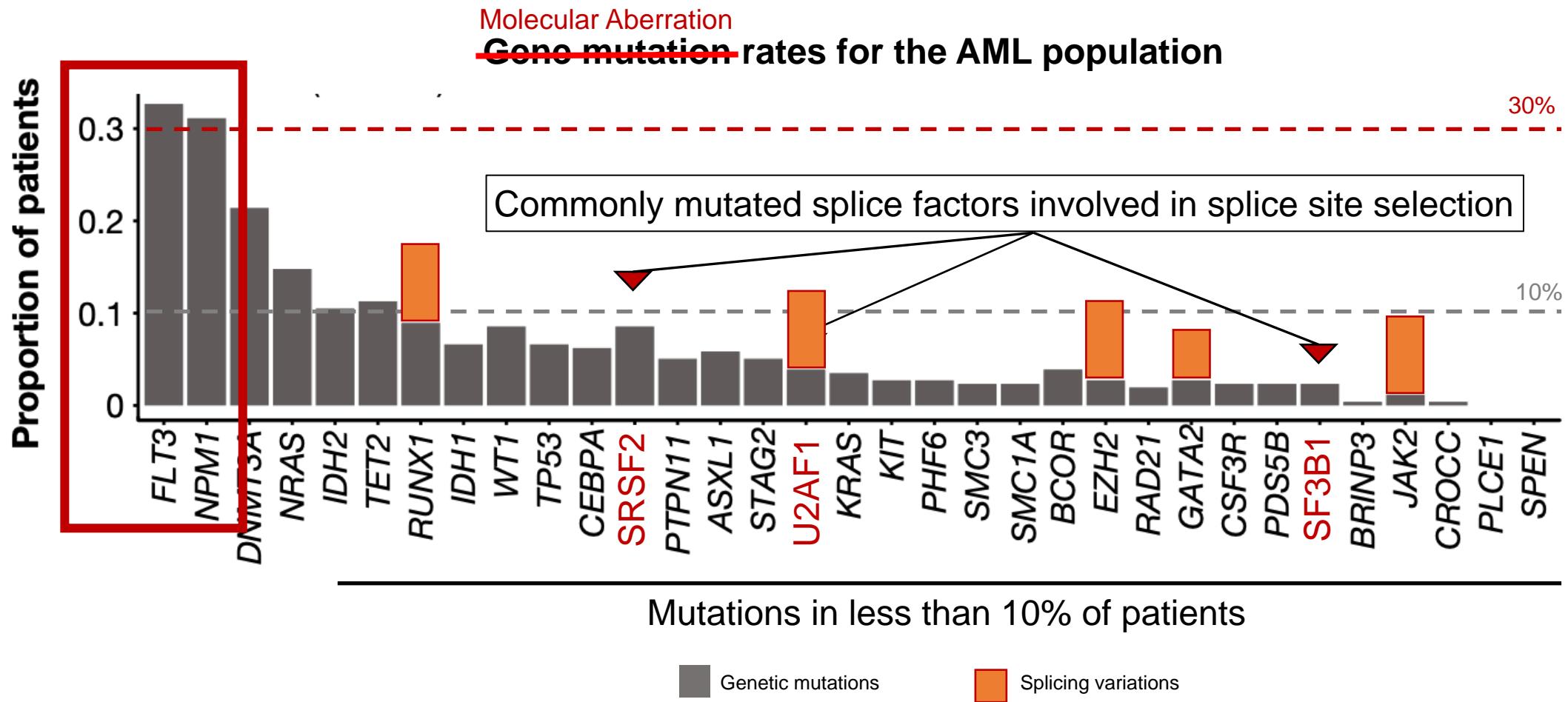
- Weakened immune system (lack of normal leukocytes, WBCs)
- hemorrhage (lack of platelets)
- anemia (lack of red blood cells, RBCs)

AML is a highly heterogeneous class of hematologic disease

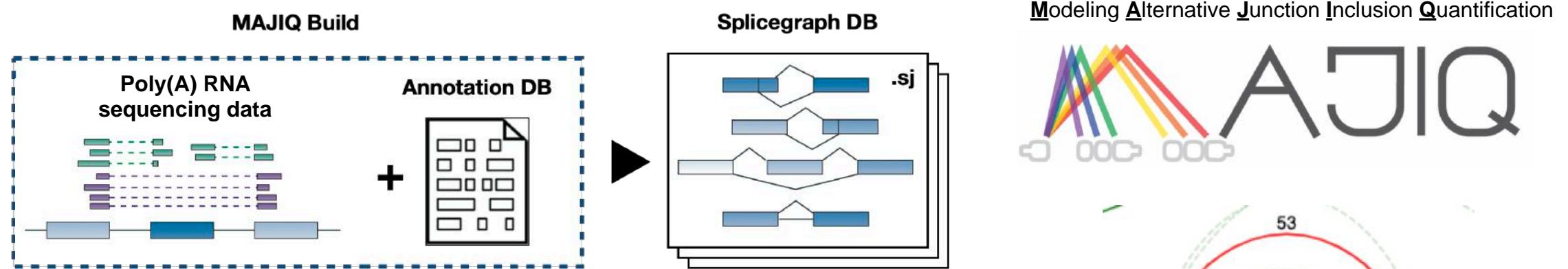


The molecular basis of this disease remains to be fully understood.

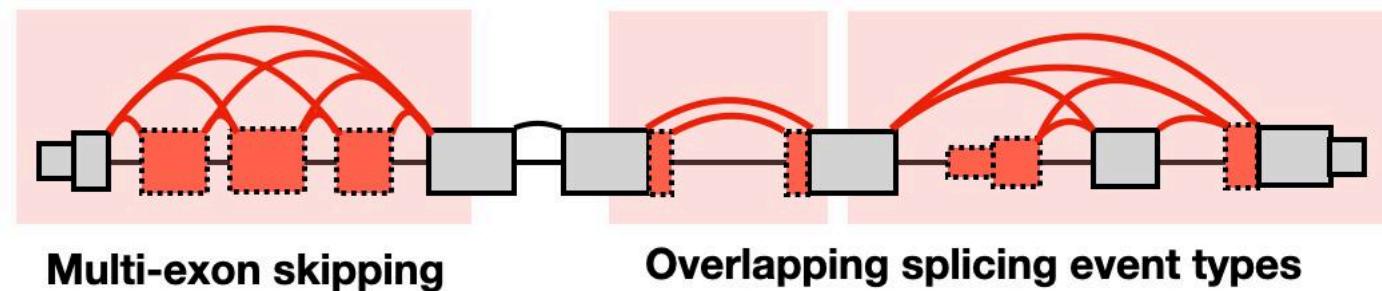
# What if alternative splicing increased the proportion of patients with loss-of-function aberrations?



# High-throughput studies of alternative splicing were only recently possible



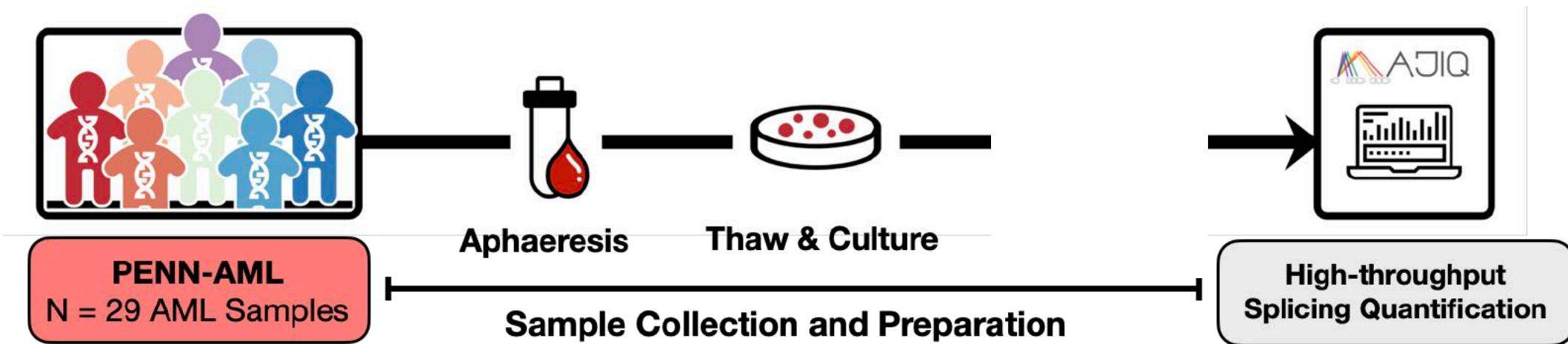
- ✓ The first algorithm to capture ***de novo*** splicing events
- ✓ The first algorithm to capture **complex** splicing events.



# Results and Discussion

# STUDY DESIGN

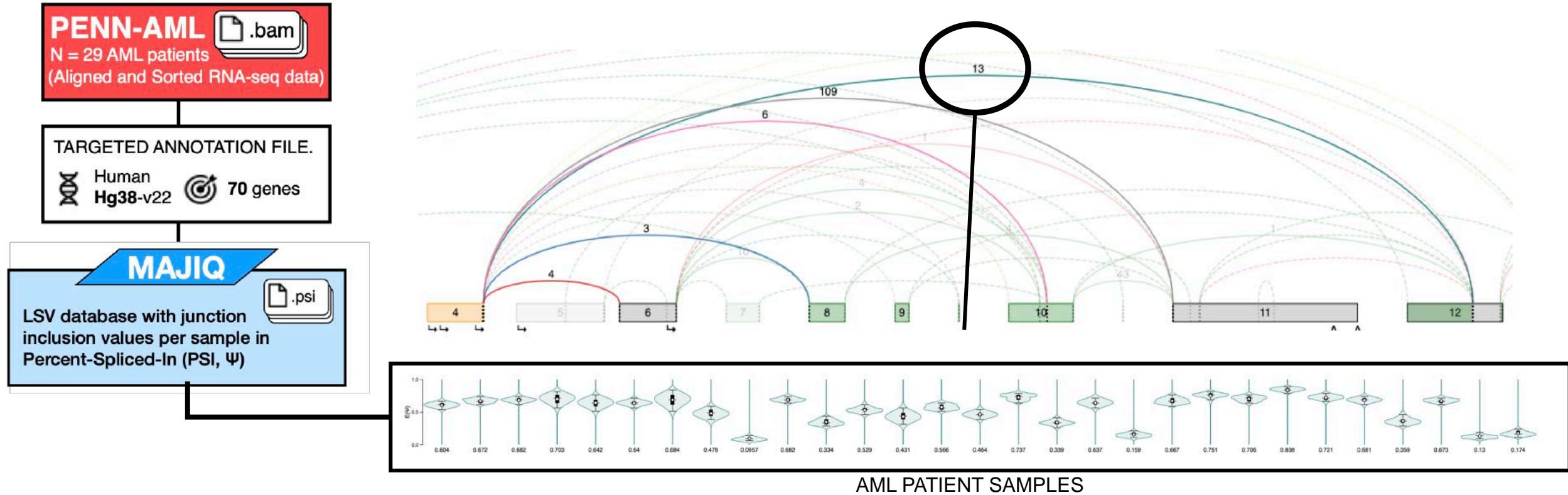
- ✓ Construct an RNA-seq dataset composed of 29 AML samples which we analyzed using the MAJIQ splicing quantification algorithm.



**Hypothesis:** Dysregulated mRNA splicing increases the proportion of AML patients with altered functional expression of proteins encoded by a particular AML-associated gene.



# Intrinsic heterogeneity of the AML patient dataset rendered the differential comparison of alternative splicing between *a priori* defined groups computationally unfeasible.

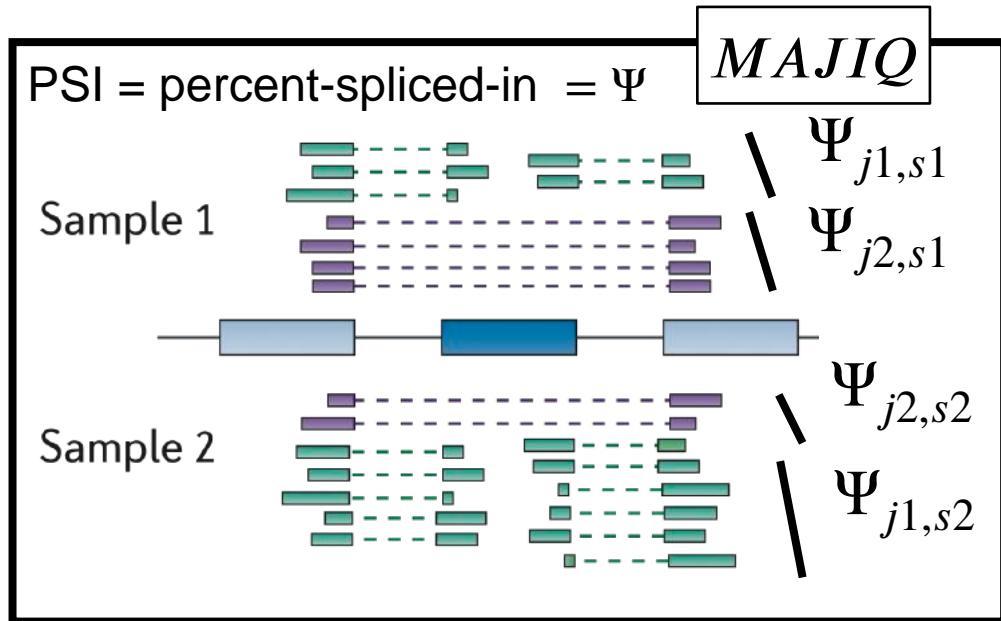


Instead of ranking events by their level of *differential* splicing, I ranked by splicing by its level of *variability*



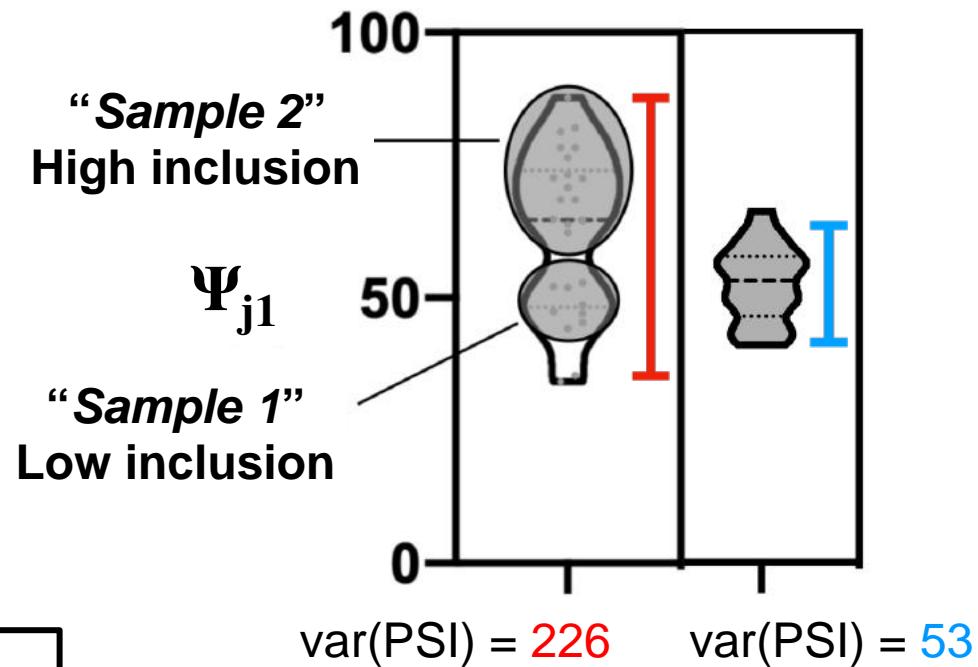
# I developed additional code to extract the most significant variations from the MAJIQ-PSI framework.

