

# Functional Iso-Transcriptomics with tappAS

---

Ángeles Arzalluz-Luque

presented by Ana Conesa



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

Genomics   
 of Gene  
Expression Lab

# Reference and links

 tappAS software: <https://github.com/ConesaLab/tappAS/releases>

Info: <https://app.tappas.org>

Fuente et al. *Genome Biology* (2020) 21:119  
<https://doi.org/10.1186/s13059-020-02028-w>

Genome Biology

**SOFTWARE** Open Access

**tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing**

Lorena de la Fuente<sup>1,2†</sup>, Ángeles Arzalluz-Luque<sup>3</sup>, Manuel Tardáguila<sup>4,5†</sup>, Héctor del Risco<sup>4</sup>, Cristina Martí<sup>1</sup>, Sonia Tarazona<sup>3</sup>, Pedro Salguero<sup>1</sup>, Raymond Scott<sup>4</sup>, Alberto Lerma<sup>1</sup>, Ana Alastrue-Agudo<sup>5</sup>, Pablo Bonilla<sup>5</sup>, Jeremy R. B. Newman<sup>6,7</sup>, Shunichi Kosugi<sup>6,8</sup>, Lauren M. McIntyre<sup>6,9</sup>, Victoria Moreno-Manzano<sup>10†</sup> and Ana Conesa<sup>4,6†</sup> 



ConesaLab / tappAS

Unwatch 2 Star 2 Fork 0

<> Code Issues 2 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights Settings

Releases Tags

Draft a new release

Latest release

v1.0.1  
8707749  
Verified  
Compare

**tappAS v1.0.1**




aarzalluz released this 27 days ago · 1 commit to master since this release

**PATCH** of version 1.0.0.

Changes:

- Fixed bug in DFI analysis for Multiple Time Course designs.
- Updated contact page.
- GO annotation has been updated (May 2, 2020).
- Miscellaneous changes.

Assets 3

 <b>tappas.jar</b>	17.4 MB
 <b>Source code (zip)</b>	
 <b>Source code (tar.gz)</b>	

Citation: de la Fuente, L., Arzalluz-Luque, Á., Tardáguila, M. et al. tappAS: a comprehensive computational framework for the analysis of the functional impact of differential splicing. *Genome Biol* 21, 119 (2020). <https://doi.org/10.1186/s13059-020-02028-w>

# Outline

---

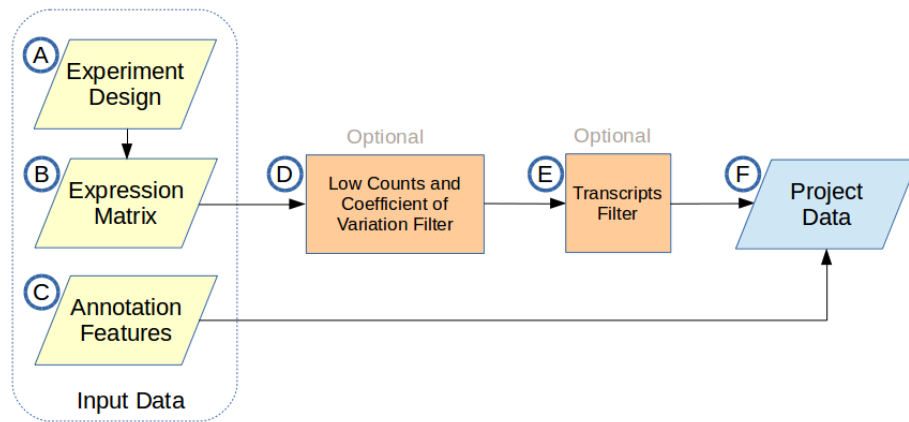
- Inputs and project creation.
- Brief guide to tappAS' interface.
- Analysis options.

