

Leveraging Long Reads Sequencing for Developing a Functional Iso-Transcriptomics Analysis Framework

Ana Conesa, Research Professor

Institute for Integrative Systems Biology

National Research Council, Spain

Courtesy Professor, University of Florida

[@anaconesa](#)

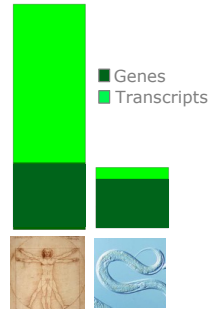
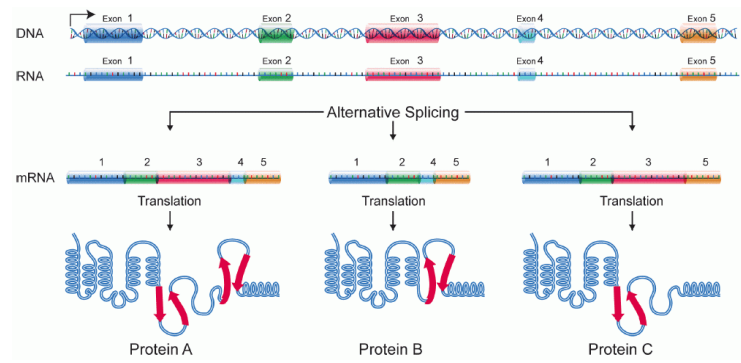
[@conesalab](#)

ana.conesa@csic.es

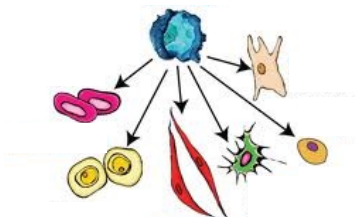


Genomics 
 of Gene
Expression Lab

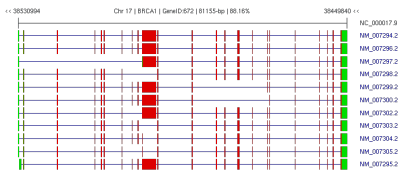
Alternative Splicing is a source of biological complexity



Organismal complexity
4x more AS in humans

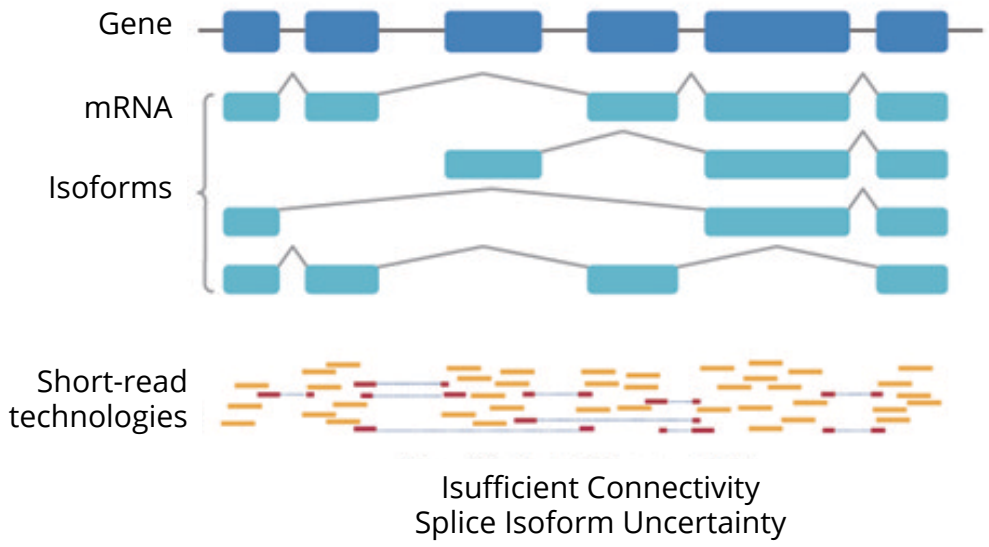


Differentiation
Tissue – specific isoforms



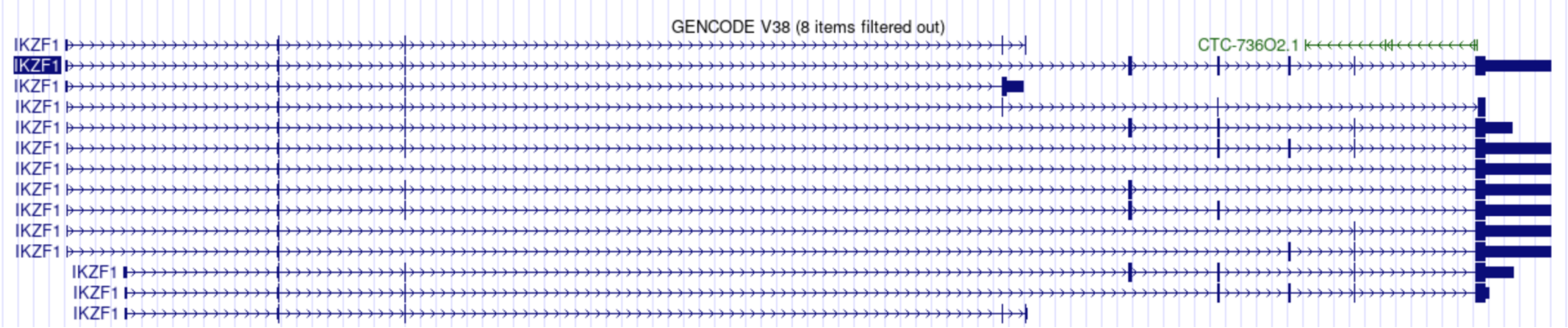
Disease
100s isoforms in BRCA1

Sequencing technologies and alternative splicing

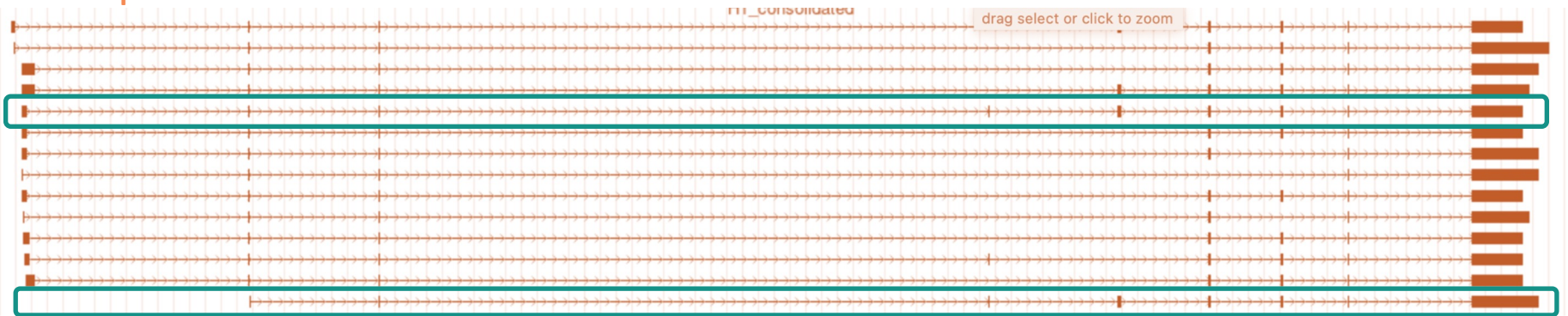


Long-reads do reveal transcriptome complexity

Reference annotation for Ikzf1




IrRNA-seq data on H1 human cell line



And what about understanding function of isoforms?

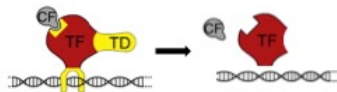
A. PTR protein-level functional impact

 Isoform modification

(1) Loss of active site



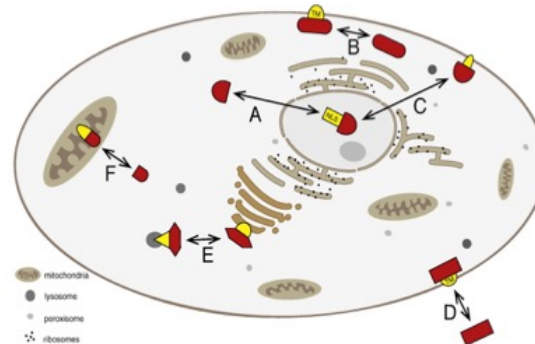
(2) Change in TFs



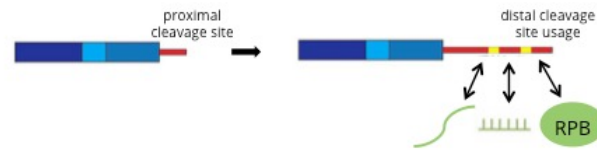
(3) Loss of PP binding motif



(4) Changes in intracellular localization



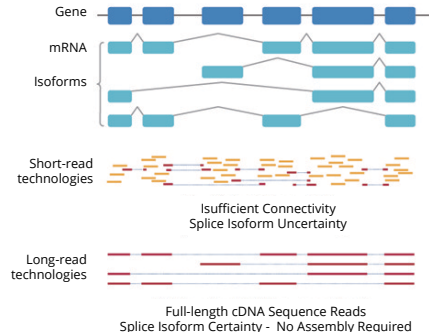
B. PTR transcript-level functional impact



- mRNA nuclear export
- mRNA stability
- mRNA translation
- Protein localisation
- mRNA localisation

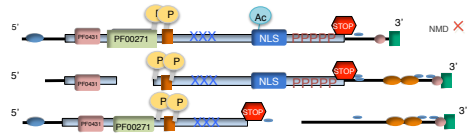
Introducing Functional Iso-Transcriptomics analysis