What is PSI of a splice junction and how is that calculated from the mRNA-seq data?



Of interest are the splicing events with the **highest level of variance**, which is representative of splicing variability across a particular cohort

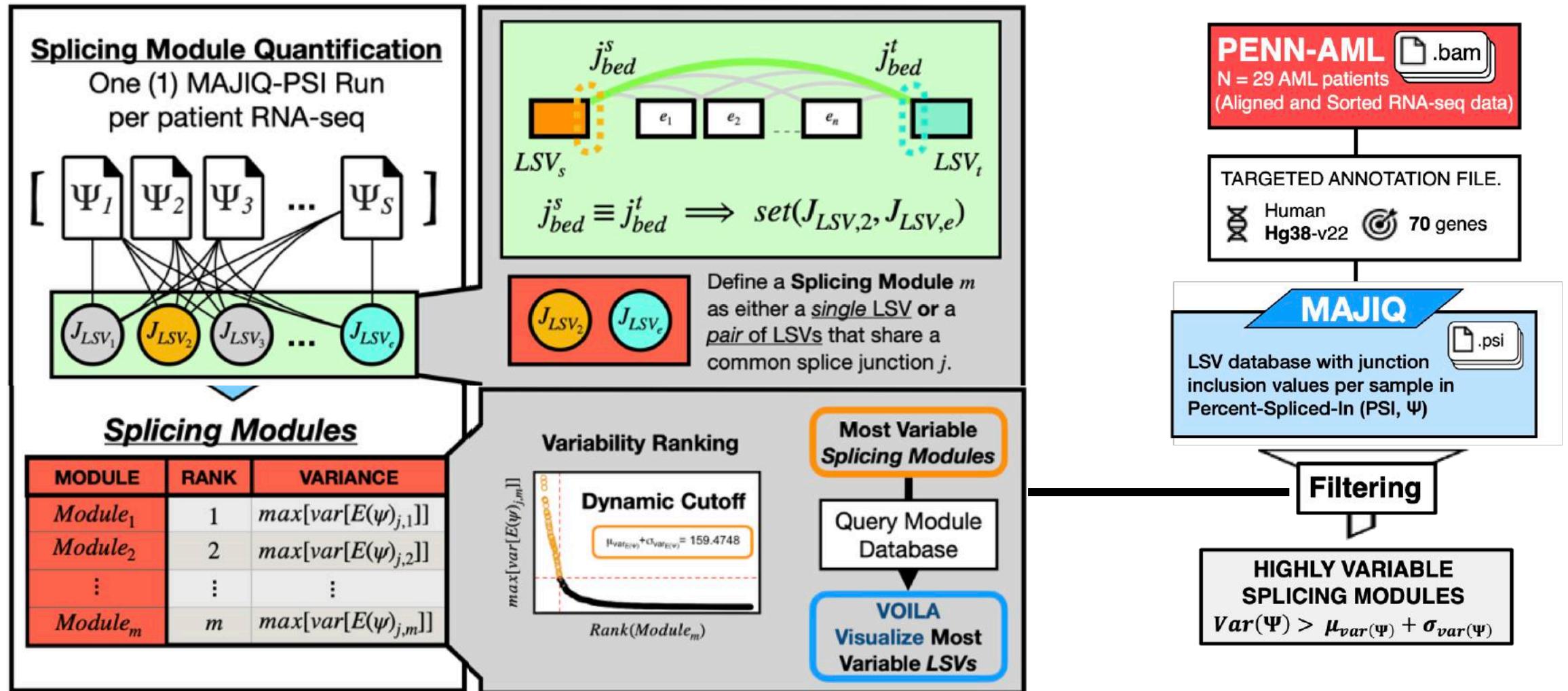
## What is Variance of PSI?



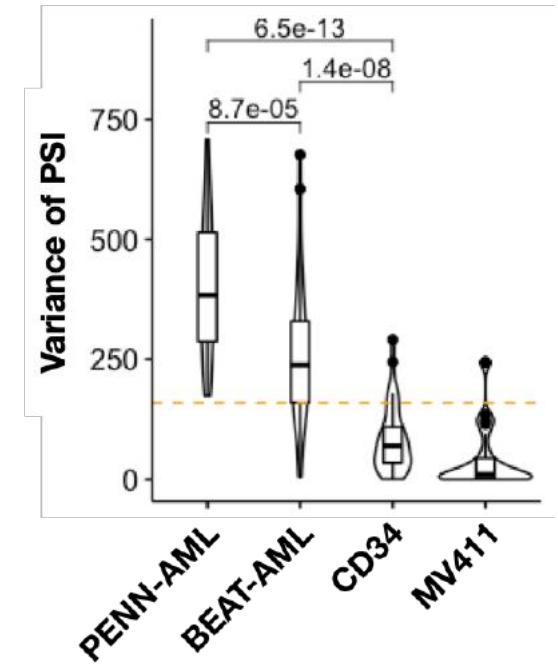
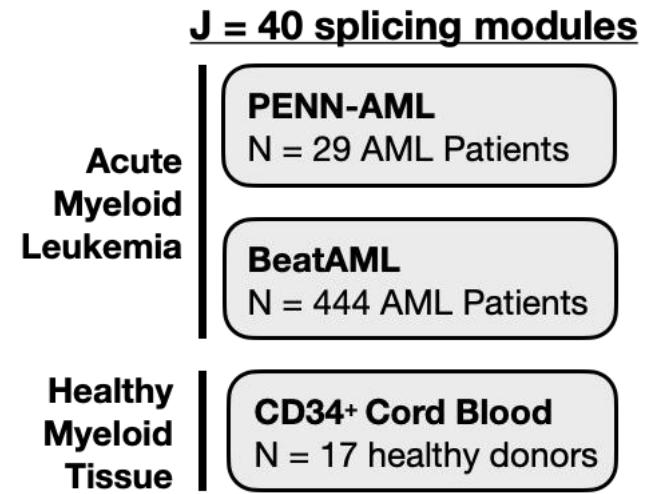
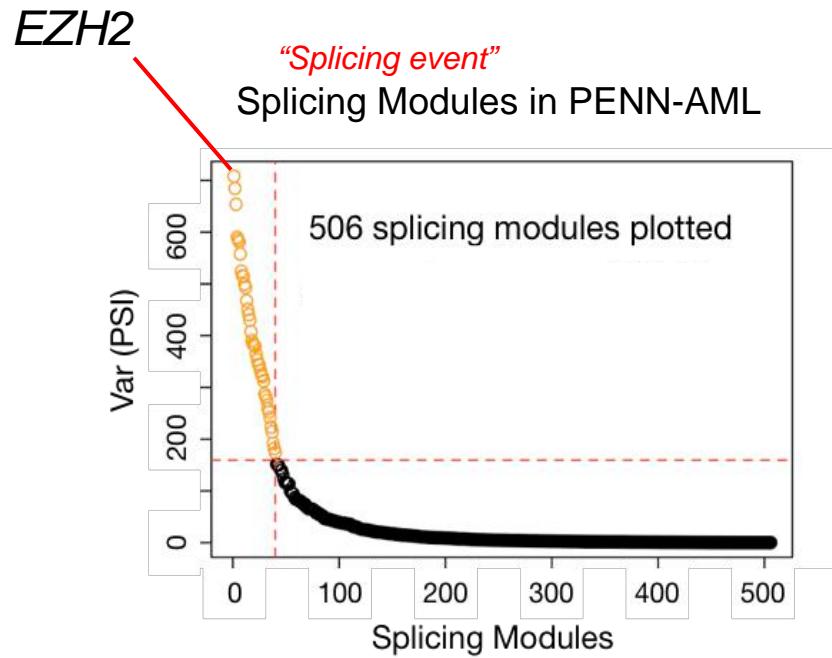
$$\text{var}(\Psi_j) = \frac{\sum (\psi_{j,s} - \bar{\psi})^2}{N-1}$$



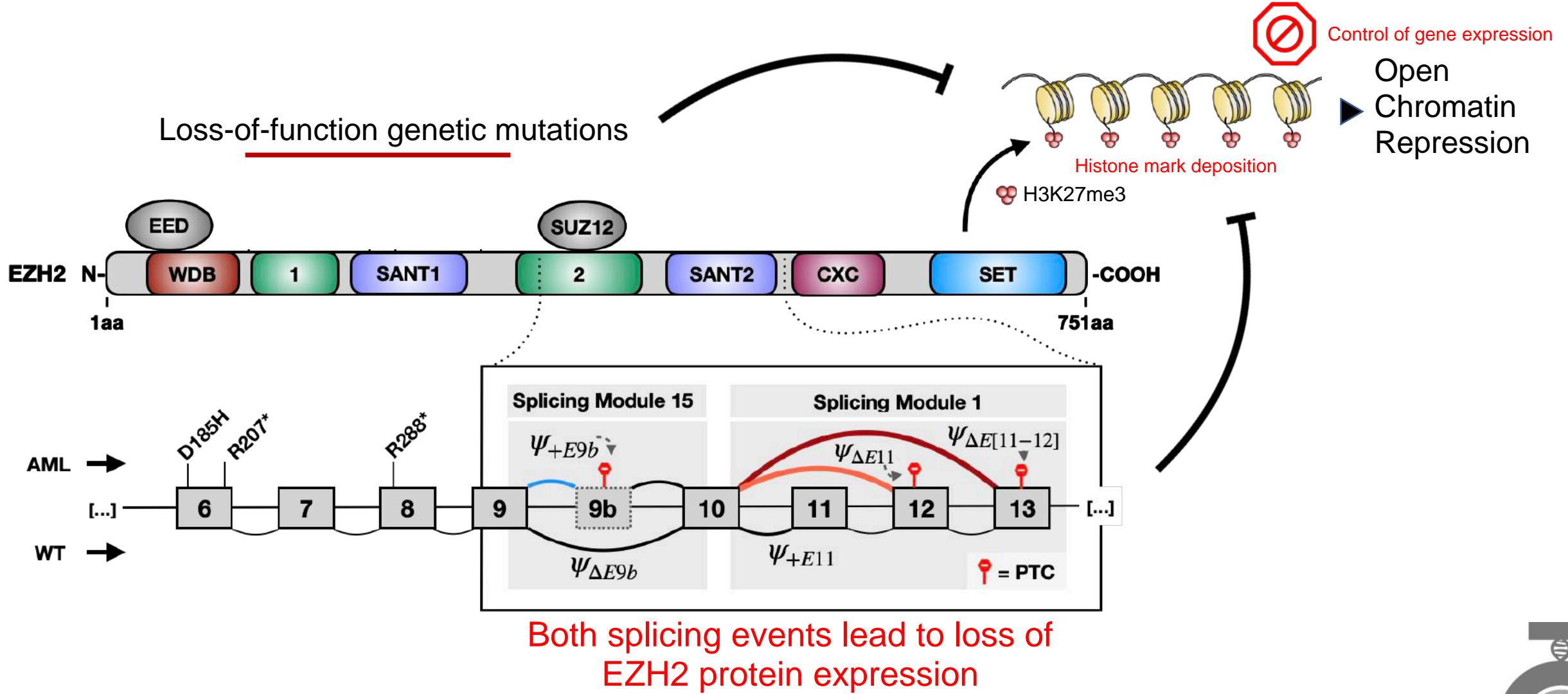
# Overlapping splicing events were compiled into distinct *splicing modules*



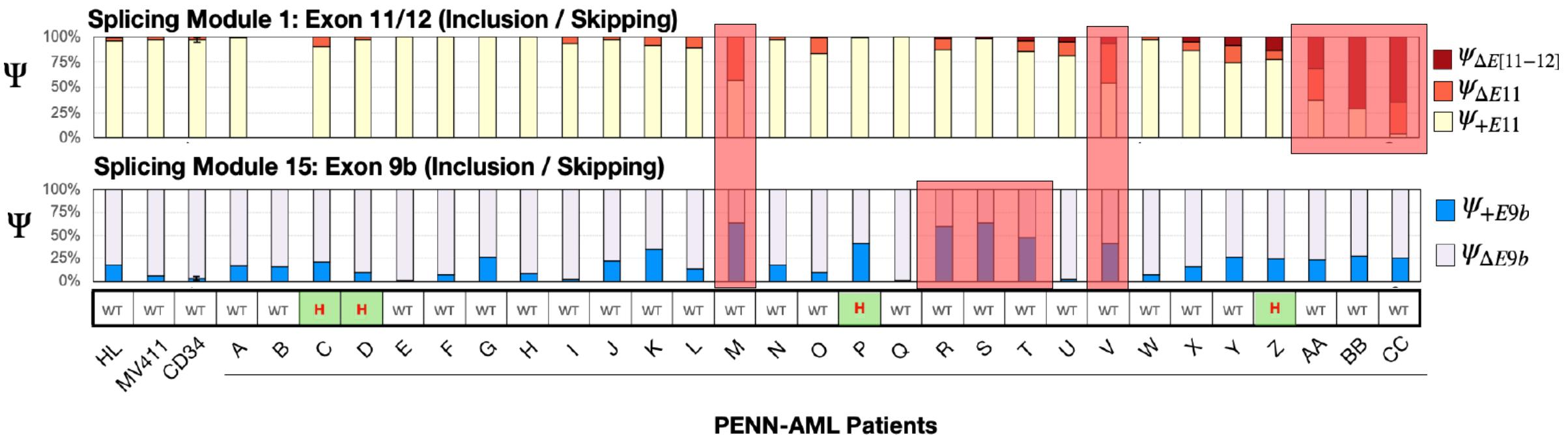
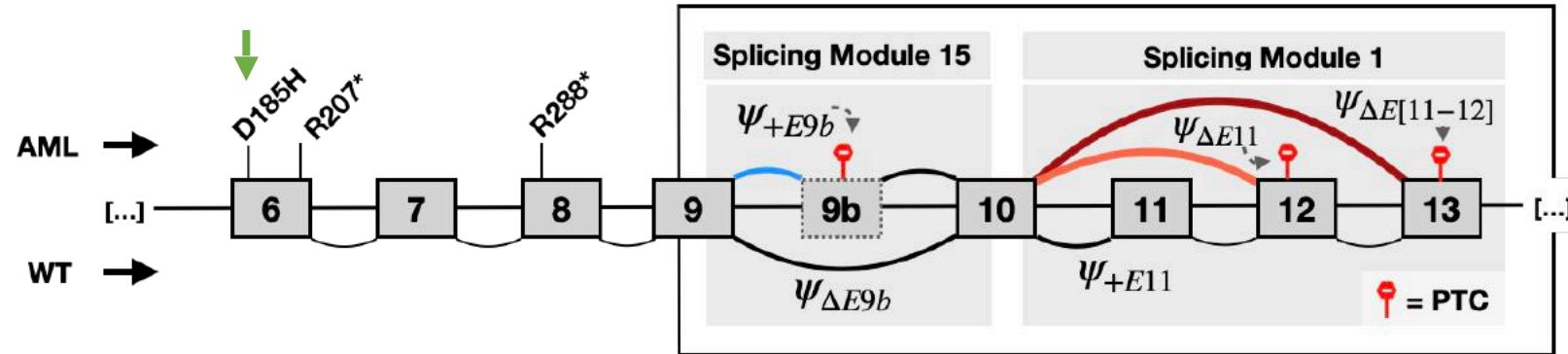
The MAJIQ pipeline quantifies 40 splicing modules in AML-related genes that are highly variable across the PENN-AML cohort.