# How do I create a project in tappAS?



# Inputs for project creation

- A. Design file.
- B. Expression matrix.
- C. Annotation file (gff3 format)\*



*\*Provided by tappAS for supported species/transcriptomes, can also be provided by user.*

# Inputs for project creation

A. Design file.

B. Expression matrix.

C. Annotation file (gff3 format)\*

Two-Group Comparison:

- **Sample names** match column names in expression matrix.
- At least **two replicates** per group.

experimental_design.tsv	input_matrix.tsv
sample	group
NSC1	NSC
NSC2	NSC
OLD1	OLD
OLD2	OLD

experimental_design.tsv		input_matrix.tsv		
NSC1	NSC2	OLD1	OLD2	
PB.1668.1	13090.25	8635.61	5911.69	6028.91
PB.972.1	2900.91	2220.34	2890.57	2836.8
PB.5017.1	3833.06	2861.24	2889.58	2544.89
PB.5017.2	691.37	353.12	81.23	148.82
PB.5017.3	3755.89	755.2	1005.15	795.3
PB.5017.4	416.69	250.44	323.03	436.99
PB.1463.1	1549	1138	459	464

\*Provided by tappAS for supported species/transcriptomes, can also be provided by user.

# Inputs for project creation

A. Design files: other supported experimental designs.

B. Expression matrix.

C. Annotation file (gff3 format)\*

design_timeCourseSingle.tsv		
sample	time	group
OLD1-1	0	OLD
OLD1-2	0	OLD
OLD2-1	1	OLD
OLD2-2	1	OLD
OLD3-1	2	OLD
OLD3-2	2	OLD

Single Time Course:

- One group, two replicates.
- Three time points.

design_timeCourseMultiple.tsv		
sample	time	group
OLD1-1	0	OLD
OLD1-2	0	OLD
OLD2-1	1	OLD
OLD2-2	1	OLD
OLD3-1	2	OLD
OLD3-2	2	OLD
NEUR1-1	0	NEUR
NEUR1-2	0	NEUR
NEUR2-1	1	NEUR
NEUR2-2	1	NEUR
NEUR3-1	2	NEUR
NEUR3-2	2	NEUR

Multiple Time Course:

- Two groups, two replicates.
- Three time points

\*Provided by tappAS for supported species/transcriptomes, can also be provided by user.

# Inputs for project creation

A.Design file.

B.Expression matrix.