Good full-length transcriptomes



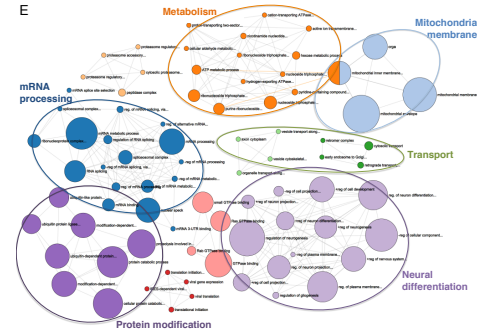
SQANTI 3

Rich isoform level functional labels



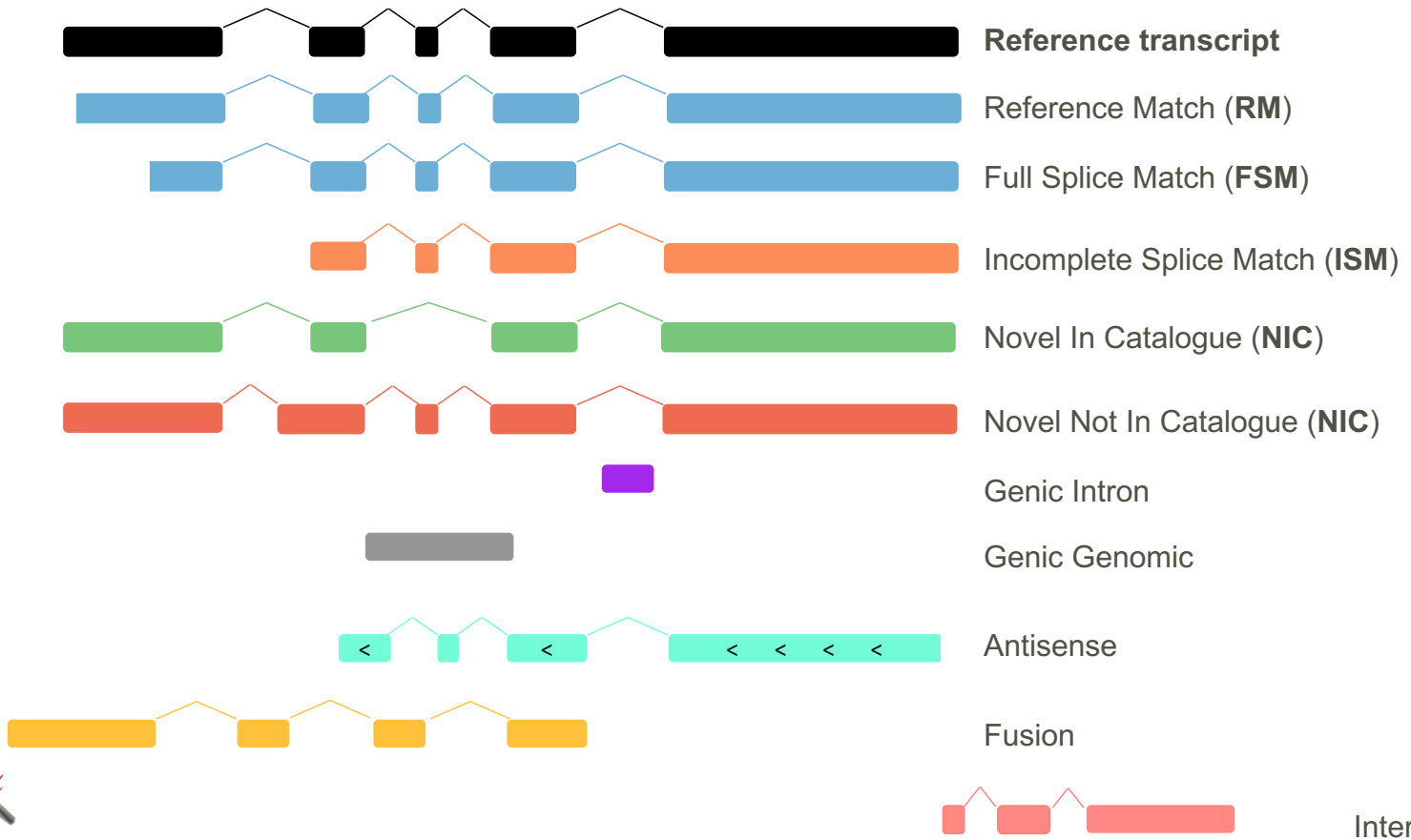
IsoAnnot

Function-oriented statistical methods

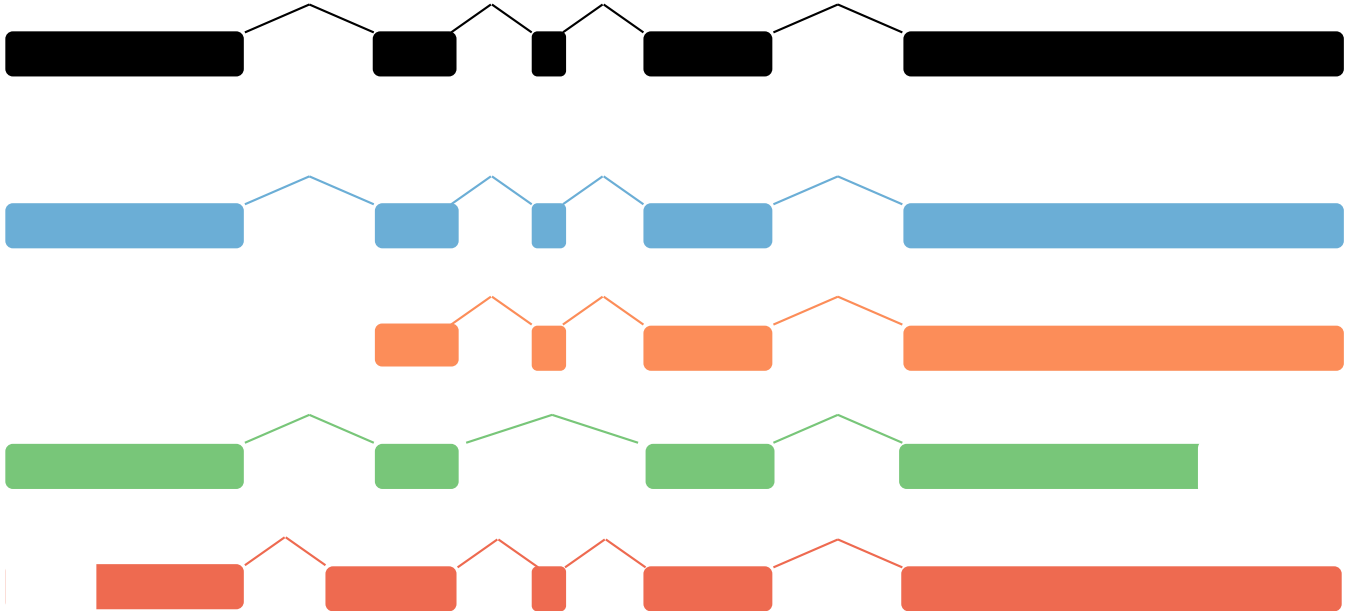


tappAS

SQANTI3 categories

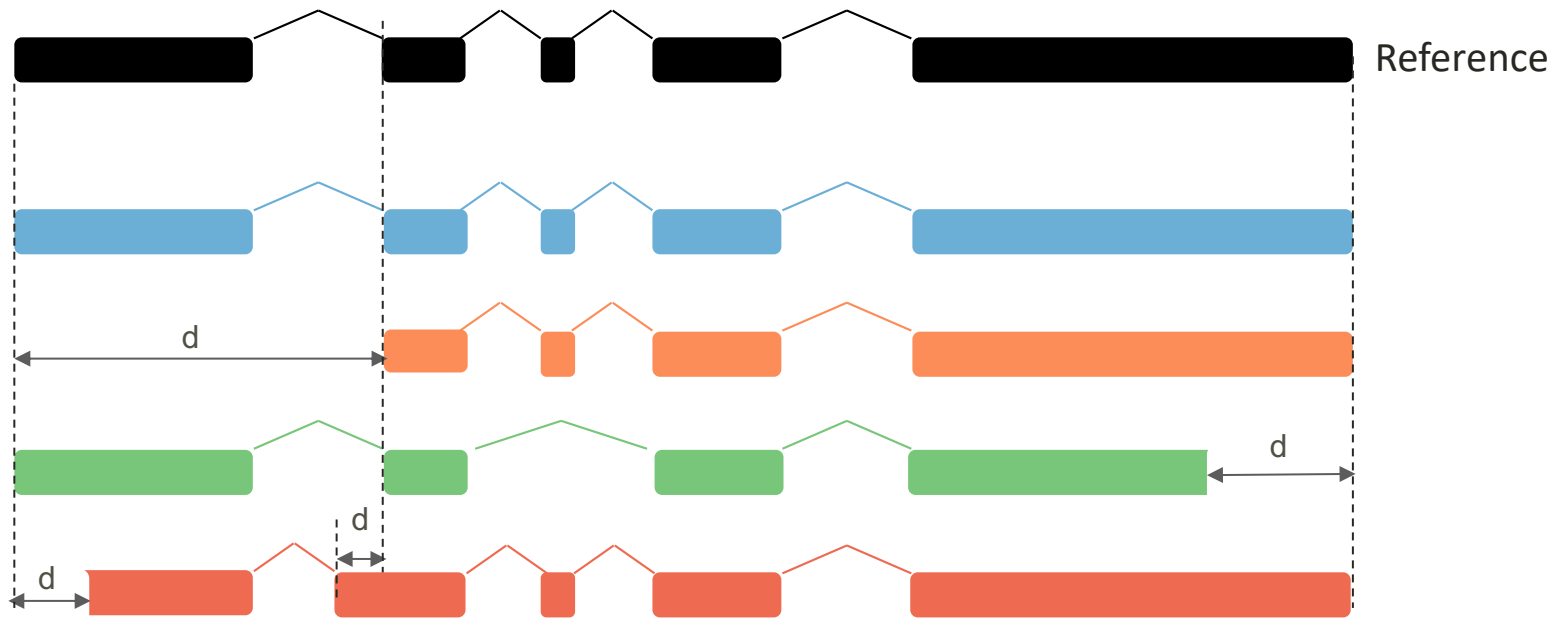


What does SQANTI3 measure?