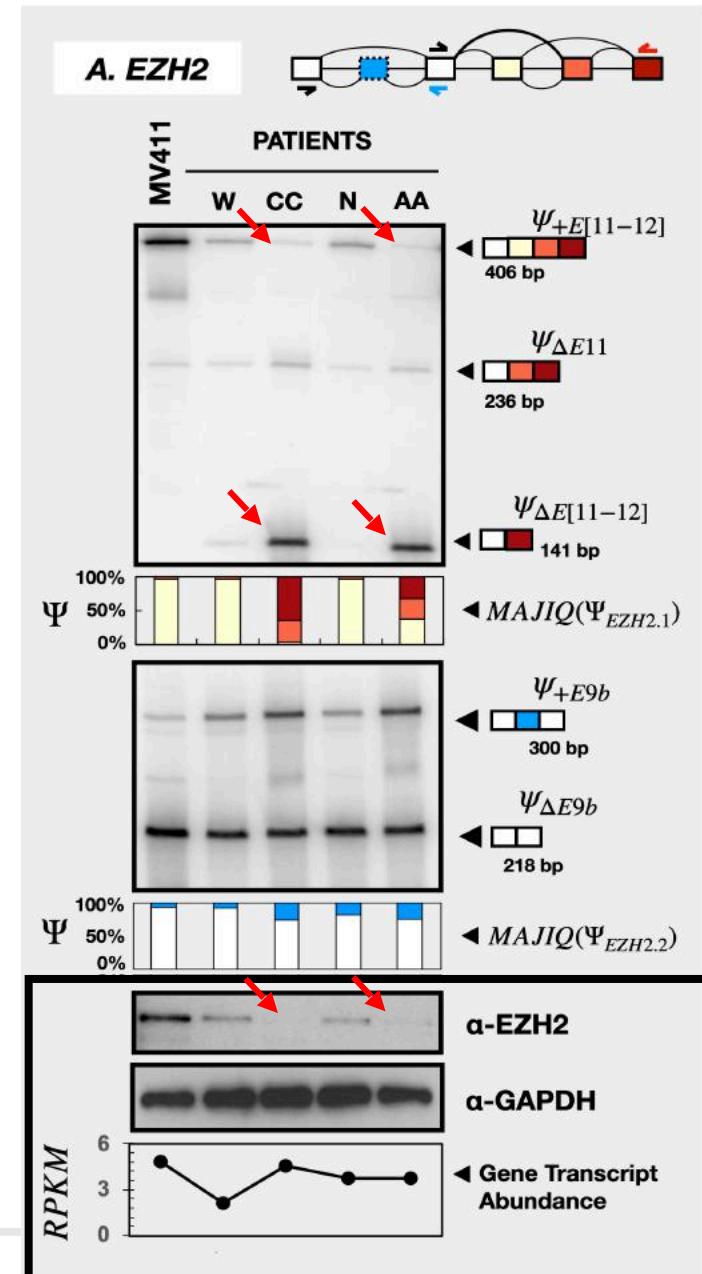
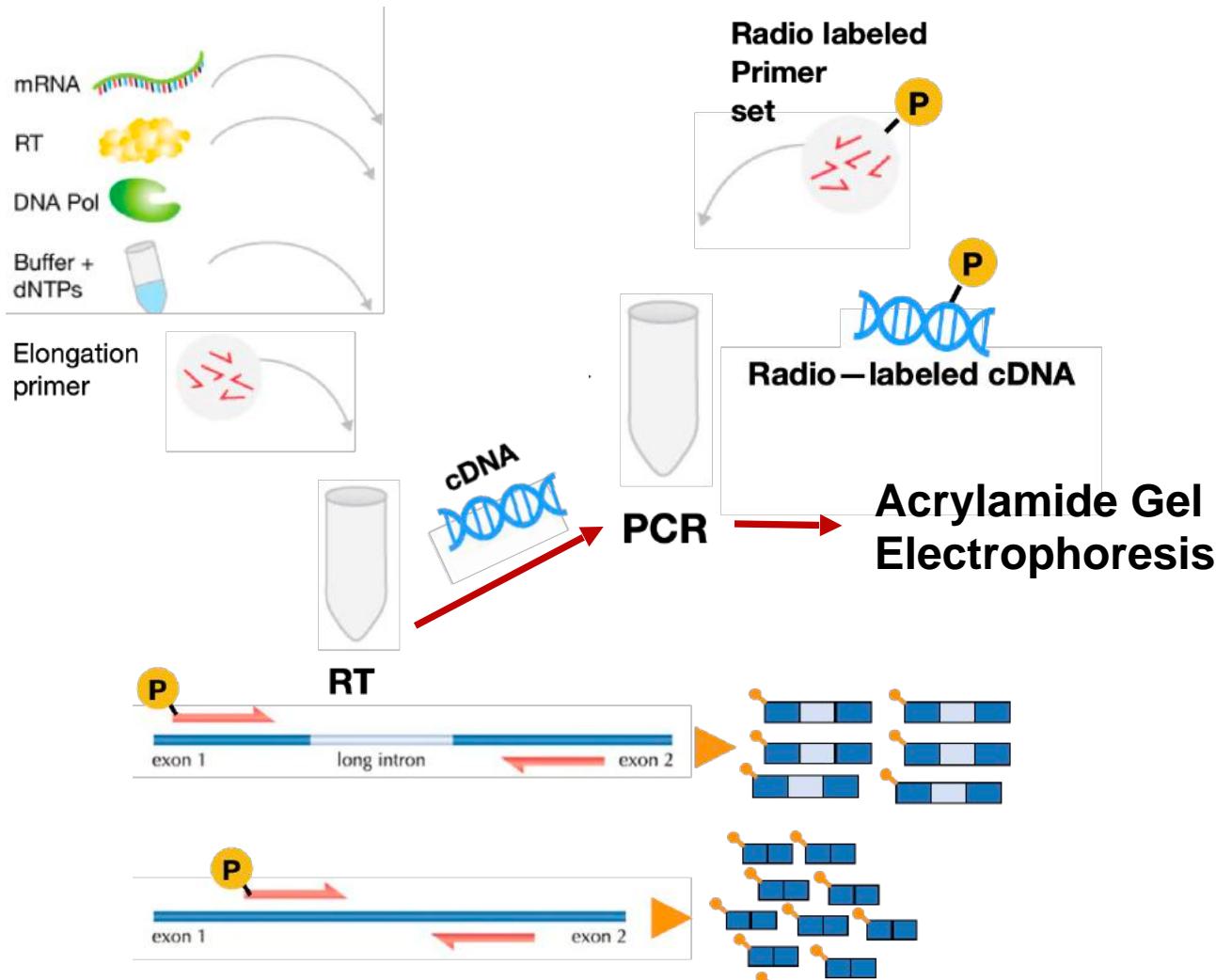
# Epigenetic factor EZH2 exhibits the highest splicing variability



Accounting for splicing triples the number of AML patients with reduced EZH2 function relative to that predicted from cis-mutations alone.

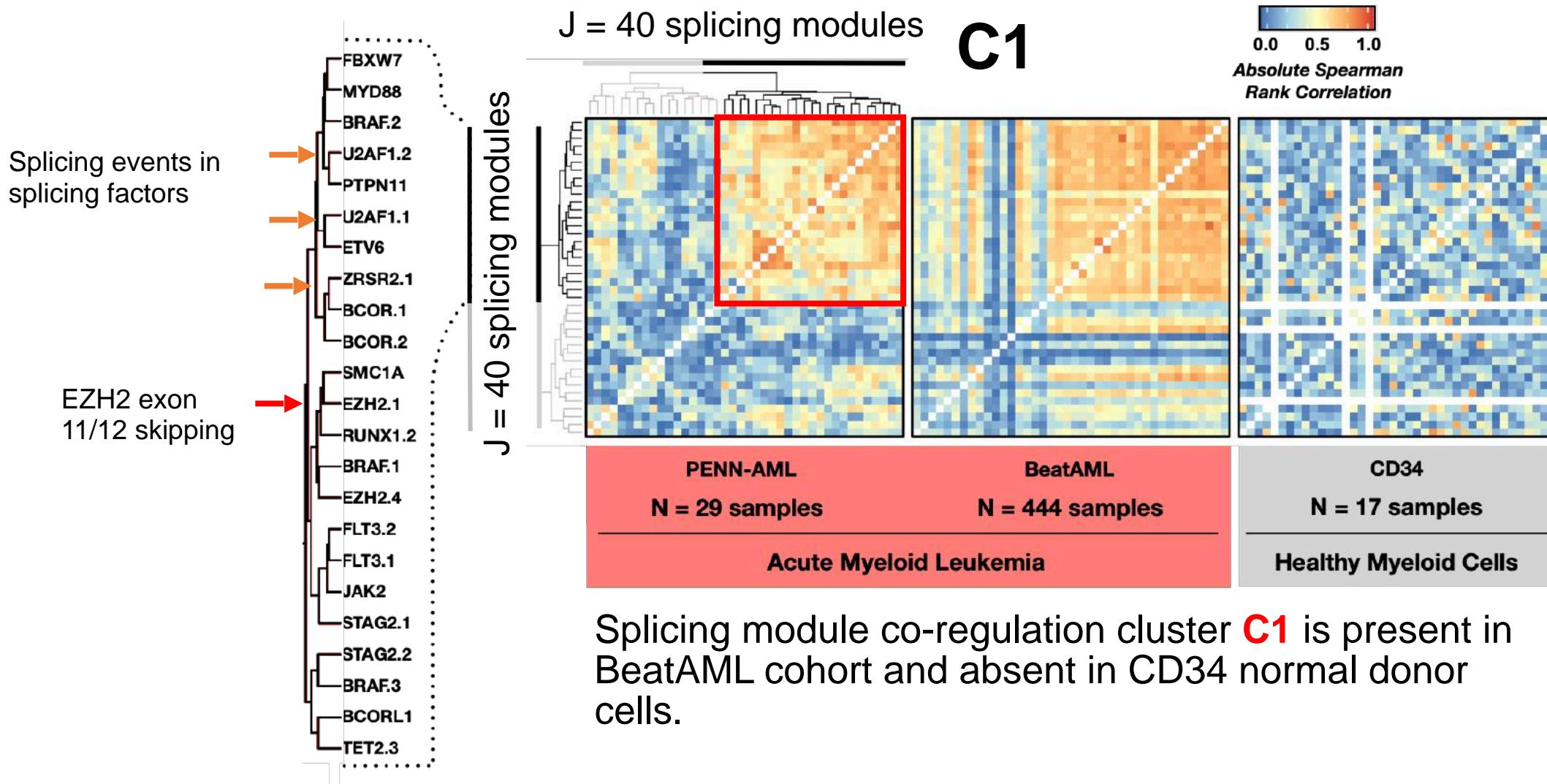


# Detection of splicing events in mRNA samples



Results and Discussion

# Splicing coregulatory signal discovered across AML cohorts



# Key Takeaway up to this point...

- Splicing variations disrupt the coding potential of AML related genes and increase the proportion of loss-of-function aberrations across AML patients.
- A subset of these splicing variations are also highly correlated with each other, suggestive of underlying biological co-regulation across the AML population.



*The identification that altered mRNA splicing patterns dysregulate a subset of AML-associated genes led me to ask a slightly different question:*

What are *other* important genes that may be functionally dysregulated by altered splicing that have not yet been associated with the *de novo* occurrence of AML?



# Recently identified pathogenic germline variants in familial AML give us a new set of genes to query for highly variable splicing events.

Published: 02/25/2020



ARTICLE

<https://doi.org/10.1038/s41467-020-14829-5>

OPEN

Check for updates

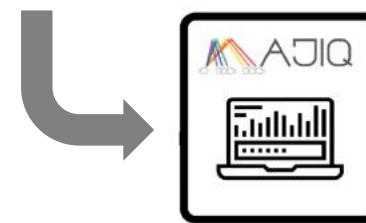
The complex genetic landscape of familial MDS and AML reveals pathogenic germline variants

Ana Rio-Machin et al.<sup>#</sup>

Rio-Machin, A. et al. The complex genetic landscape of familial MDS and AML reveals pathogenic germline variants. *Nat Commun* 11, 1044 (2020).