C.Annotation file (gff3 format)\*

*Arabidopsis thaliana*  
Ensembl



*Drosophila melanogaster*  
Ensembl



*Homo sapiens*  
Ensembl, RefSeq



*Mus musculus*  
Demo, Ensembl, RefSeq



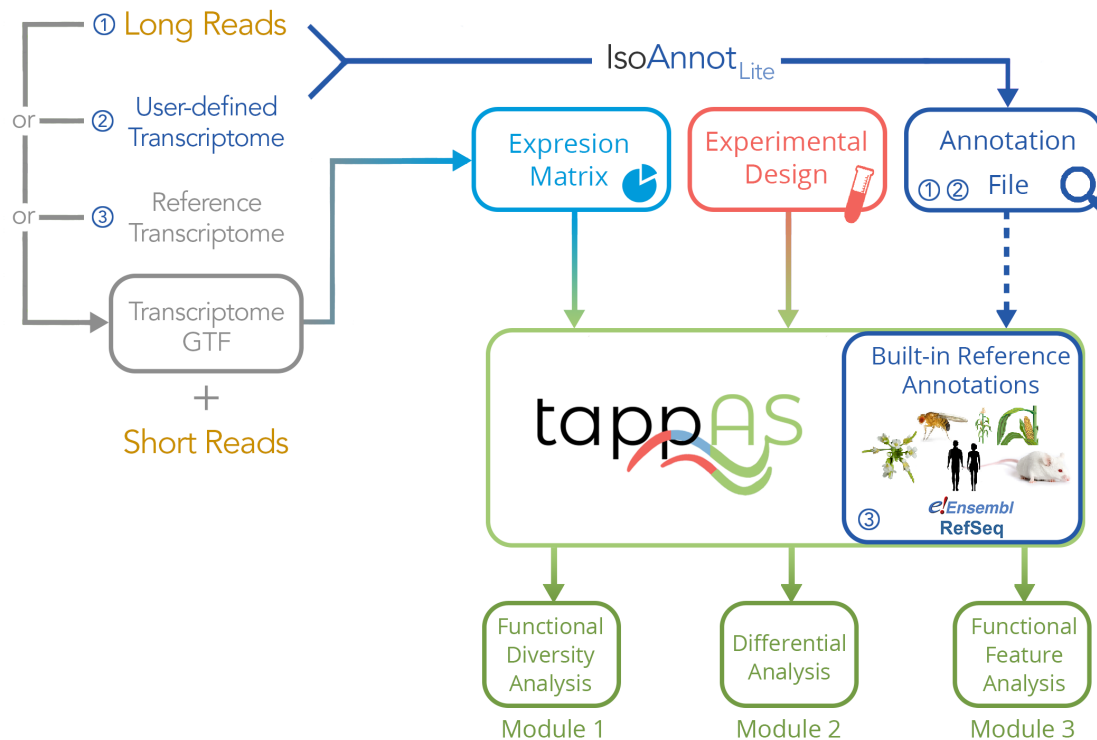
*Zea Mays*  
Ensembl



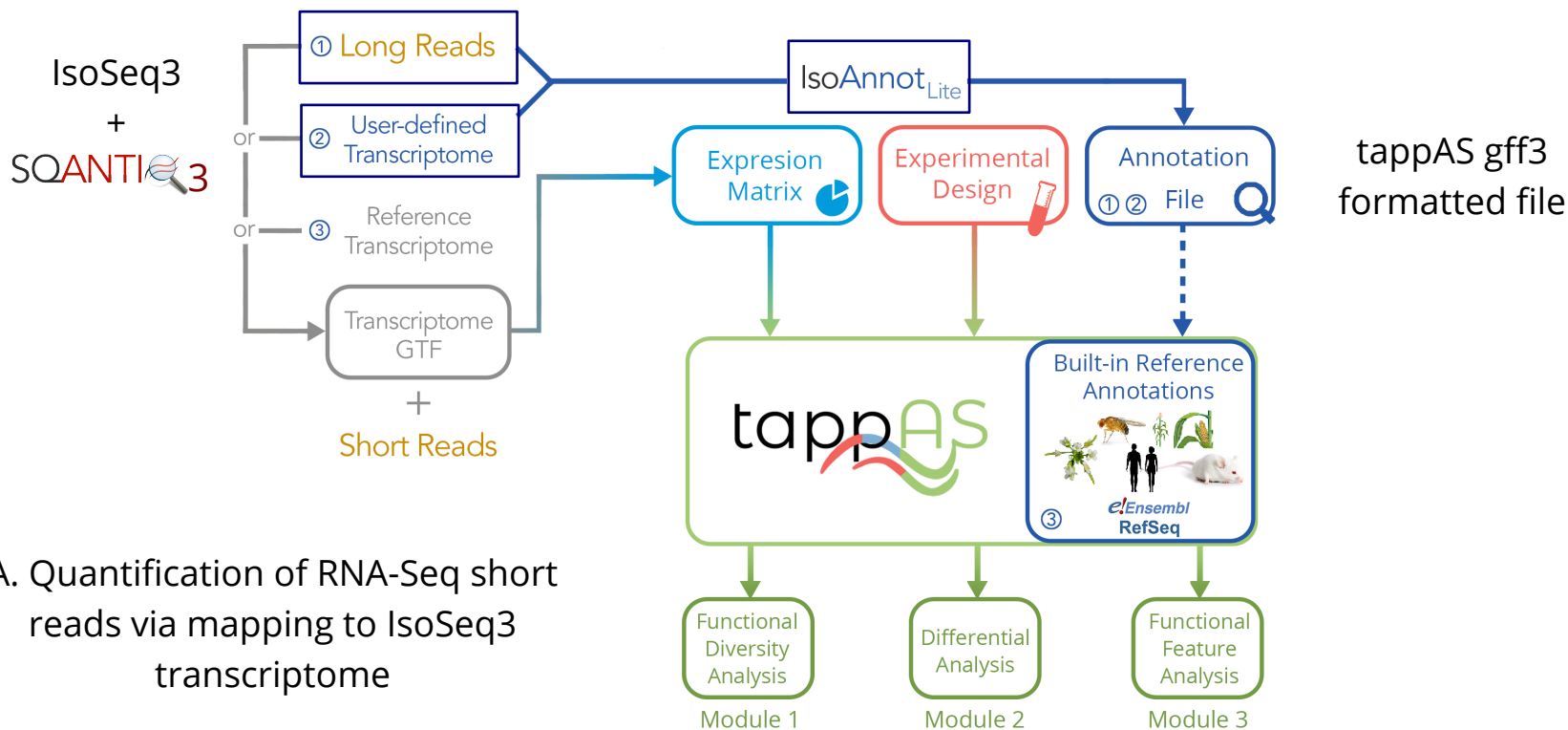
```
PB.3189.4 tappAS transcript 1 1399 . + . ID=XM_006524897; primary_class=full-splice_match; PostType=T
PB.3189.4 tappAS gene 1 1399 . + . ID=Opct; Name=Opct; Desc=glutaminyI-peptide cyclotransferase (glutaminyI cyclase); PostType=T
PB.3189.4 tappAS CDS 10 948 . + . ID=NP_001303658.1; PostType=T
PB.3189.4 scanForMotifs PAS 1369 1398 . + . ID=MammalianPolyadenylationElement; Name=MammalianPolyadenylationElement; Desc=Mammalian Polyadenylation Element; PostType=T
PB.3189.4 scanForMotifs 3UTRmotif 1148 1222 . + . ID=tRNAlikestructure; Name=tRNAlikestructure; Desc=tRNA like structure; PostType=T
PB.3189.4 miRWalk miRNA Binding 1090 1097 . + . ID=mmu-miR-139-5p; Name=mmu-miR-139-5p; Desc=3UTR; PostType=T
PB.3189.4 miRWalk miRNA Binding 1100 1107 . + . ID=mmu-miR-376b-3p; Name=mmu-miR-376b-3p; Desc=3UTR; PostType=T
PB.3189.4 miRWalk miRNA Binding 1294 1301 . + . ID=mmu-miR-374b-5p; Name=mmu-miR-374b-5p; Desc=3UTR; PostType=T
PB.3189.4 tappAS genomic 1 1 . + . Chr=chr17; PostType=G
PB.3189.4 tappAS exon 79052257 79052388 . + . Chr=chr17; PostType=G
PB.3189.4 tappAS exon 79070673 79070951 . + . Chr=chr17; PostType=G
