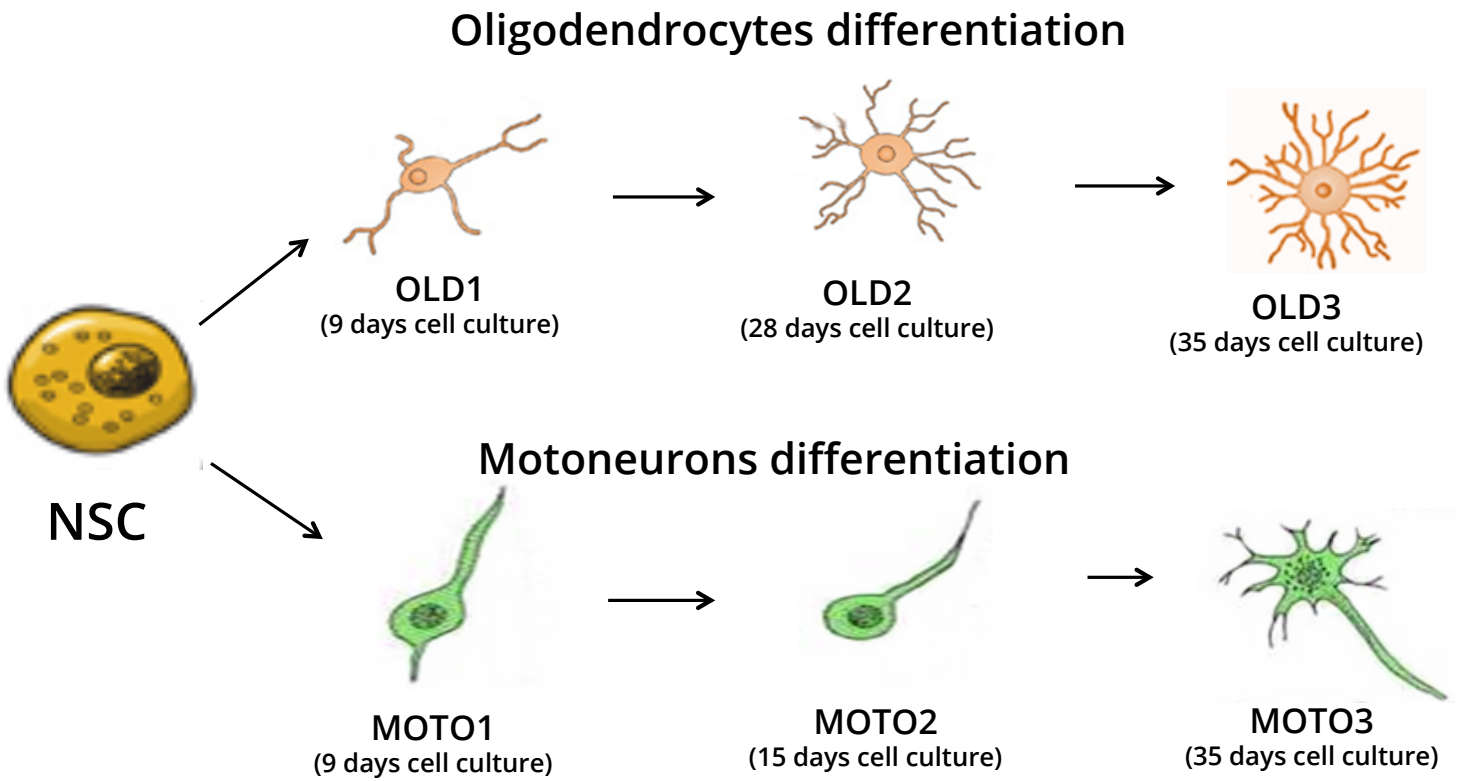
SQANTI and TAPP**AS**: Making Sense of Iso-Seq Data

Ana Conesa, PhD
Genomics of Gene Expression Lab
CIPF/UF

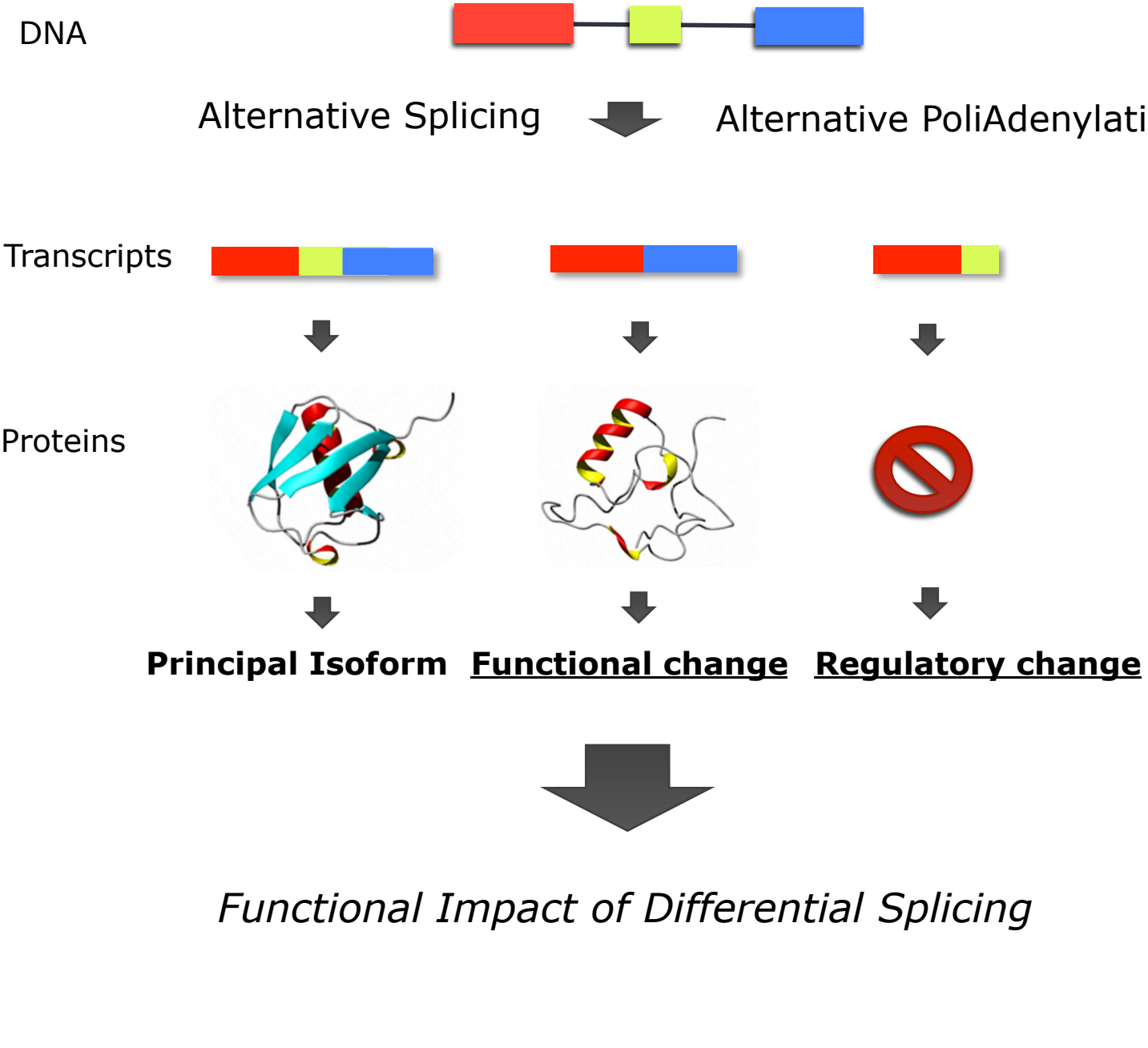
PRINCIPE FELIPE
CENTRO DE INVESTIGACION

UF
UNIVERSITY of FLORIDA

Functional Implications of Differential Splicing?



Functional Implications of Differential Splicing



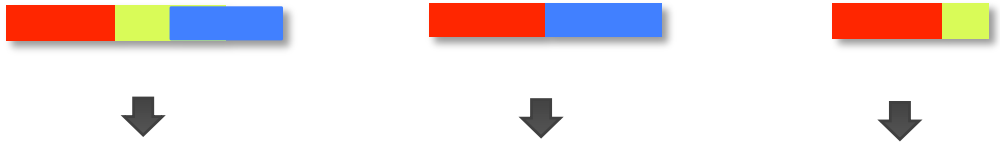
Functional Impact of Differential Splicing

Functional Implications of Differential Splicing

DNA

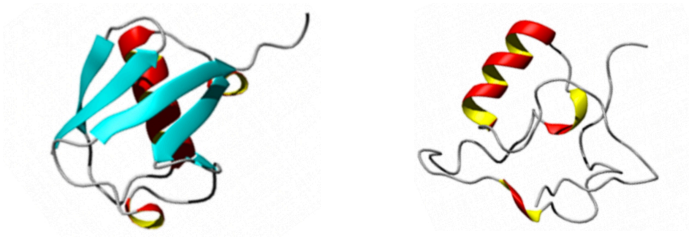
Alternative Splicing → Alternative PoliAdenylation

Transcripts



PACIFIC BIOSCIENCES™

Proteins



Principal Isoform **<u>Functional change</u>** **<u>Regulatory change</u>**

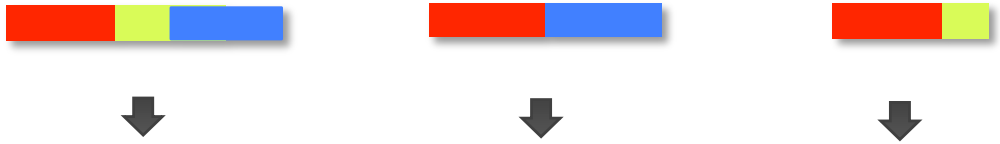
Functional Impact of Differential Splicing

Functional Implications of Differential Splicing

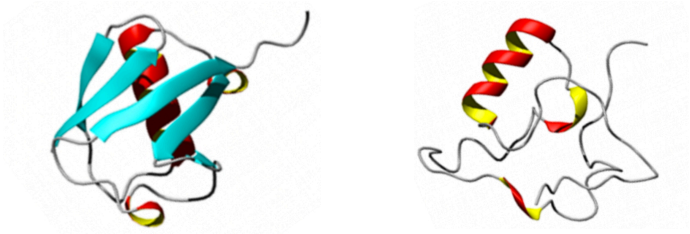
DNA

Alternative Splicing → Alternative PoliAdenylation

Transcripts



Proteins



Principal Isoform **Functional change** **Regulatory change**

Functional Impact of Differential Splicing

Transcriptome definition

SQanti

Functional Annotation of Splicing variants

Iso Annot

AS oriented Statistical Methods

tappAS

Structural and **Q**uality **An**notation of **T**ranscript Isoforms

https://bitbucket.org/ConesaLab/sqanti

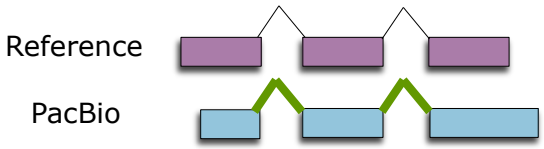
Tardaguila et al. *SQANTI: extensive characterization of long read transcript sequences for quality control in full-length transcriptome identification and quantification.* **Preprint at BiorXiv.** ***2017 Genome Biology, in press.***