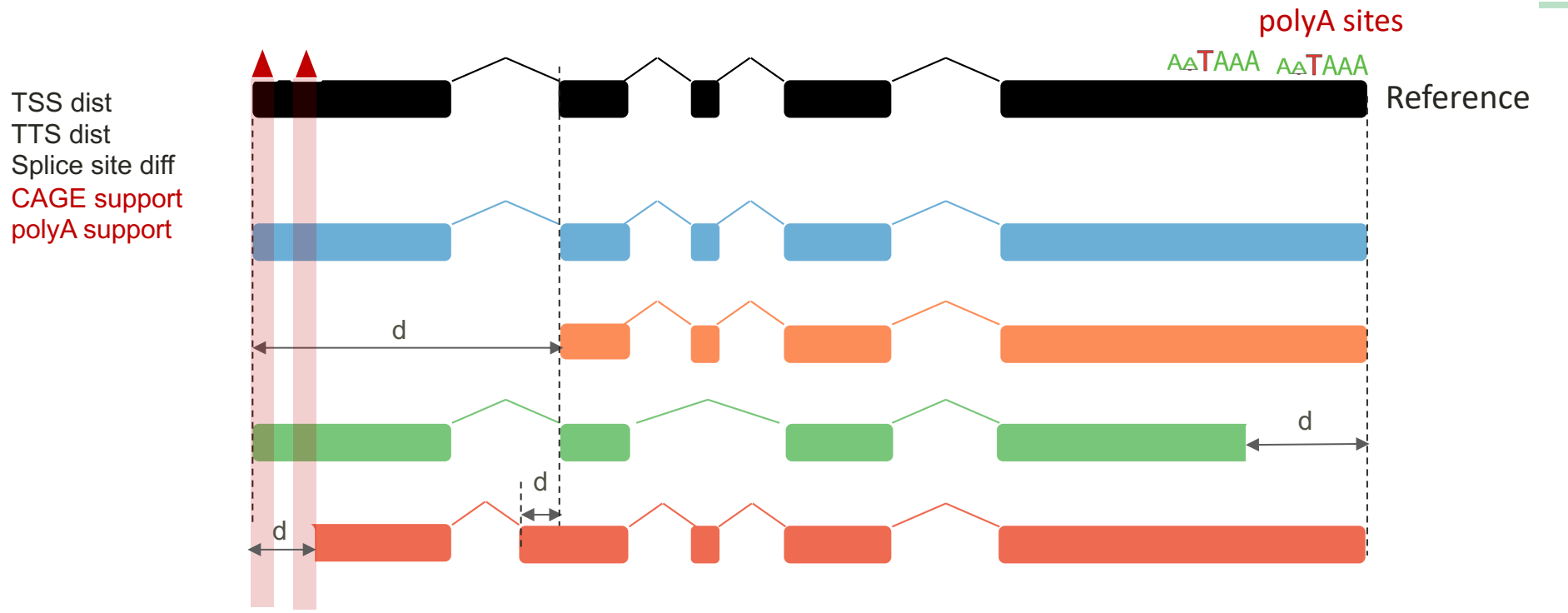


Distance to TSS, TTS and to known junctions

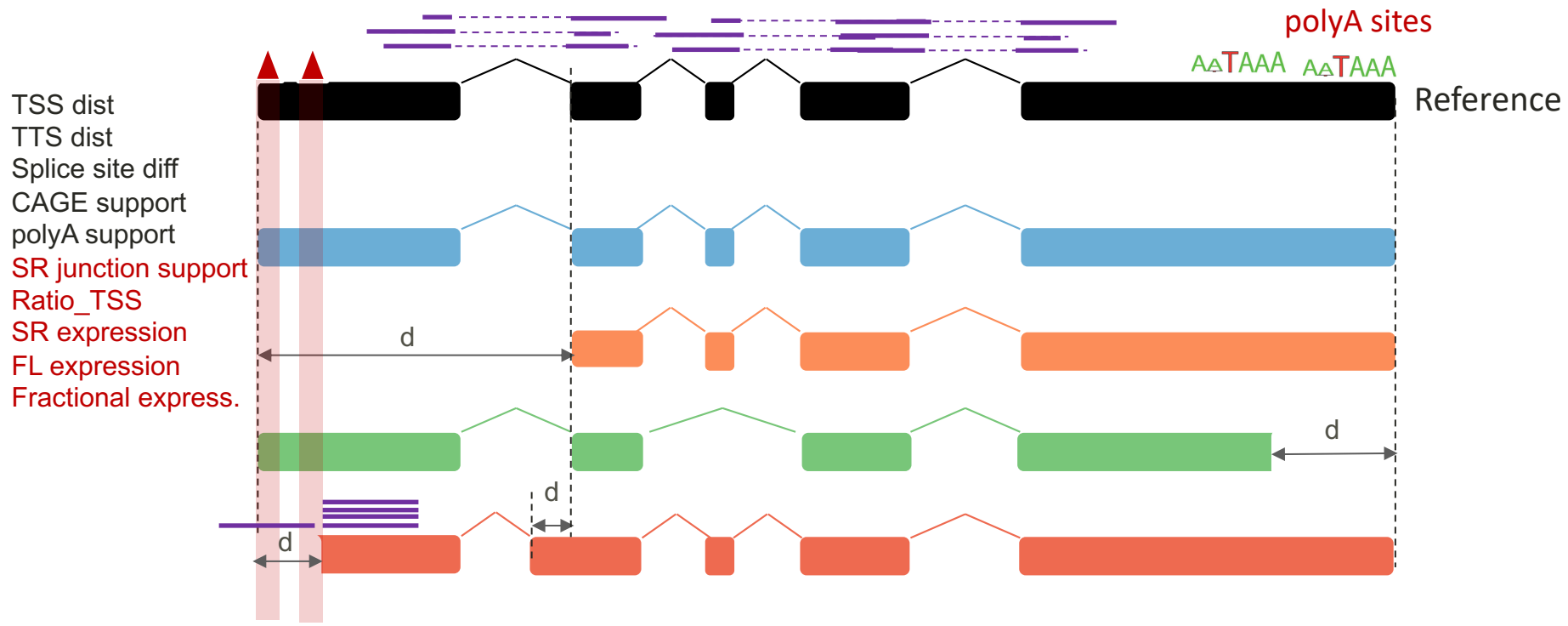
TSS dist
TTS dist
Splice site diff



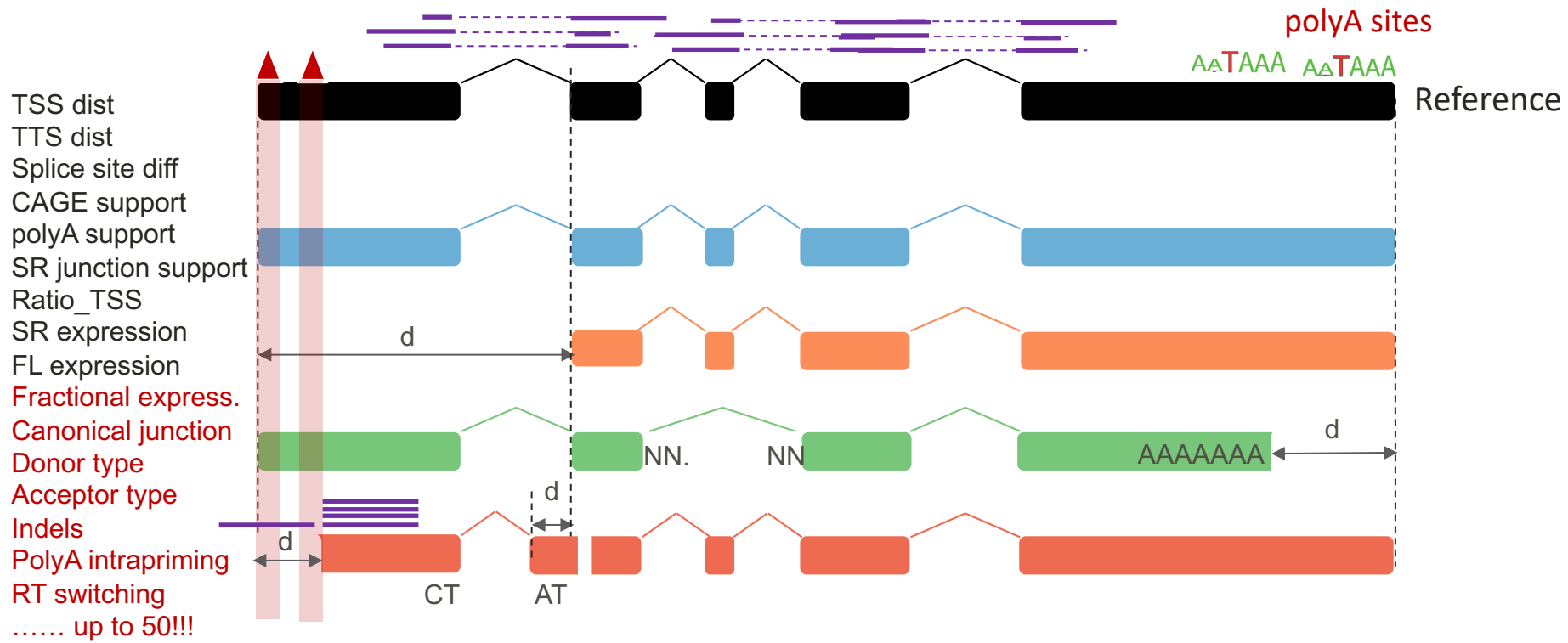
CAGE and PolyA support



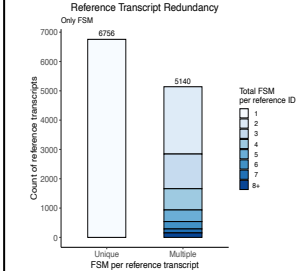
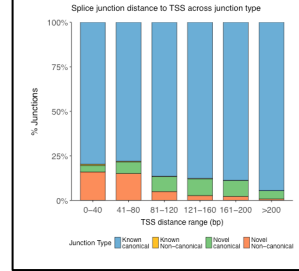
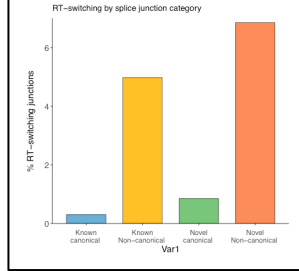
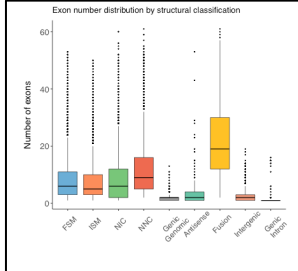
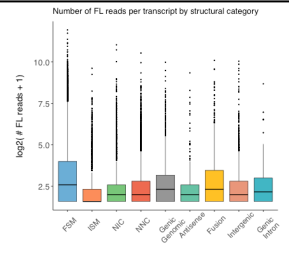
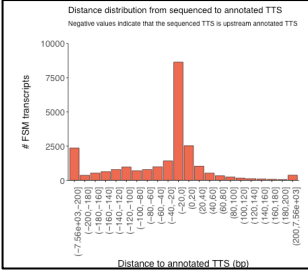
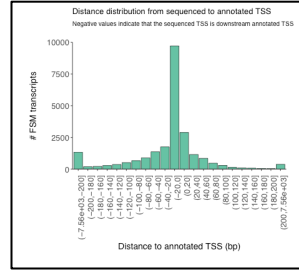
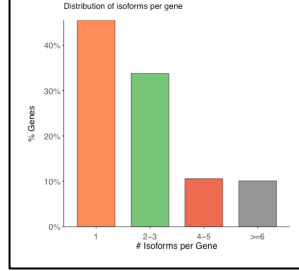
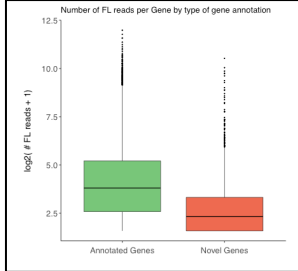
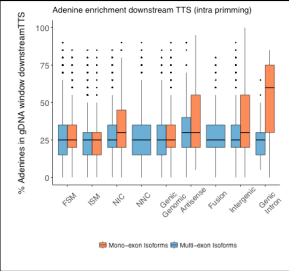
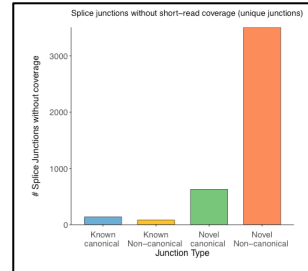
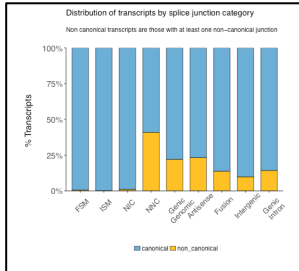
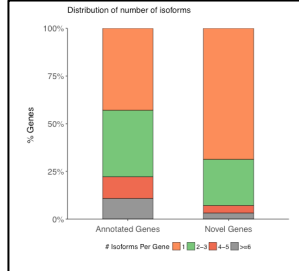
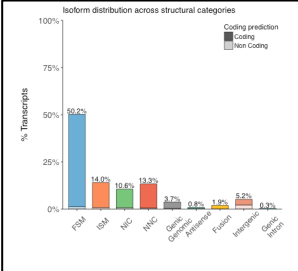
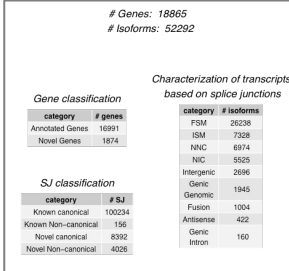
Short reads and lrrRNA-seq support



Canonical junction, Intrapriming, RT switching....

