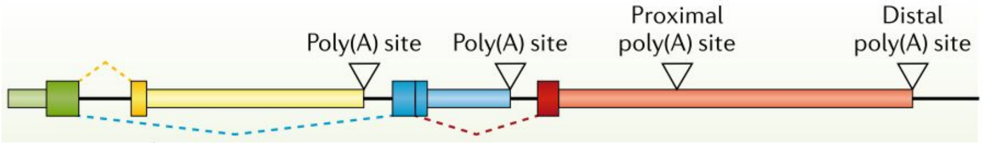
Transcriptome sequencing suggests that pre-mRNA splicing counteracts premature intronic polyadenylation



Maria Vlasenok
D. Pervouchine Lab
Skoltech

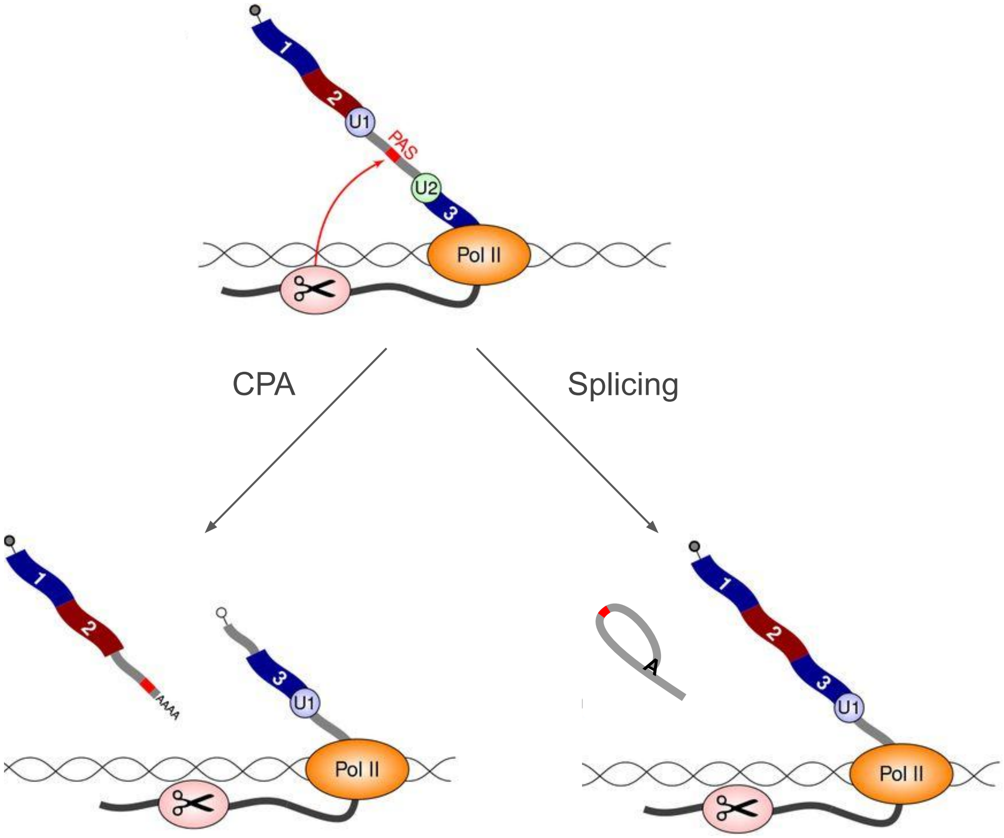
Benasque
7-21 Aug 2022

Figure: Gruber, A. J. & Zavolan, M. Nat. Rev. Genet. (2019)

Background

- Nearly all transcripts generated by Pol II are cleaved and polyadenylated (CPA).
- >50% of human genes undergo alternative polyadenylation (APA),
    - >20% - intronic polyadenylation (IPA).
- CPA machinery may operate in all introns[1].
- Splicing and polyadenylation compete or co-occur at intronic polyadenylation sites (PAS) ?

Goal: Study the interplay between splicing and intronic polyadenylation.



---

[1] Yao, C. et al. Proc. Natl Acad. Sci. USA (2012).

Background

- Nearly all transcripts generated by Pol II are cleaved and polyadenylated (CPA).
- >50% of human genes undergo alternative polyadenylation (APA),
    - >20% - intronic polyadenylation (IPA).
- CPA machinery may operate in all introns[1].
- Splicing and polyadenylation compete or co-occur at intronic polyadenylation sites (PAS) ?

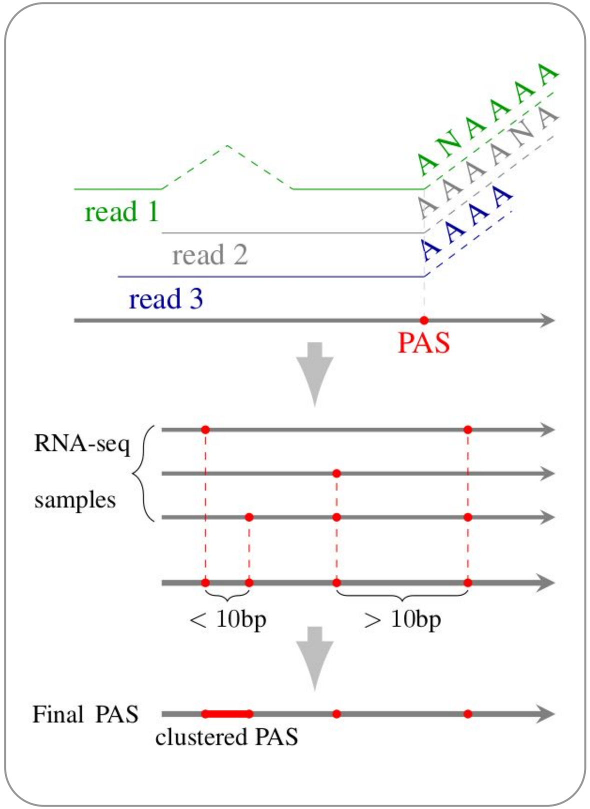
Goal: Study the interplay between splicing and intronic polyadenylation.

How?

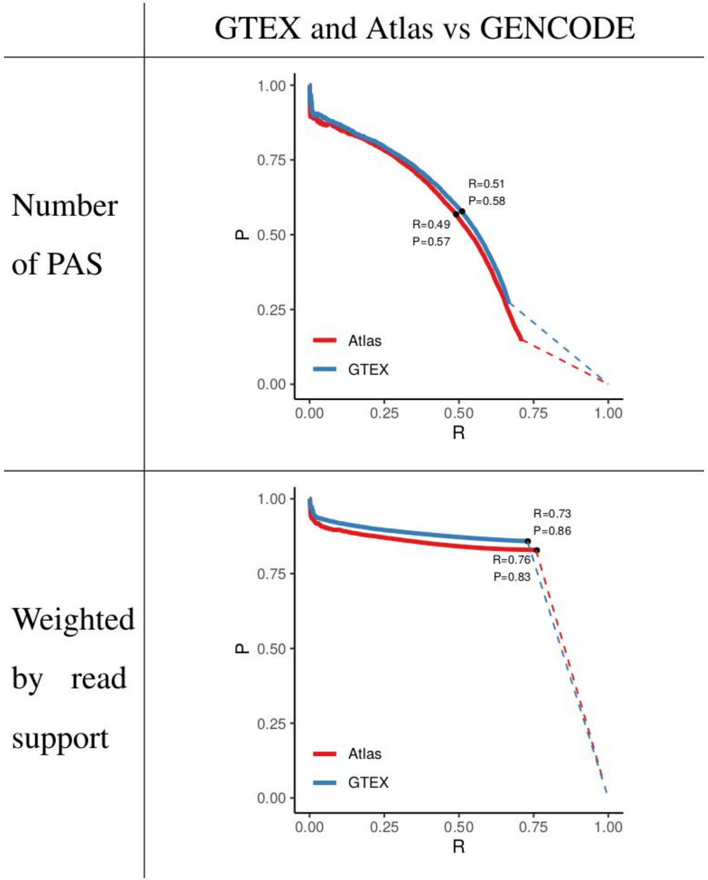
- We need data to simultaneously assess alternative splicing (AS) and IPA.
- Matched datasets are limited
    - RNA-seq for splicing,
    - 3’-seq (3’RACE) for PA.
- Use poly(A)+ RNA-seq to study polyadenylation.
- No methods to identify IPA sites from RNA-seq*.

*a bit later came mountainClimber. Cass, A. A. & Xiao, X. Cell Syst. (2020)

PAS can be identified *de novo* from large scale RNA-seq data

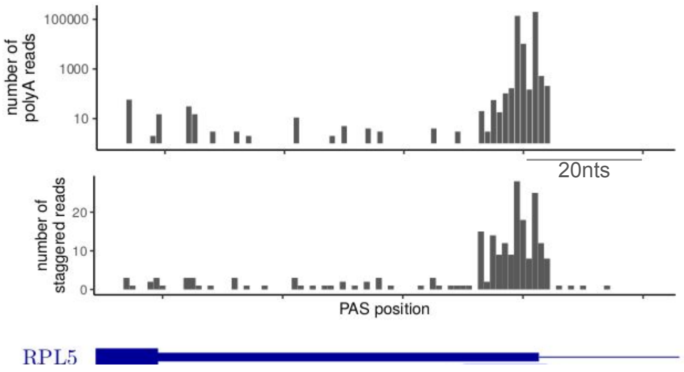


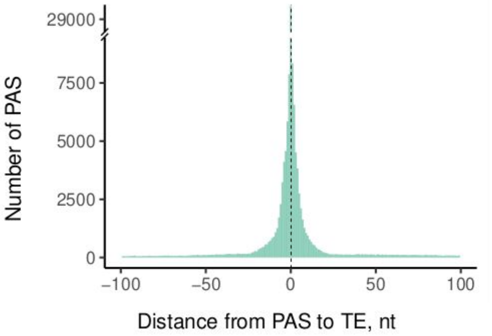
ContextMap 2.0 Bonfert, T. & Friedel, C. C. PLoS One (2017).
BIROL, I. et al. KLEAT: Biocomputing 2015

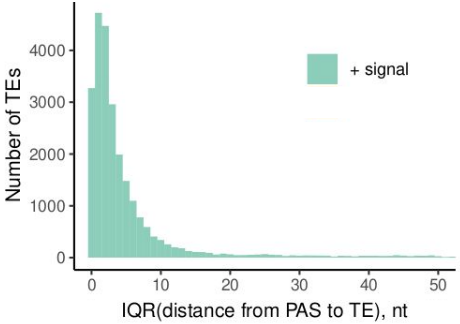


PolyASite 2.0. Herrmann, C. J. *et al.* Nucleic Acids Res.(2020)

PASs cluster around transcript ends (TEs)