PB.3189.4 tappAS exon 79077482 79077658 . + . Chr=chr17; PostType=G
PB.3189.4 tappAS exon 79079467 79079566 . + . Chr=chr17; PostType=G
PB.3189.4 tappAS exon 79081747 79081863 . + . Chr=chr17; PostType=G
PB.3189.4 tappAS exon 79089623 79090216 . + . Chr=chr17; PostType=G
PB.3189.4 tappAS splice_junction 79052389 79070672 . + . ID=known_canonical; Chr=chr17; PostType=G
PB.3189.4 tappAS splice_junction 79070952 79077481 . + . ID=known_canonical; Chr=chr17; PostType=G
PB.3189.4 tappAS splice_junction 79077659 79079466 . + . ID=known_canonical; Chr=chr17; PostType=G
PB.3189.4 tappAS splice_junction 79079567 79081746 . + . ID=known_canonical; Chr=chr17; PostType=G
PB.3189.4 tappAS splice_junction 79081864 79089622 . + . ID=known_canonical; Chr=chr17; PostType=G
PB.3189.4 tappAS protein 1 313 . + . ID=NP_001303658.1; PostType=P
PB.3189.4 Provean FunctionalImpact . . . ID=lessor; Score=-111.14; PostType=N
PB.3189.4 GeneOntology C . . . ID=GO:0070062; Name=extracellular exosome; PostType=N
PB.3189.4 GeneOntology F . . . ID=GO:0008270; Name=zinc ion binding; PostType=N
PB.3189.4 GeneOntology F . . . ID=GO:0016603; Name=glutaminyI-peptide cyclotransferase activity; PostType=N
PB.3189.4 GeneOntology P . . . ID=GO:0017186; Name=peptidyl-pyroglutamic acid biosynthetic process, using glutaminyI-peptide cyclotransferase; PostType=N
```

\*Provided by tappAS for supported species/transcriptomes, can also be provided by user.