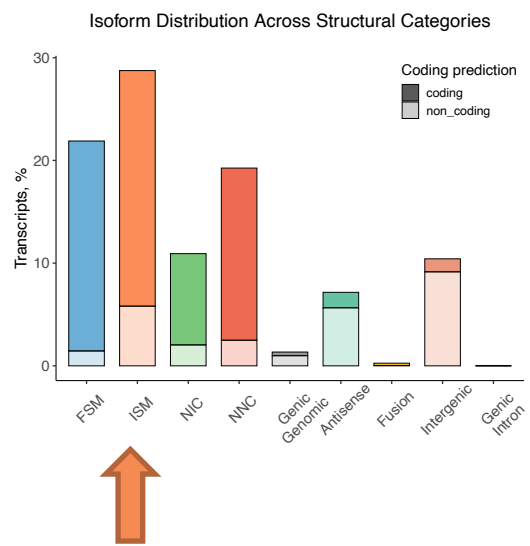


SQANTI3

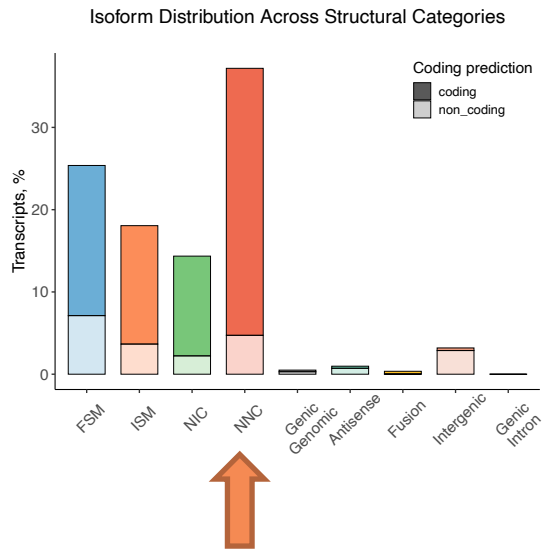


Important conclusions from SQANTI3 output

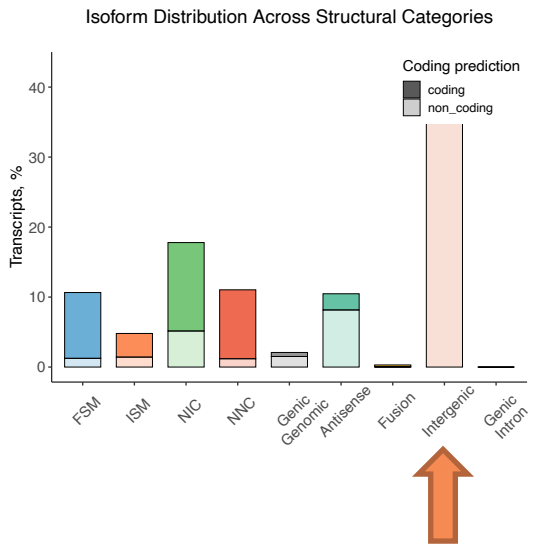
Pipeline 1



Pipeline 2



Pipeline 3



LRGASP: systematic assessment of lrrRNA-seq methods



3 species



human



mouse



manatee

3 sequencing
platforms



4 sample types



one cell type



cell mixtures



spike-ins



synthetic data

4 library protocols



cDNA



direct RNA



R2C2



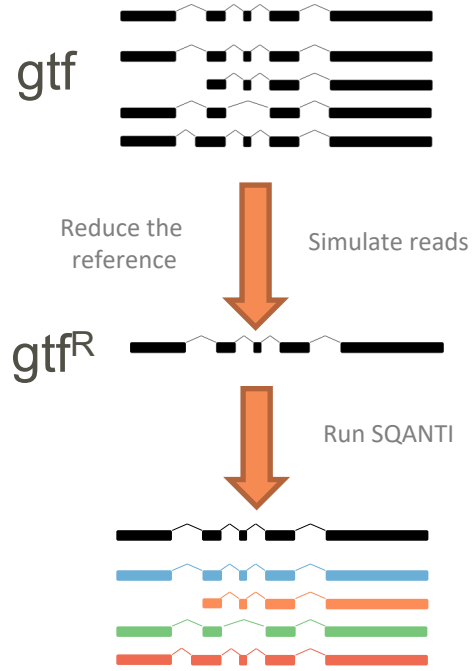
CapTrap

In a nutshell...

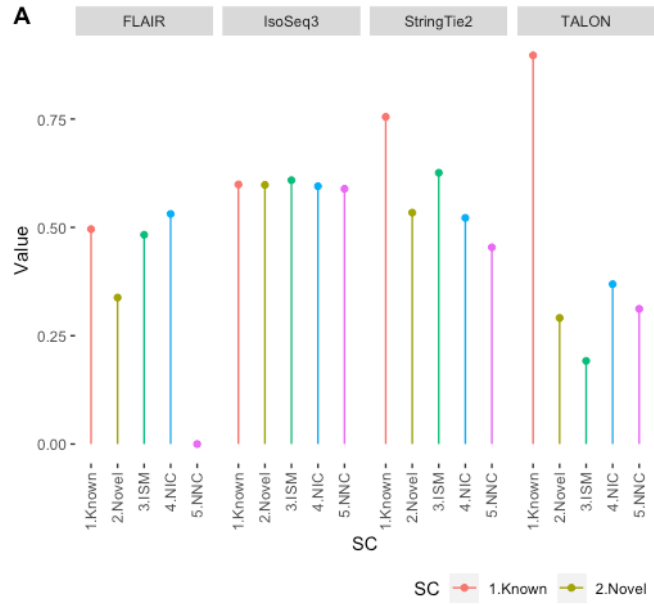
One big upcoming challenge in lrrRNA-seq is to identify which of the detected transcript models are real, worth to annotate and how to describe this complexity



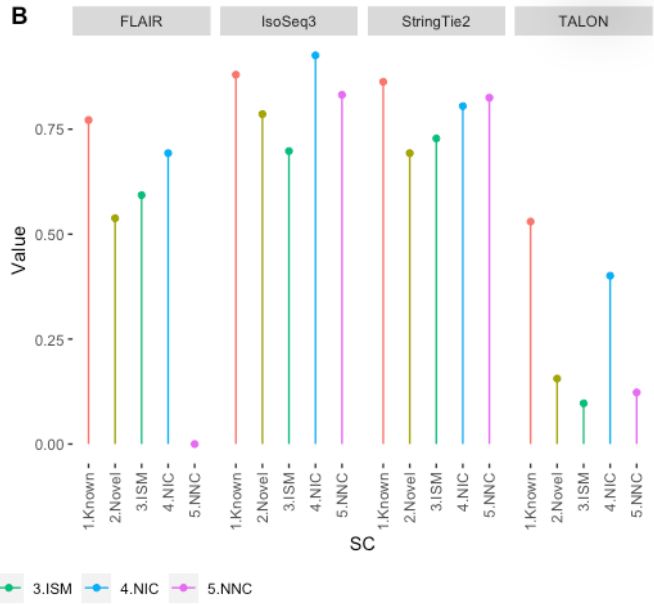
Assessing the novelty



SENSITIVITY



PRECISION

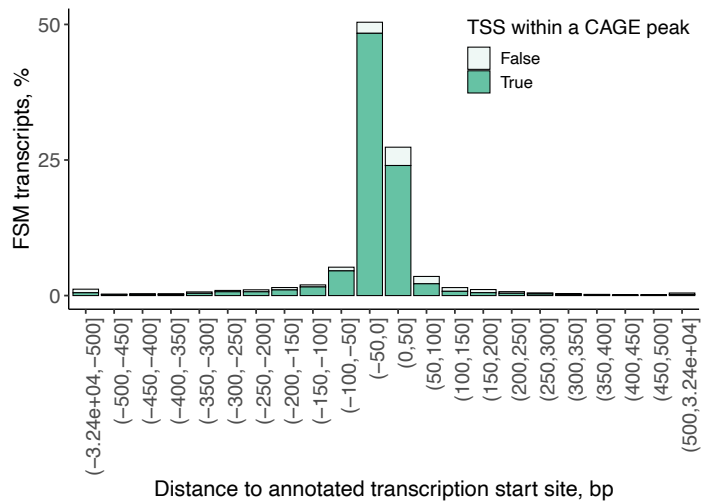


SC 1.Known 2.Novel 3.ISM 4.NIC 5.NNC

CAGE data may support FSM!

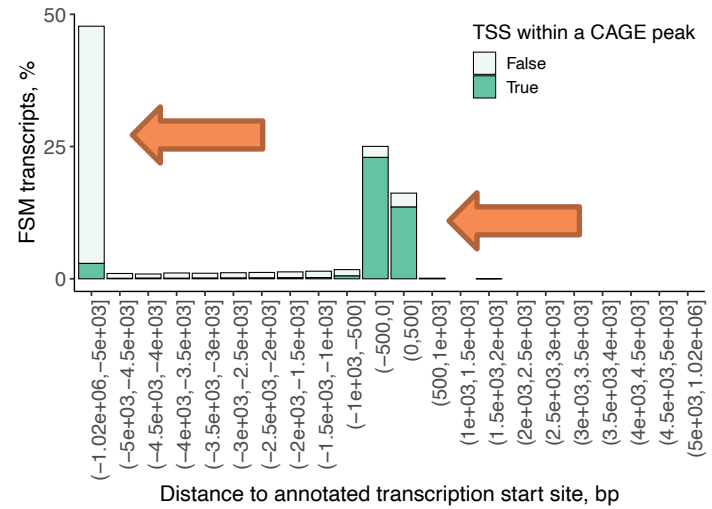
Distance to Annotated Transcription Start Site for FSM

Negative values indicate downstream of annotated TSS



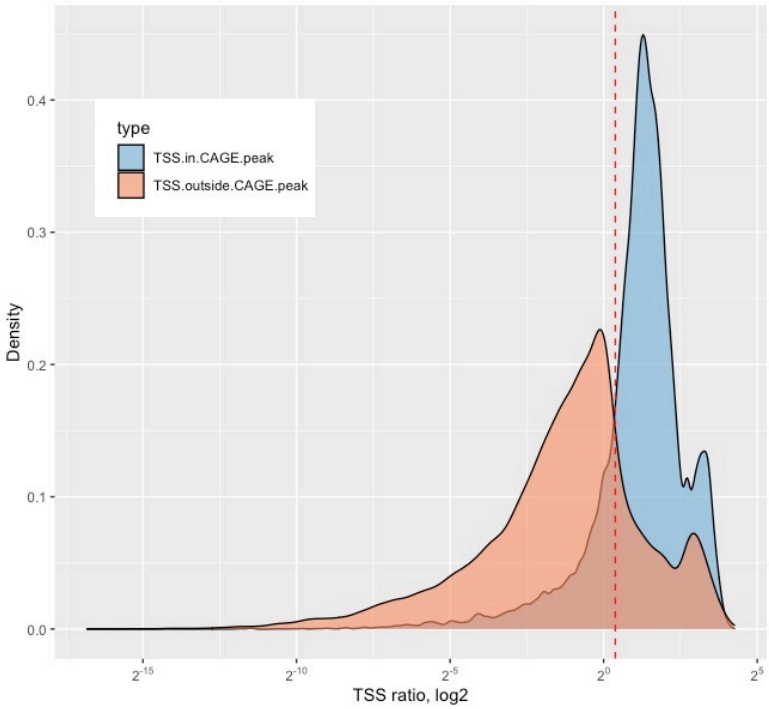
Distance to Annotated Transcription Start Site for ISM