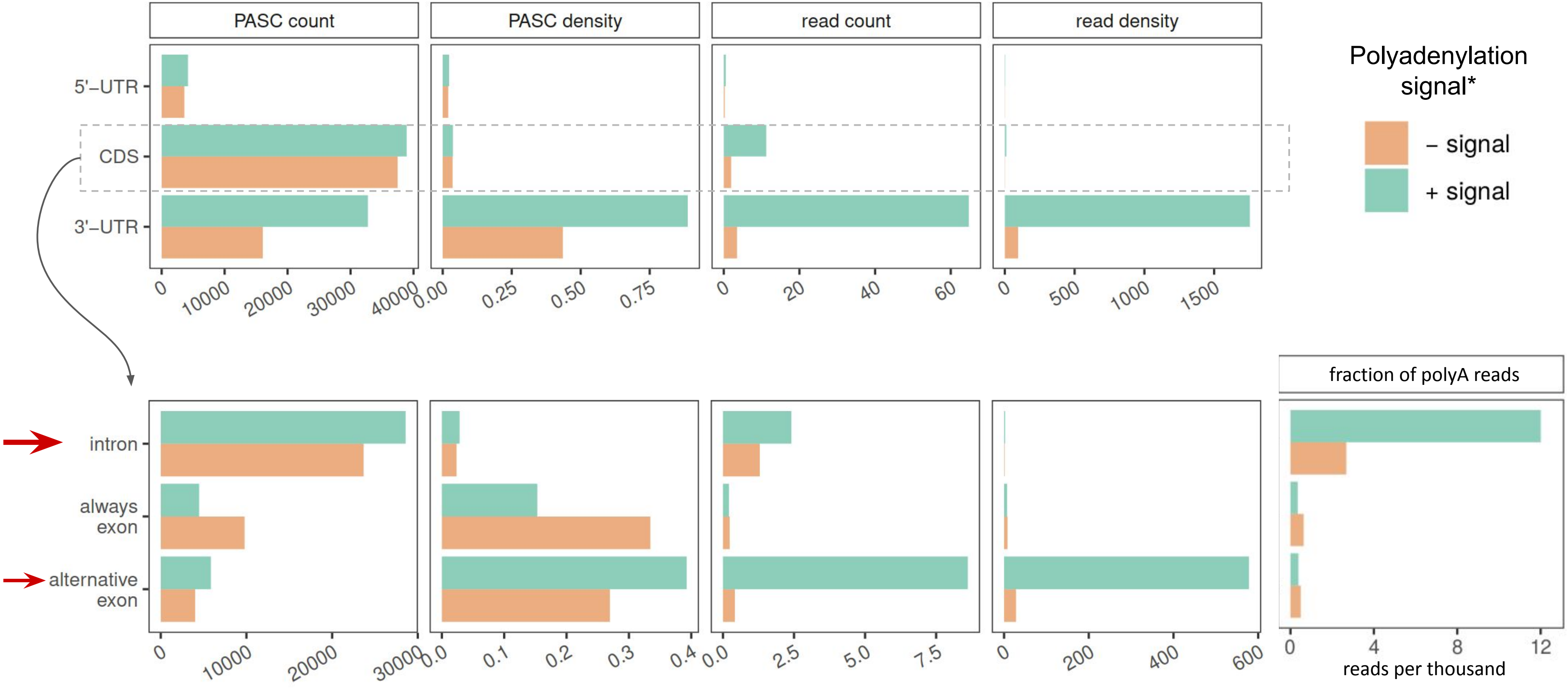




- The method shows fine structure of PAS distribution around TEs
- CPA machinery is not precise*
  - most PAS are located within 10nt of the TE

* Tian, B., Hu, J., Zhang, H. & Lutz, C. S. *Nucleic Acids Res.* (2005).
Derti, A. et al. Genome Res. (2012).

IPA is widespread

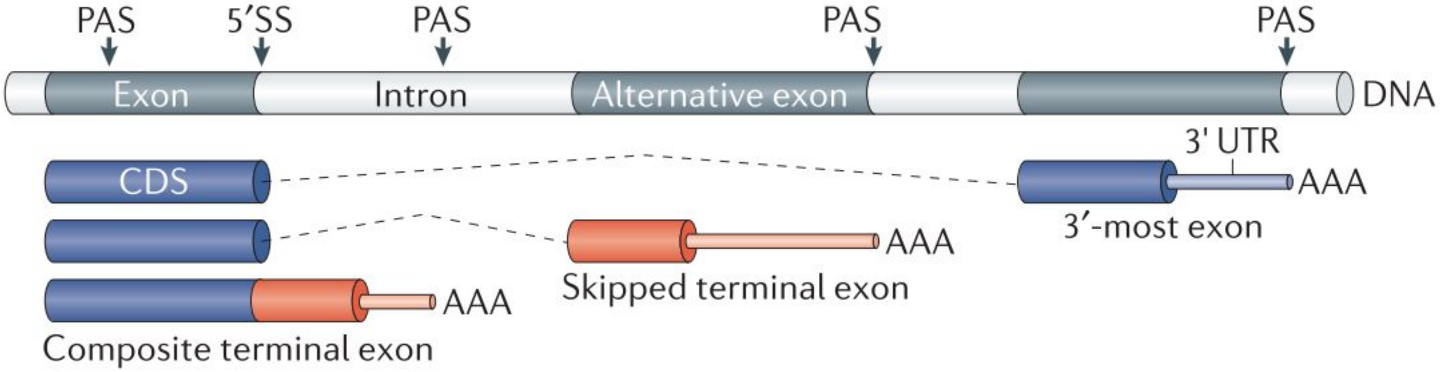


* 1 mismatch from AAUAAA

IPA leads to the expression of alternative terminal exons.

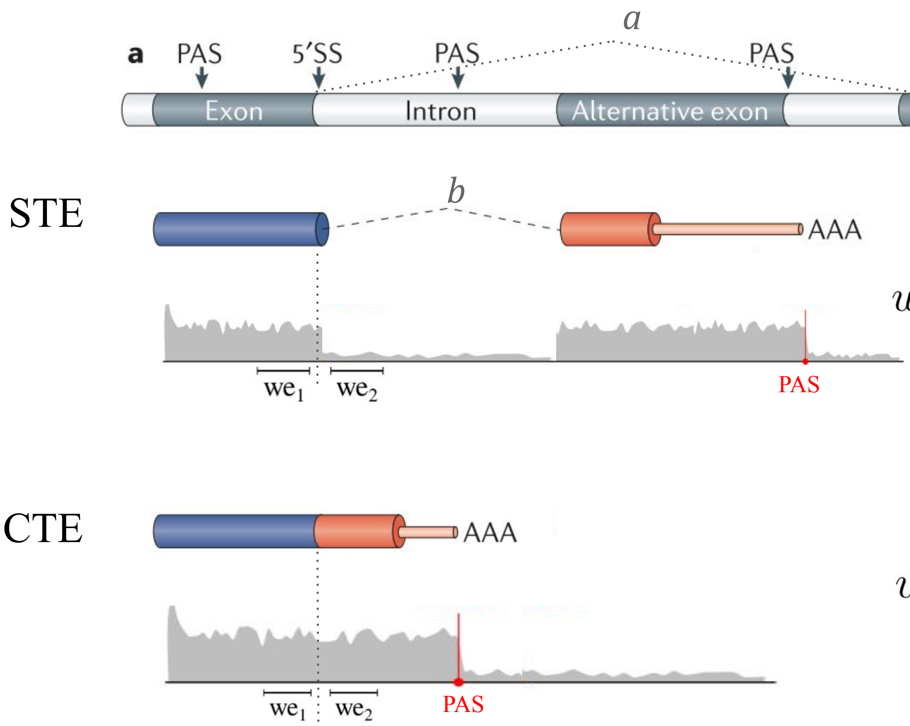
By Bin Tian the resulting alternative terminal exons (TEs) can be

- skipped TE - alternative internal exons selected through splicing to be the TE (cryptic PAS in cassette exon),
- composite TE - extension of an internal exon through inhibition of the 5´SS (cryptic PAS in retained intron)



Tian, B. & Manley, J. L. Nat. Rev. Mol. Cell Biol. 2016

Terminal exon type can be determined from coverage and split reads.



$$\psi = \frac{a}{a+b}$$

- fraction of split reads supporting canonical splicing.

STE

$$we_1 \gg we_2$$

$\psi < 1$ - unannotated alternative splicing event before the intronic PAS.

CTE

$$we_1 \simeq we_2$$

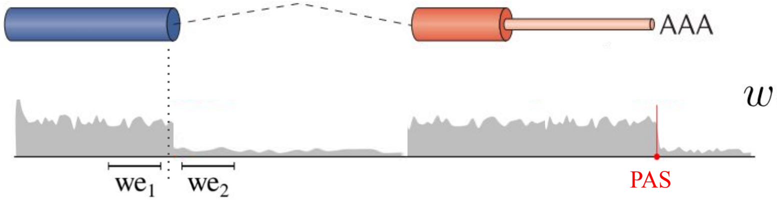
$\psi \sim 1$ - intron is spliced canonically

From the GTEx data we computed coverage fold change at PAS, we_1, we_2 and ψ^* for each intronic PAS (iPAS) and each tissue.

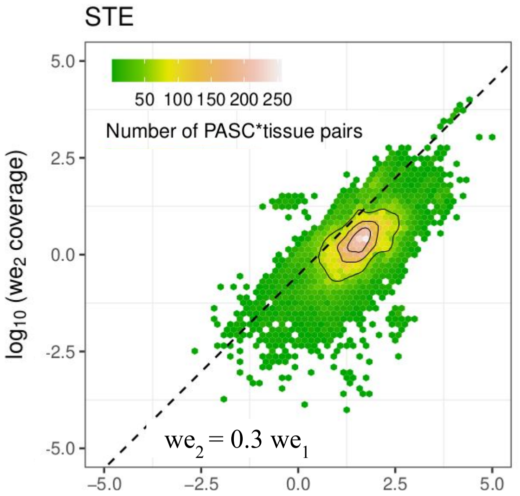
*IPSA pipeline. Pervouchine D.D, Knowles DG, Guigó R.Bioinformatics 2013

The two types of annotated TEs have distinct coverage and splicing profiles in GTEx.

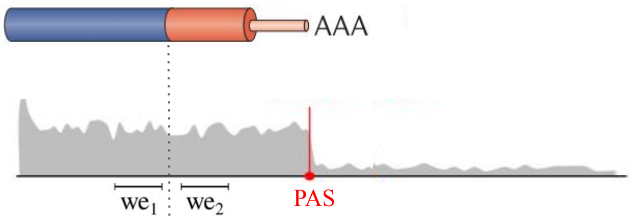
STE



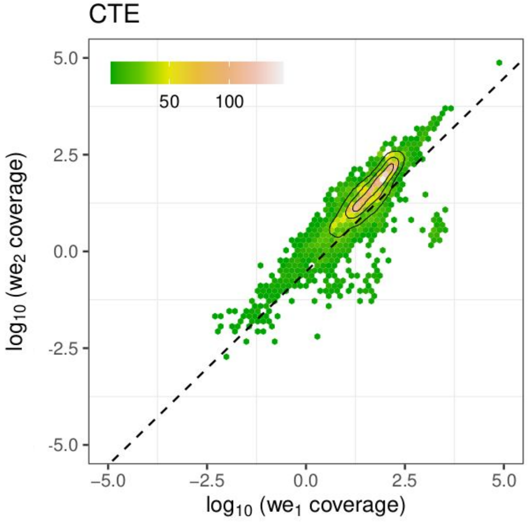
$we_1 \gg we_2$



CTE

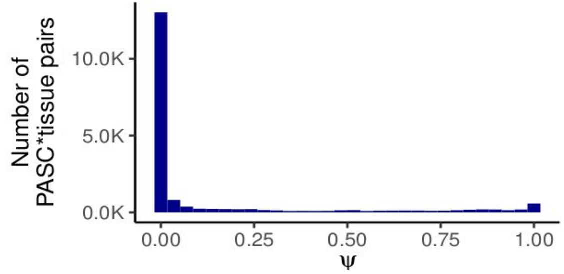


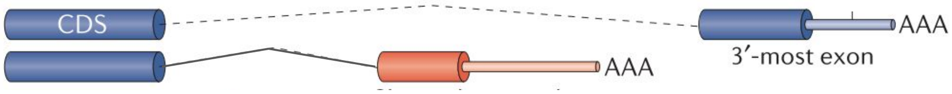
$we_1 \simeq we_2$



The two types of annotated TEs have distinct coverage and splicing profiles in GTEx.