## Familial AML-related genes

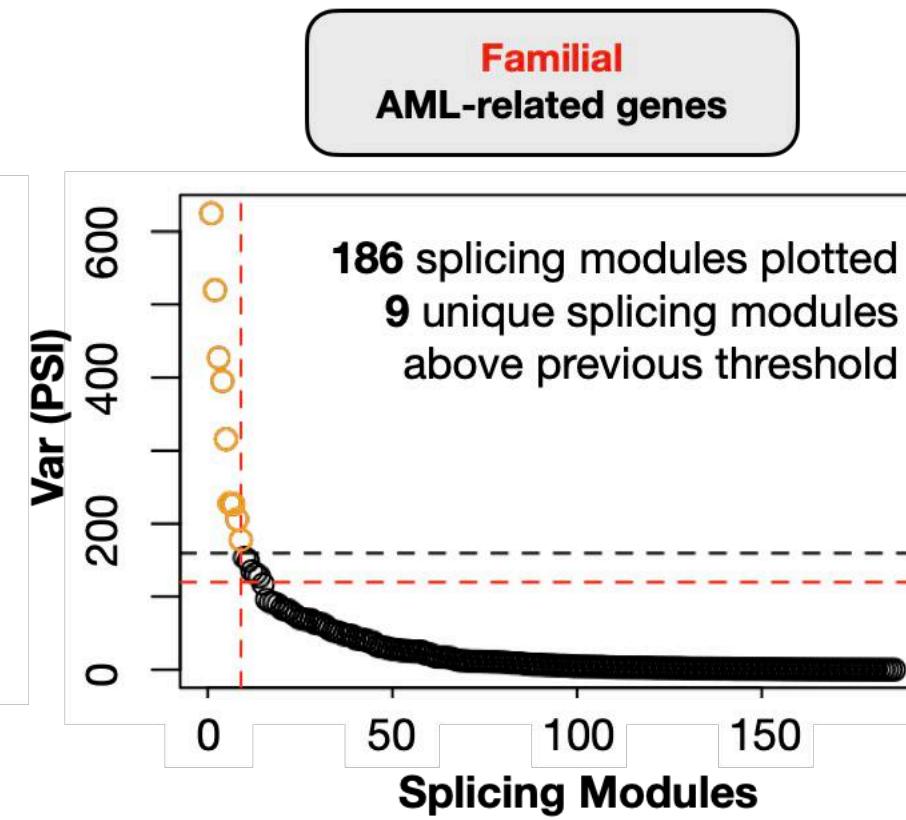
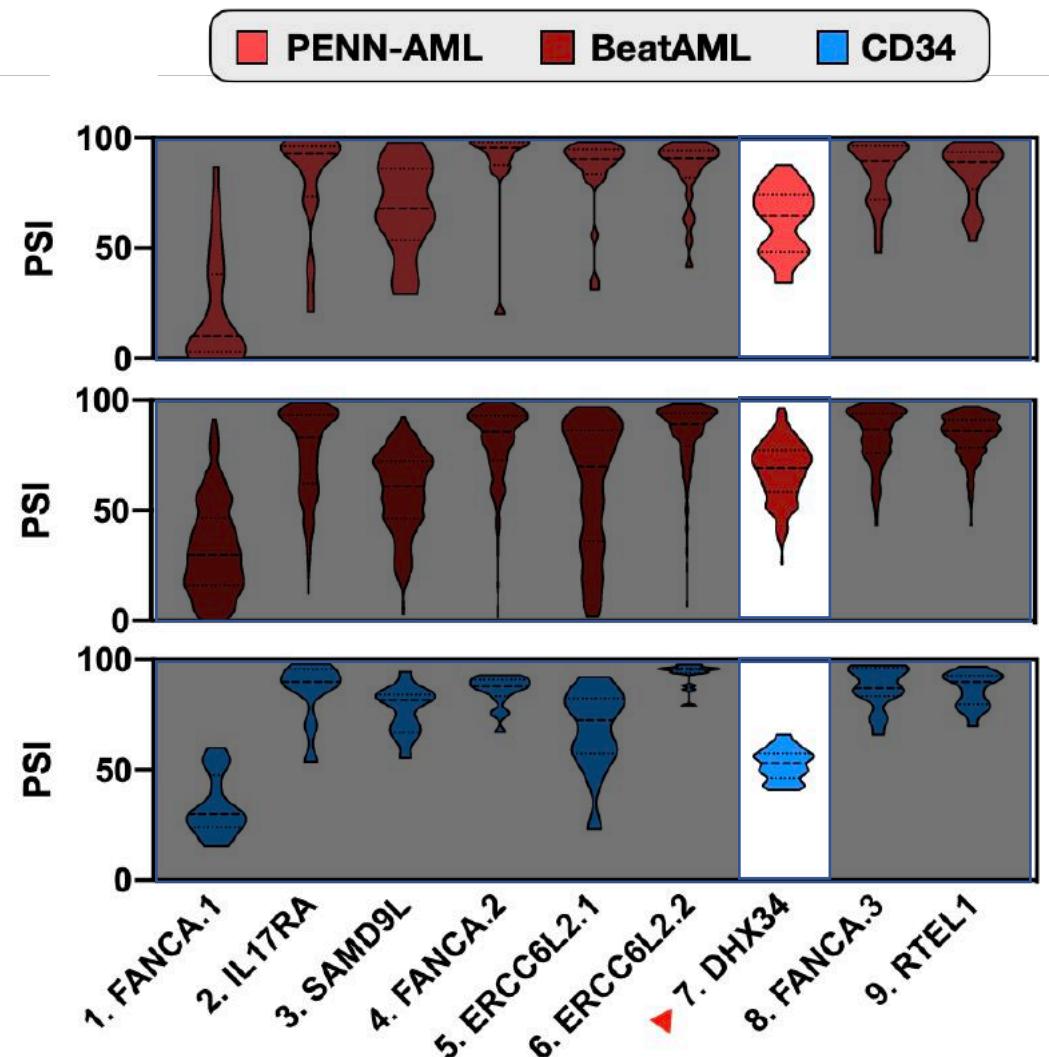
DNAH9	DHX34	ERCC6L2
WAS	NAPRT	DDX41
MECOM	TERT	FANCA
GP6	SRP72	ADA
SEC23B	SAMD9L	RTEL1
SH2B3	IL17RA	TERC
SBDS	PRF1	



High-throughput  
Splicing Quantification



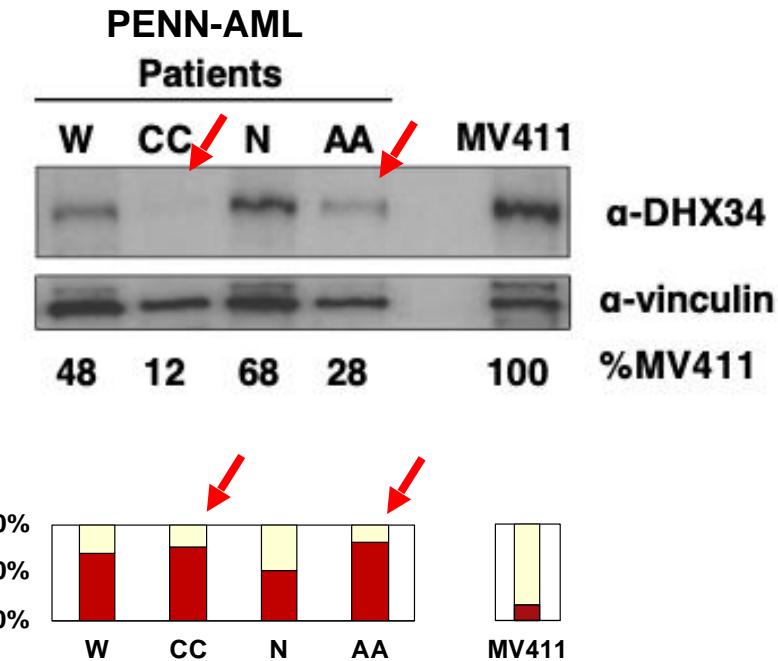
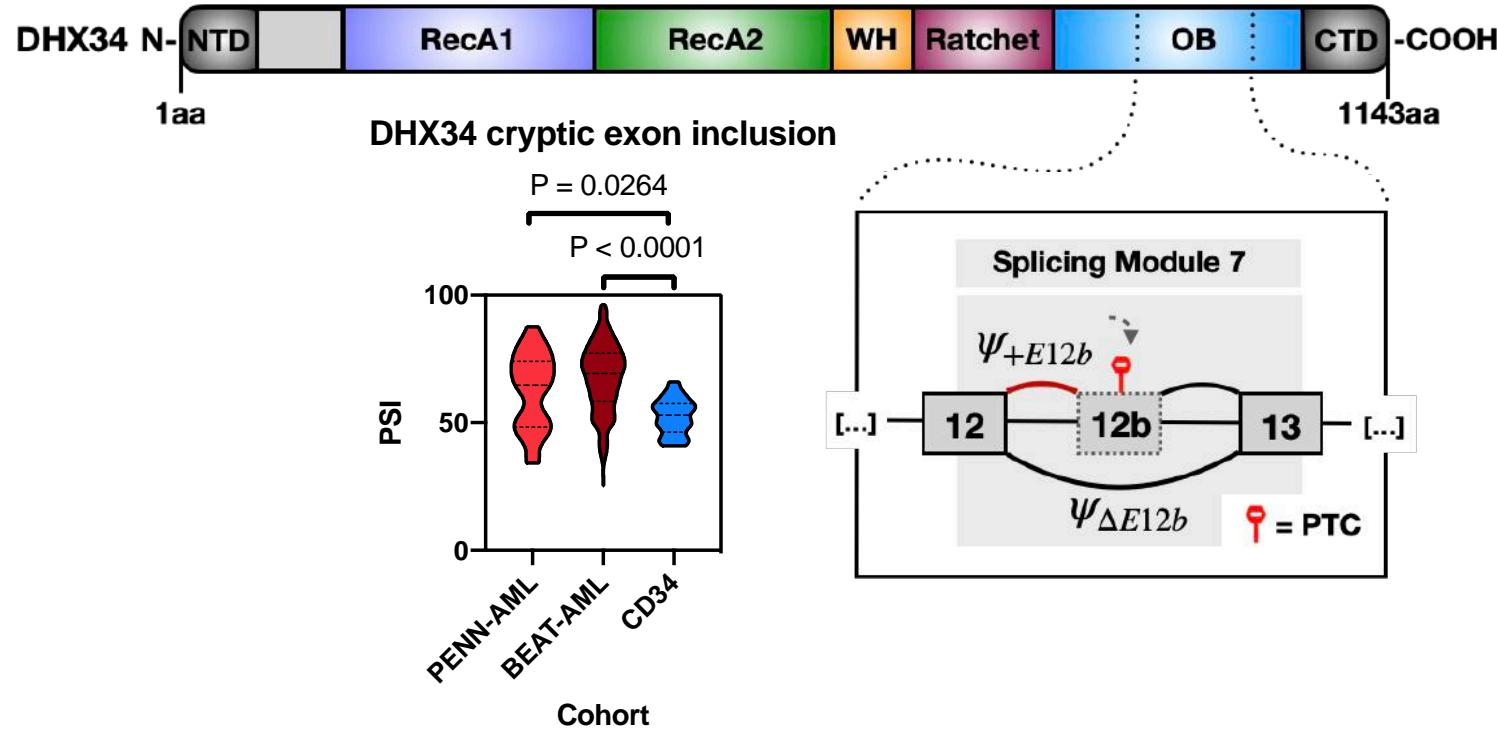
# Splicing variations were also found in genes related to familial cases of AML



Of particular interest is the splicing module in the transcripts of gene *DHX34*



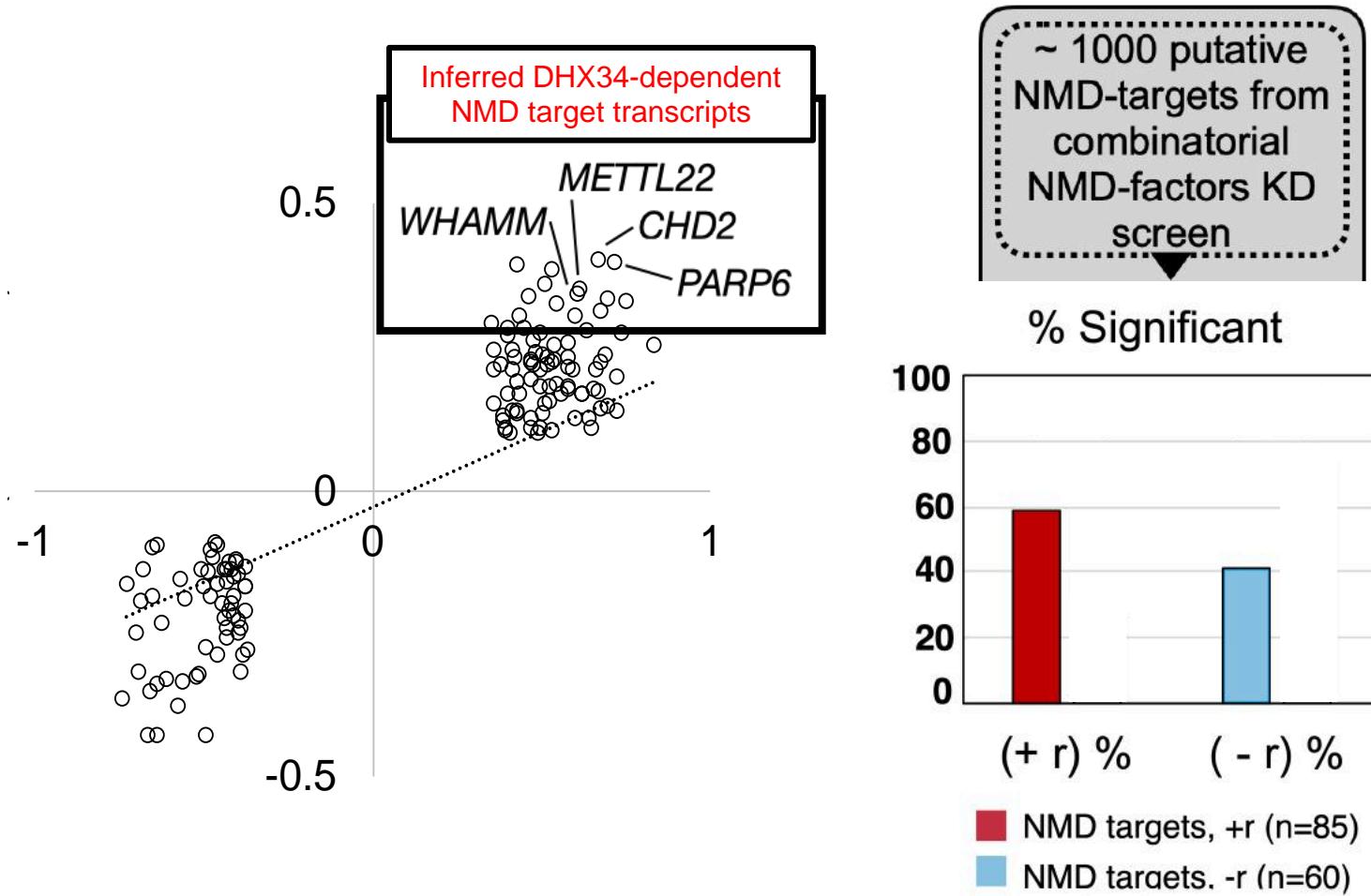
# Poison exon inclusion in *DHX34* inclusion results in significantly higher loss of DHX34 protein expression in the AML population.