Negative values indicate downstream of annotated TSS

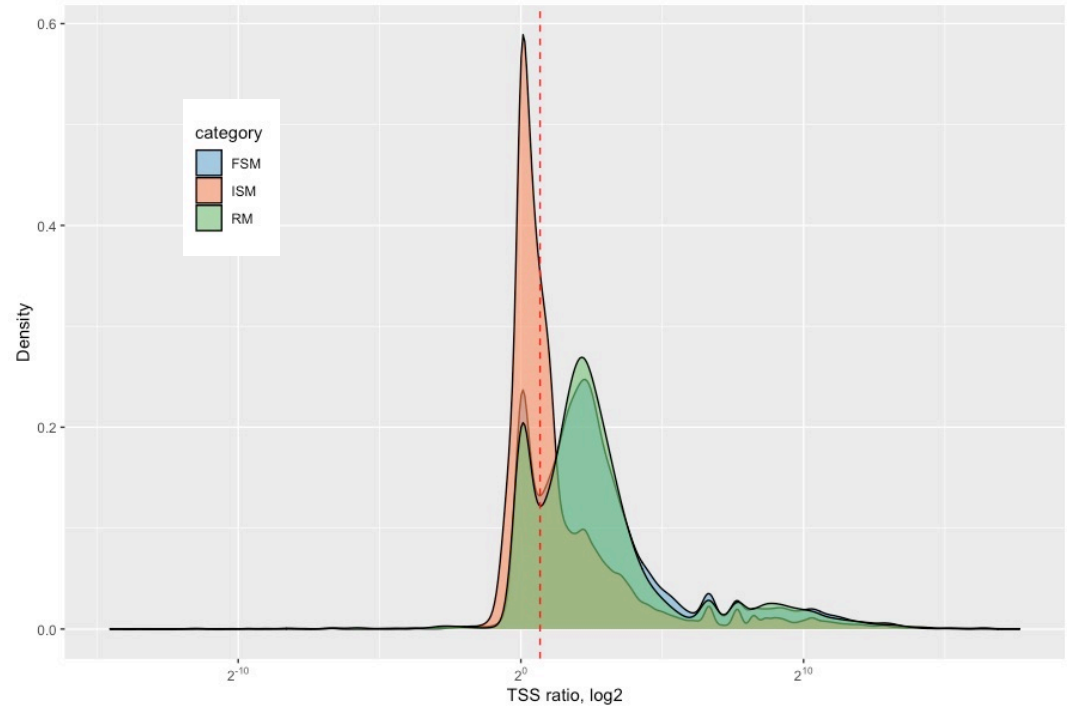


TSS_ratio is informative of category and CAGE support

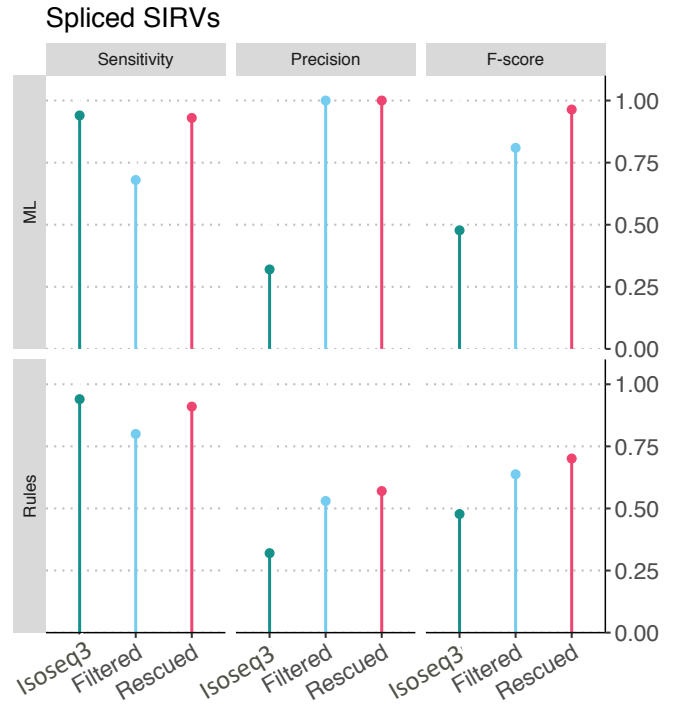
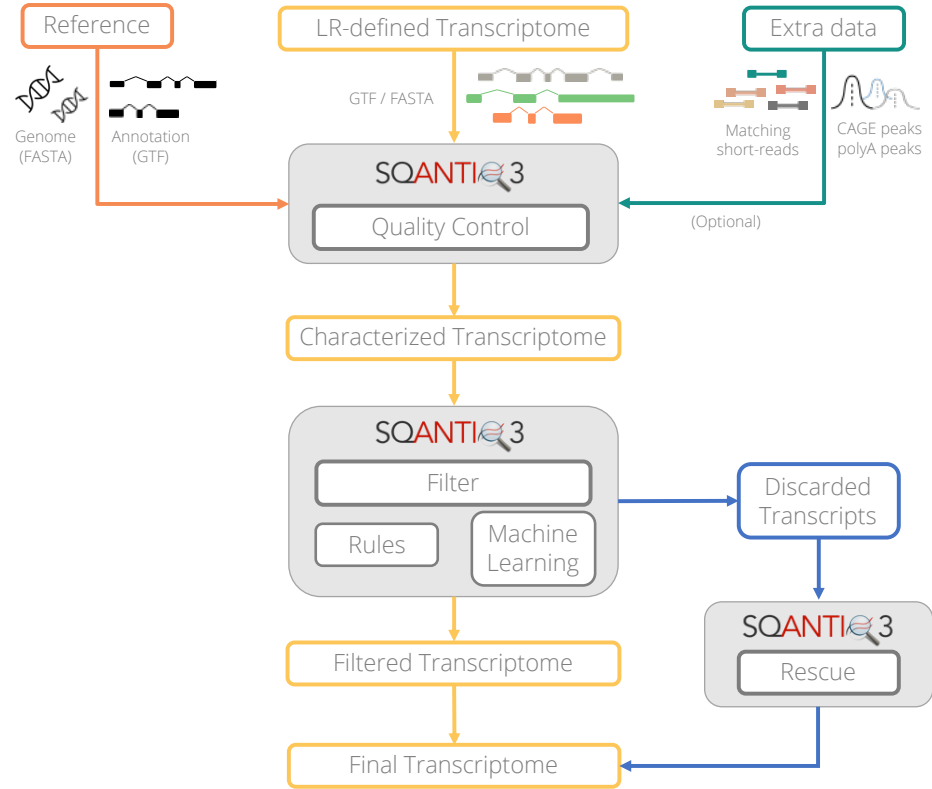
TSS ratio and CAGE support



TSS ratio for FSM, ISM and RM



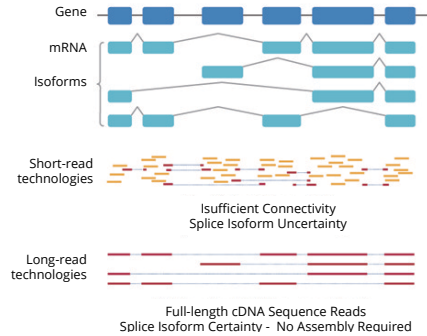
SQANTI3 uses orthogonal data to curate lrrna-seq data



Pardo Palacio et al., in preparation

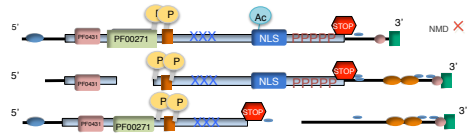
Introducing Functional Iso-Transcriptomics analysis