STE

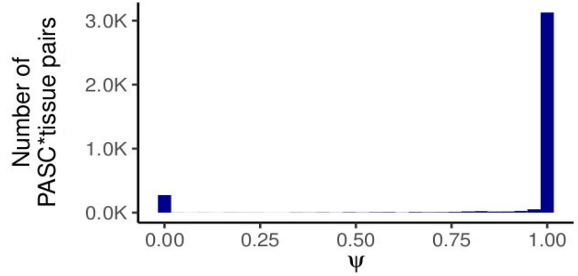




ψ < 1

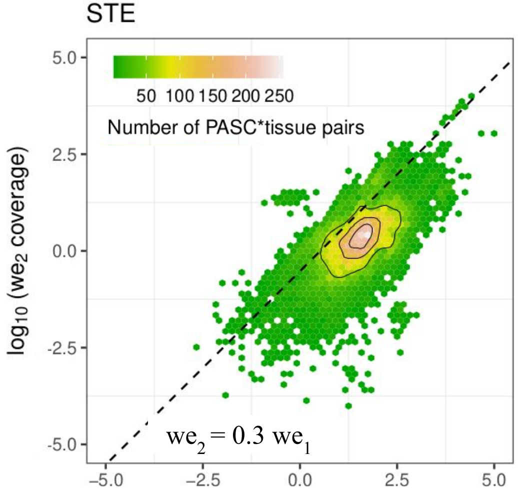
- unannotated AS event before the intronic PAS.

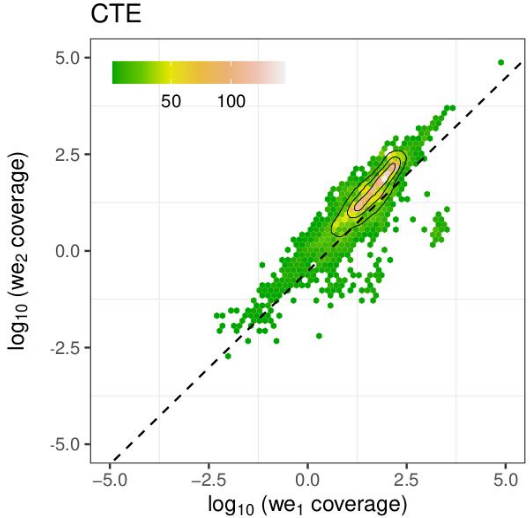
CTE



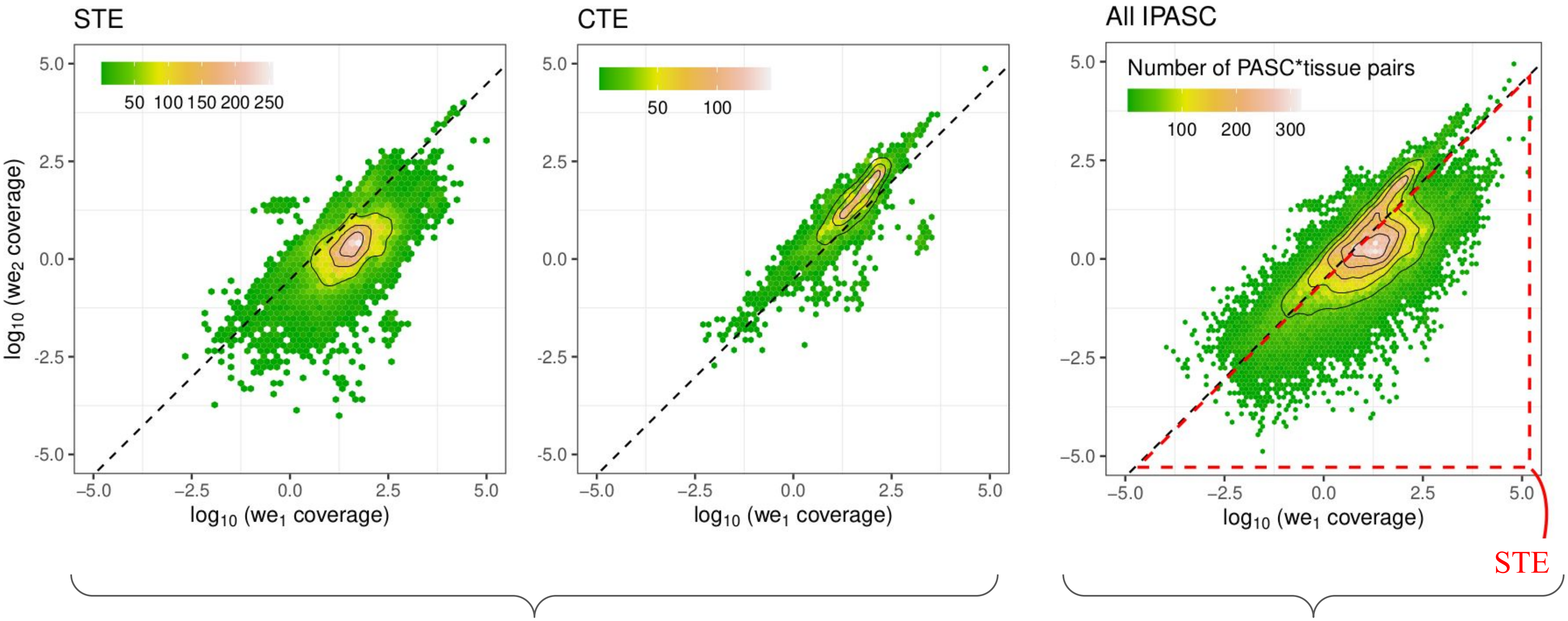
ψ ~ 1

- intron is spliced canonically.





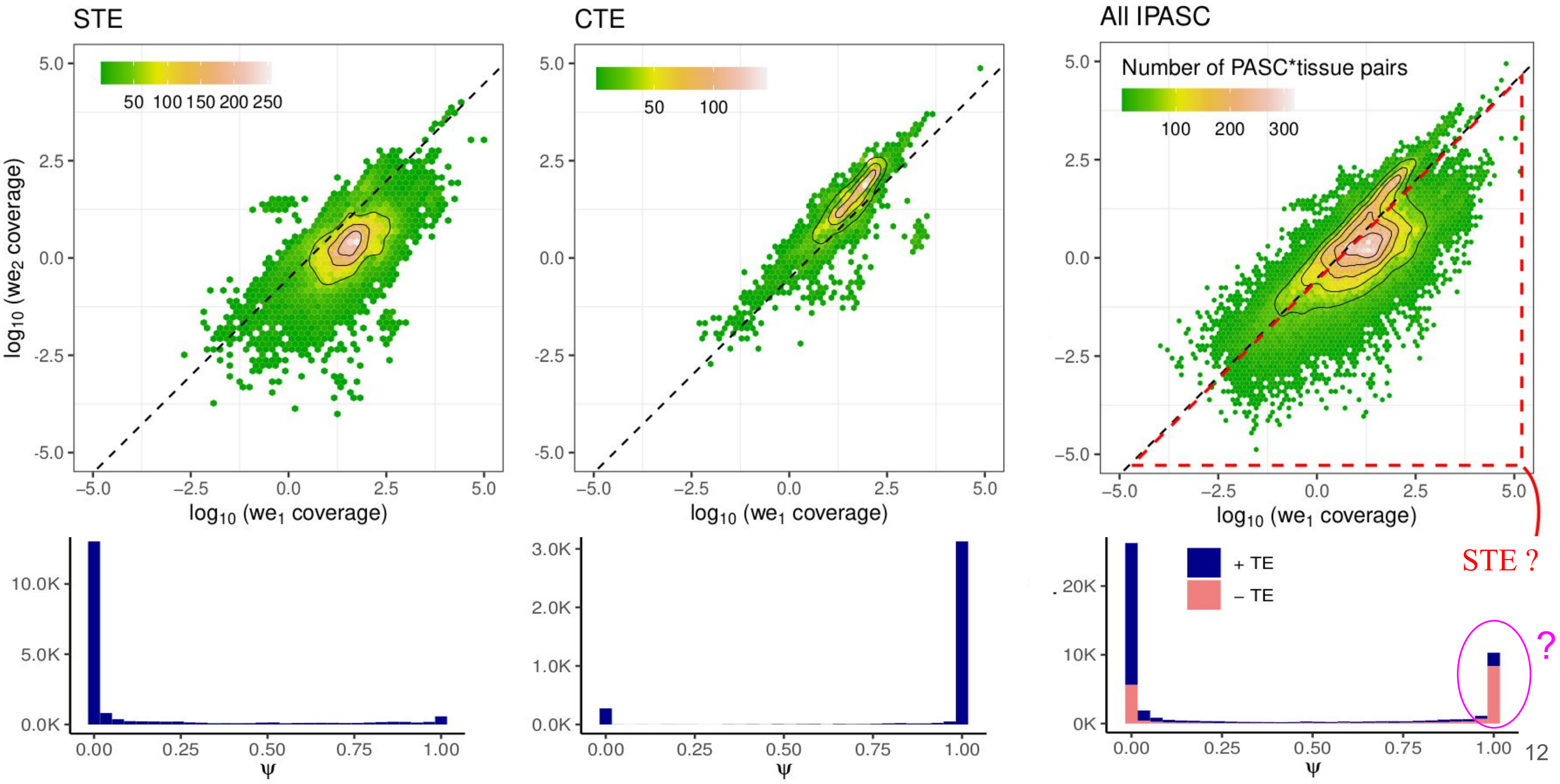
Terminal exon type can be determined from coverage and split reads.