**Q:** Does loss of DHX34 protein expression due to poison exon 12b inclusion decrease NMD activity?



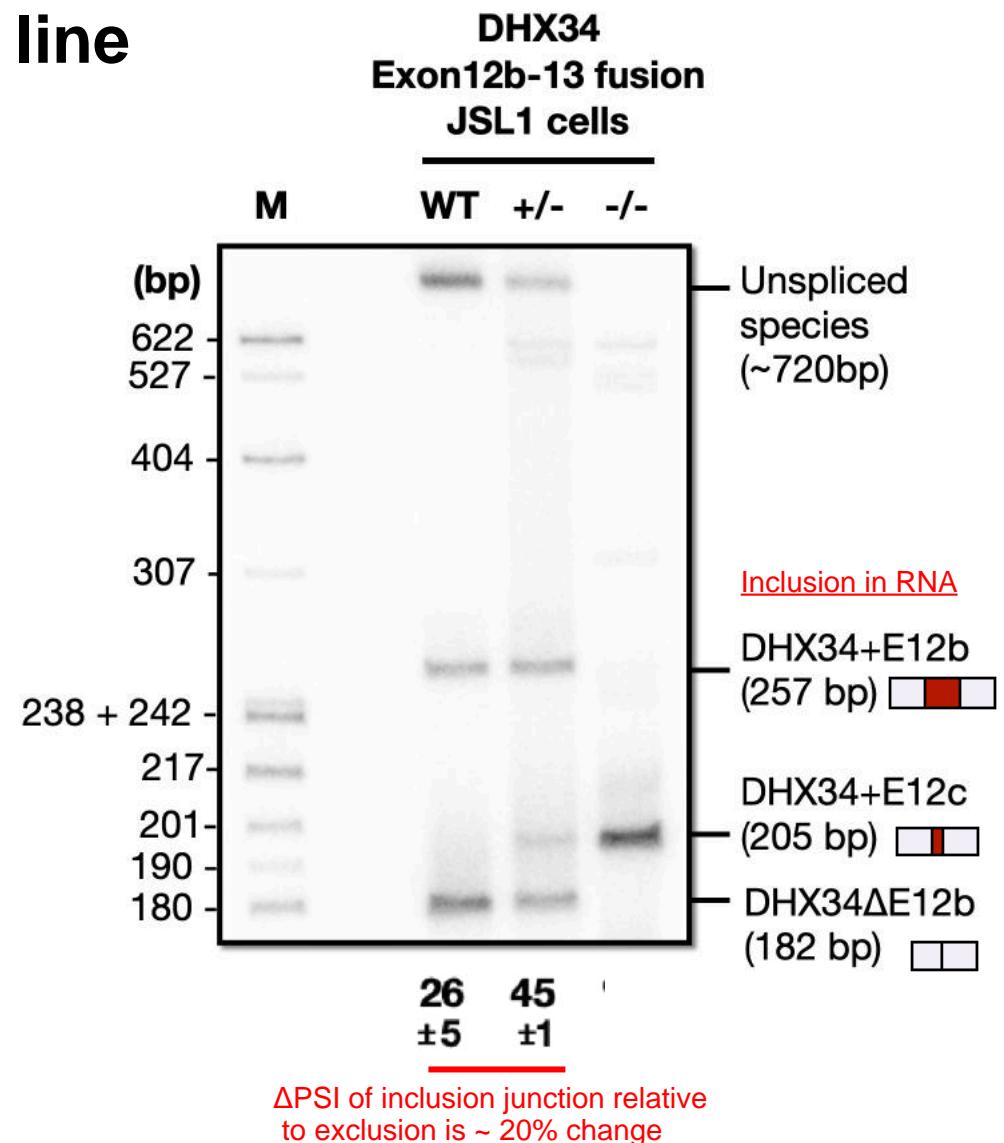
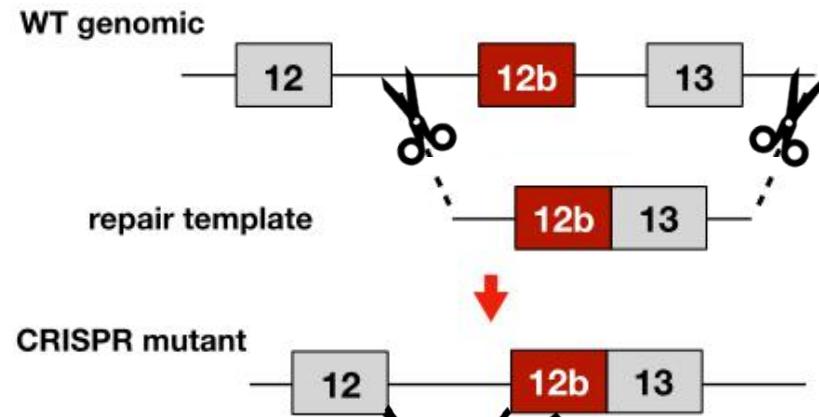
# Inclusion of *DHX34* poison exon 12b significantly correlates with gene expression of a subset of putative NMD targets.



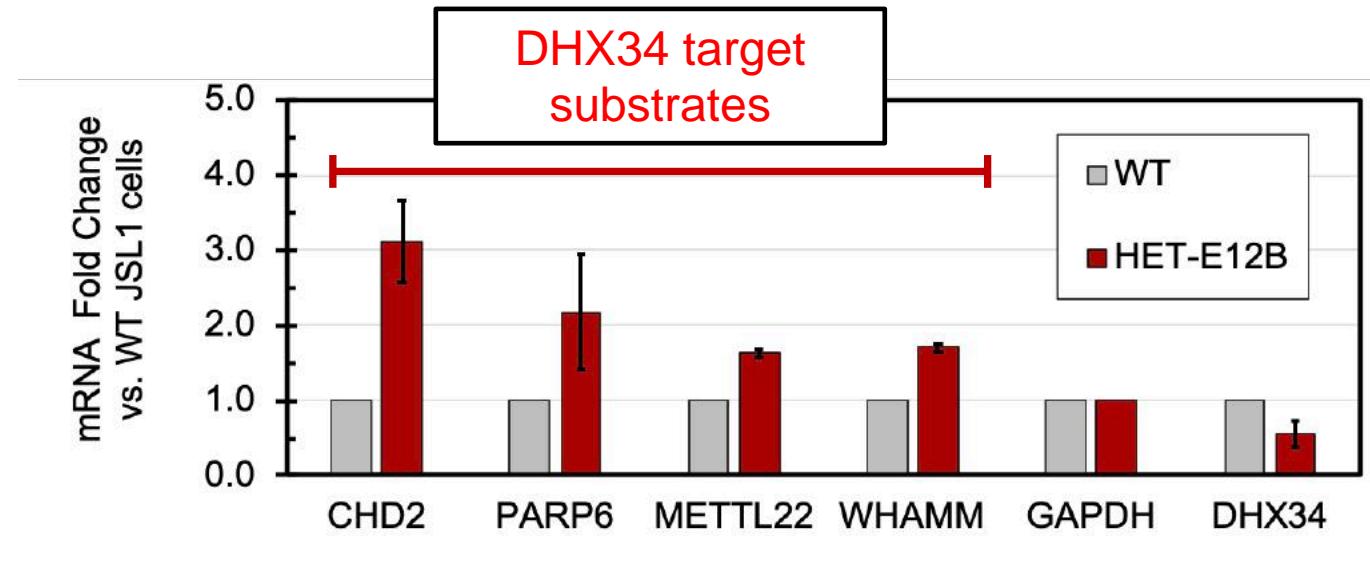
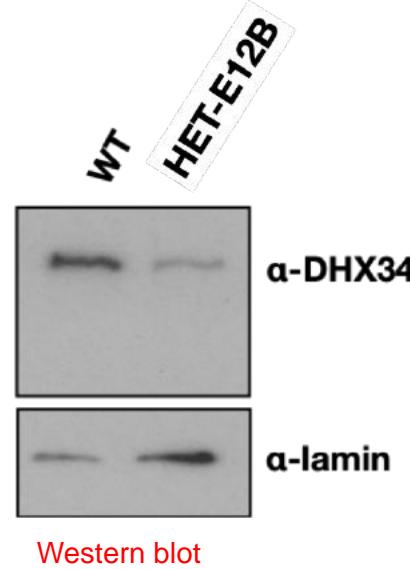
Therefore, I had now a confident **endpoint** for which to test the consequent effects of forced *DHX34* poison 12b inclusion in a model cell line.



# Using CRISPR, we forced higher inclusion of DHX34 poison exon 12b within a JSL1 cell line



# Forced inclusion of DHX34 exon 12b in JSL1 cell lines leads to reduced DHX34 protein expression and increased target substrate abundance.

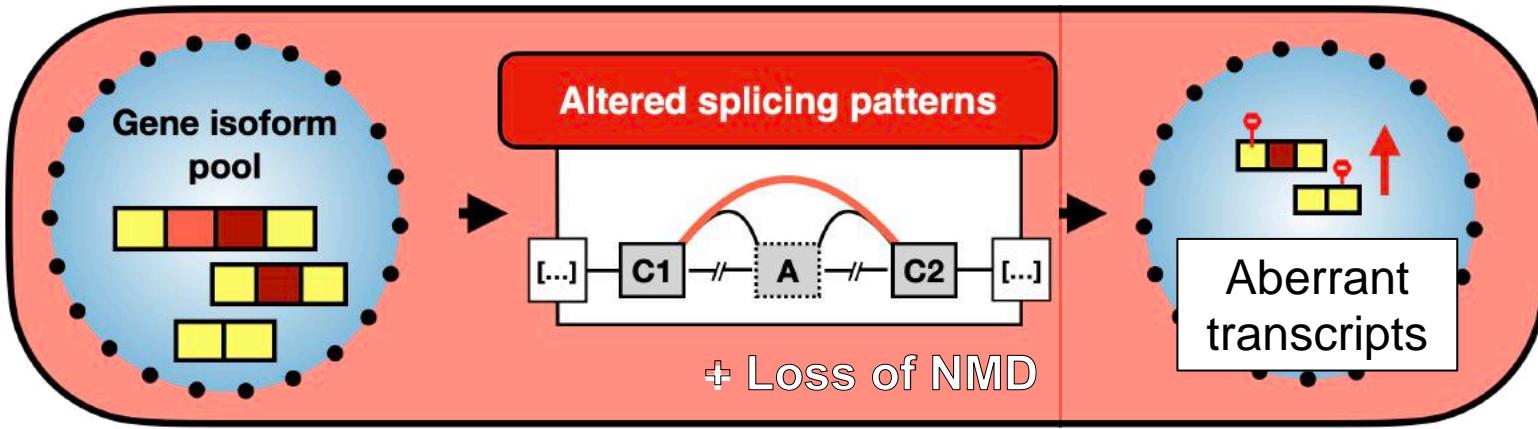


This is suggestive that NMD inactivation via reduction of DHX34 protein expression is a potential contributor to AML molecular pathology.





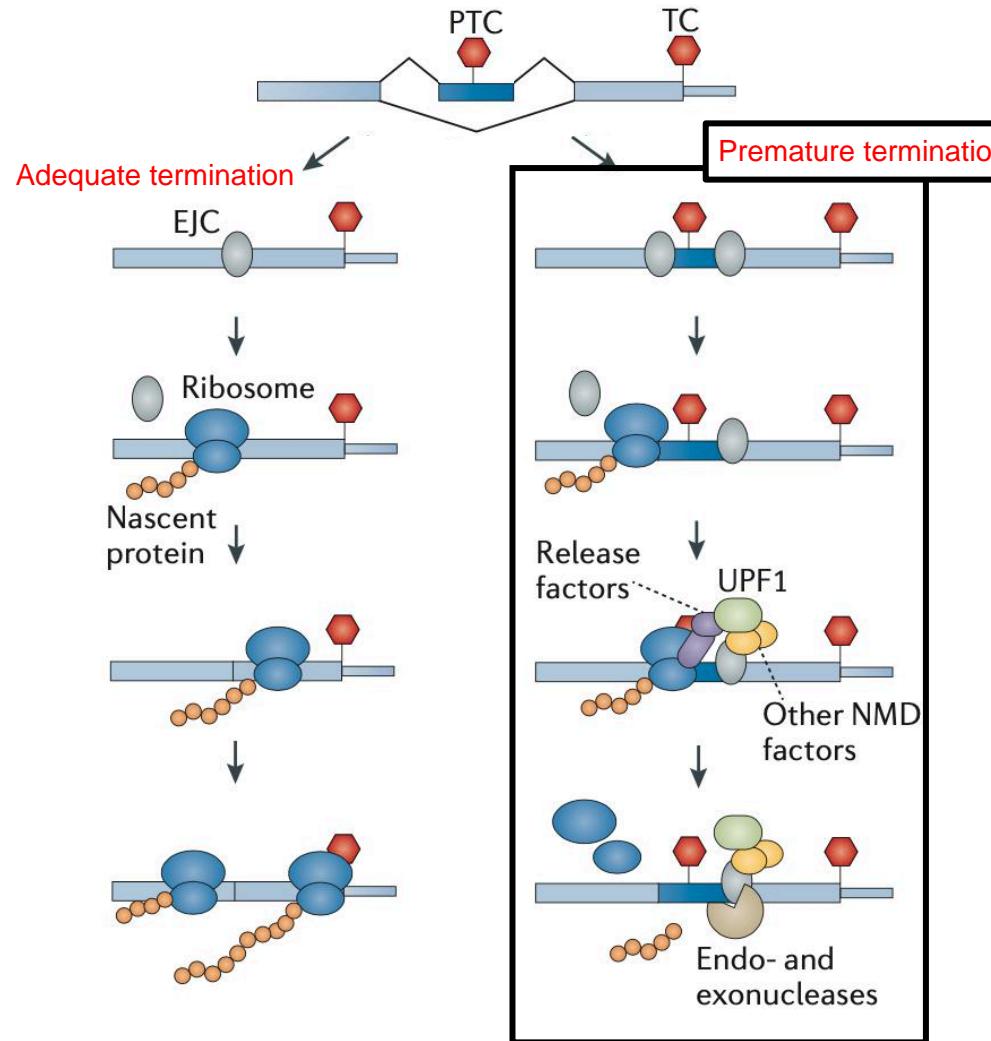
The higher inclusion of DHX34 poison exon 12b suggests that dysregulation of NMD may be leveraged by the AML blast cell to augment oncogenicity.