1. Classification

Known Isoforms

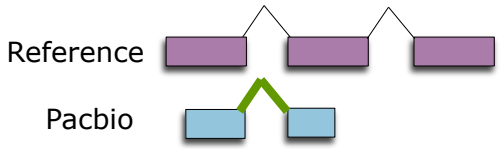
Full-Splice Match

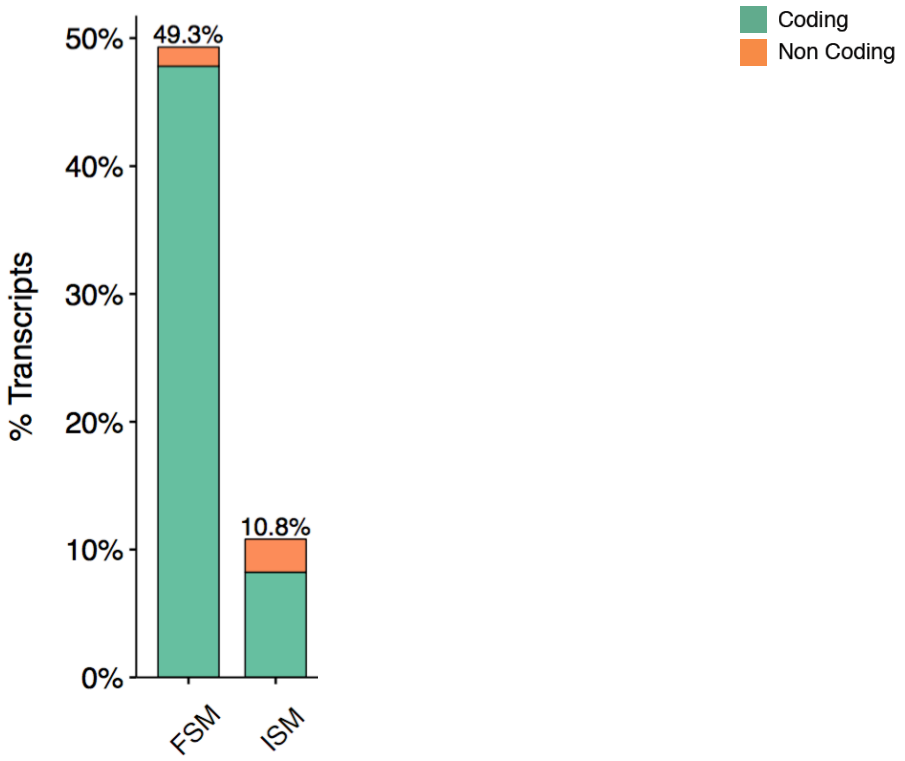
FSM



Incomplete-Splice Match

ISM



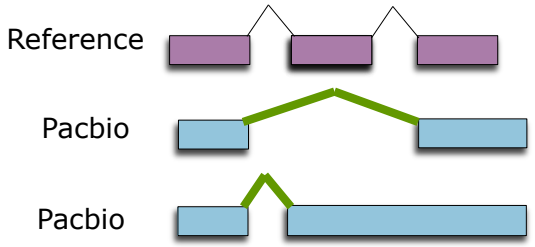


1. Classification

Novel Isoforms – Known genes

Novel In catalog

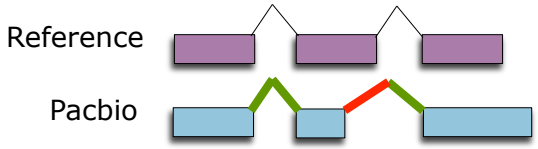
NIC



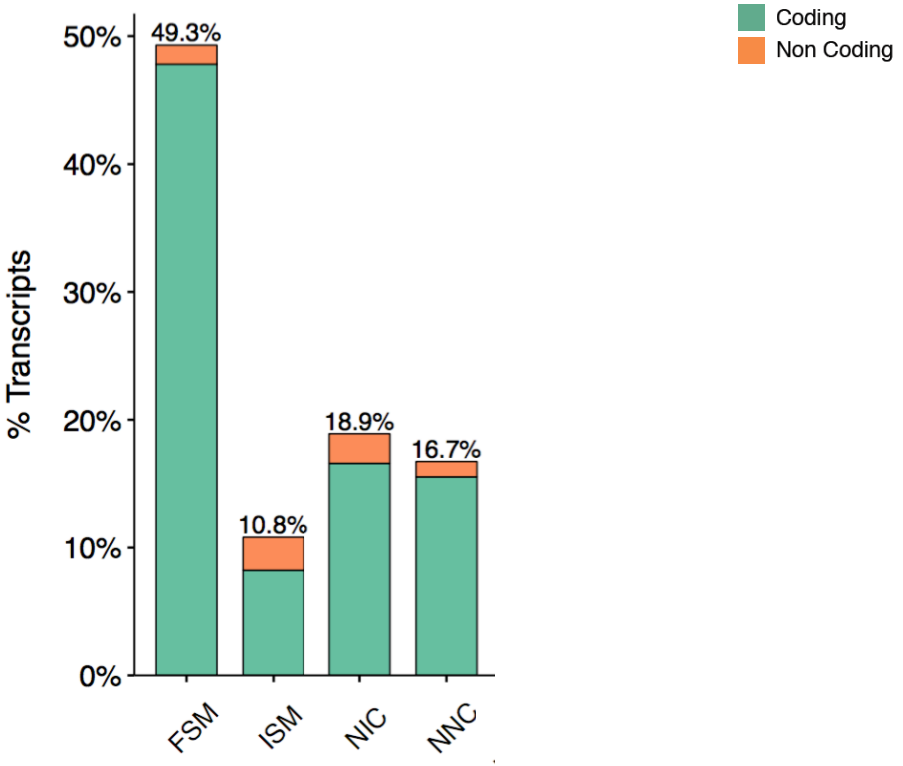
Known donors and acceptors

Novel Not in Catalog

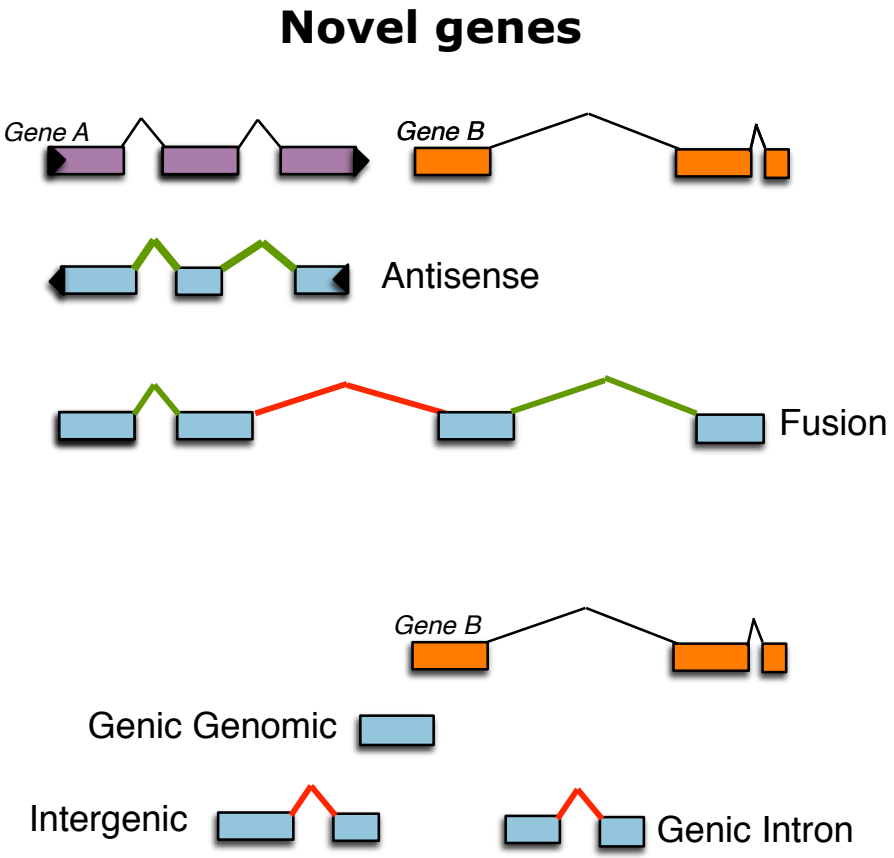
NNC

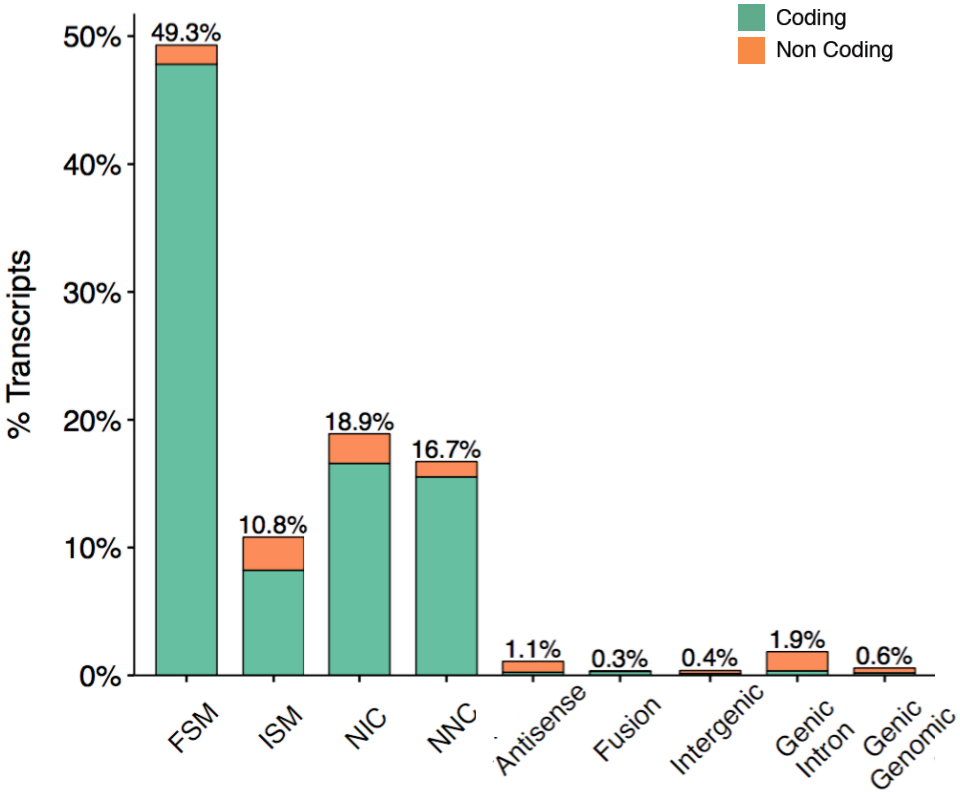


Novel donor or acceptor



1. Classification

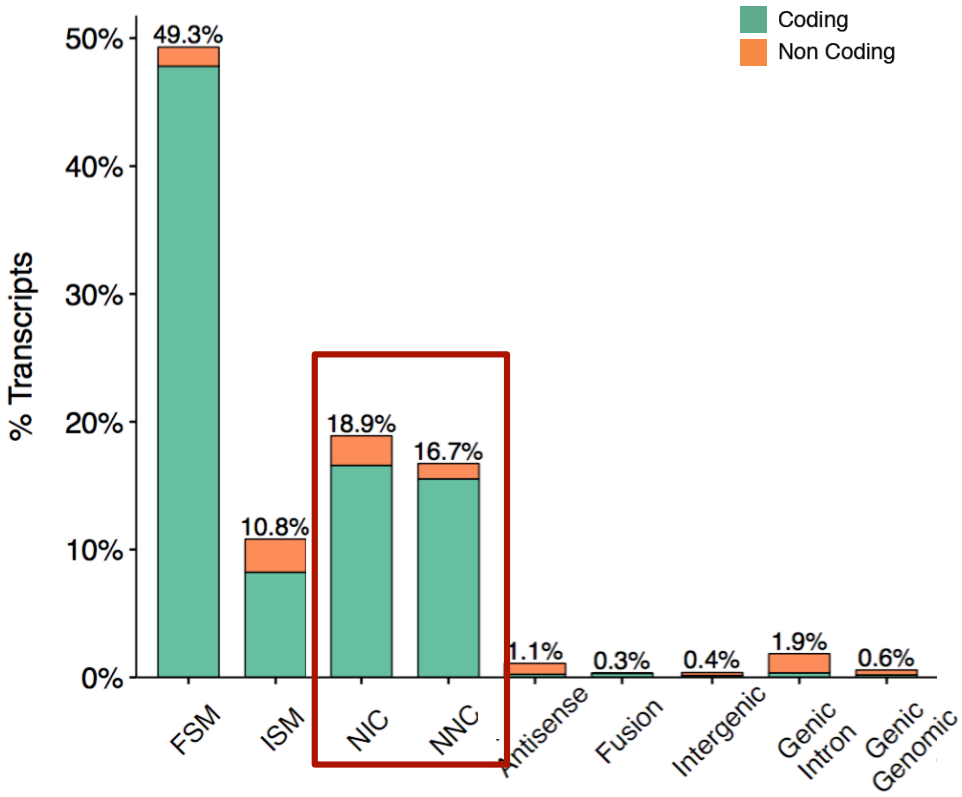




1. Classification

35 % of novel isoforms in mouse…

Are all of them real?



Transcriptome characterization

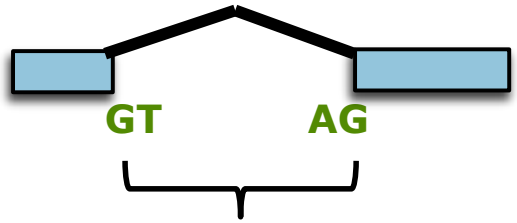
1. Classification

2. QC descriptors

Transcriptome characterization

1. Classification

2. **QC descriptors: SJ canonical status**