Good full-length transcriptomes



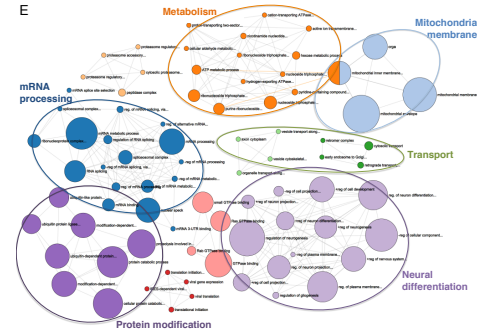
SQANTI 3

Rich isoform level functional labels



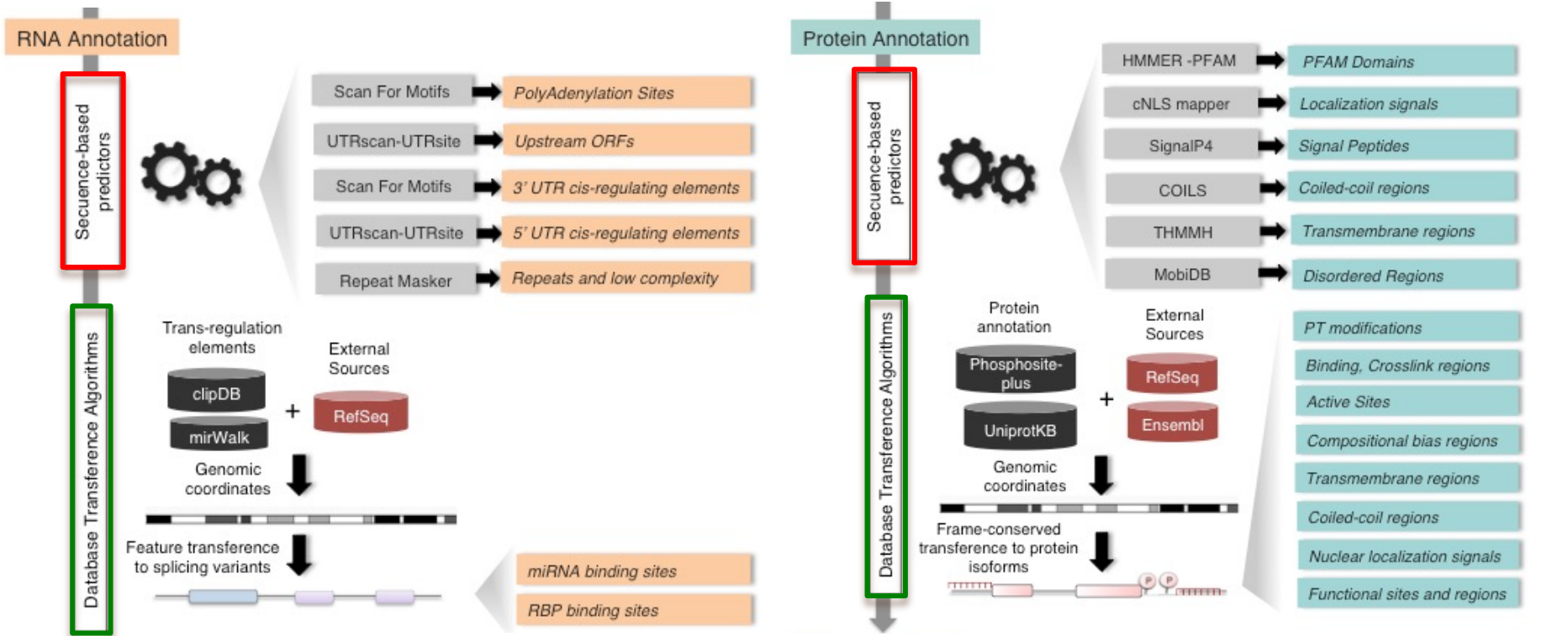
IsoAnnot

Function-oriented statistical methods



tappAS

Rich functional Annotation of isoforms with IsoAnnot



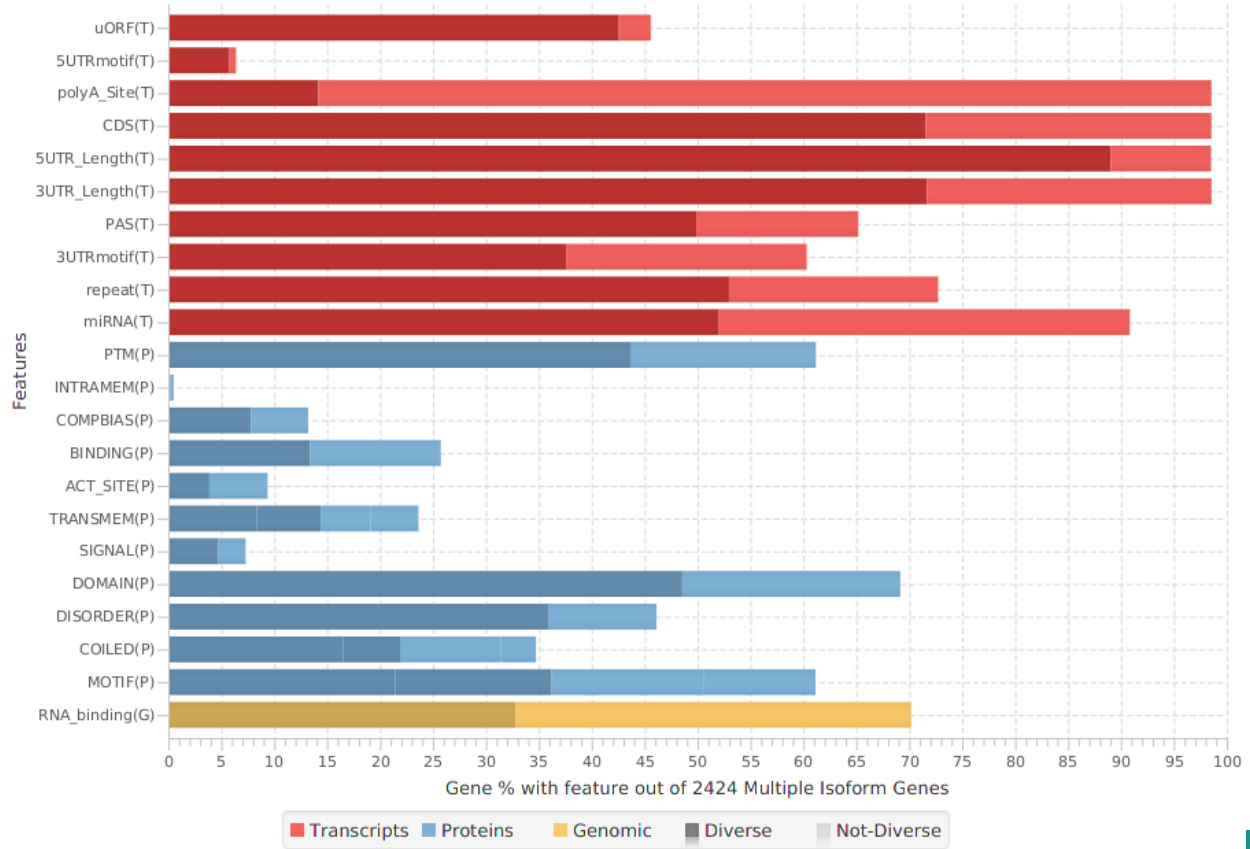
Annotation coverage across species

Organism	Reference	# Isoform Resolved Functional Features	% Isoforms with Functional Features		
			Transcript Level	Protein Level	Total
<i>Mus musculus</i>	PacBio (Tardaguila et al., 2018)	386,114	98.08	98.20	99.73
	RefSeq78	2,628,525	90.24	96.38	94.66
	Ensembl86	1,314,089	59.17	92.70	68.11
<i>Homo sapiens</i>	RefSeq78	5,977,941	94.42	98.19	96.40
	Ensembl86	2,888,409	65.03	95.38	71.88
<i>Drosophila melanogaster</i>	Flybase617	1,023,087	80.85	83.85	85.86
<i>Arabidopsis thaliana</i>	Ensembl34	457,825	54.56	90.43	85.35
<i>Zea mays</i>	Ensembl34	1,099,124	65.29	86.24	91.08

Functional diversity of isoforms

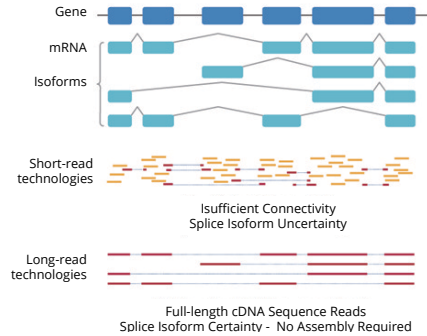


Gene Level Functional Diversity



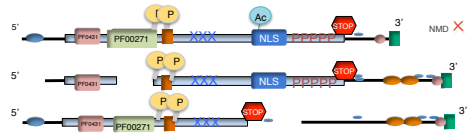
Introducing Functional Iso-Transcriptomics analysis

Good full-length transcriptomes



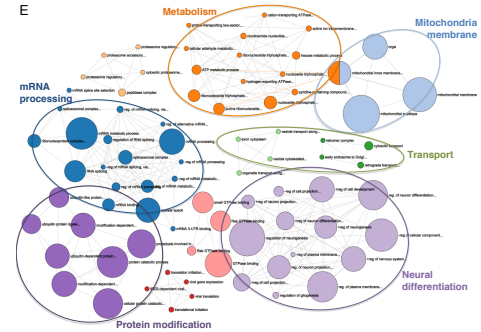
SQANTI 3

Rich isoform level functional labels

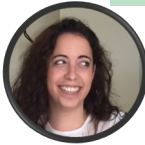
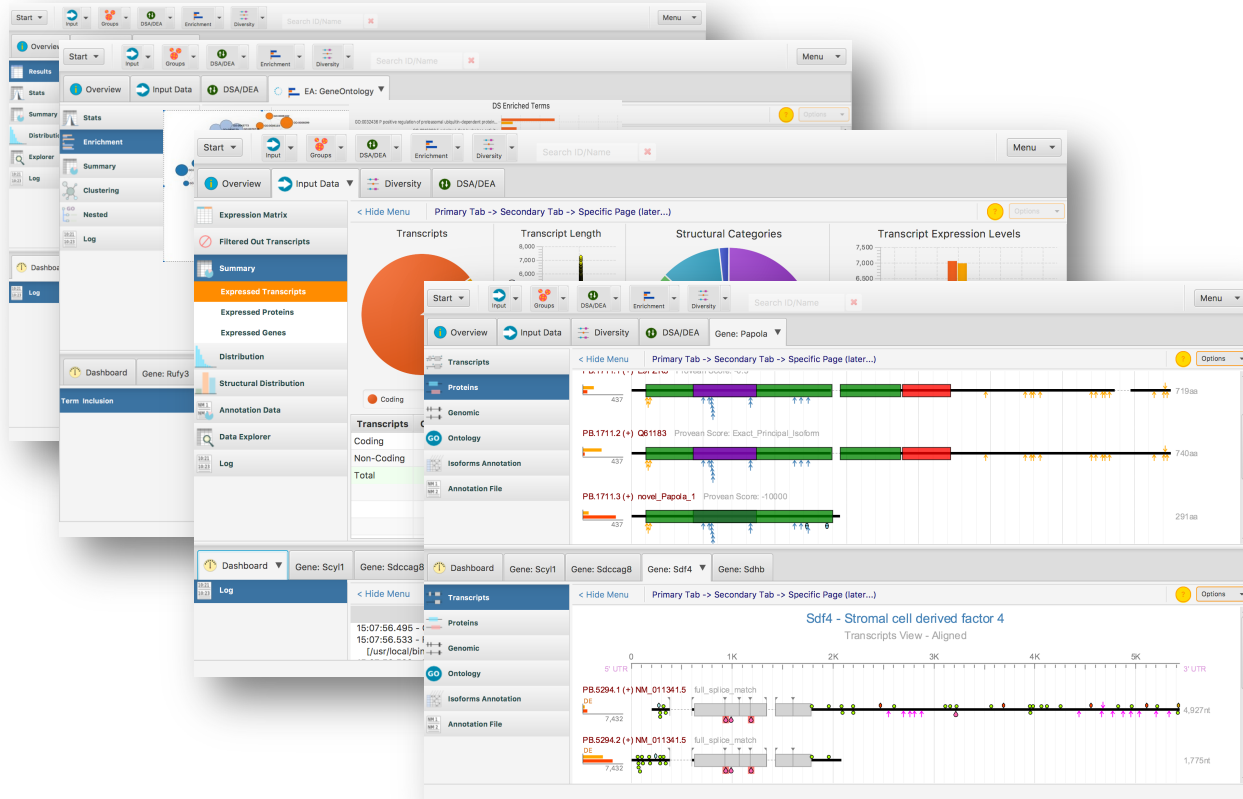


IsoAnnot

Function-oriented statistical methods



tappAS



INPUT DATA

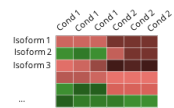
Gene Models



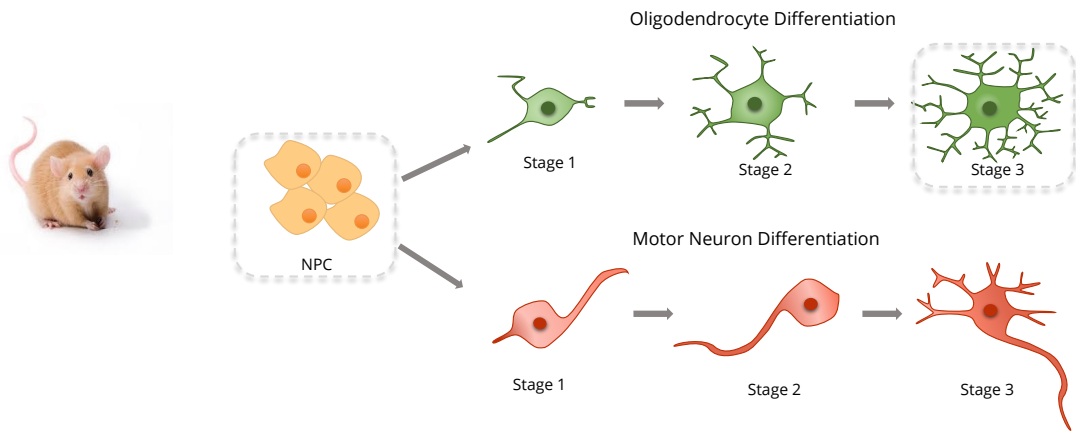
Isoform-resolved functional annotation



Isoform quantification



Dataset for tappAS development

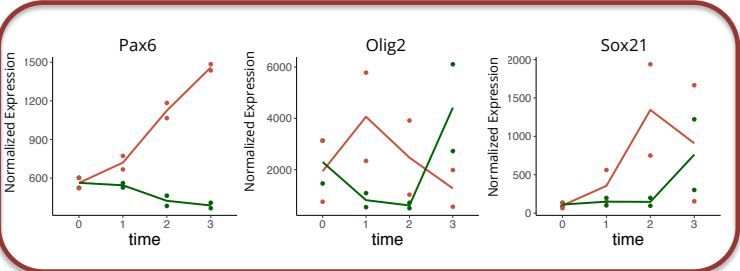
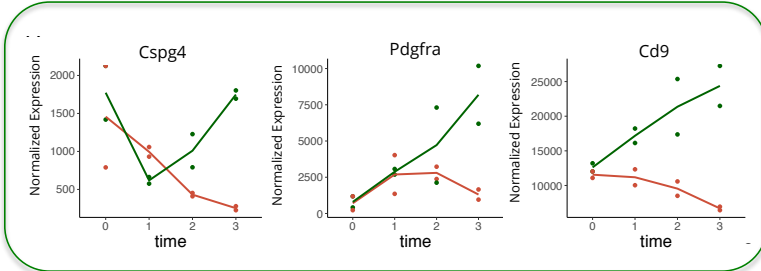