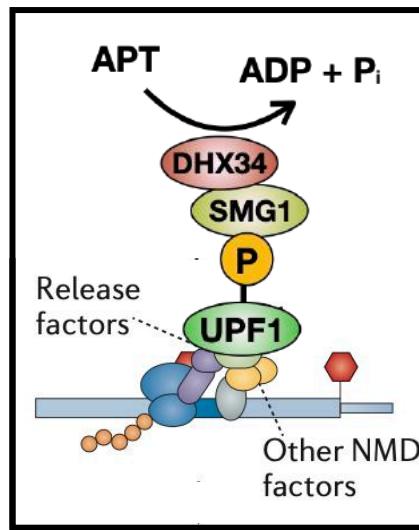
**Are there any other potentially deleterious splicing events occurring in NMD factors across AML patients?**



# NMD removes non-productive transcript that may interfere with cell functions.



# NMD removes non-productive transcript that may interfere with cell functions.



I curated a gene list

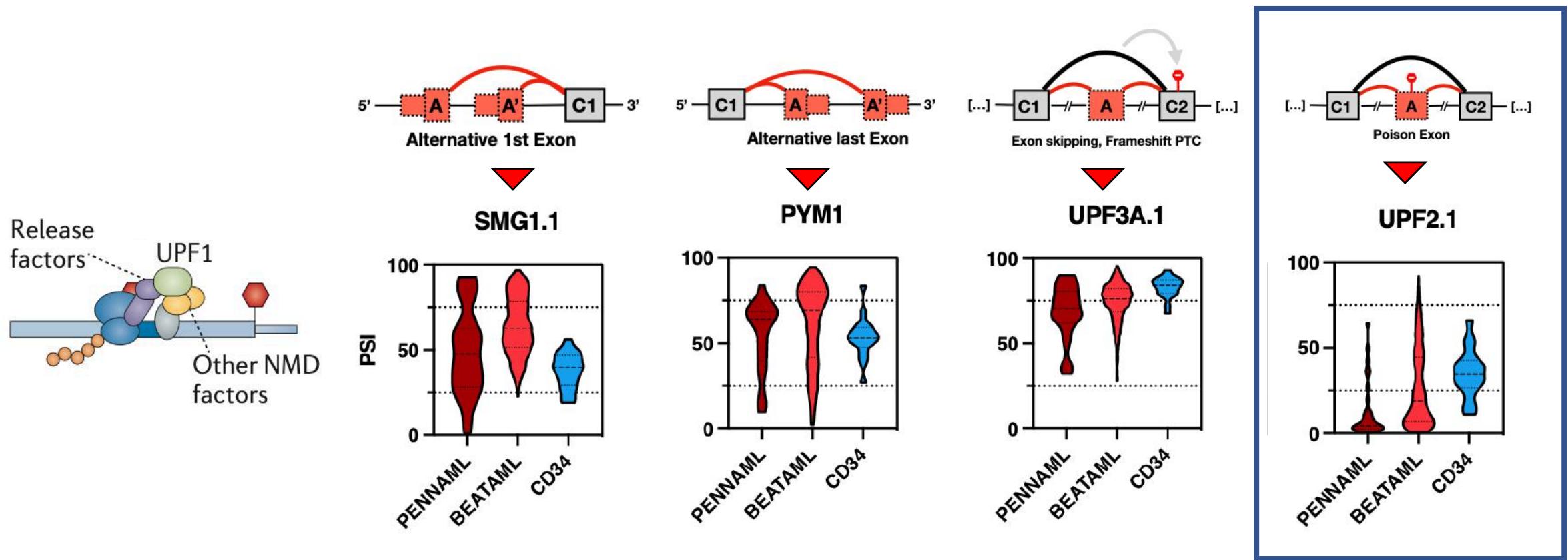
## NMD factor genes

UPF1	EIF4A3
SMG6	UPF2
GSPT1	SMG1
CASC3	MAGOH
SMG7	UPF3A
ETF1	PYM1
UPF3B	SMG5
DHX34	RBM8A

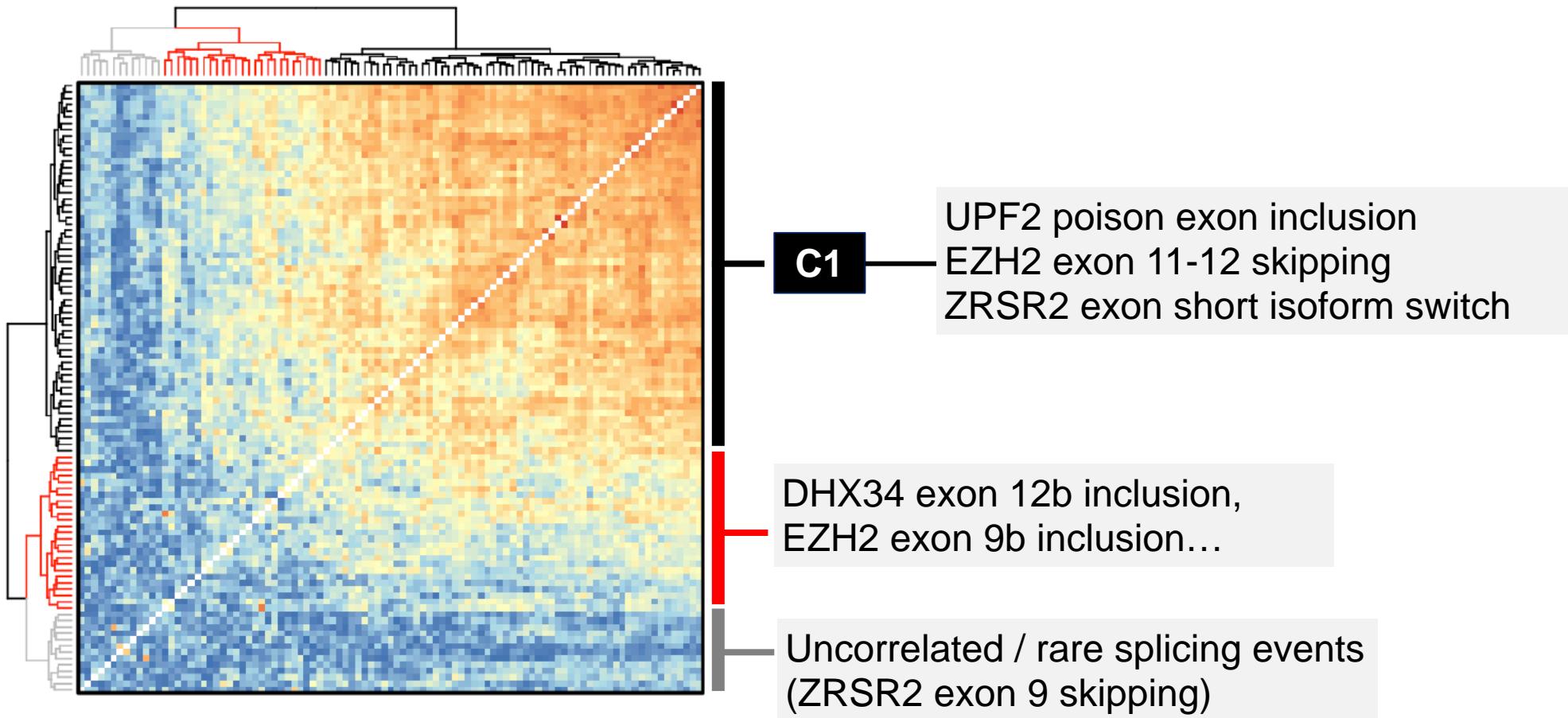
...and I ran MAJIQ again!



# Splicing events across multiple NMD factors indicate potentially broad dysregulation of mRNA degradation

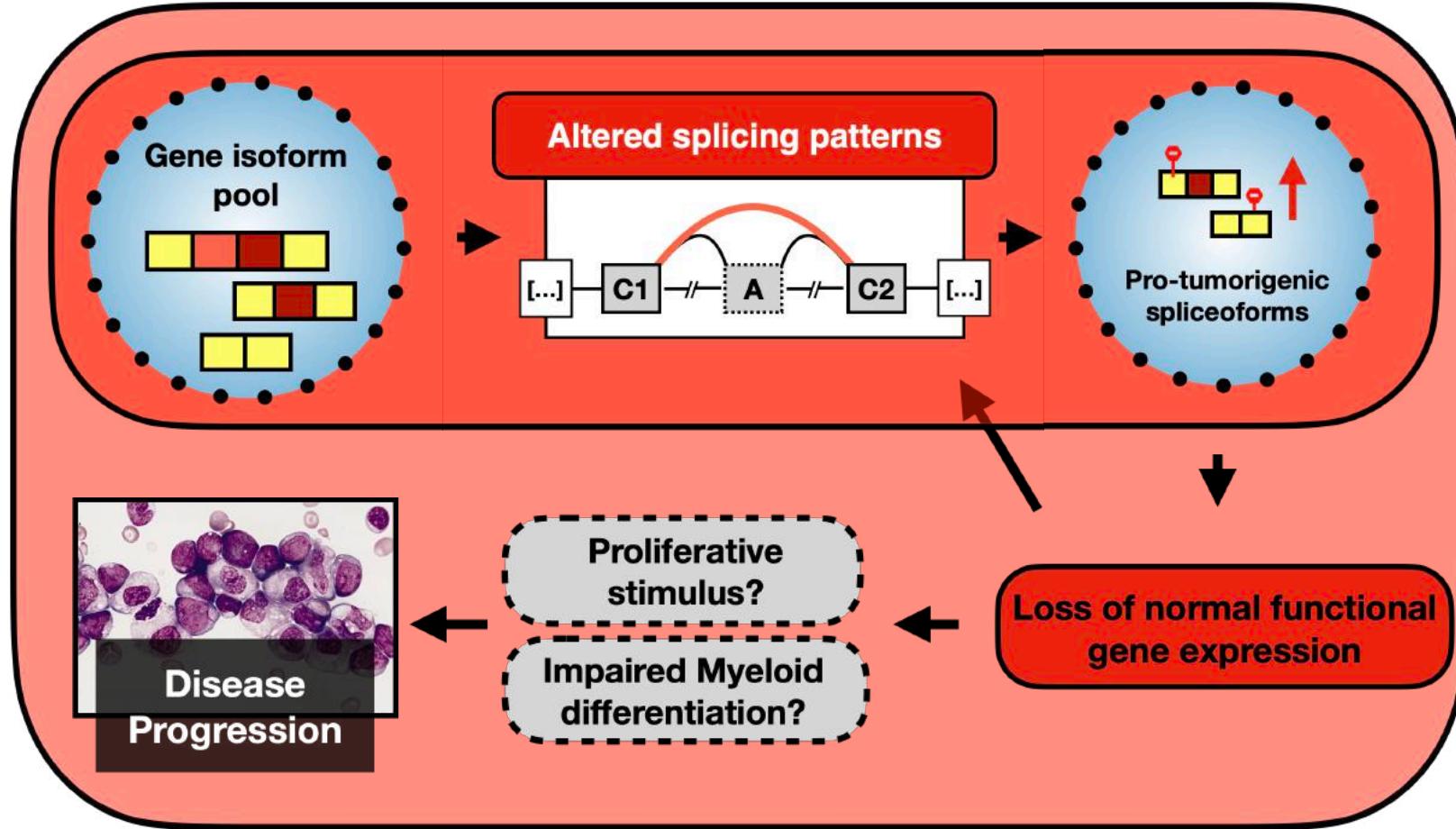


# Clustering reveals patterns of co-occurrence between splicing modules in AML-associated genes and NDM factors



# Conclusion and Closing Remarks

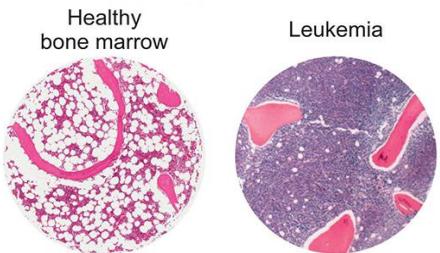
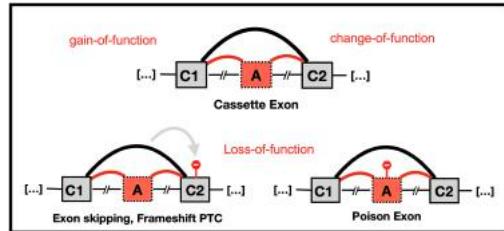
Alternative mRNA splicing is a highly plastic cellular process that can be dynamically leveraged by the cancer cell to dysregulate a wider set of genes than those altered by genetic mutations alone.



# Conclusions

## Future Directions

### How are coordinated splicing patterns manifested?



- Coordinated dysregulation of alternative splicing produces significant molecular variations that is widely overlooked by most studies of cancer.
- Alternative mRNA splicing redefines the landscape of commonly dysregulated genes in AML. **What is the functional consequence of most of these splicing variations?**
- Short-read sequencing data originating from the RNA-seq methods used in this study is rich in information that can be leveraged to decode multiple layers of biological regulation.

**What is the penetrance of the perturbations to the splicing-NMD regulatory axis across distinct cancers?**



# Acknowledgements



## Barash Lab (BioCiphers)

<https://www.biociphers.org/>

- Yoseph Barash
- Jordi Vaquero
- Paul Jewell
- Mathew Gazzara
- Anupama Jha
- Caleb Radens
- David Lee
- Joseph Aicher
- Mathieu Quesnel-Vallieres
- Kevin Yang
- David Wang



## “Barynch” Folk

## Lynch Lab

<http://www.kwlynchlab.org/>

- Kristen W. Lynch
- Michael Mallory
- Laura Agosto
- Krystal Haislop
- Rakesh Chatrikhi
- Davia Blake
- Max Ferretti