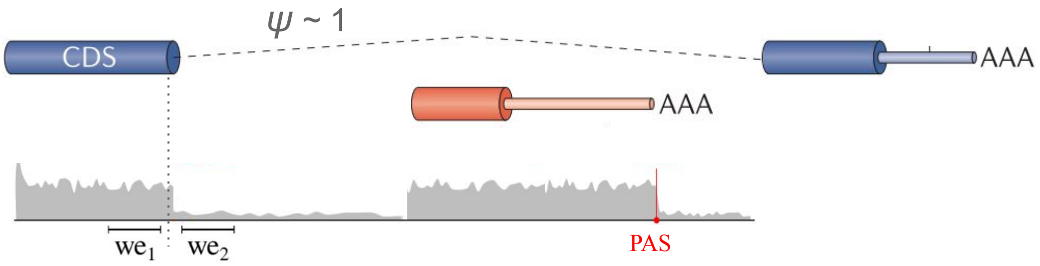
annotated

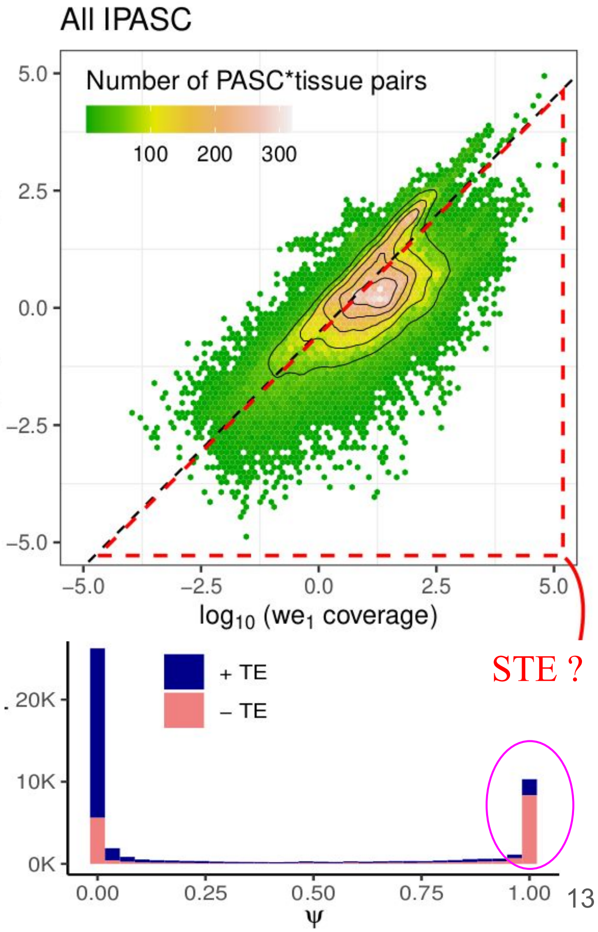
all

PSI distribution does not correspond to STE for a subset of STE-candidates.

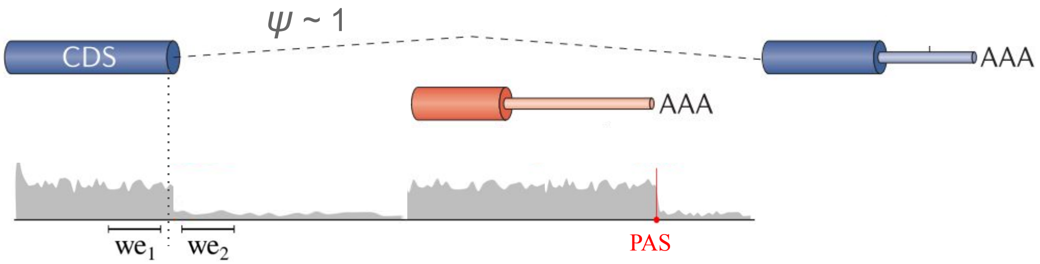


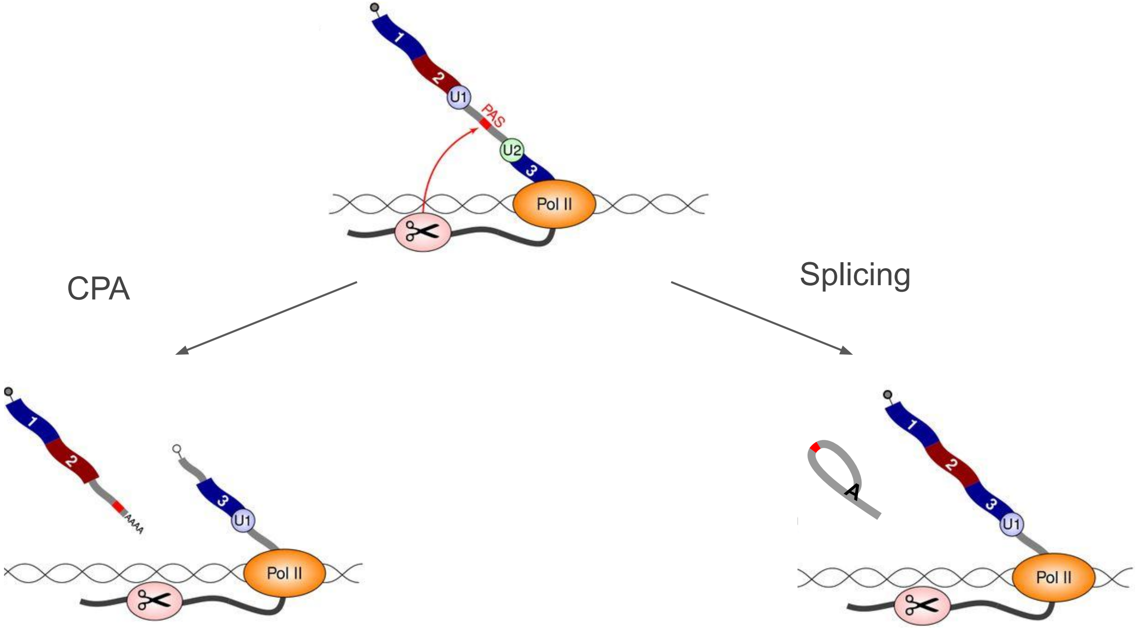
These RNAs are characterized by STE-like coverage and canonical splicing.





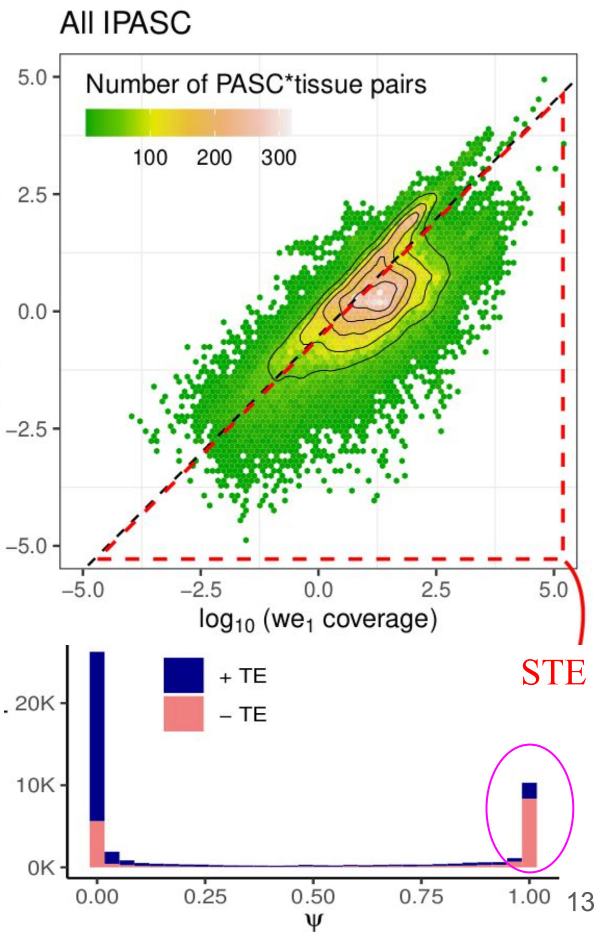
How could they arise?