Canonical splice junction

≈ 98,7 % of canoncal SJ in mammalian*

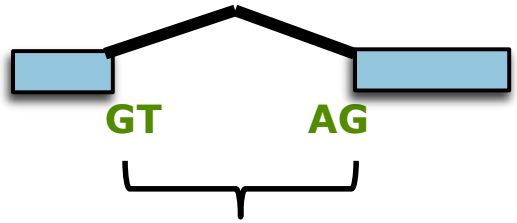
97,7 % of total splice junctions in our neural transcriptome are canonical

*Burset et al, 2000

Transcriptome characterization

1. Classification

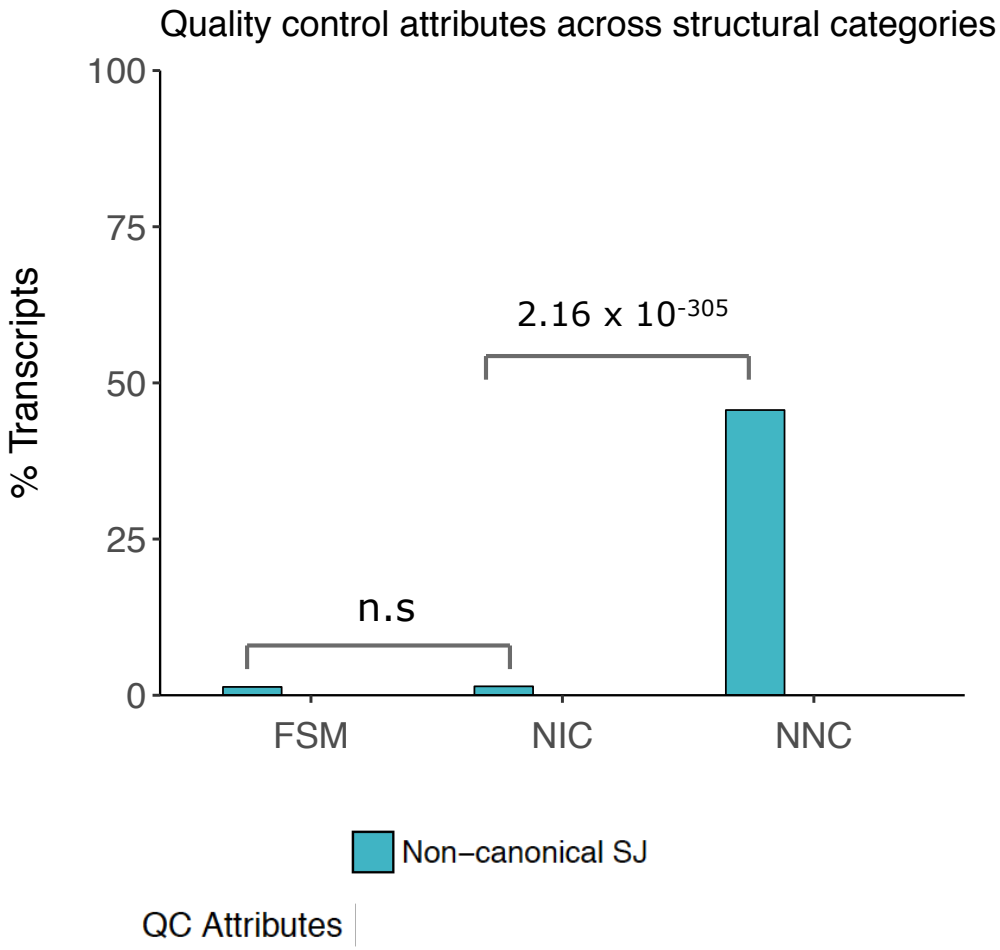
2. **QC descriptors: SJ canonical status**



Canonical splice junction

≈ 98,7 % of canoncal SJ in mammalian*

97,7 % of total splice junctions in our neural transcriptome are canonical

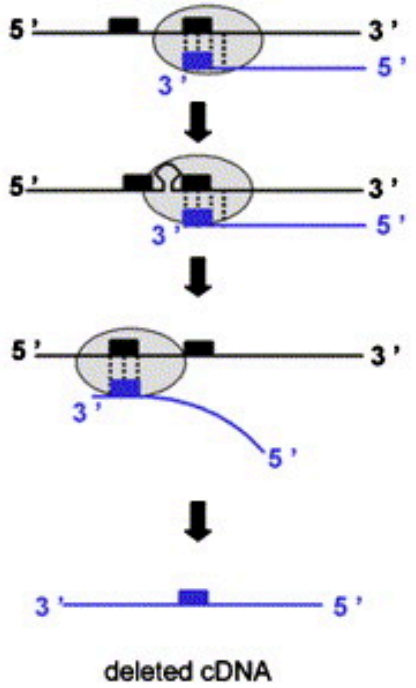


*Burset et al, 2000

Transcriptome characterization

1. Classification

2. QC descriptors: RT-switching

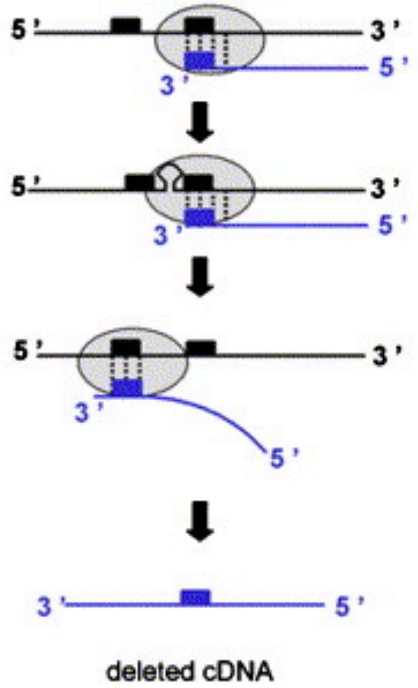


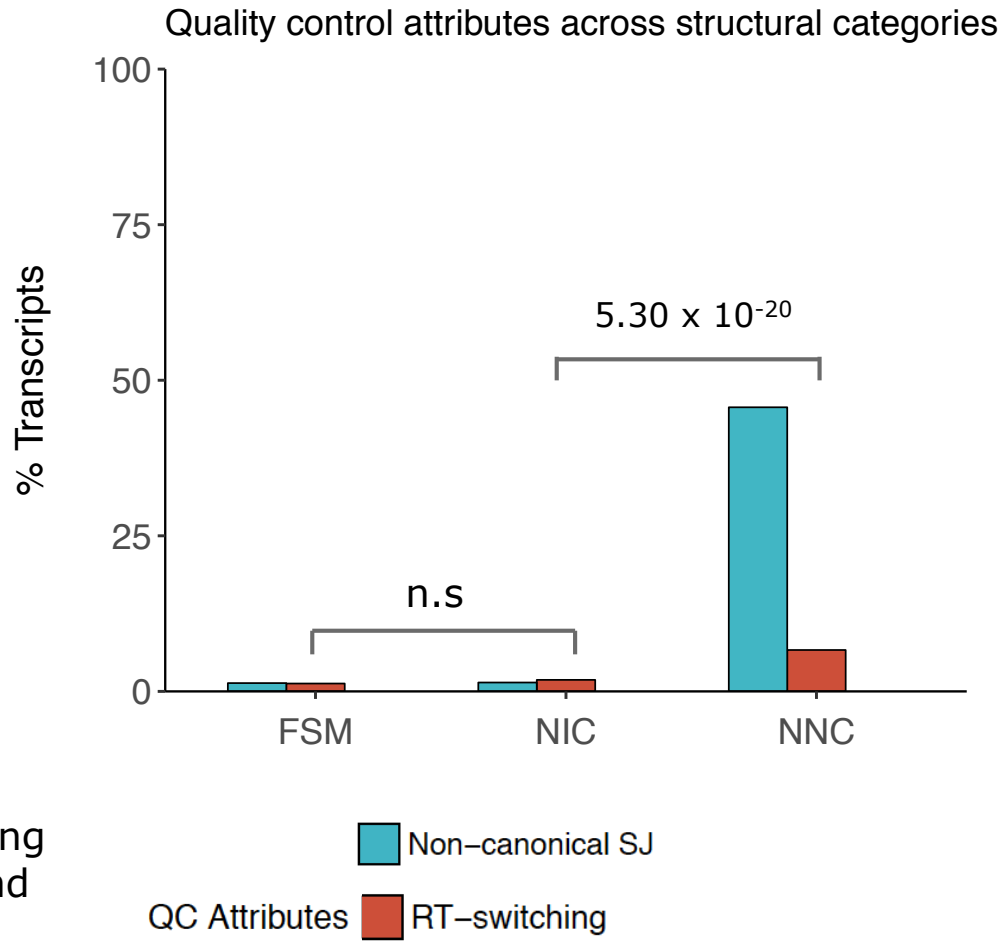
- Reverse transcriptase template switching
- Caused by RNA secondary structure and repeated regions.
- Appears as novel splice junctions