Perelman  
School of Medicine  
UNIVERSITY OF PENNSYLVANIA



Institute for  
Biomedical  
Informatics



## Thesis Committee

- Martin Carroll (Chair)
- Robert Babak Faryabi
- Peter Choi
- Stephen Liehaber

## Collaborators

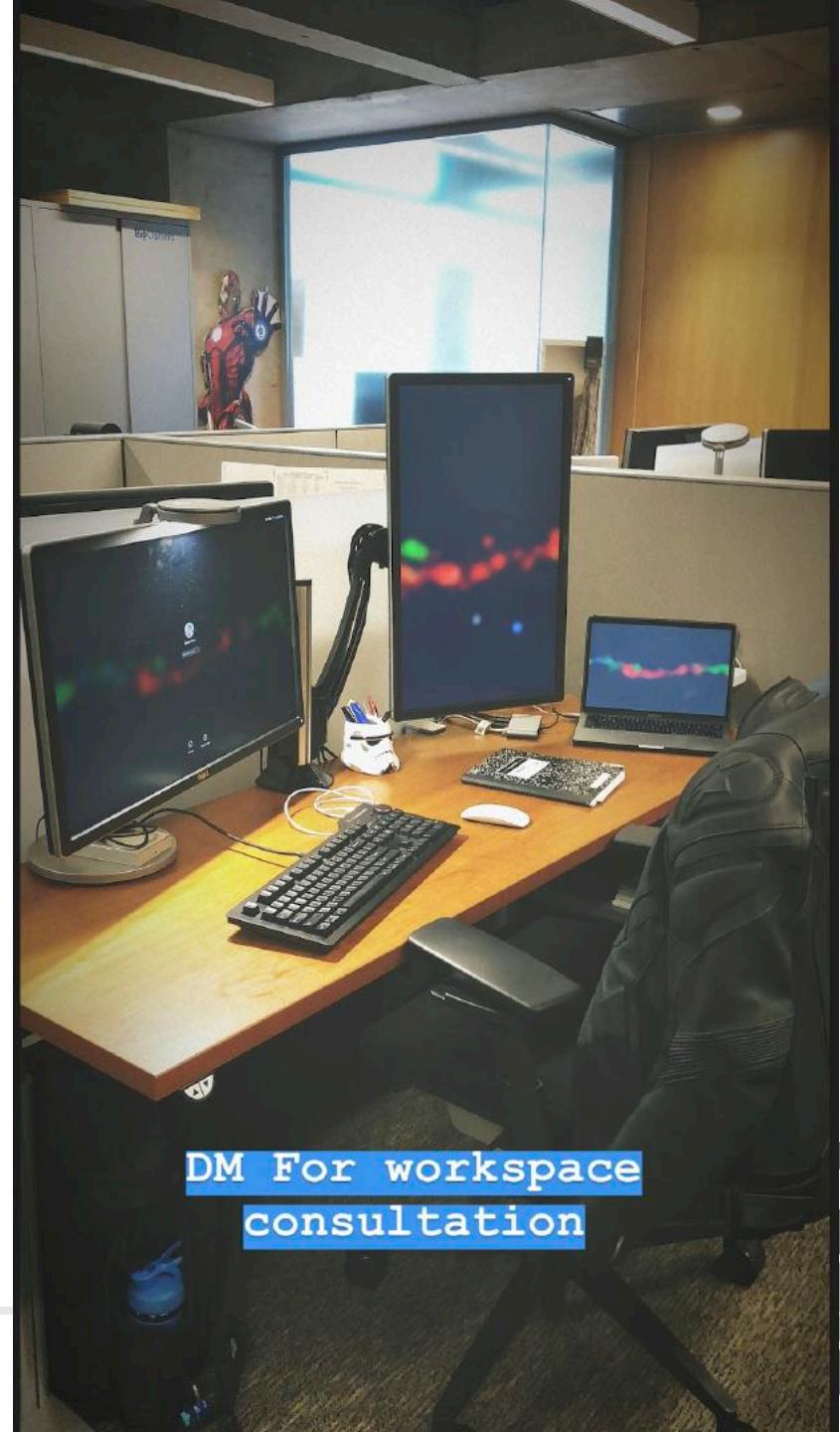
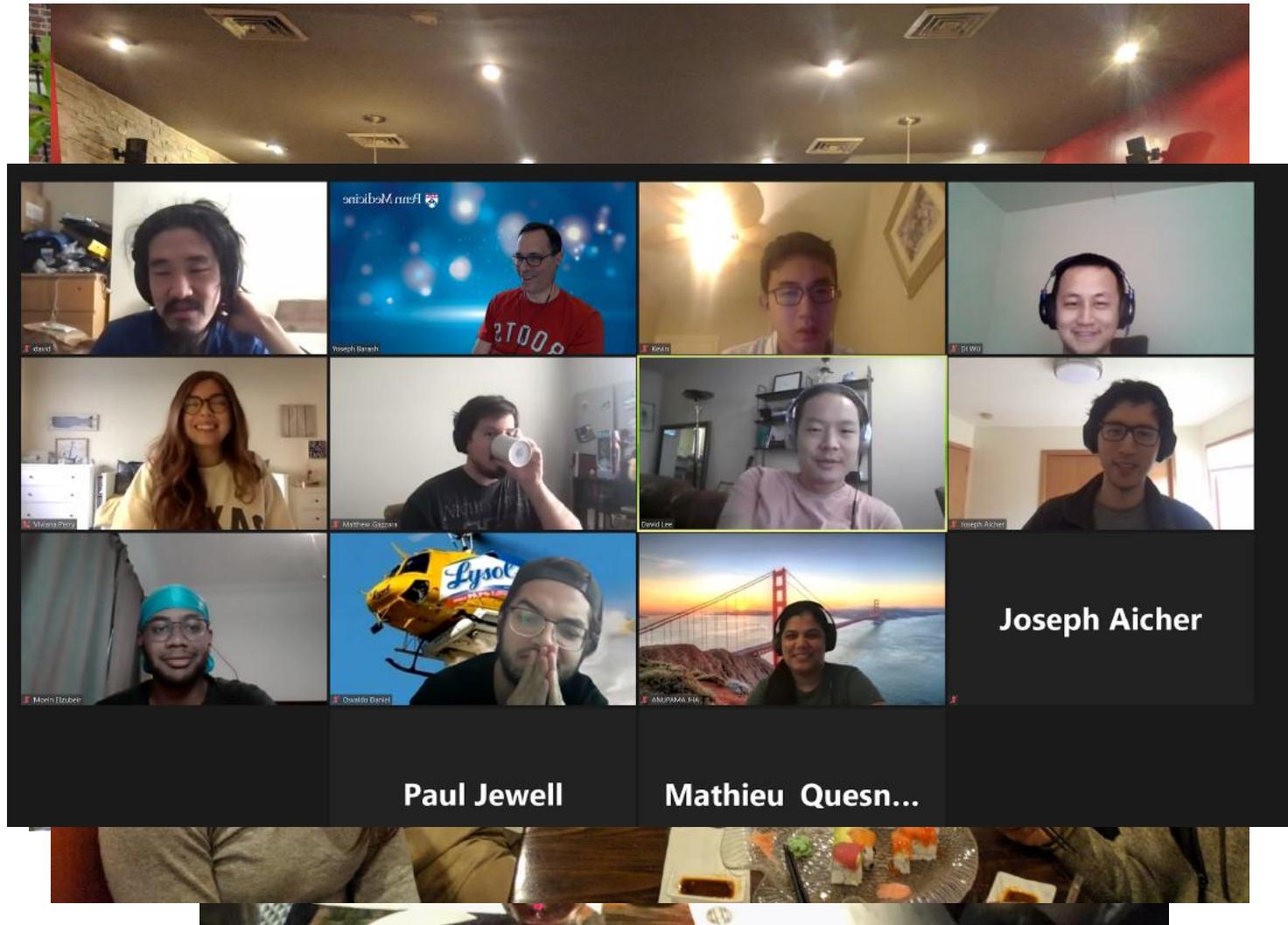
- Sara Cherry
- Martin Carroll

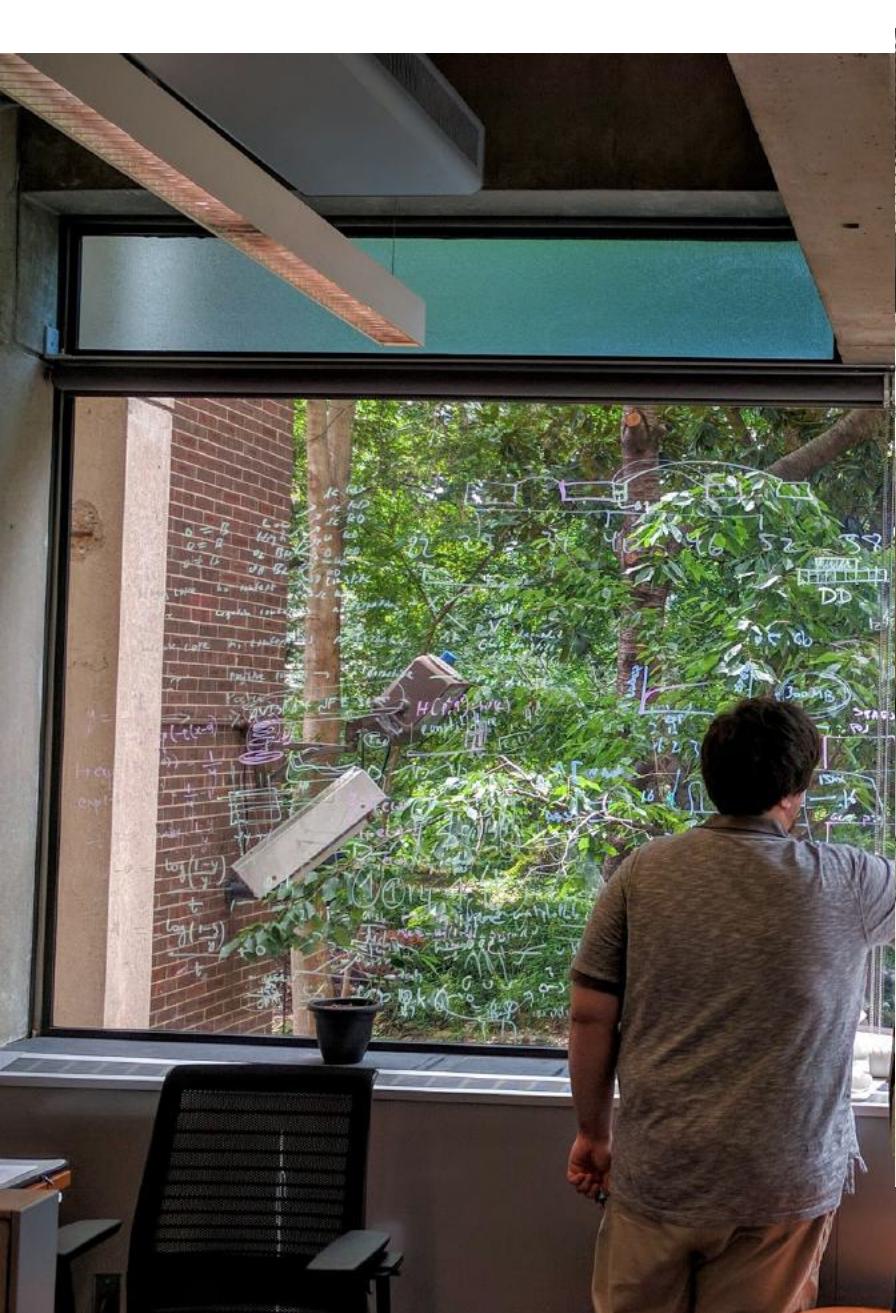


## Funding

NSF GRFP : Life Sciences Bioinformatics and Computational Biology

# The “Barynch Experience”





# The “Barynch Experience”





Molecular Biology Un



Talking about Sotillo et al. MAJIQ in  
CART-CD19 paper in senior year  
presentation



Turned into a bioinformatician  
#GDM #everyday



# Acknowledgements



## Barash Lab (BioCiphers)

<https://www.biociphers.org/>

- Yoseph Barash
- Jordi Vaquero
- Paul Jewell
- Mathew Gazzara
- Anupama Jha
- Caleb Radens
- David Lee
- Joseph Aicher
- Mathieu Quesnel-Vallieres
- Kevin Yang
- David Wang



## Lynch Lab

<http://www.kwlynchlab.org/>

- Kristen W. Lynch
- Michael Mallory
- Laura Agosto
- Krystal Haislop
- Caleb Radens
- Rakesh Chatrikhi
- Davia Blake
- Max Ferretti

## Thesis Committee

- Martin Carroll (Chair)
- Robert Babak Faryabi
- Peter Choi
- Stephen Liehaber

## Collaborators

- Sara Cherry
- Martin Carroll

## Funding

NSF GRFP : Life Sciences Bioinformatics and Computational Biology

Thank You for tuning in my talk!

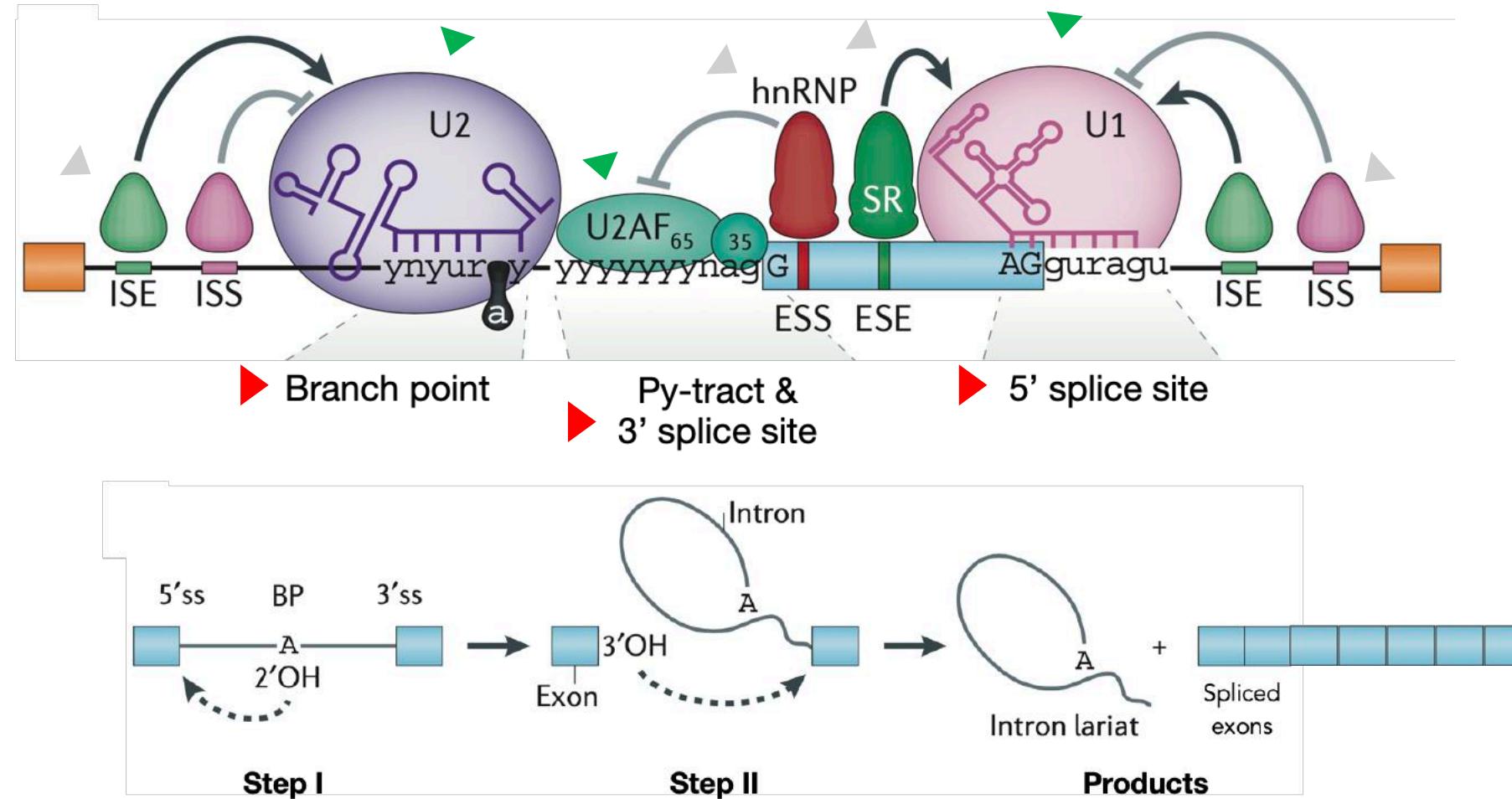
Send inquiries to: [orivera@pennmedicine.upenn.edu](mailto:orivera@pennmedicine.upenn.edu)