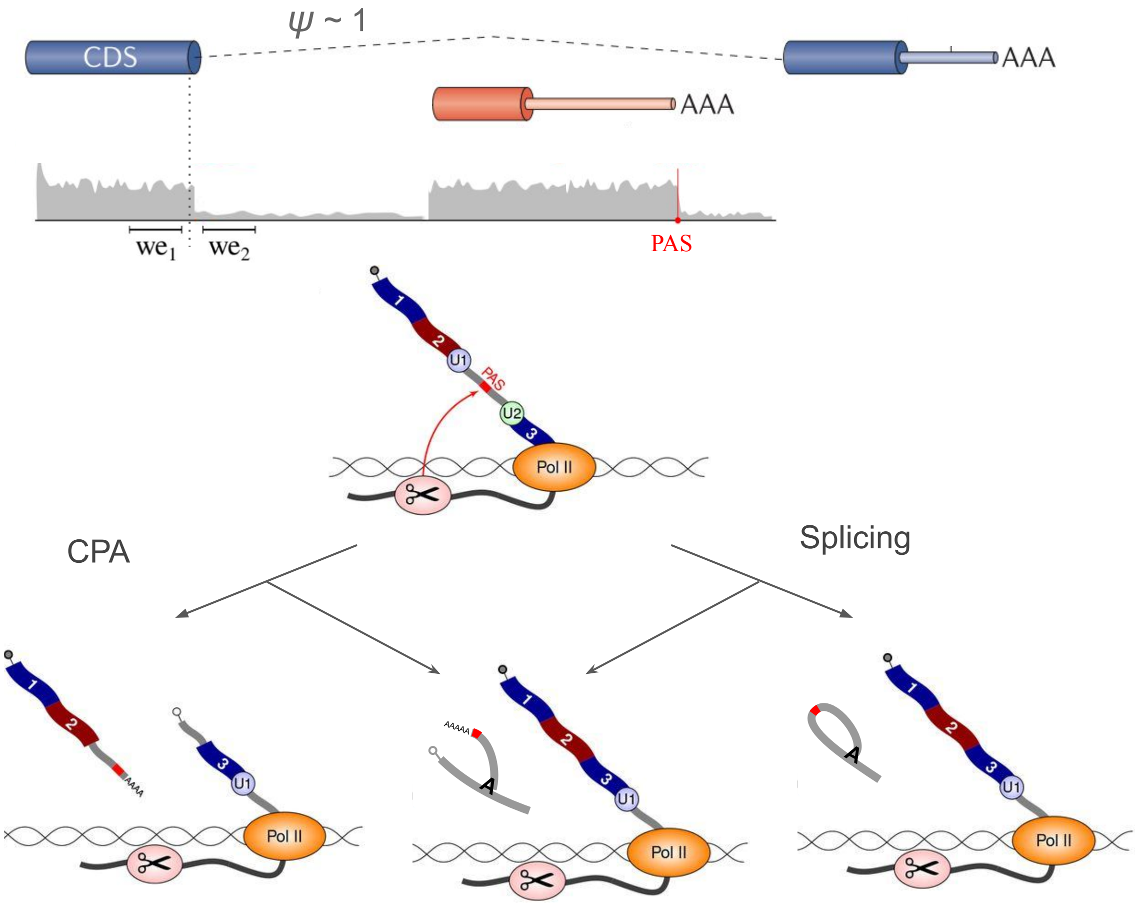


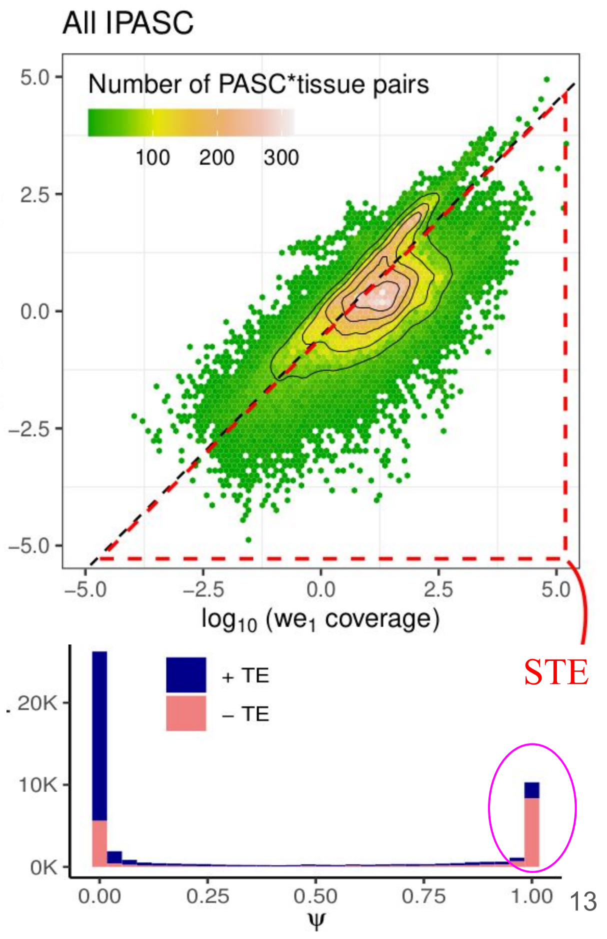
CPA

Splicing



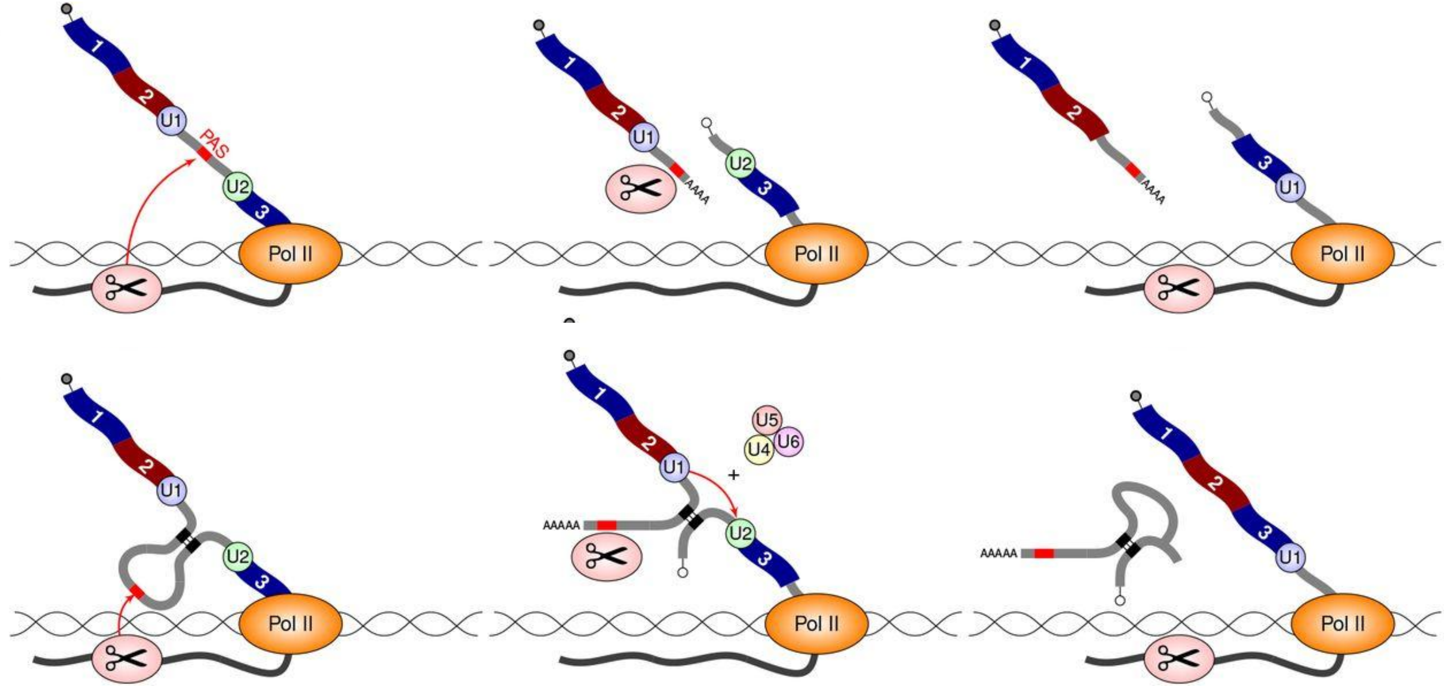
How could they arise? From interaction between splicing and PA.





Hypothesis: Pre-mRNA splicing counteracts premature intronic polyadenylation.

These interactions can result in cleaved and polyadenylated lariats detectable in large-scale RNA-seq datasets.



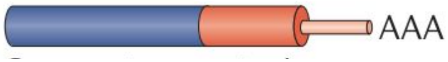
Polyadenylation “prevails”

Splicing “prevails”

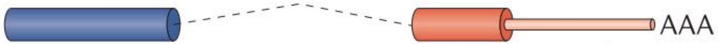
Figure: Kalmykova, S. et al. Nat. Commun. (2021).

Examples

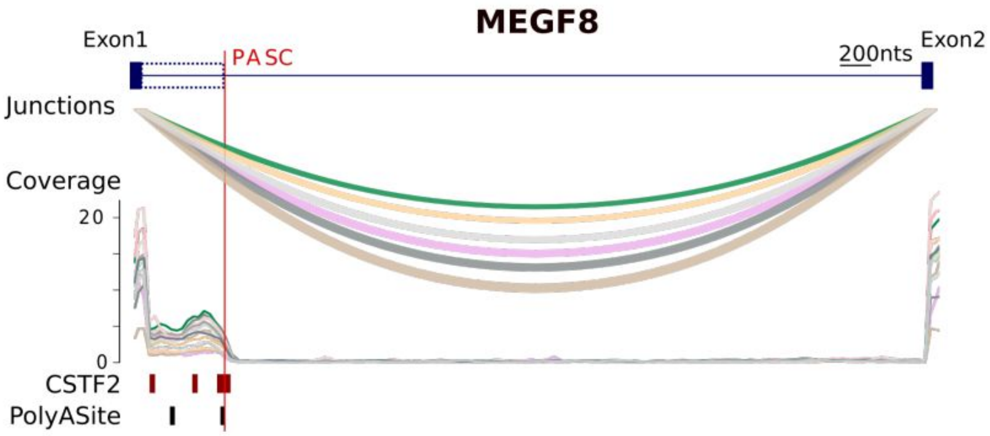
Composite terminal exon

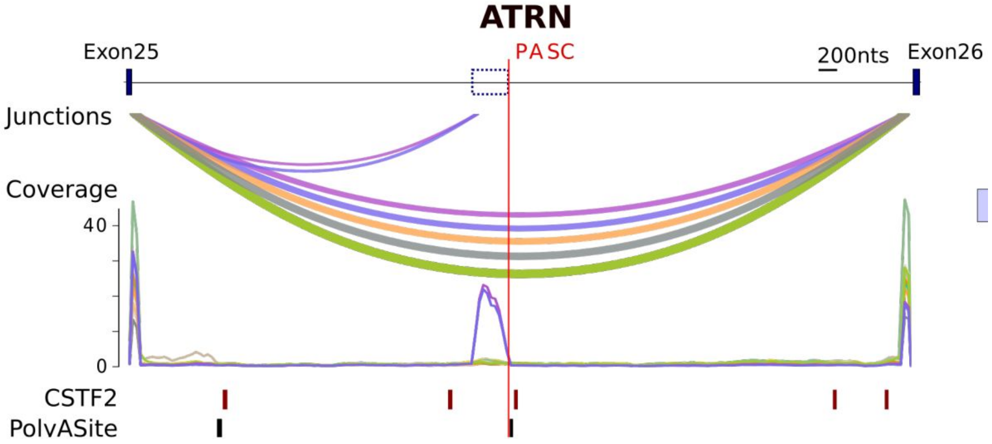


Skipped terminal exon



● Adipose
● Adrenal Gland
● Artery
● Bladder
● Lung
● Minor Salivary Gland
● Muscle
● Nerve
● Breast
● Cervix
● Colon
● Esophagus
● Ovary
● Pancreas
● Pituitary
● Prostate
● Fallopian Tube
● Heart
● Kidney
● LIver
● Skin
● Small Intestine
● Spleen
● Stomach
● Testis
● Thyroid
● Uterus
● Vagina

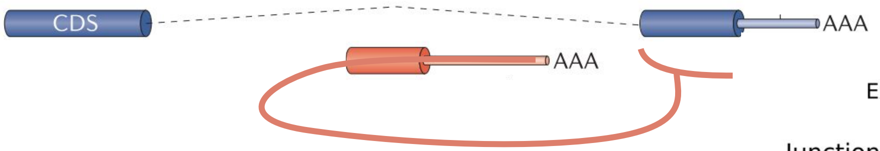


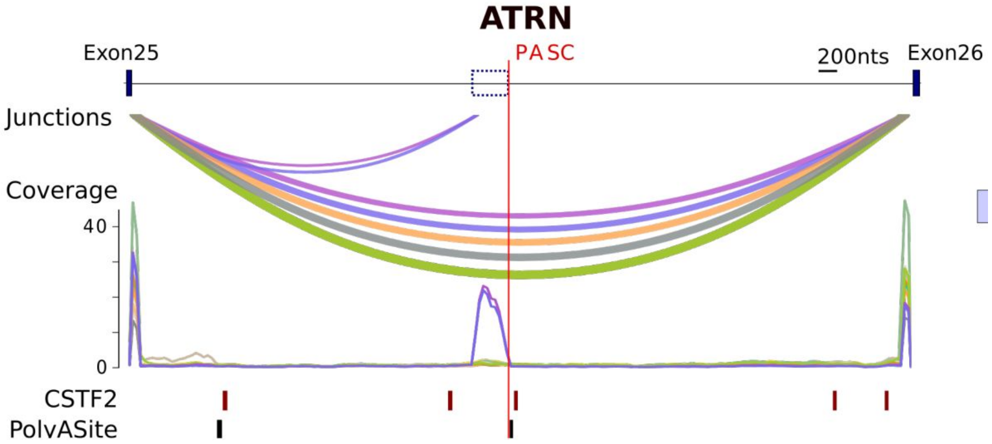


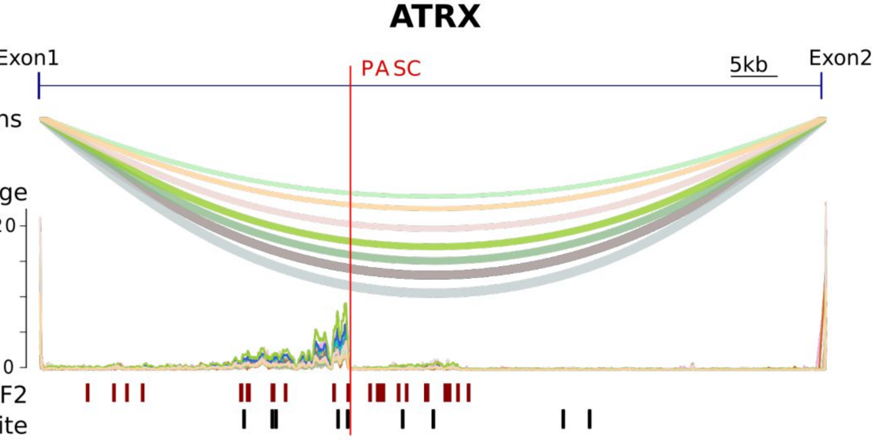
Examples

Skipped terminal exon

Lariat polyadenylation







- ● Adipose
- ● Adrenal Gland
- ● Artery
- ● Bladder
- ● Lung
- ● Minor Salivary Gland
- ● Muscle
- ● Nerve
- ● Breast
- ● Cervix
- ● Colon
- ● Esophagus
- ● Ovary
- ● Pancreas
- ● Pituitary
- ● Prostate
- ● Fallopian Tube
- ● Heart
- ● Kidney
- ● LIver
- ● Skin
- ● Small Intestine
- ● Spleen
- ● Stomach
- ● Testis
- ● Thyroid
- ● Uterus
- ● Vagina