Transcriptome characterization

1. Classification

2. QC descriptors: RT-switching



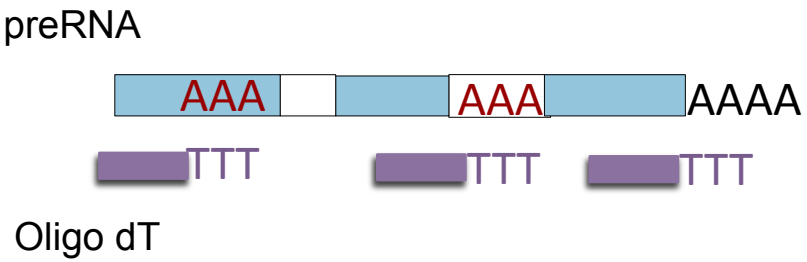


- Reverse transcriptase template switching
- Caused by RNA secondary structure and repeated regions.
- Appears as novel splice junctions

Transcriptome characterization

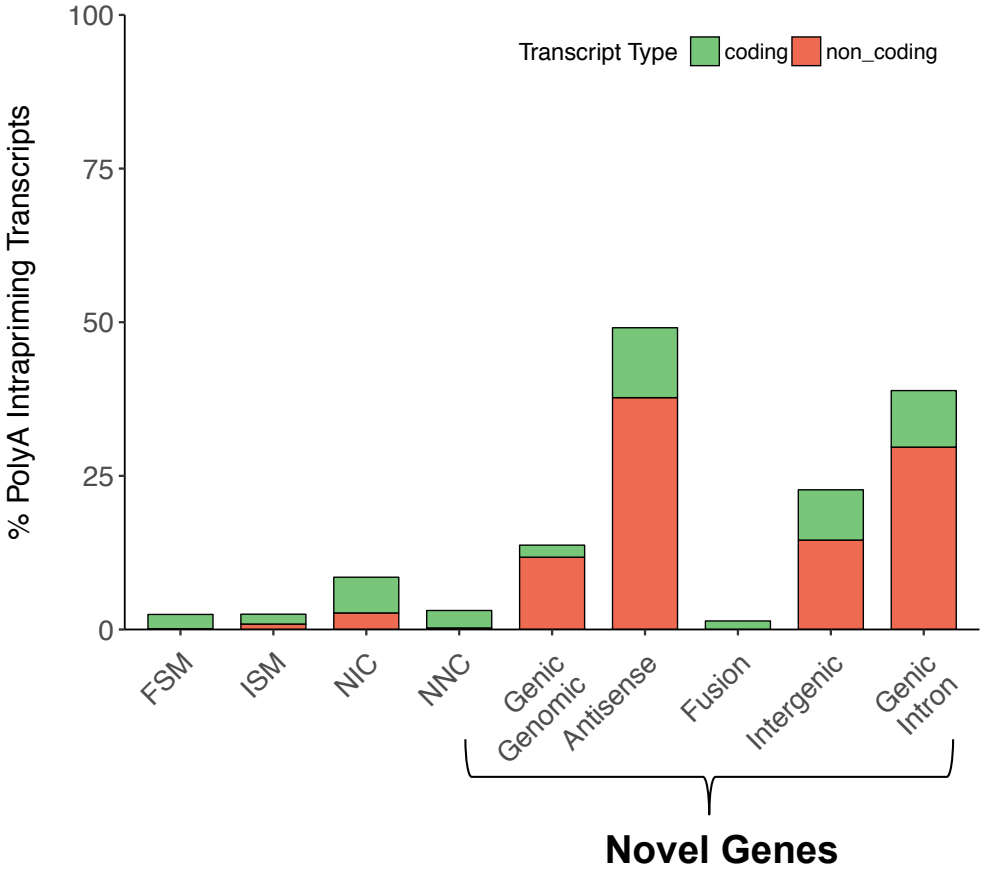
1. Classification

2. **QC descriptors: PolyA intra-priming**



- oligodT can prime outside polyA tail in A rich regions inside transcribed regions.

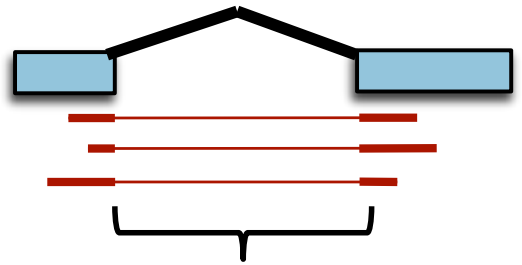
- We looked for transcripts showing >= 80% Adenines in the 20 nts downtream “detected” 3’ end



Transcriptome characterization

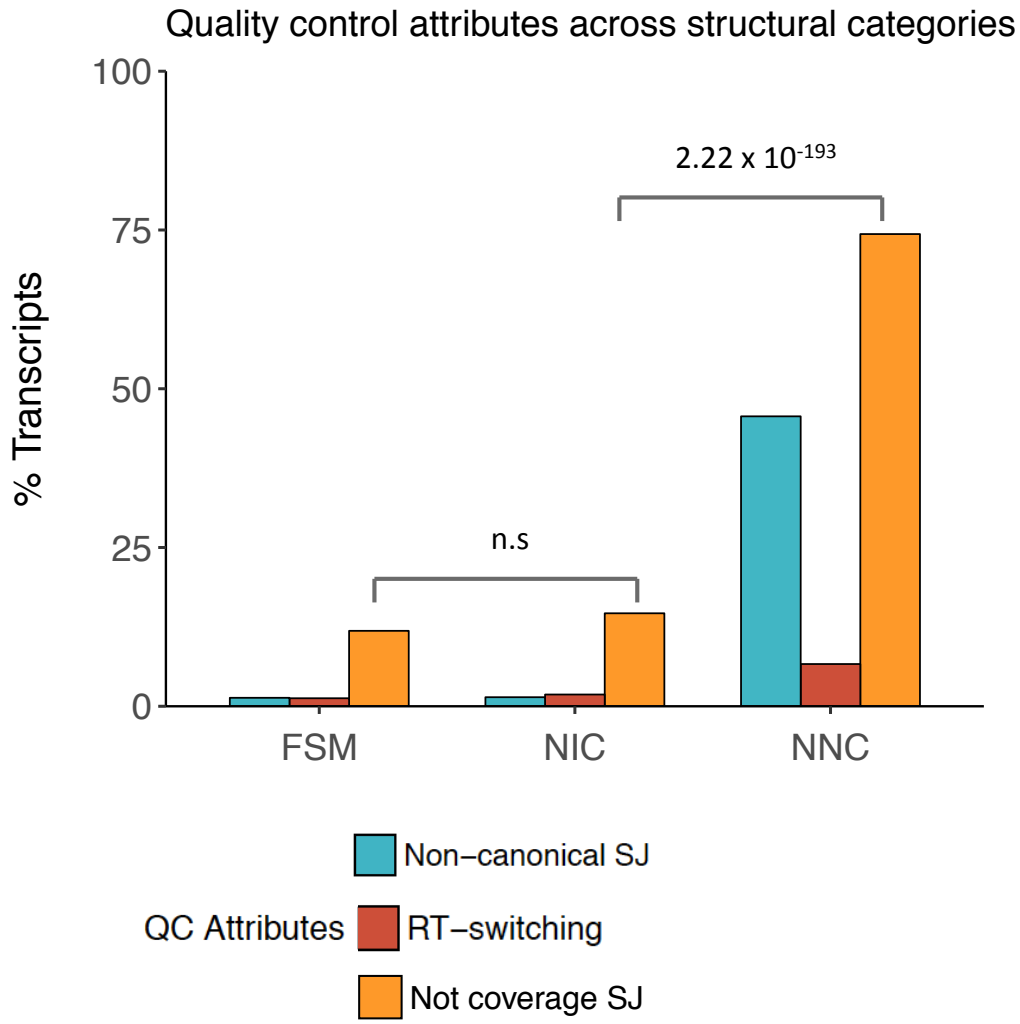
1. Classification

2. **QC descriptors: SJ support**



Supported splice junction

Illumina Reads from same cDNA sequenced by PacBio



Transcript level attributes

1. Transcript Classification
    1. Reference Gene match
    2. Reference Transcript match
    3. Structural Category

2. Structural characteristics
    1. Detected/Reference Length
    2. Detected/Reference number of exons
    3. Distance to nearest annotated TSS
    4. Distance to nearest annotated TTS
    5. Bite

3. Quality Control attributes
    1. RT-switching
    2. PolyA Intrapriming
    3. Canonical status
    4. Indels near SJ

4. Support
    1. Minimum splice junction coverage
    2. Minimum sample coverage
    3. Minimum coverage position
    4.. Number of Full-length reads supporting the transcript

5. Expression levels:
    1. Transcript level
    2. Gene level

6. Coding potential
    1. Coding/non coding
    2. ORF/CDS length
    3. CDS start and end positions

...

Junction level attributes

1. Junction Classification
    1. Novel/Known
    2. Splice site motif

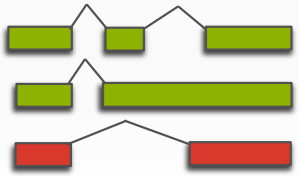
2. Structural characteristics
    1. Diffterence to nearest ref. donor
    2. Diffterence to nearest ref. acceptor
    3. Bite

3. Quality Control attributes
    1. Canonical
    2. Rts_junction
    3. Indel near junc

4. Support
    1. Samples with cov
    2. Total coverage
    3. Coverage per sample

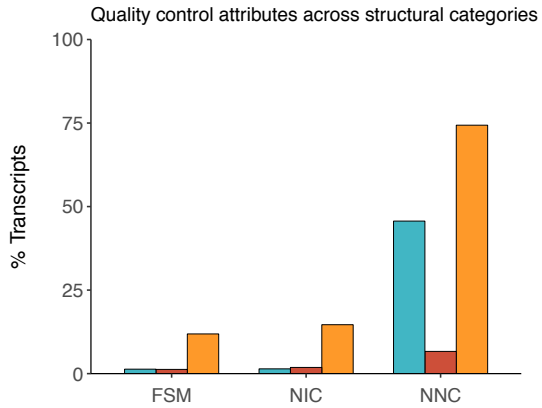
Filtering out artifact isoforms

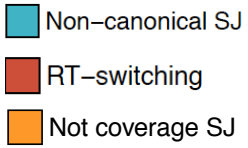
PacBio Transcriptome

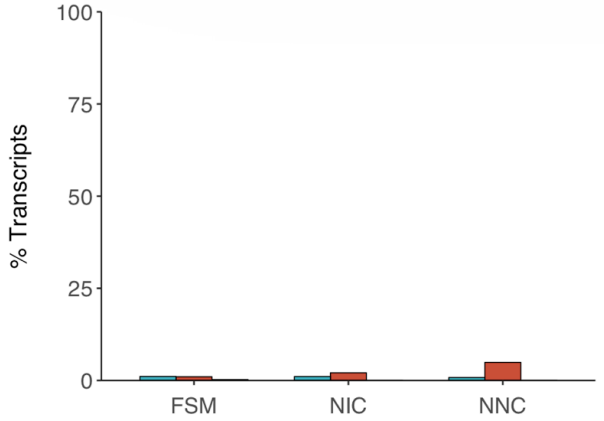


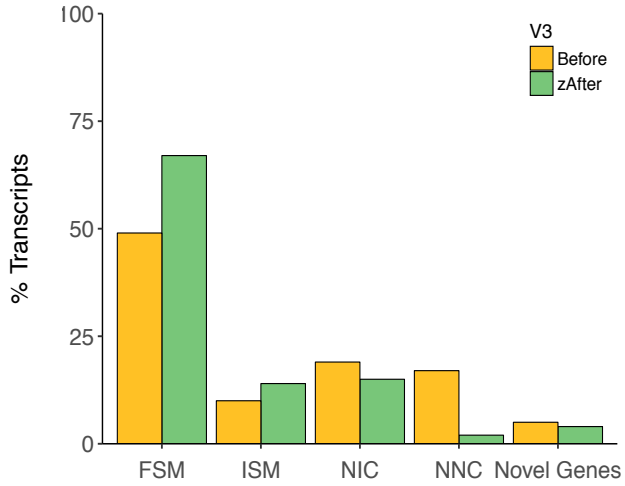
Machine learning with SQANTI descriptors