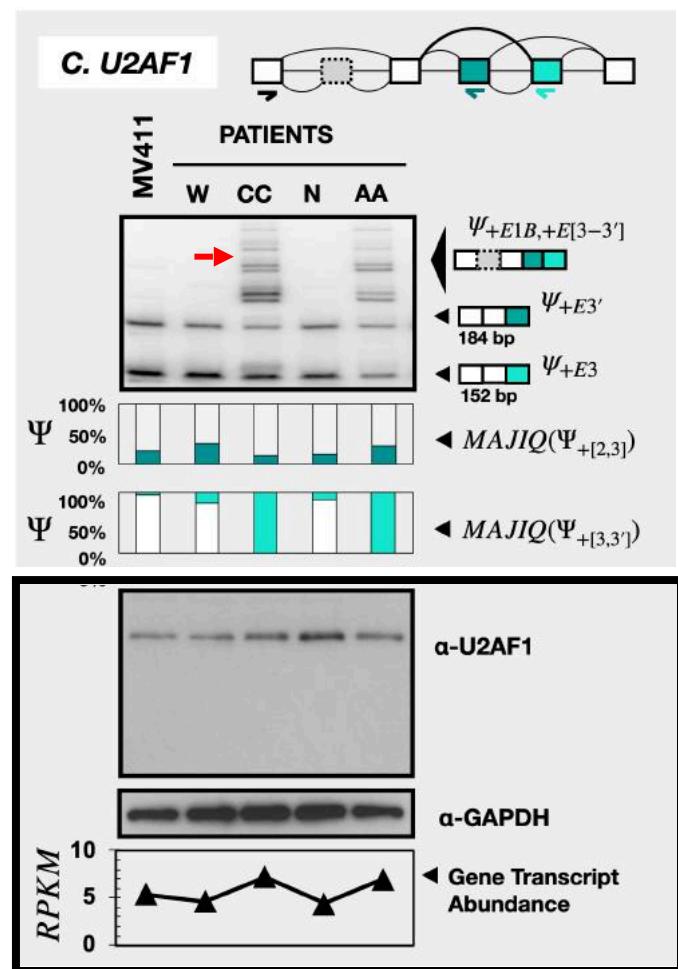
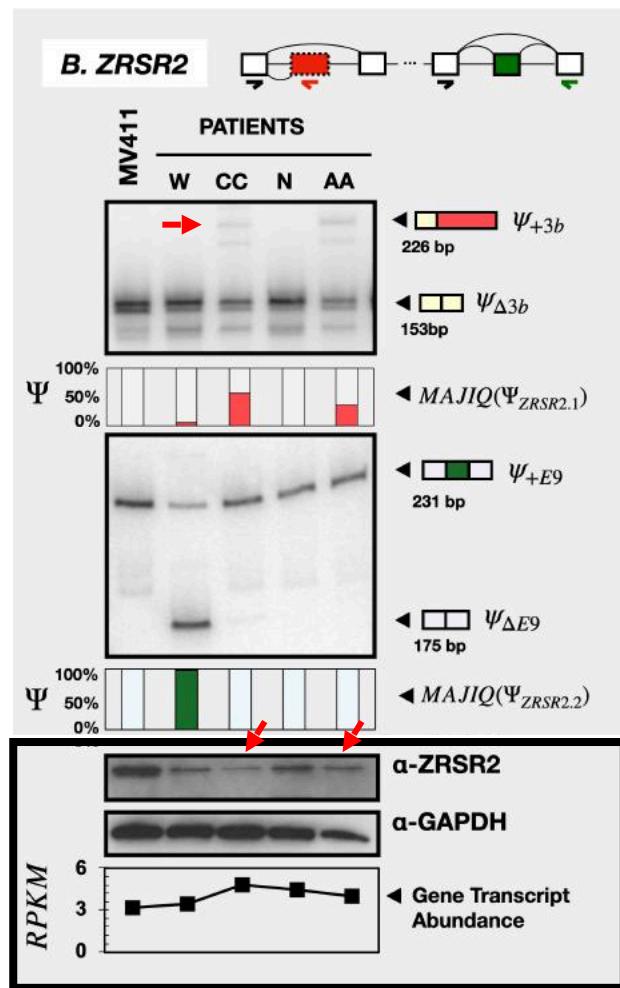
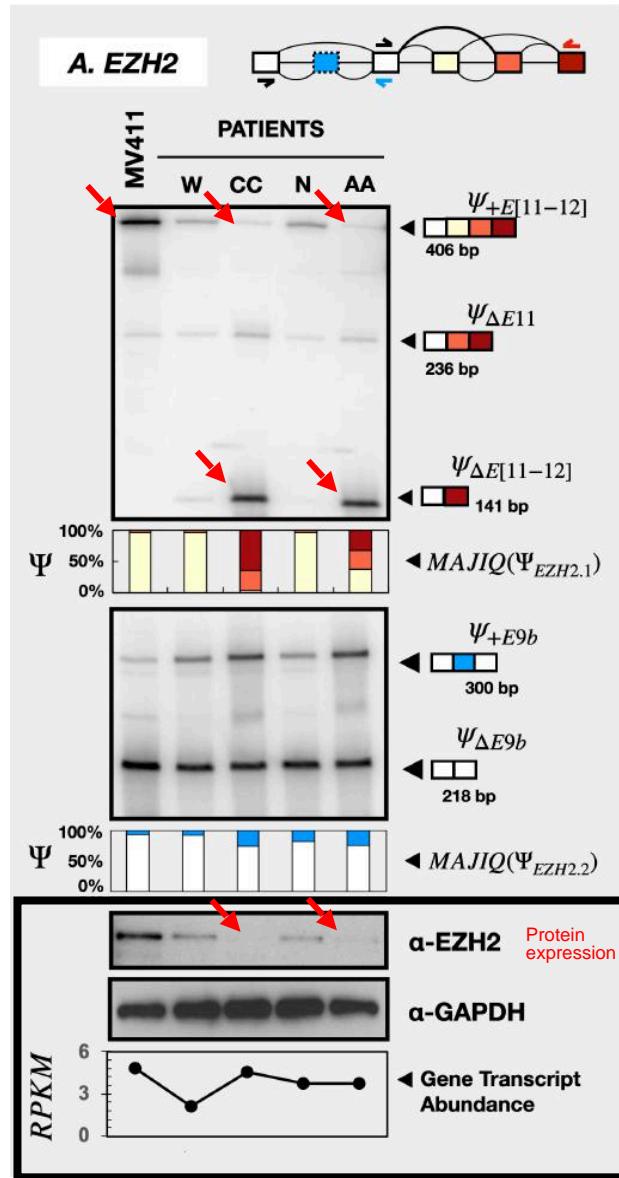
# Thank you!



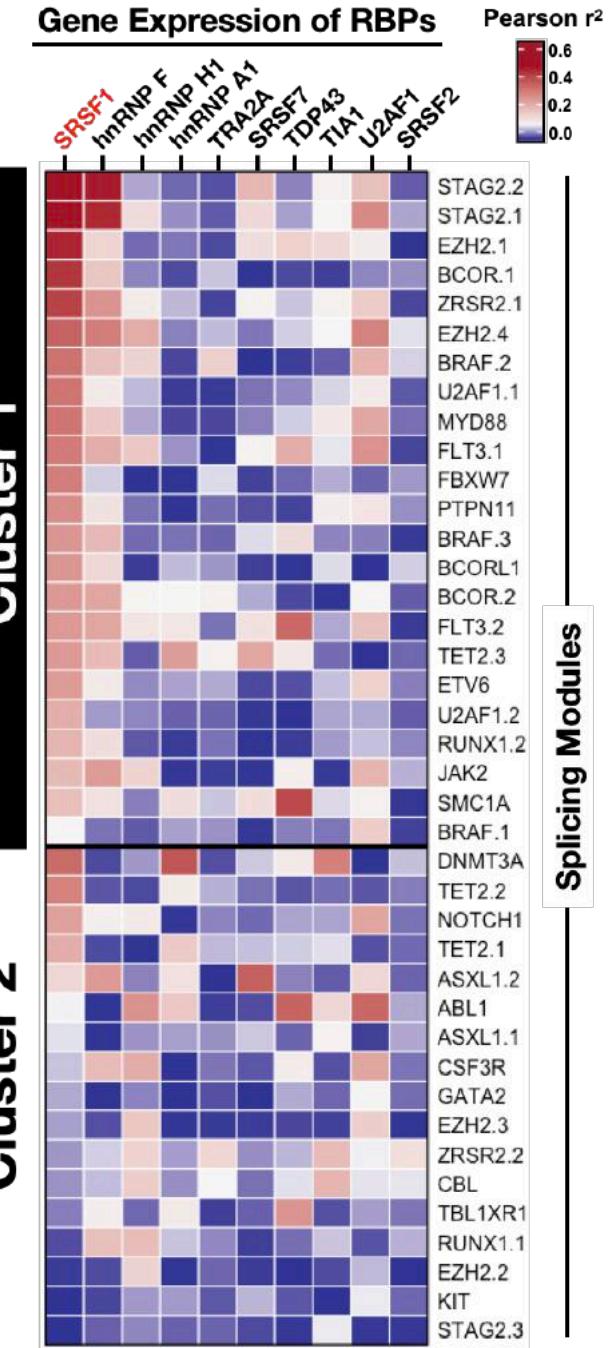
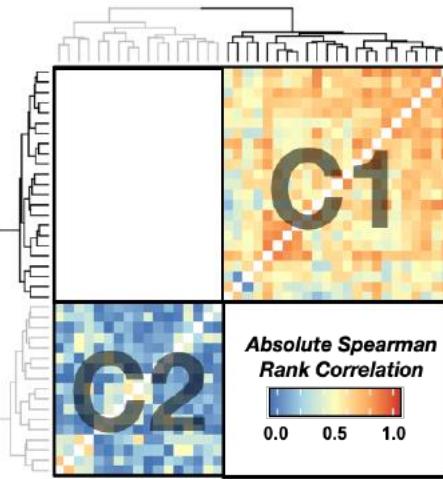
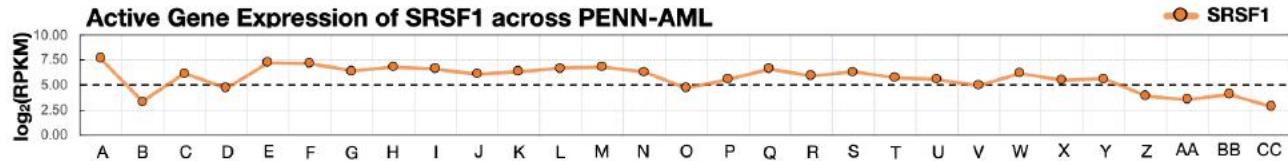
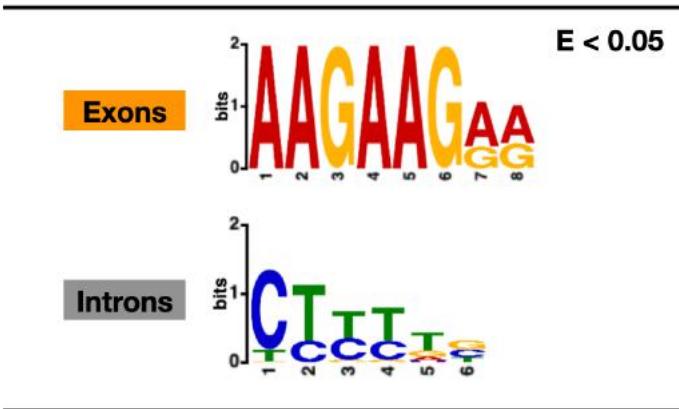
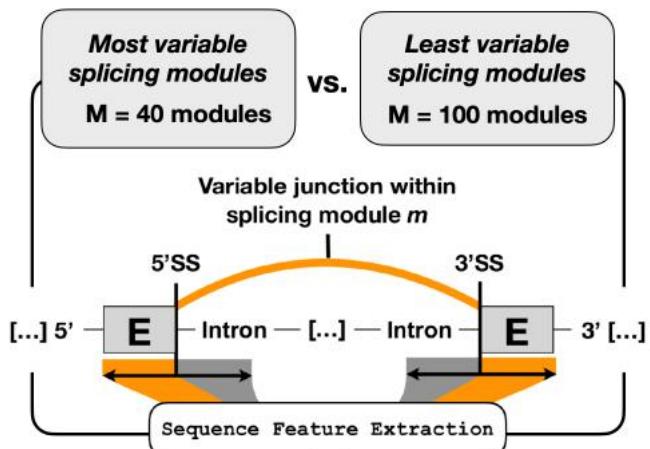
# The mRNA Splicing Reaction is coordinated by RNA-binding proteins that bind to “splicing signals” in the mRNA sequence



## Splicing variations in splicing factors themselves



# Gene expression of SRSF1 strongly correlates the cluster of co-regulated splicing modules.



# Sample processing obscures cancer-specific alterations in leukemic transcriptomes

Heidi Dvinge<sup>a,b</sup>, Rhonda E. Ries<sup>c</sup>, Janine O. Ilagan<sup>a,b</sup>, Derek L. Stirewalt<sup>c</sup>, Soheil Meshinchi<sup>c,d</sup>, and Robert K. Bradley<sup>a,b,1</sup>

<sup>a</sup>Computational Biology Program, Public Health Sciences Division, <sup>b</sup>Basic Sciences Division, and <sup>c</sup>Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; and <sup>d</sup>Division of Pediatric Hematology/Oncology, School of Medicine, University of Washington, Seattle, WA 98195

Edited\* by Robert N. Eisenman, Fred Hutchinson Cancer Research Center, Seattle, WA, and approved October 14, 2014 (received for review July 14, 2014)

No correlation of DHX34 poison exon inclusion and sample storage method.

geneName	PBMC.none.4h	PBMC.none.8h	PBMC.none.24h	PBMC.none.48h
DHX34	0.15	1.65	0.22	-0.15
WHAMM	0.52	0.65	2.4	3.8
METTL22	0.09	0.52	0.18	0.17
CHD2	0.38	0.87	1.25	2.8
PARP6	0.39	0.48	0.97	1.5

geneName	PBMC.cryo.24h	PBMC.cryo.48h	PBMC.ice.24h	PBMC.ice.48h
DHX34	-0.55	-0.62	0.69	0.44
WHAMM	3.2	3.95	-0.02	0.03
METTL22	-0.1	0.22	0.34	0.41
CHD2	1.9	2.95	0.05	-0.03
PARP6	0.82	1.2	0.1	0.13