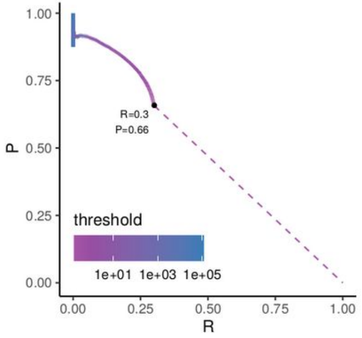
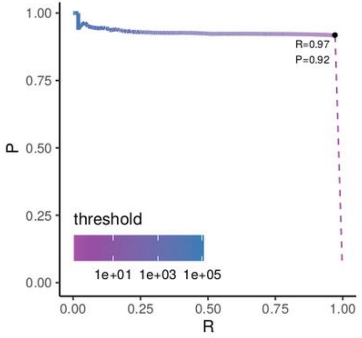
Conclusions

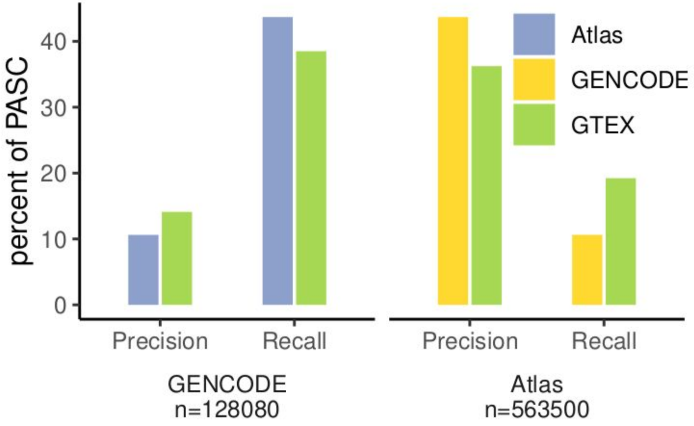
- PAS, including intronic ones, can be identified *de novo* from large scale RNA-seq data.
- A set of 318k PAS clusters from GTEx dataset can be used to study tissue-specific APA.
- A set of 70k iPASs with computed splicing characteristics can be used to examine interplay between AS and IPA.
- Data suggests existence of spliced out polyadenylated RNAs
  - We hypothesise that they are a result of splicing counteracting premature IPA.
  - Next step would be to look at splicing factor KD data to see how disruption of splicing affects the iPAS.

The publication is under revision. Preprint: M. Vlasenok, S. Margasyuk, D. Pervouchine. Transcriptome sequencing suggests that pre-mRNA splicing counteracts premature intronic polyadenylation. bioRxiv 2022.05.27.493724

Supplementary

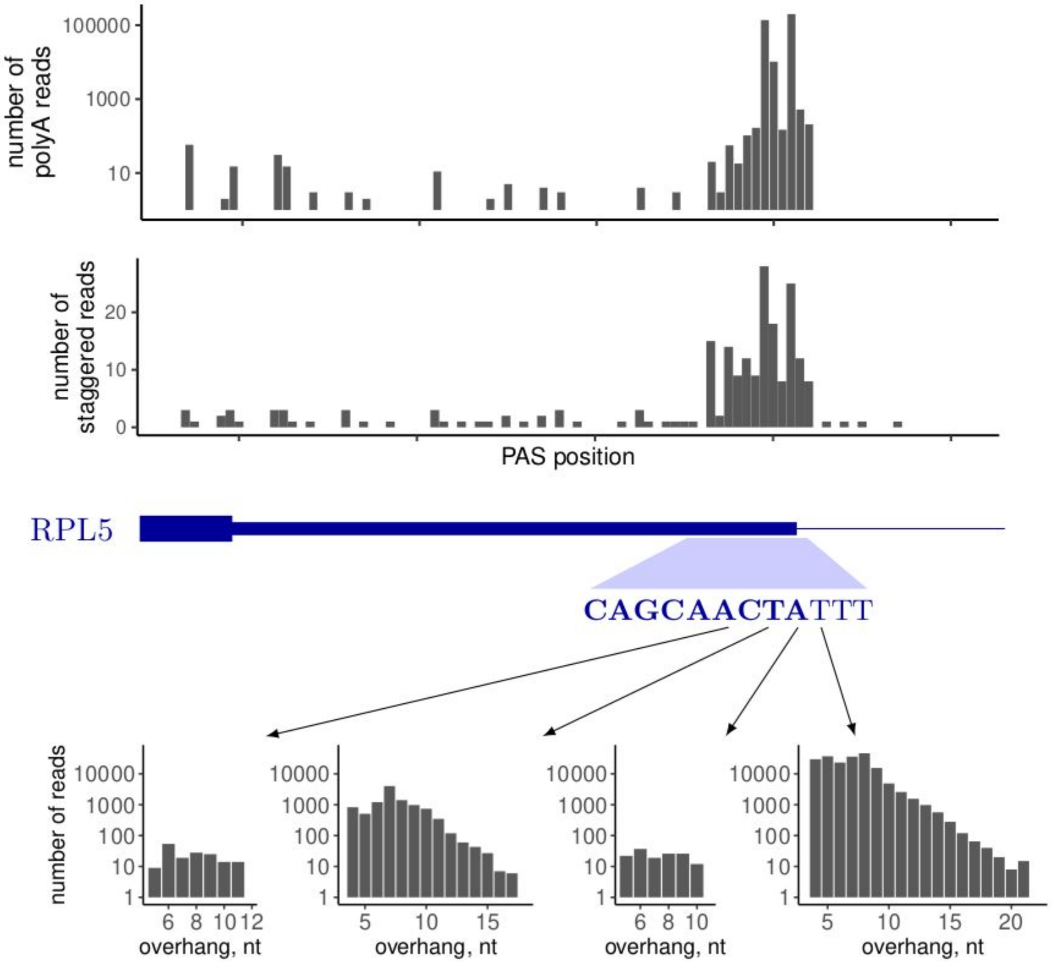
PAS set from GTEx is comparable with PolyASite 2.0

	GTEX vs Atlas
Number of PAS	
Weighted by read support	

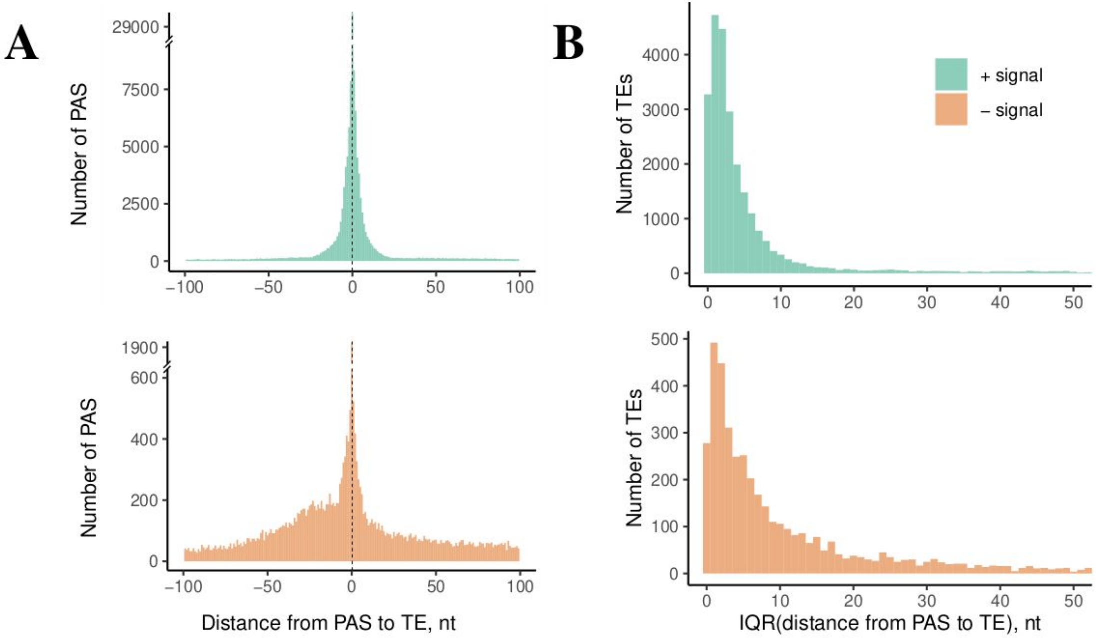


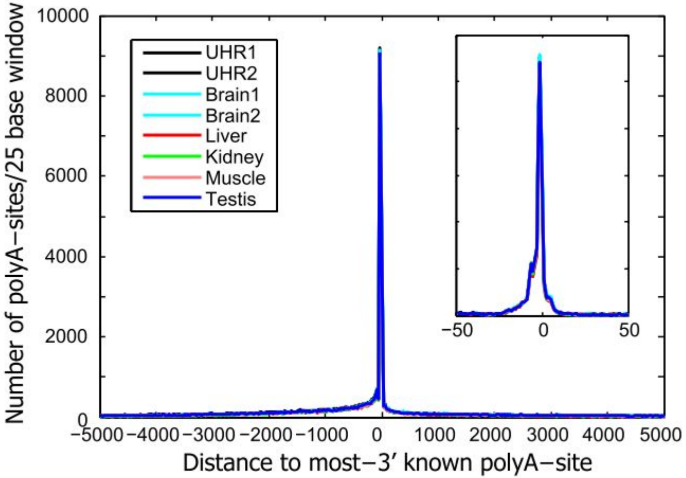
Precision (GTEx compared to Atlas) ~ 40%
Recall (GTEx compared to Atlas) ~ 20%

RPL5


number of polyA reads
100000
1000
10
number of staggered reads
20
10
0
PAS position
RPL5
CAGCAACTATTT
number of reads
10000
1000
100
10
1
6 8 10 12
overhang, nt
5 10 15
overhang, nt
6 8 10
overhang, nt
5 10 15 20
overhang, nt

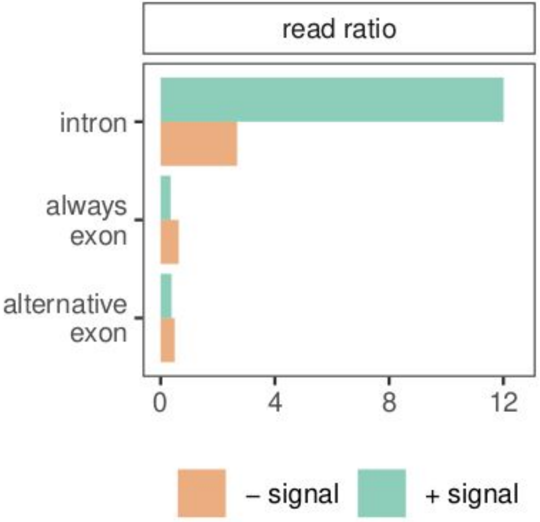
PAS around transcript ends (TEs)