Curated PacBio Transcriptome









Functional Annotation of Splicing Variants

Transcriptome definition

Functional Annotation of Splicing variants

Functional Profiling at isoform level

SQANTI

Iso Annot

tappAS

Splicing Variants (known and novel)

Sequenced-based predictors

Coordinate-based transference algorithms

TMHMM
SignalP
HMMER-Pfam

UTRscan

RepeatMasker

NMD algorithm

Uniprot

Phospho sitePlus

TargetScan

CLIPdb

PFAM Domains

NMD prediction

UTR motifs

Transmembranes regions

Repeats

Signal Peptides

AU rich regions

Post-translational modifications

Motifs

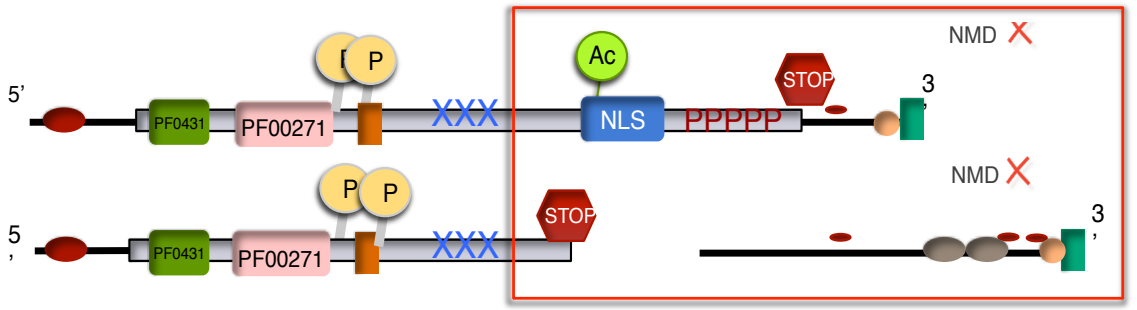
Binding
CA-bind
Crosslink
Np-bind

miRNA binding

Active Sites

Compositional bias

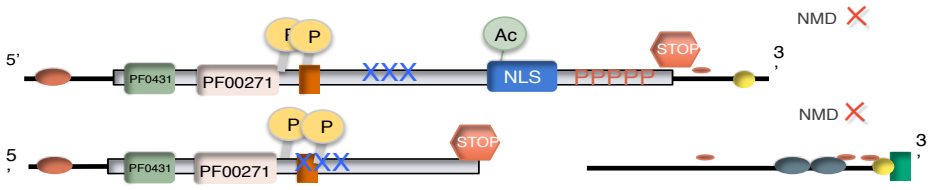
RNA binding proteins



Domain and Motif annotation at Isoform Resolution

Functional Profiling at Isoform Level

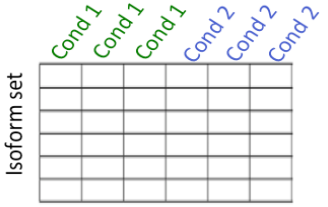
Structural Annotation and Functional Annotation



Reference Annotation provided

OPTIONAL: User-defined

Isoform quantification



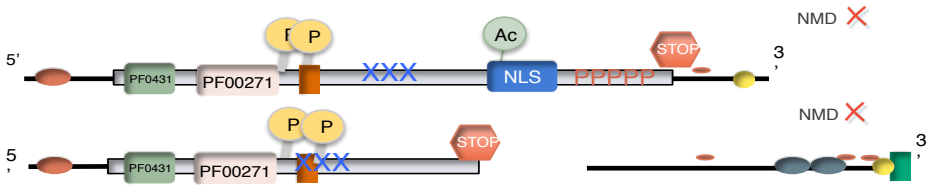
User-defined

INPUT

tappAS

Functional Profiling at Isoform Level

Structural Annotation and Functional Annotation

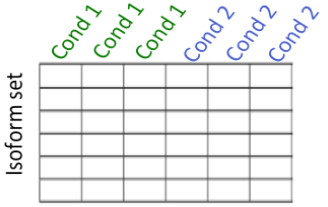


Reference Annotation provided

OPTIONAL: User-defined

tappAS

Isoform quantification

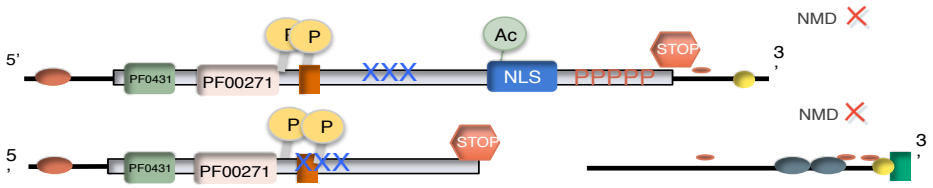


User-defined

INPUT

Functional Profiling at Isoform Level

Structural Annotation and Functional Annotation



Reference Annotation provided

OPTIONAL: User-defined

Module 1

Visualization Interface

Module 2

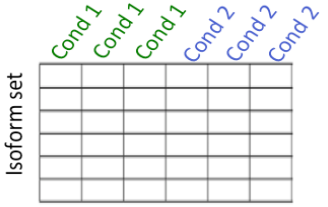
Motif Diversity

CDS/UTRs Diversity

Functional Diversity Analysis (FDA)

tappAS

Isoform quantification

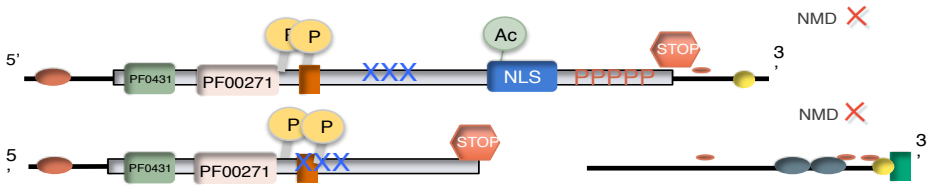


User-defined

INPUT

Functional Profiling at Isoform Level

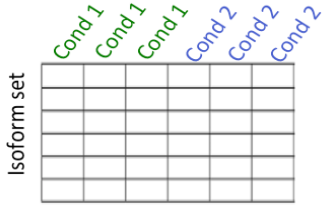
Structural Annotation and Functional Annotation



Reference Annotation provided

OPTIONAL: User-defined

Isoform quantification

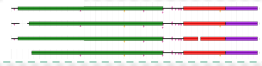


INPUT

User-defined

Module 1

Visualization Interface



Module 2

Motif Diversity

CDS/UTRs Diversity

Functional Diversity Analysis (FDA)

tappAS

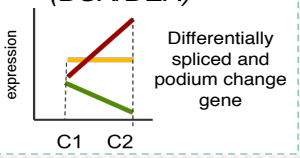
Module 3

Differentially Spliced genes

Mayor Isoform Switching

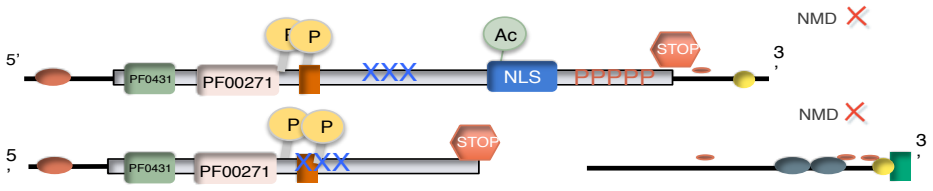
Differentially expressed genes/transcripts/ORFs

Differential Analysis (DSA/DEA)



Functional Profiling at Isoform Level

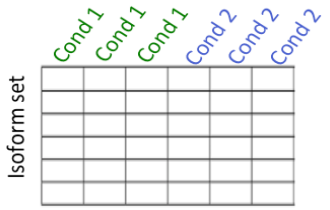
Structural Annotation and Functional Annotation



Reference Annotation provided

OPTIONAL: User-defined

Isoform quantification

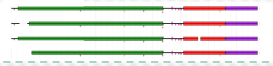


INPUT

User-defined

Module 1

Visualization Interface

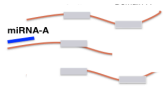


Module 2

Motif Diversity

CDS/UTRs Diversity

Functional Diversity Analysis (FDA)



tappAS

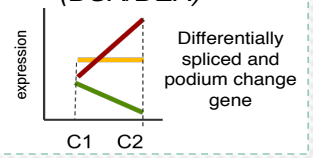
Module 3

Differentially Spliced genes

Mayor Isoform Switching

Differentially expressed genes/transcripts/ORFs

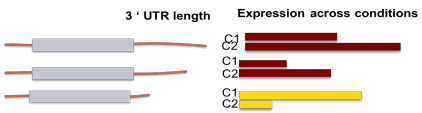
Differential Analysis (DSA/DEA)



Integrative methods

Module 4

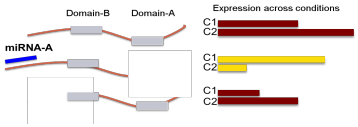
UTR length regulation



UTR Differential Splicing (UtrDS)

Regulation of annotated elements

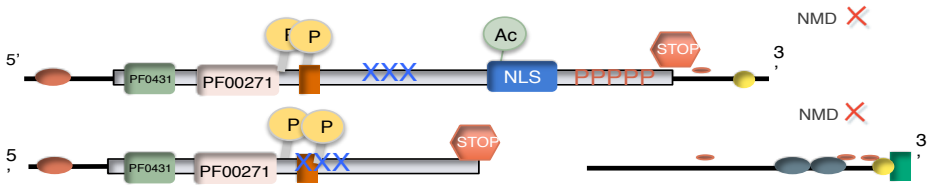
Motifs,Domains, miRNA binding sites



Motif Differential Splicing (mDS)

Functional Profiling at Isoform Level

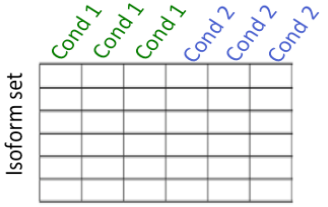
Structural Annotation and Functional Annotation