PACBIO®
135 SMRT-cells

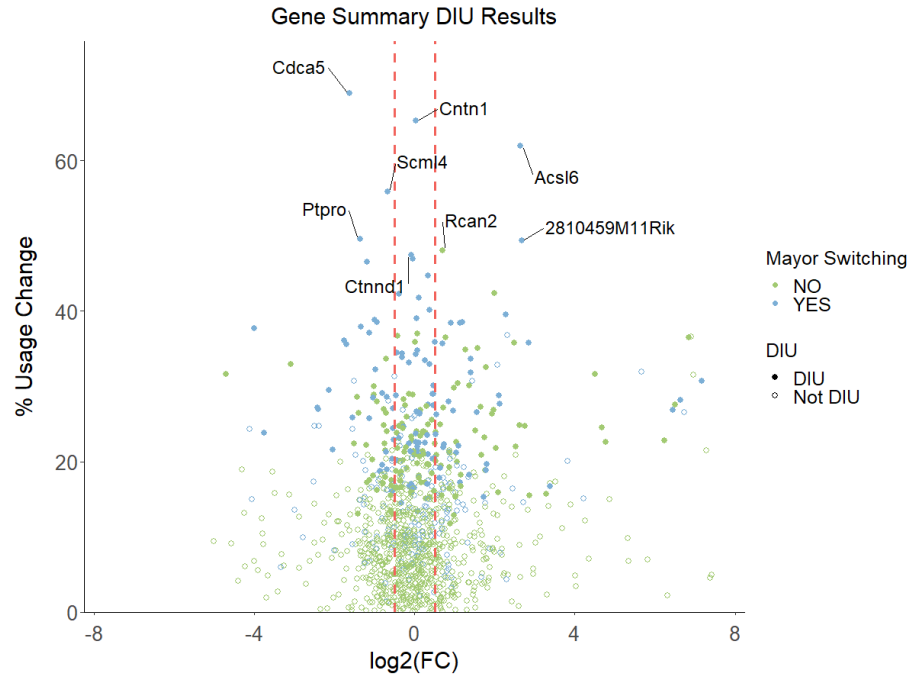
illumina®
60 M reads/sample
2x75 bp

Oligodendrocyte markers

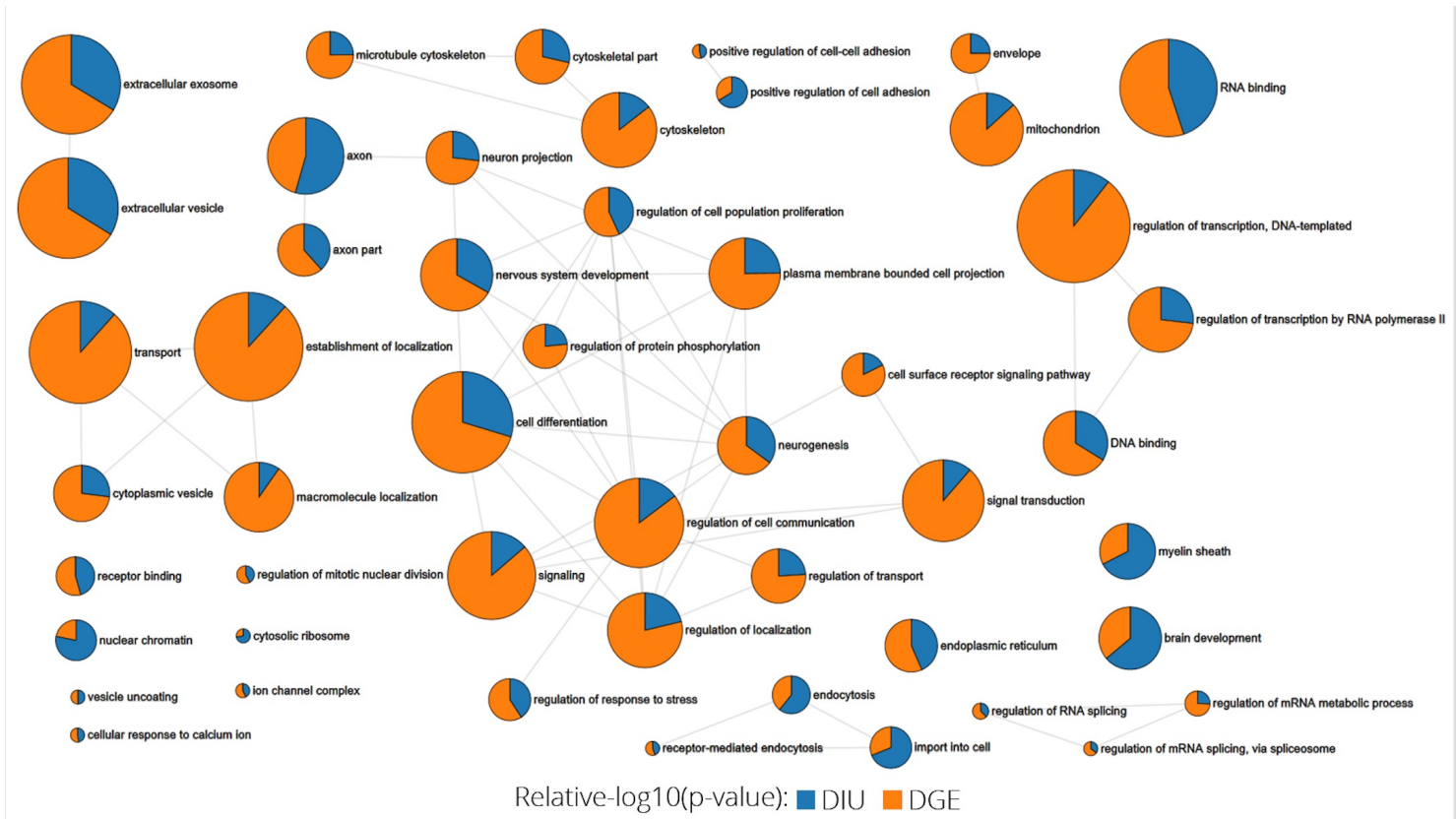
Motoneuron markers



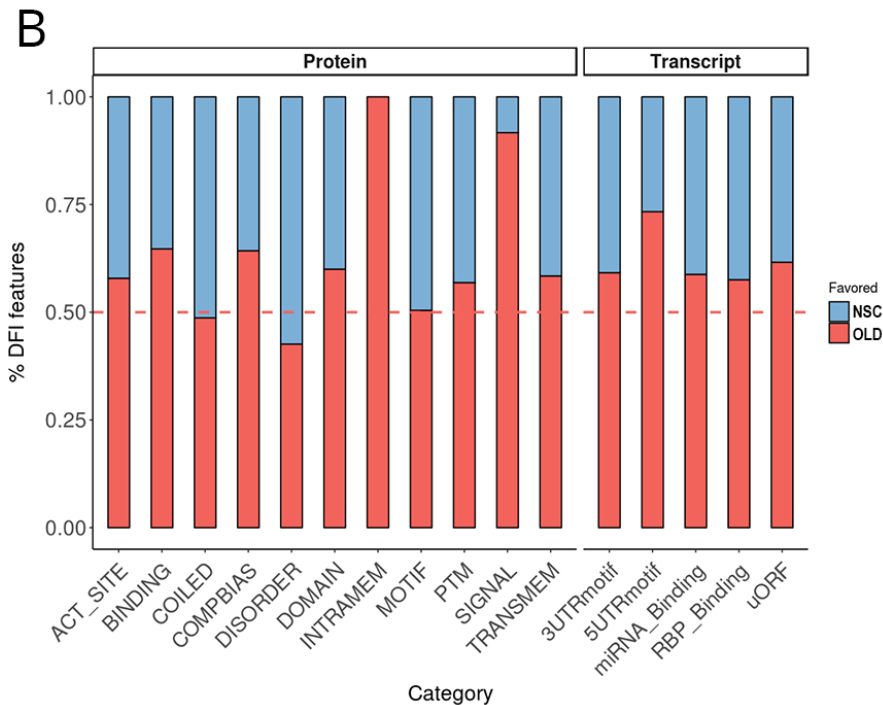
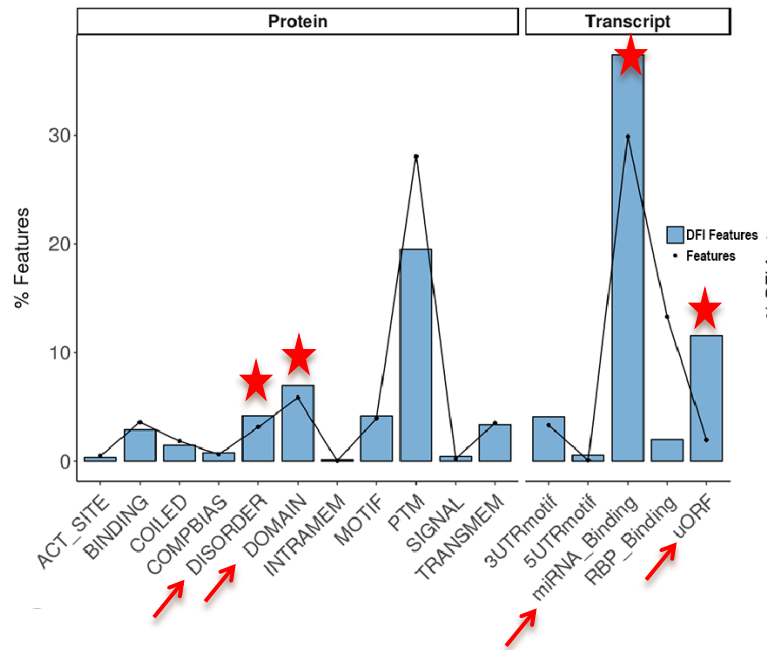
Many genes with AS are not differentially expressed



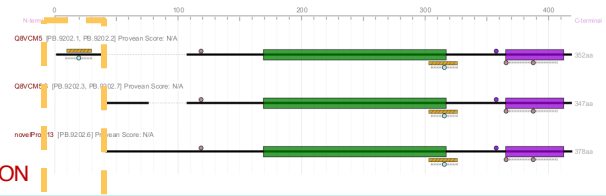
DIU and DGE functional interplay



Distinct motifs are regulated post-transcriptionally

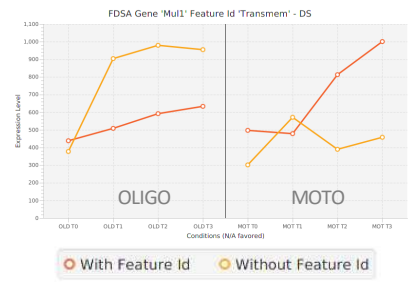


Alternative Splicing of MUL1 at protein level



TRANSMEMBRANE REGION

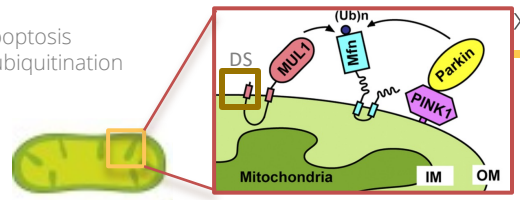
Significant differential Splicing of the TM region across lineages



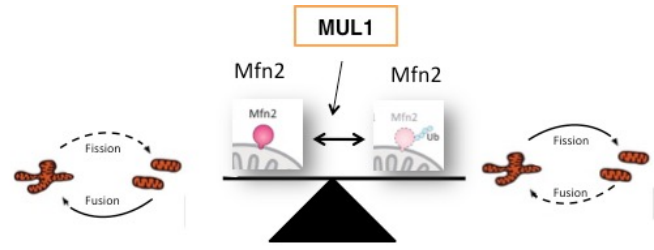
TM DS in Mul1
Podium switch

- Mul1 suppresses mitochondrial fusion through ubiquitination of Mfn proteins.
- Mul1 involves in apoptosis or mitophagy through ubiquitination of AKT

MUL1 function



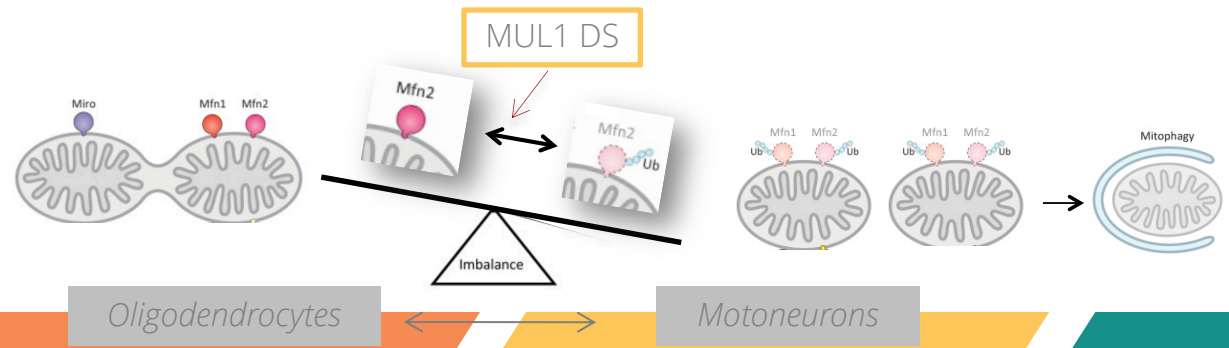
Implication in the system



In mitochondria, ubiquitination controls organellar dynamics as fission and fusion.

Ling et al. Trends in Cell Biology. 2013

Mul1 function regulation by DS of a TM region?