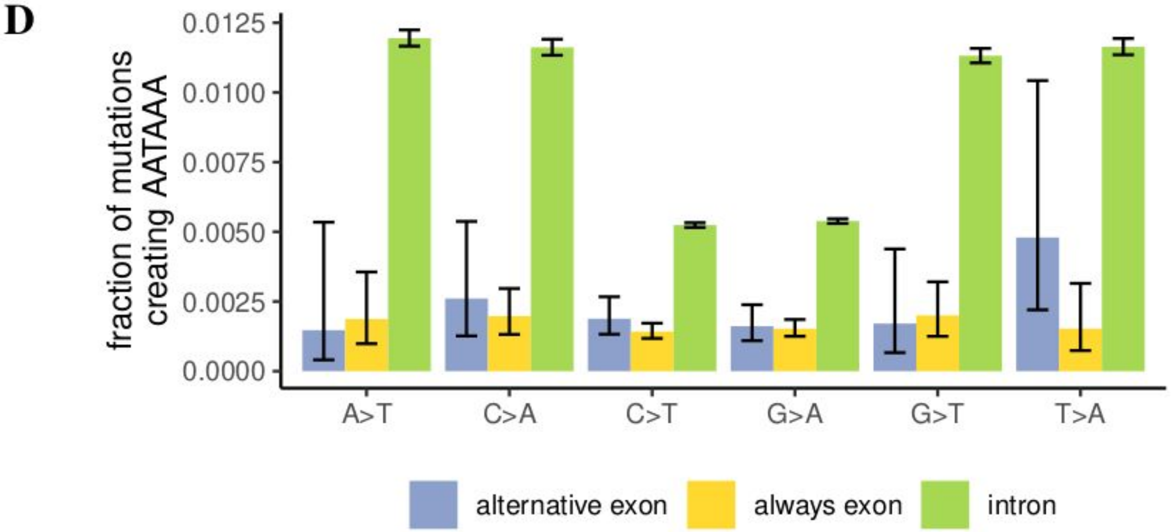




Derti, A. et al. A quantitative atlas of polyadenylation in five mammals. Genome Res. 22, 1173–83 (2012).

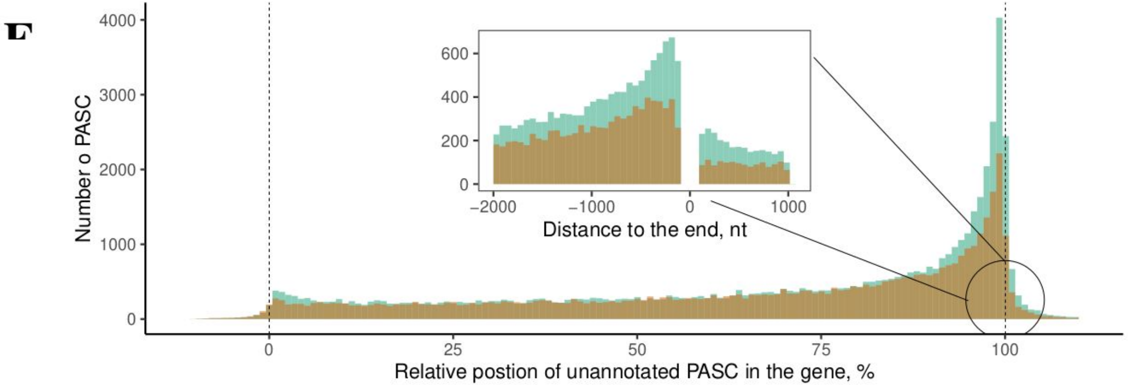
Cryptic intronic polyadenylation is tolerated by the system