Reference Annotation provided

OPTIONAL: User-defined

Isoform quantification

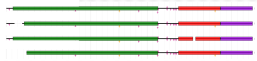


INPUT

User-defined

Module 1

Visualization Interface

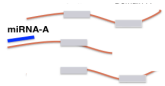


Module 2

Motif Diversity

CDS/UTRs Diversity

Functional Diversity Analysis (FDA)



tappAS

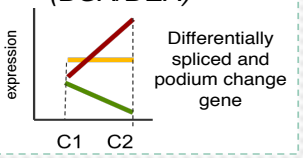
Module 3

Differentially Spliced genes

Mayor Isoform Switching

Differentially expressed genes/transcripts/ORFs

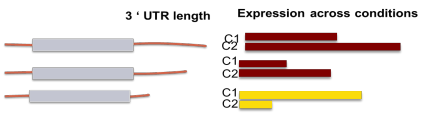
Differential Analysis (DSA/DEA)



Integrative methods

Module 4

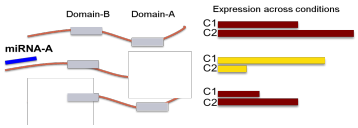
UTR length regulation



UTR Differential Splicing (UtrDS)

Regulation of annotated elements

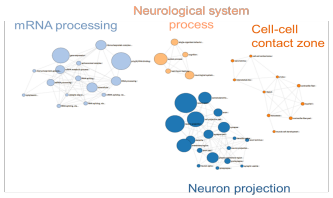
Motifs,Domains, miRNA binding sites



Motif Differential Splicing (mDS)

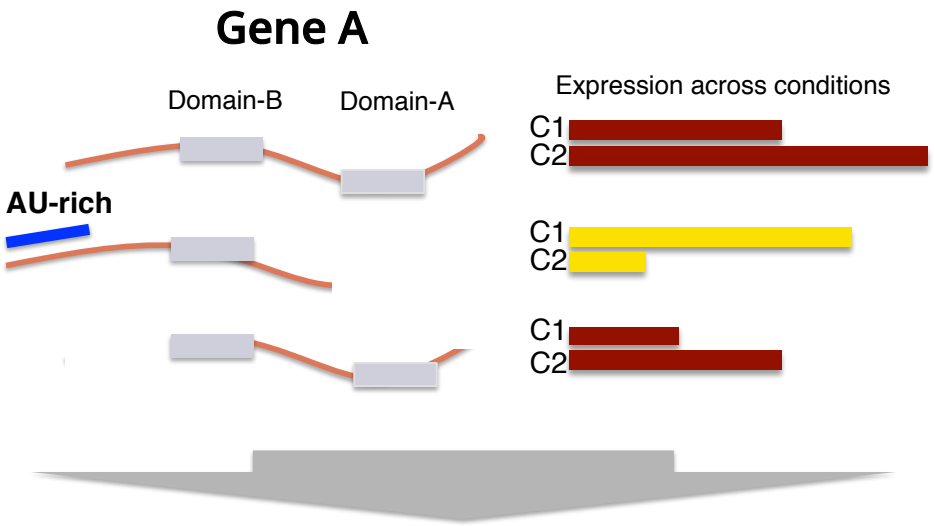
Module 5

Functional enrichment over any annotated category

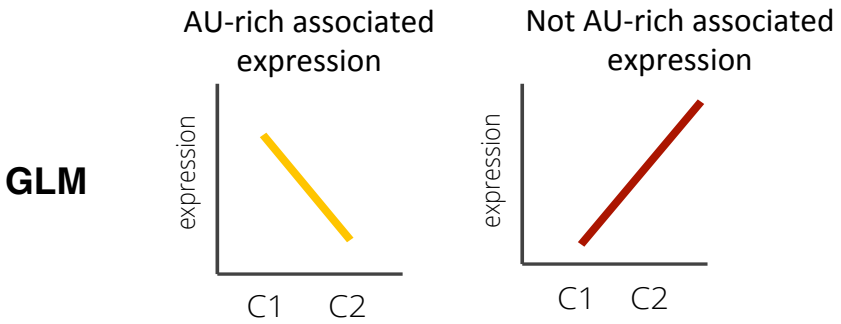


Functional Enrichment and Gene Set Analysis (FEA/GSA)

Motif and Feature Differential Splicing



Significant differential usage of AU-rich motif in Gene A?



DS AU-rich region

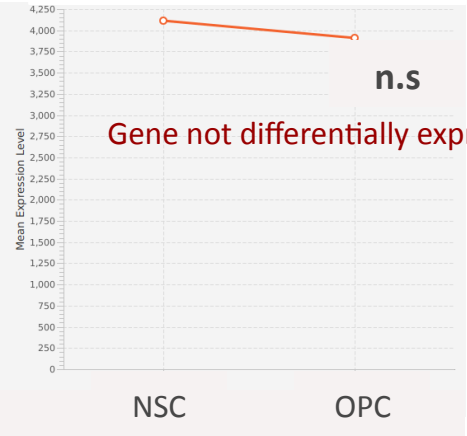
AU-rich element favored in condition 1 by DS

Ctnnd1 – p120 – catenin delta 1

Involved in cell-cell adhesion and signal transduction

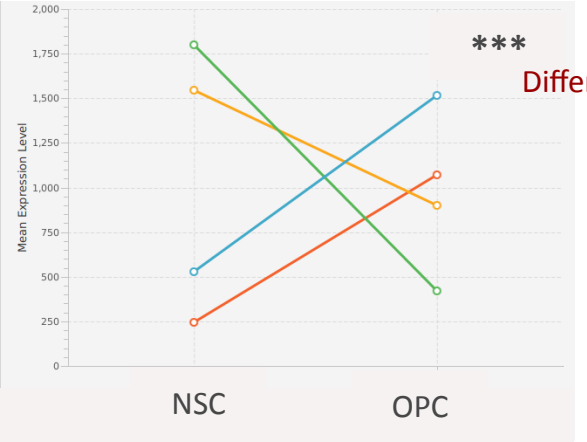
Regulation of protein motifs by differential splicing

Gene Expression



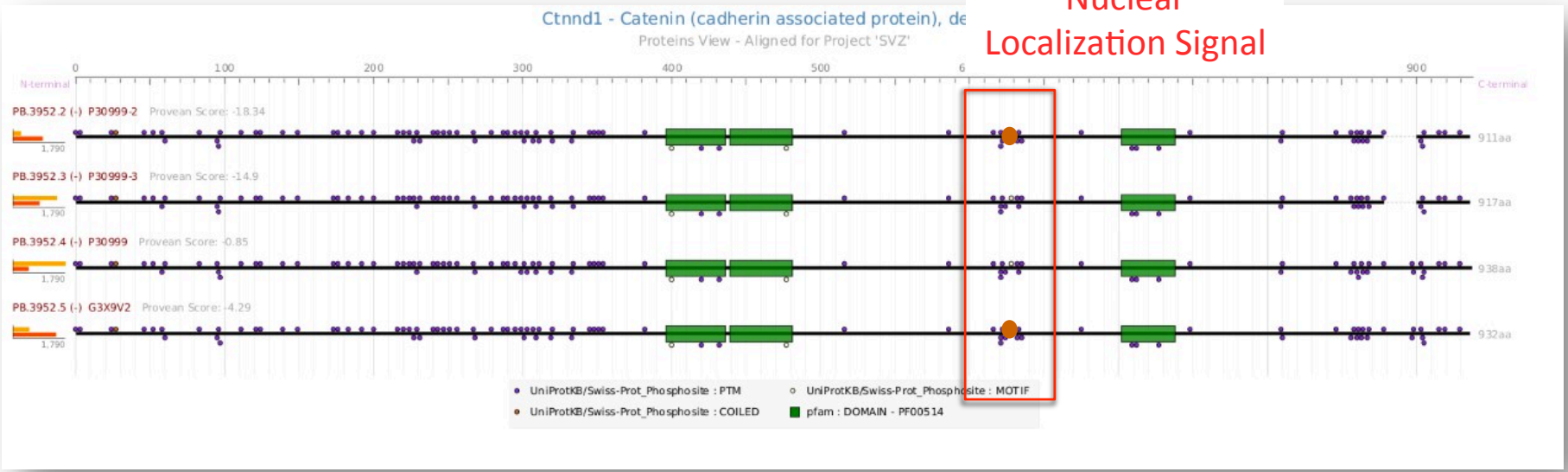
Gene not differentially expressed

Transcripts Expression



Differentially Spliced Gene

Functional impact?