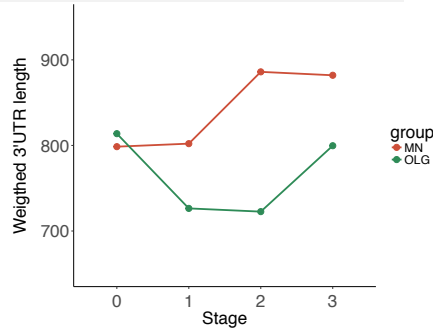


Oligodendrocytes

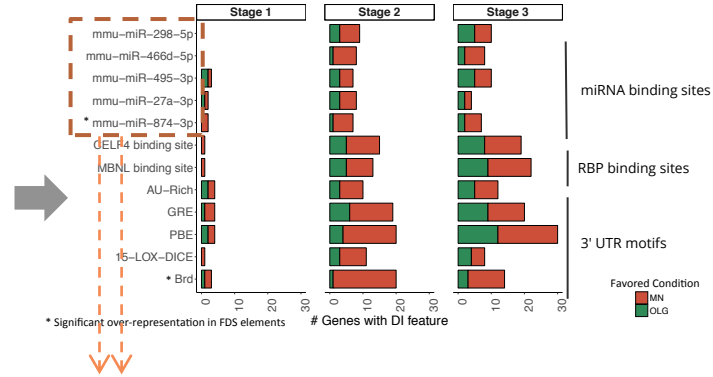
Motoneurons

Differential PolyAdenylation Analysis (DPA)

Analysis of UTR length

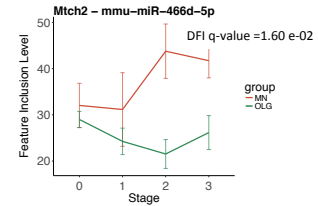
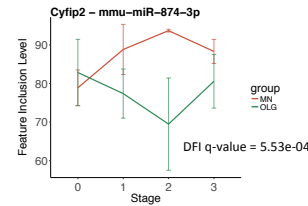
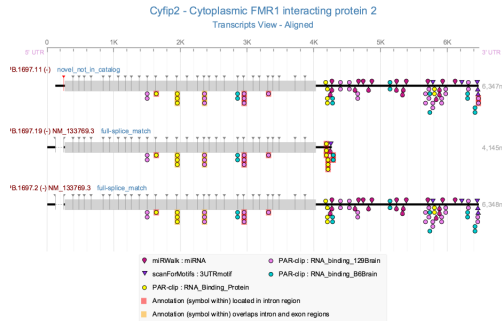


DFI of UTR functional features



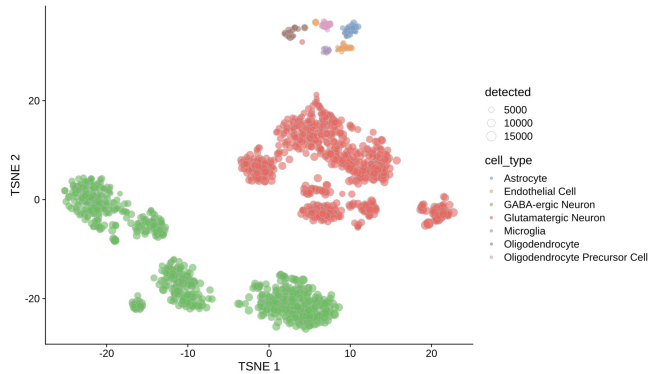
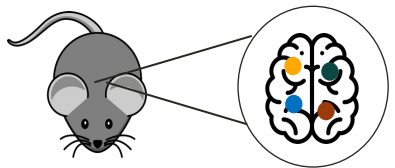
miRNAs involved in axon regeneration

Motti et al. 2017



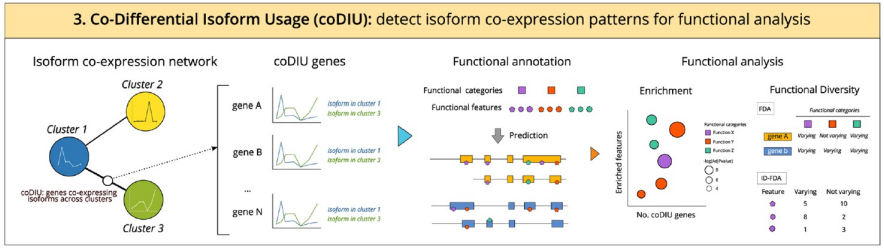
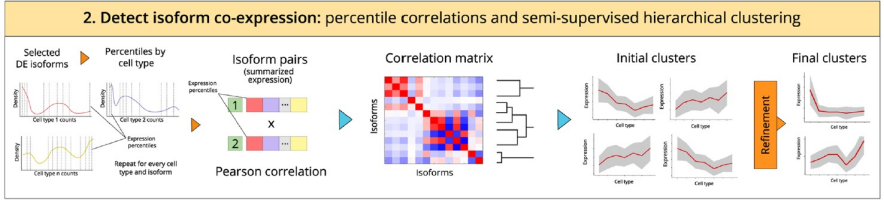
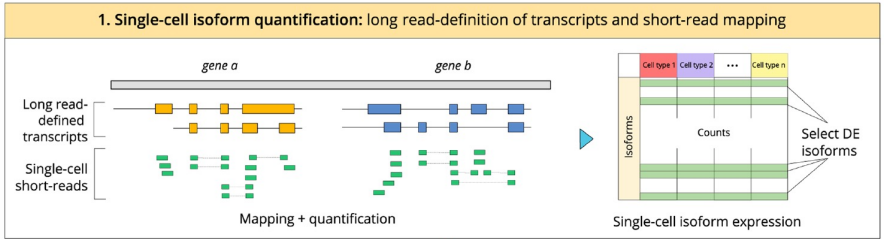


Functional Iso-Transcriptomics at the single-cell level



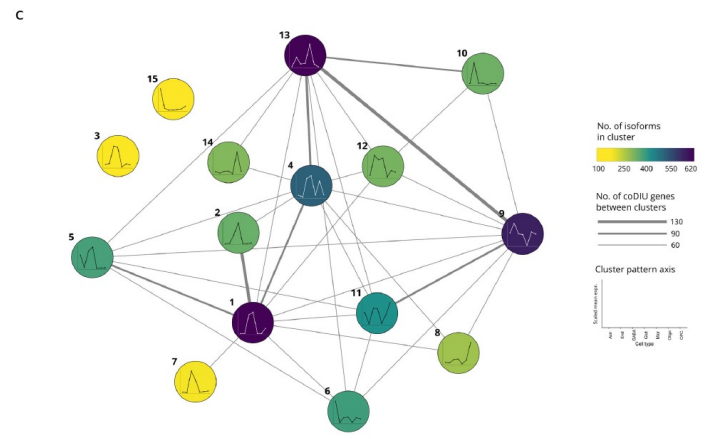
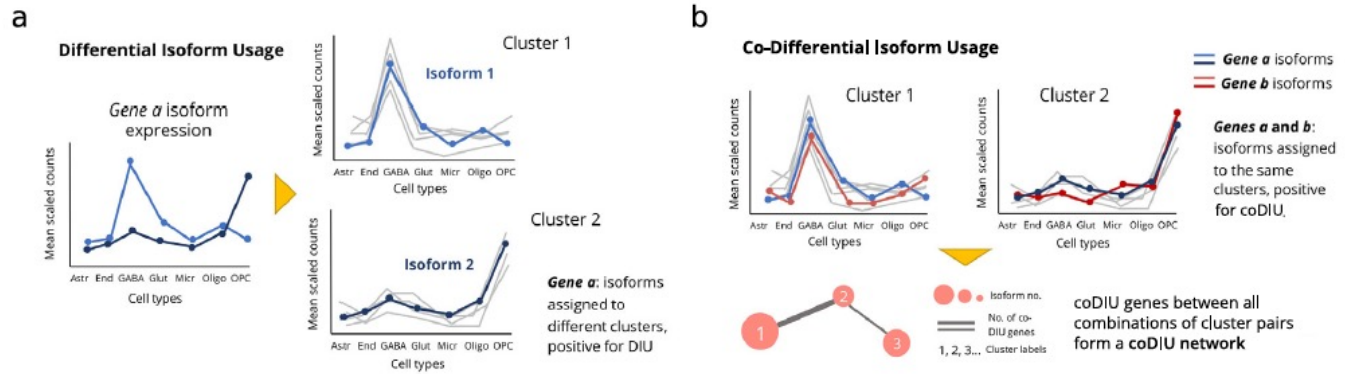
Single-cell
Smart-Seq2,
mouse visual cortex
Tasic et al. 2016

Long reads
ENCODE bulk PacBio
reads from hippocampus
and cortex

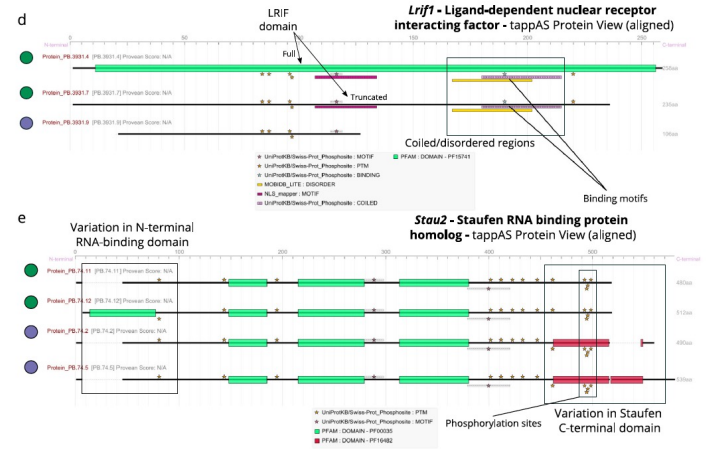
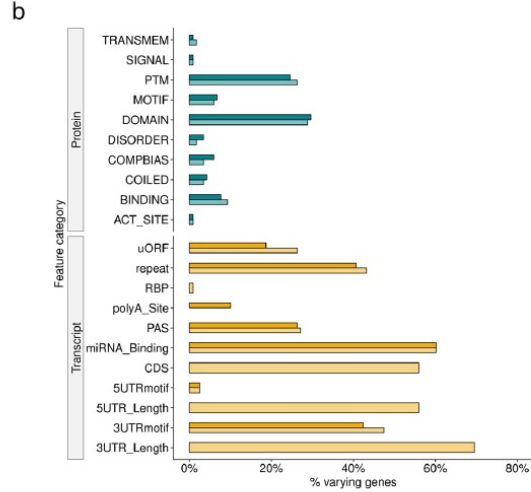
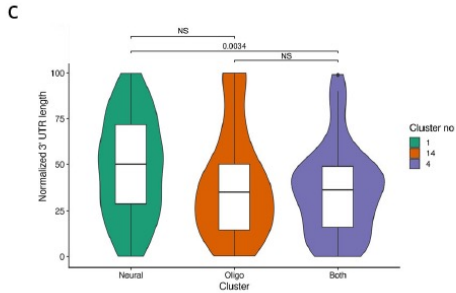
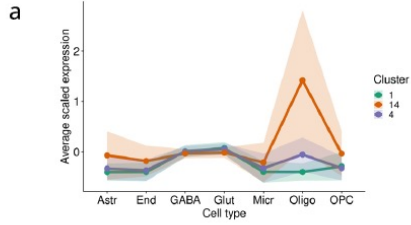


Arzalluz et al., Nature Communications, 2022

Functional Iso-Transcriptomics at the single-cell level



Functional Iso-Transcriptomics at the single-cell level



Concluding remarks

- Third Generation Sequencing technologies are a new and amazing magnification lens in the study the transcriptome but also pose important challenges
 - Identification the true transcriptomes
 - Description and annotation of the complexity
 - Understanding of the functional consequences of this complexity

Acknowledgements

SQANTI Team @ ConesaLab



Lorena de la Fuente



Manu Tardaguila



Hector del Risco



Fran Pardo



Angeles Arzalluz



Pedro Salguero



Jorge Mestre



Tianyuan Liu



Myla Kondratova



Rocio Amarin



Alessandra Martínez

Pacbio

Elizabeth Tseng

University Virginia

Gloria Sheynkman

UC Irvine

Ali Mortazavi

Fairlie Reese

University of Florida

Lauren McIntyre

Adalena Nani

The ConesaLab