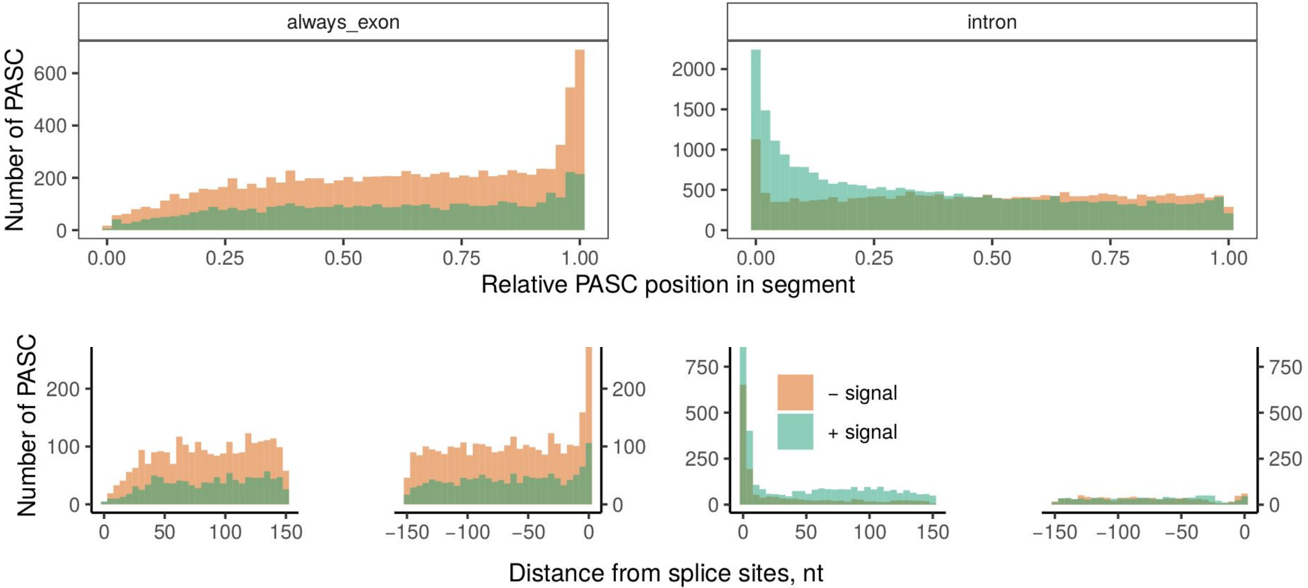


PASC distribution in gene and gene regions

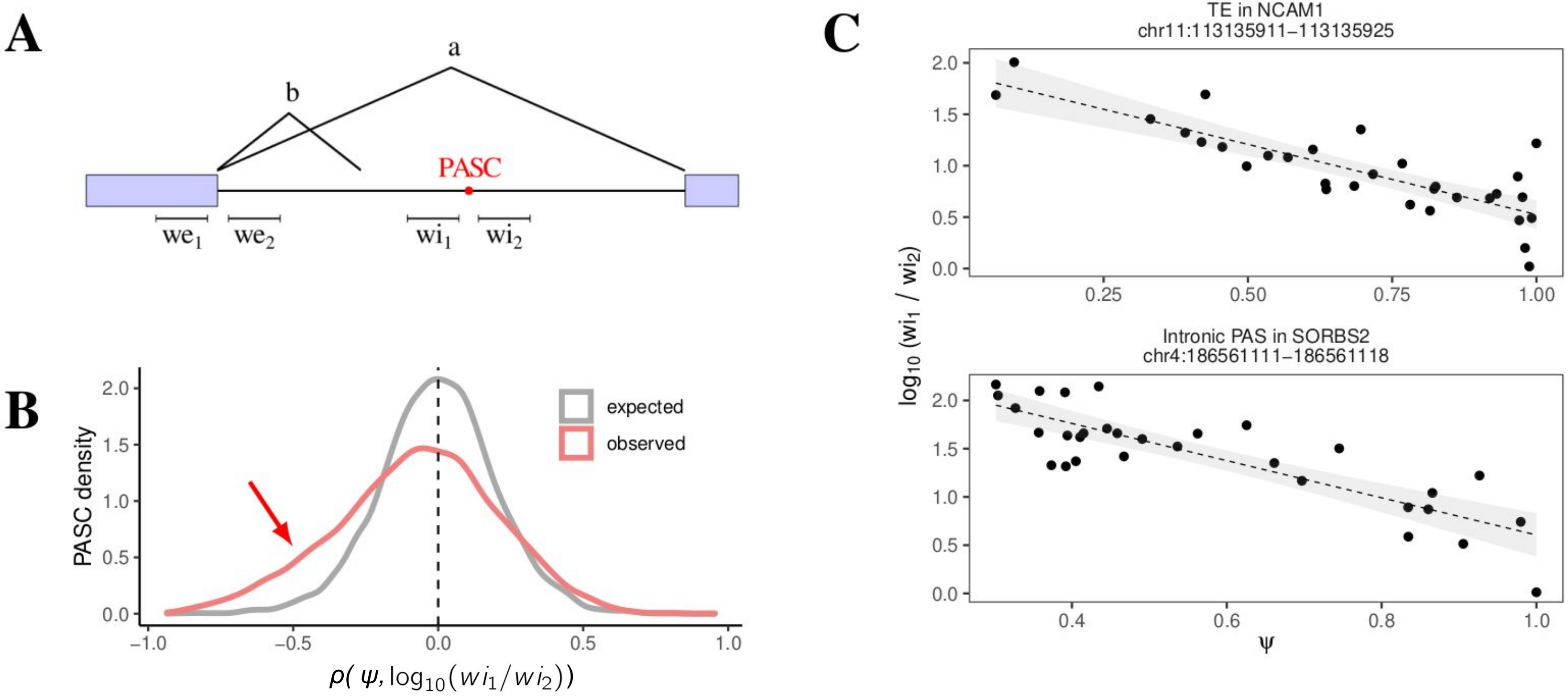
A



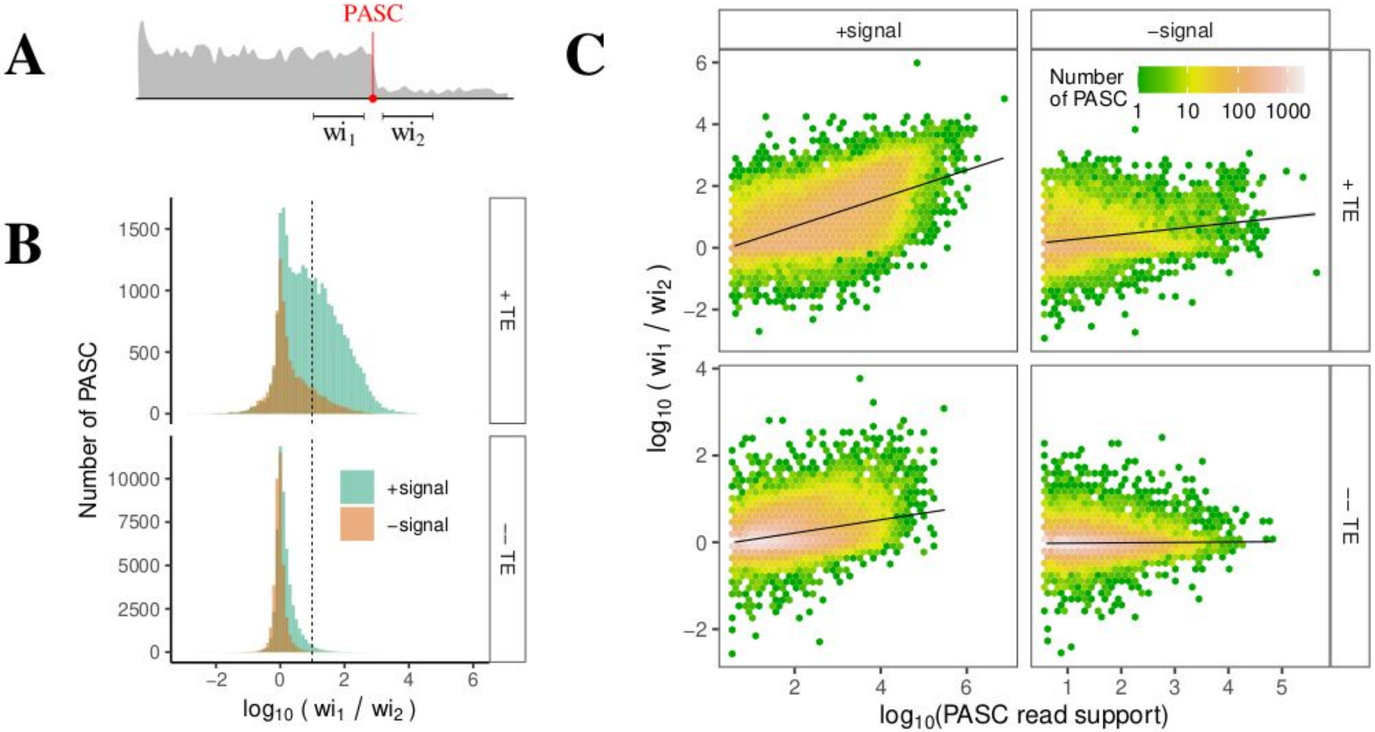
B



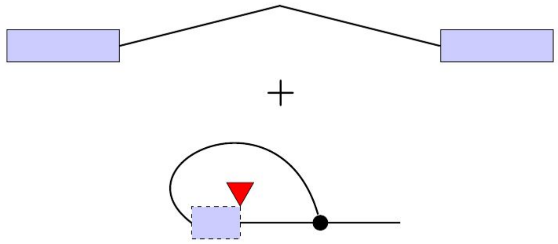
Intronic polyadenylation. Splicing vs polyadenylation

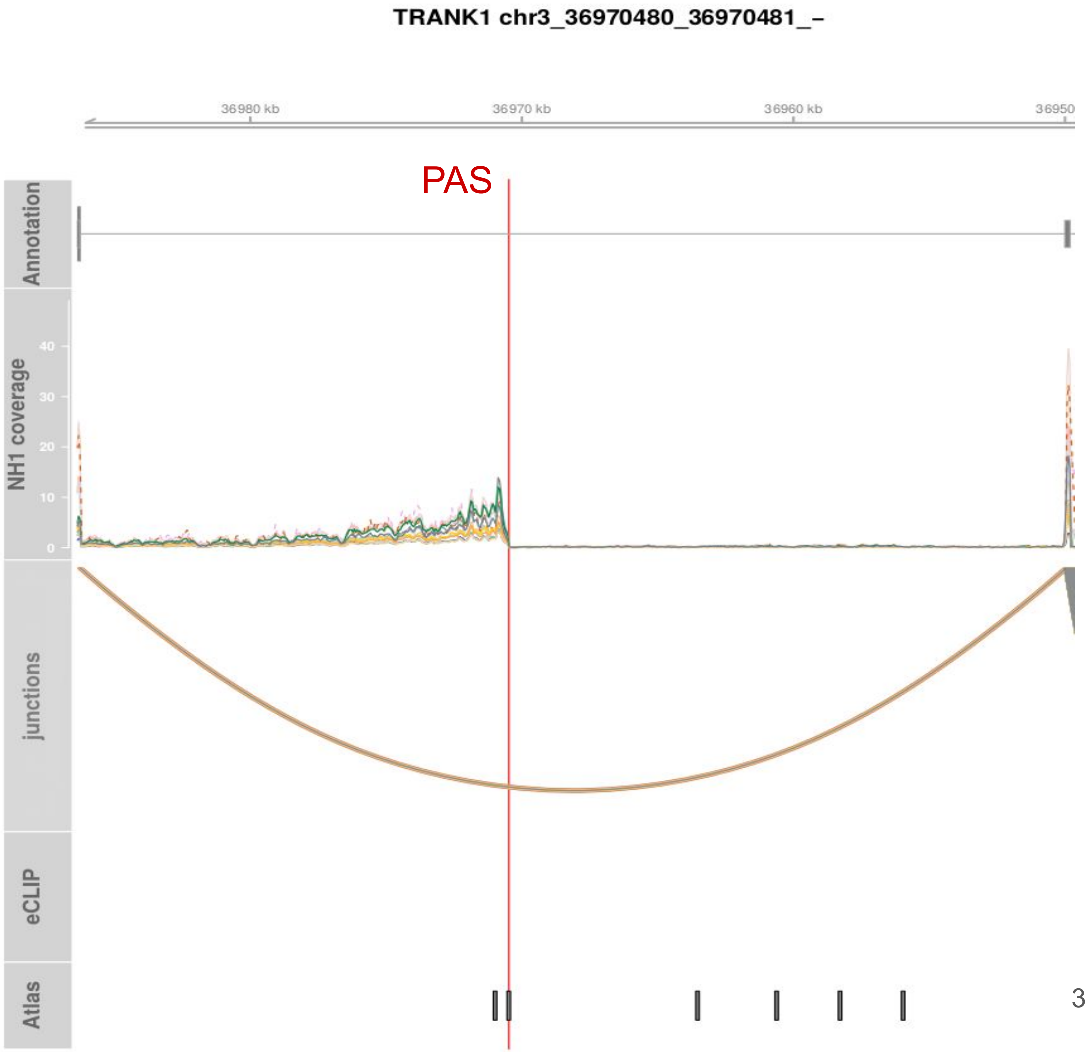


Coverage fold change at PASC



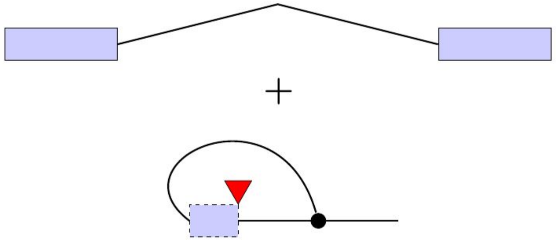
Lariat polyadenylation examples

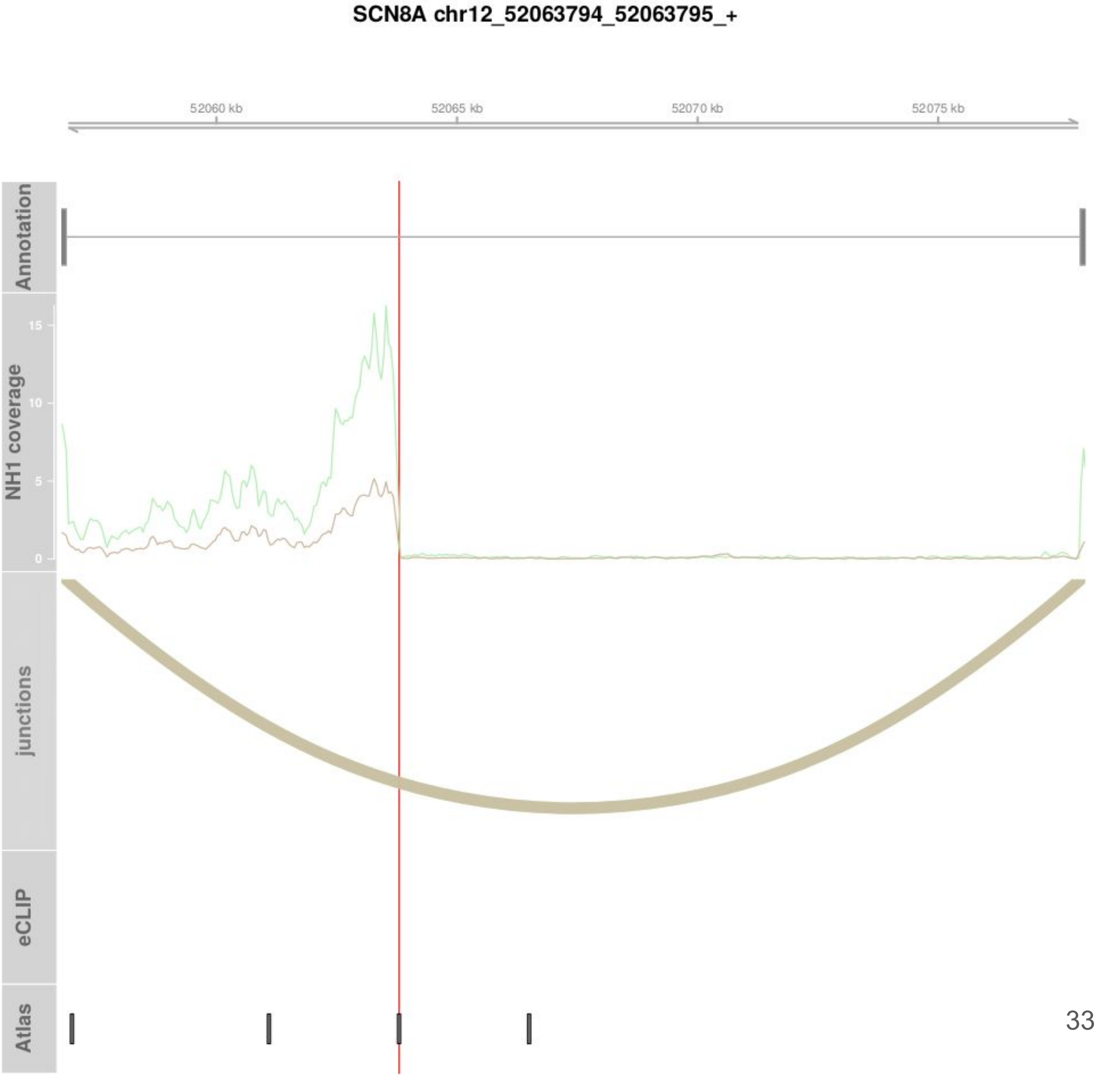



TRANK1 chr3_36970480_36970481_–

- ● Adipose
- ● Adrenal_Gland
- ● Artery
- ● Bladder
- ● Breast
- ● Cervix
- ● Colon
- ● Esophagus
- ● Fallopian_Tube
- ● Heart
- ● Kidney
- ● Liver
- ● Lung
- ● Minor_Salivary_Gland
- ● Muscle
- ● Nerve
- ● Ovary
- ● Pancreas
- ● Pituitary
- ● Prostate
- ● Skin
- ● Small_Intesti
- ● Spleen
- ● Stomach
- ● Testis
- ● Thyroid
- ● Uterus
- ● Vagina

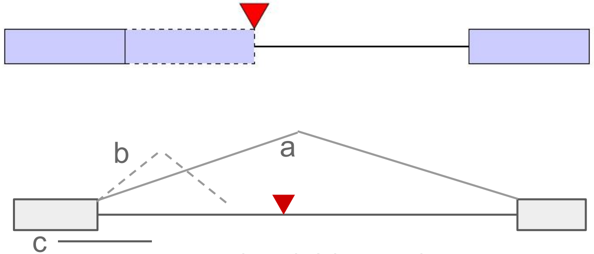
Examples. Lariat polyadenylation





● Liver ● Pituitary

Additional examples. CTE



$\psi = (b+c) / (a+b+c)$