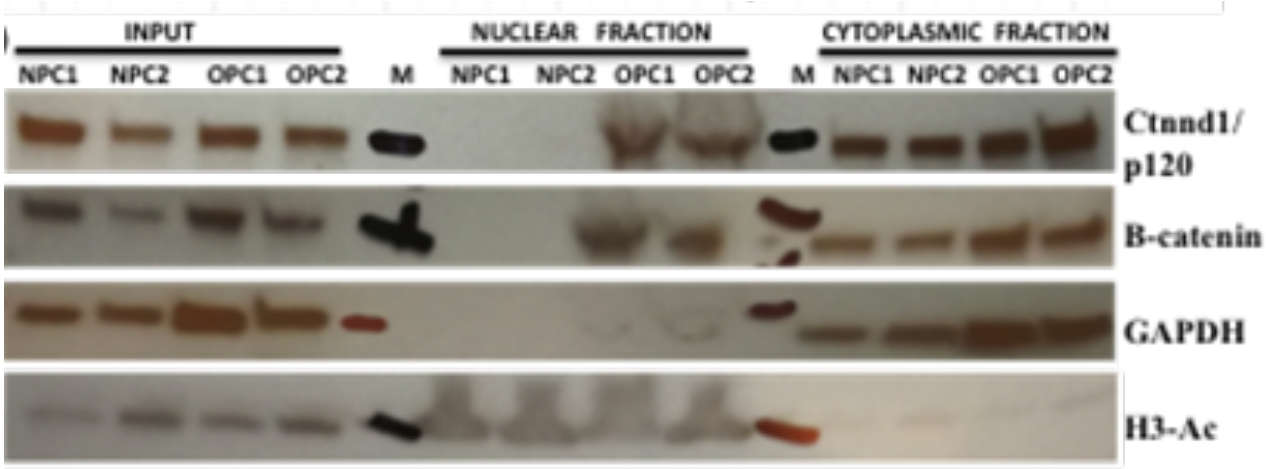


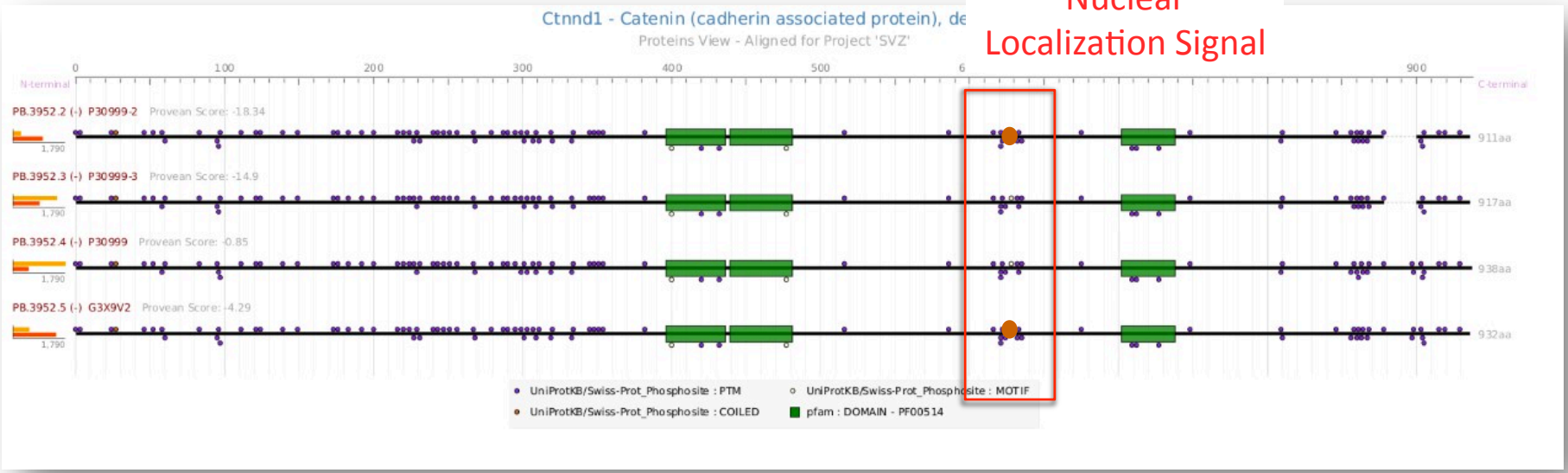
Ctnnd1 – p120 – catenin delta 1

Involved in cell-cell adhesion and signal transduction

Regulation of protein motifs by differential splicing

Experimental validation



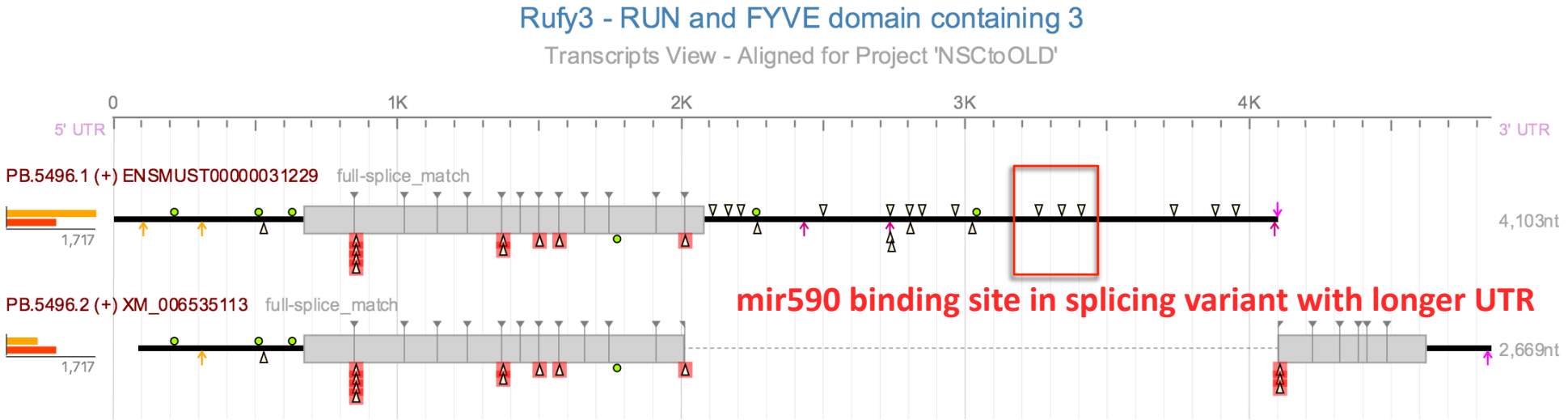


Rufy3

Generation of neuronal polarity formation and axon growth

Regulation of UTR motifs by differential splicing

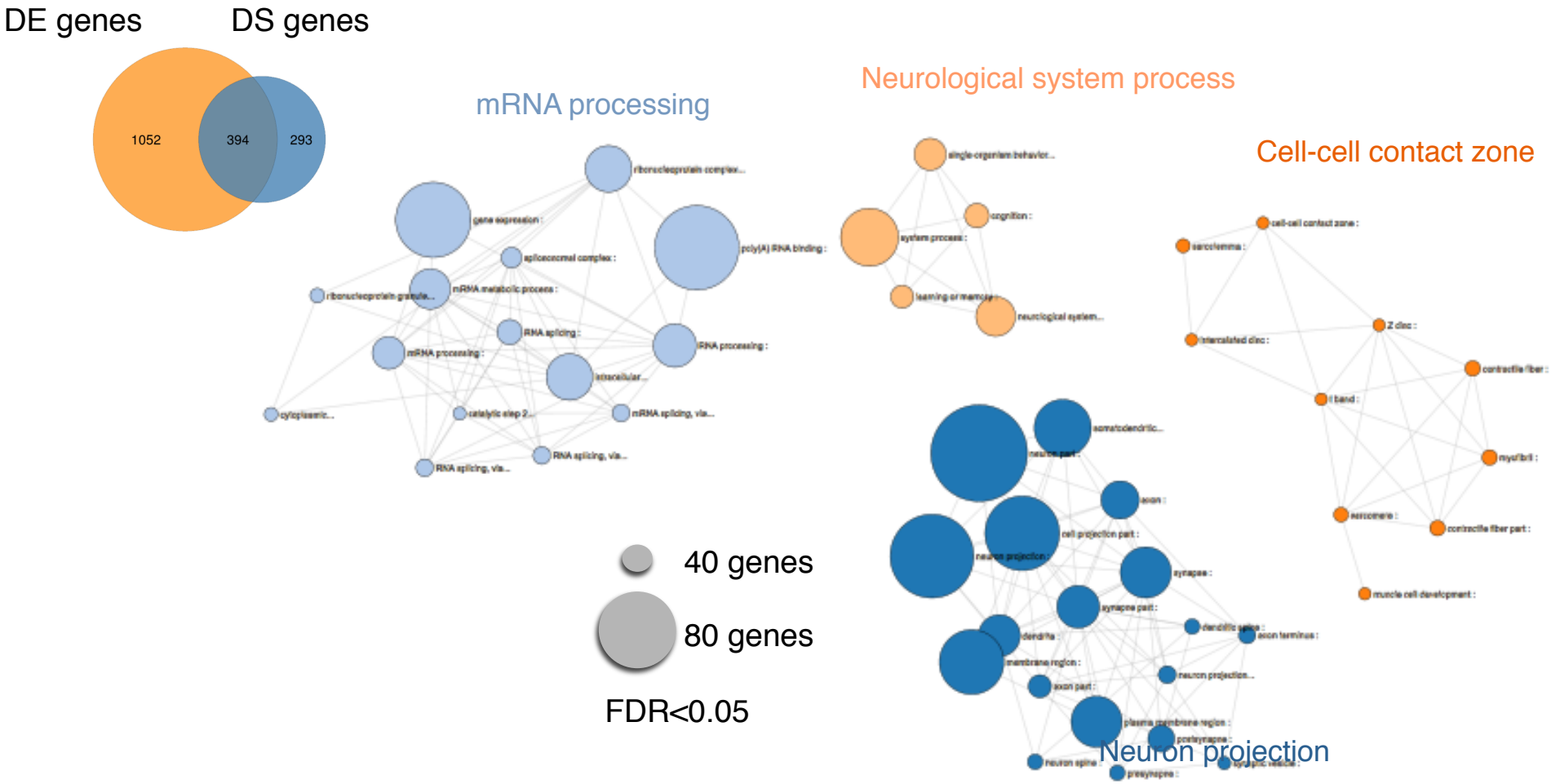
#	Gene	Feature	Feature Id	Position	FDSA Result	Q-Value	Favored Condi...	PodiumChg	TotalChg
1	Rufy3	miRNA	mmu-miR-590-3p	T88641599-88...	DS	5.2611E-9	NSC	NO	48.79



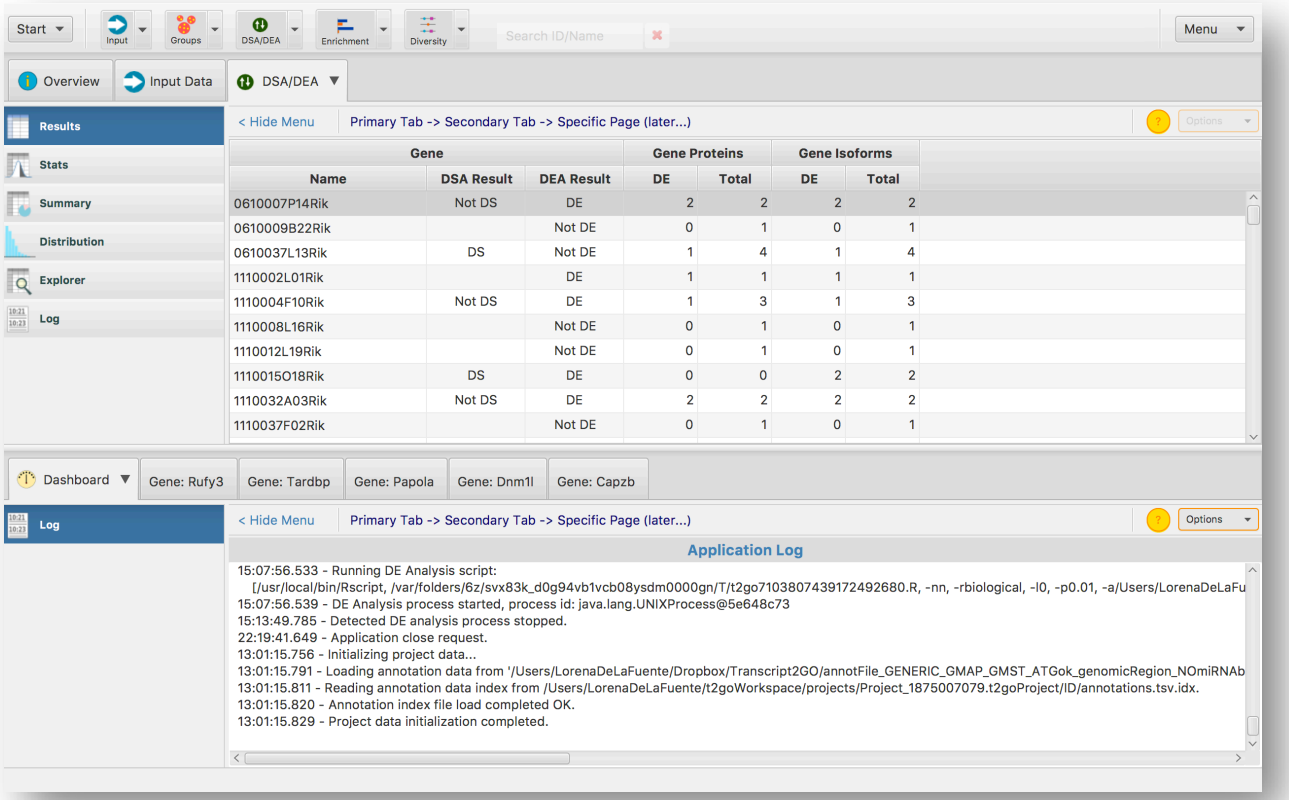
*Experimental validation **ongoing**:*

- *Analysis of miRNA 590 expression.*
- *Validation of the mirna binding site in Isoform 1 by miRNA pull-down assays.*

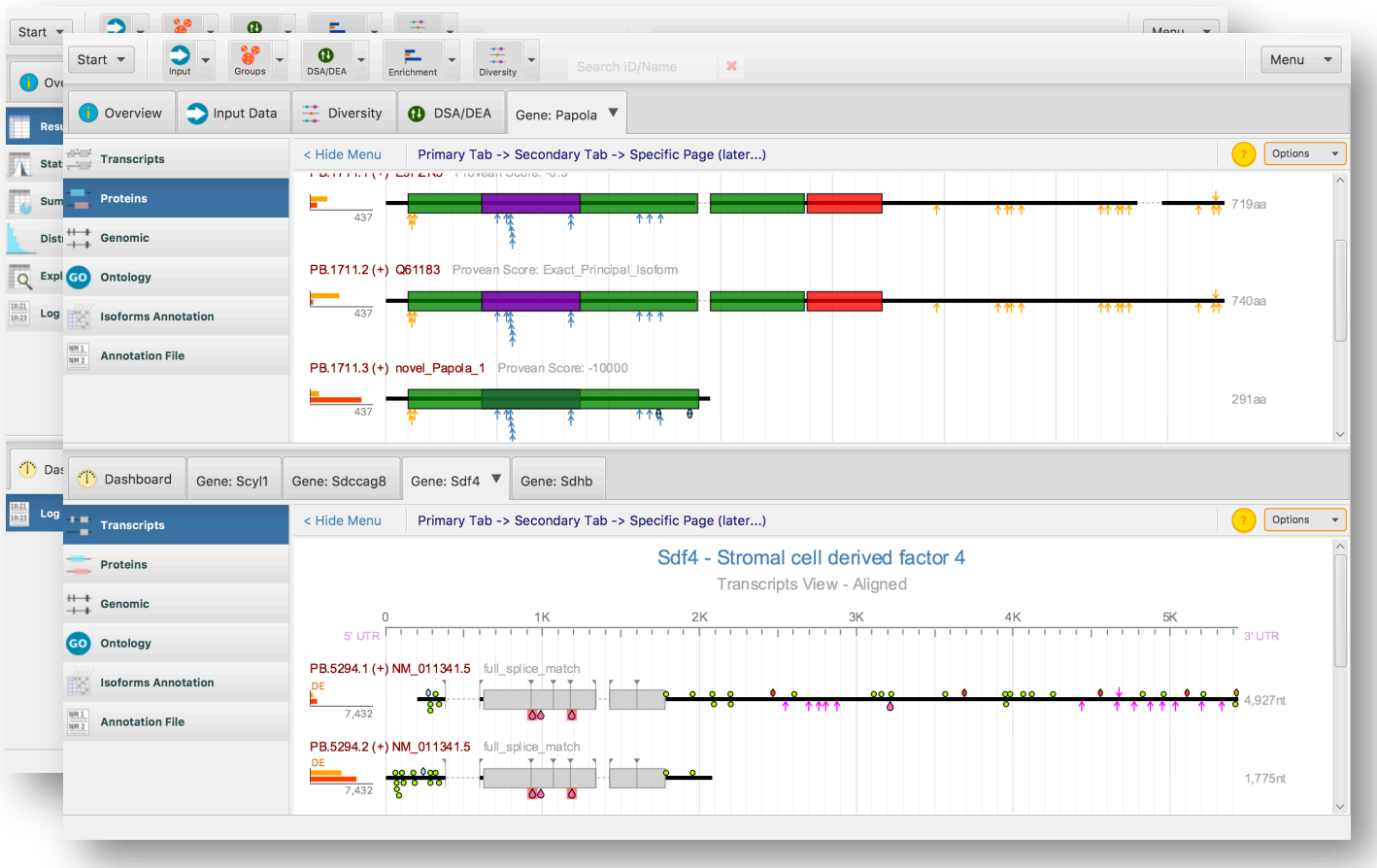
Functions enriched in Differential Spliced genes



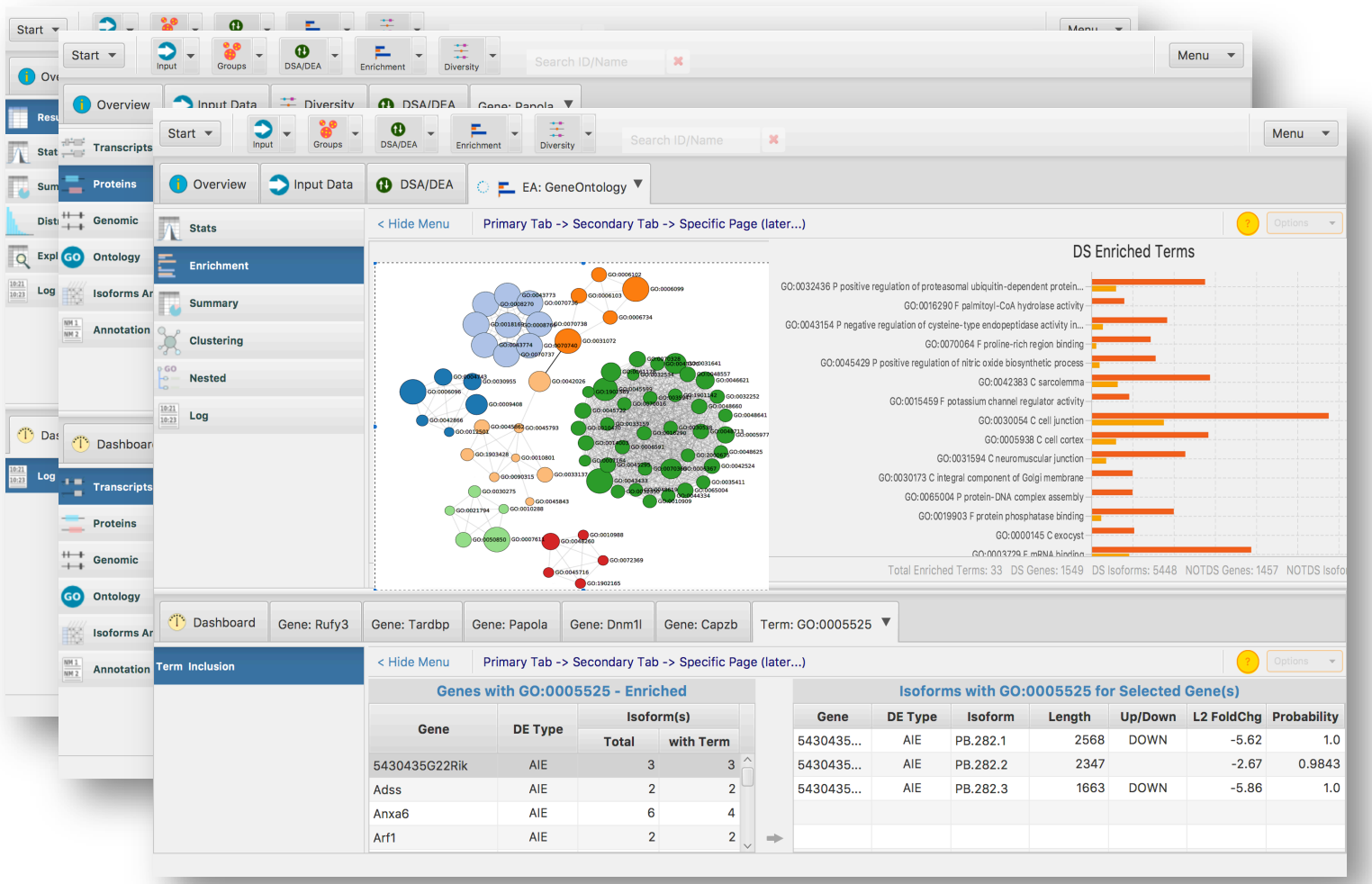
http://tappas.org



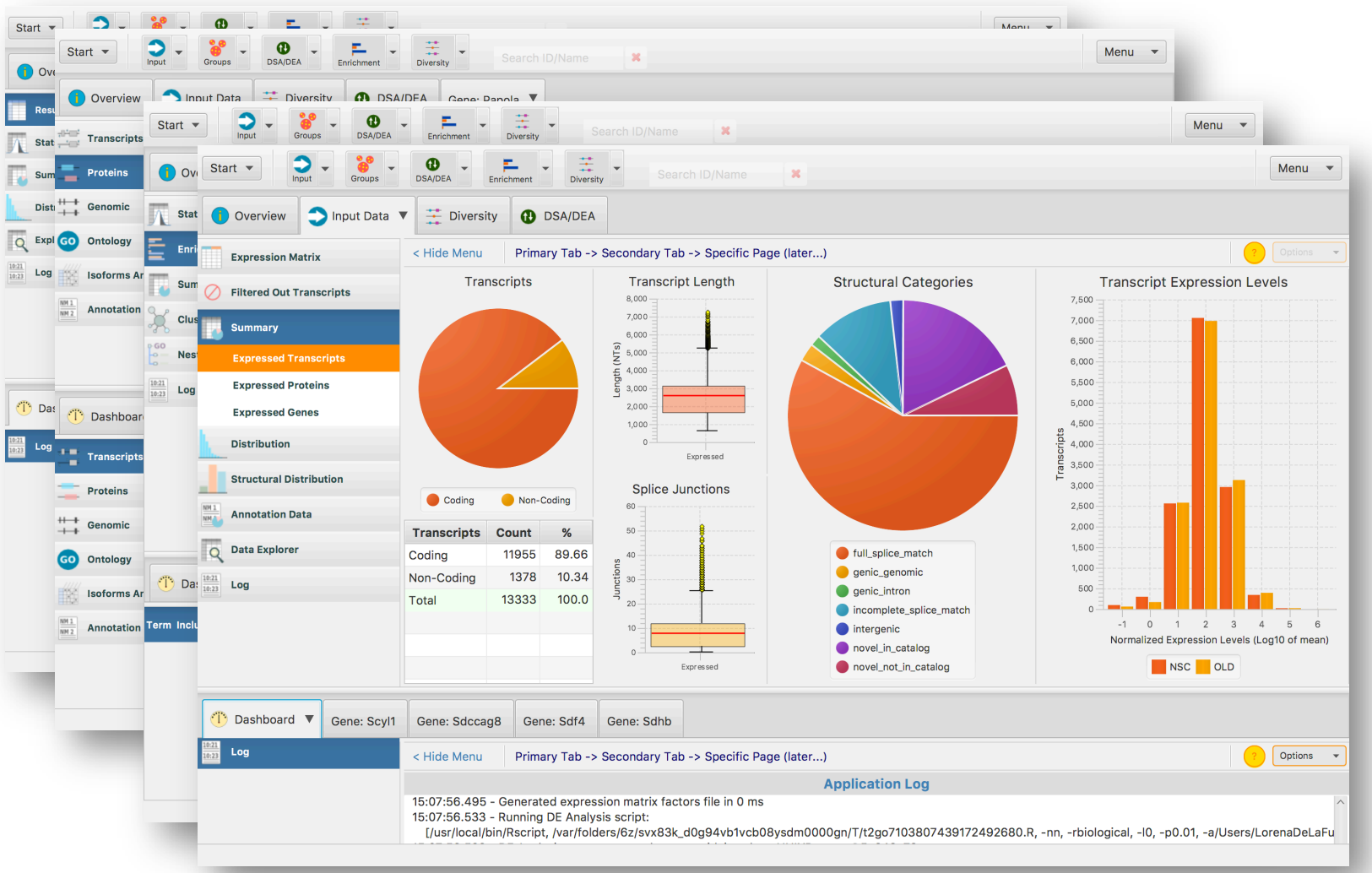
http://tappas.org



http://tappas.org



http://tappas.org



Acknowledgements

UF
William Farmerie
Eric Triplett
Lauren McIntyre

UCI
Ali Mortazavi

Pacbio
Liz Tseng

CIPF
Victoria Moreno
Susana Rodriguez