# How do I obtain an isoform-level expression matrix?

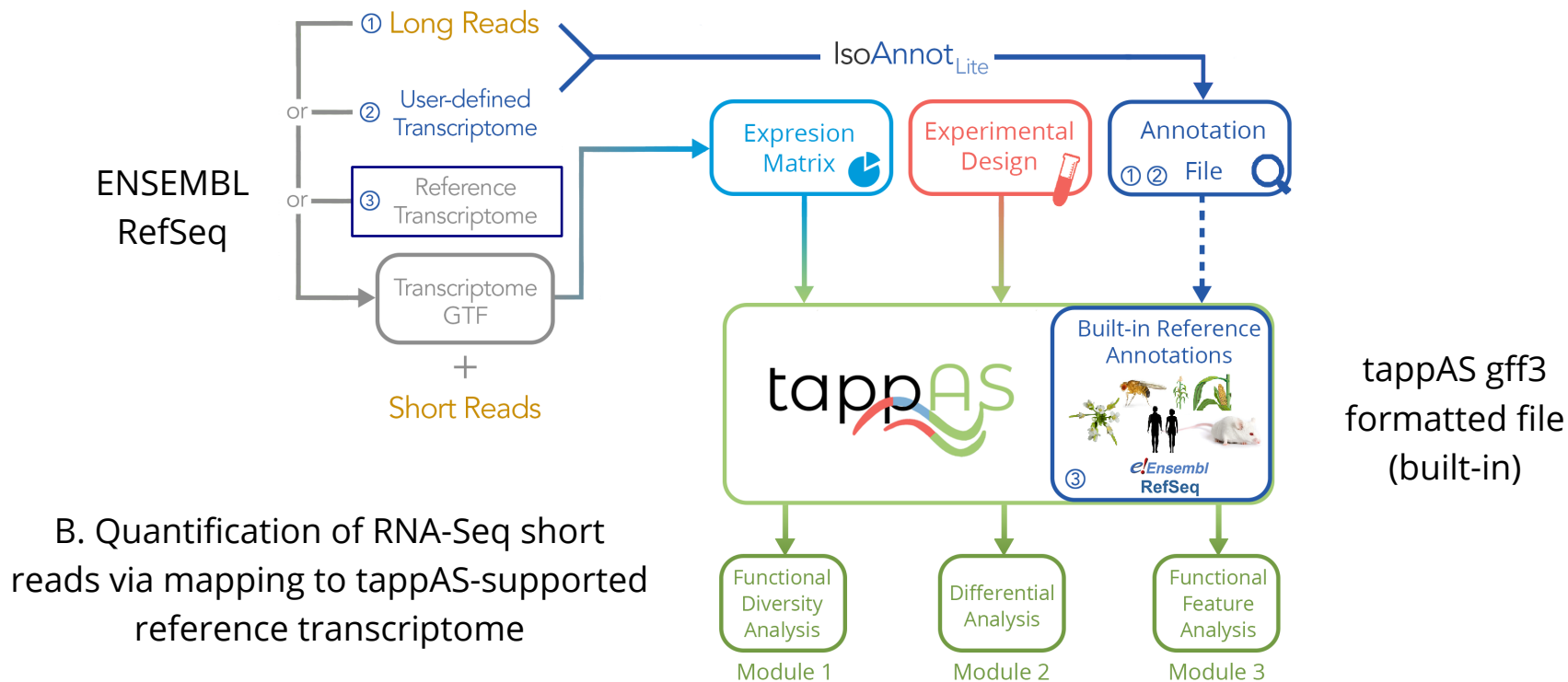


# How do I obtain an isoform-level expression matrix?

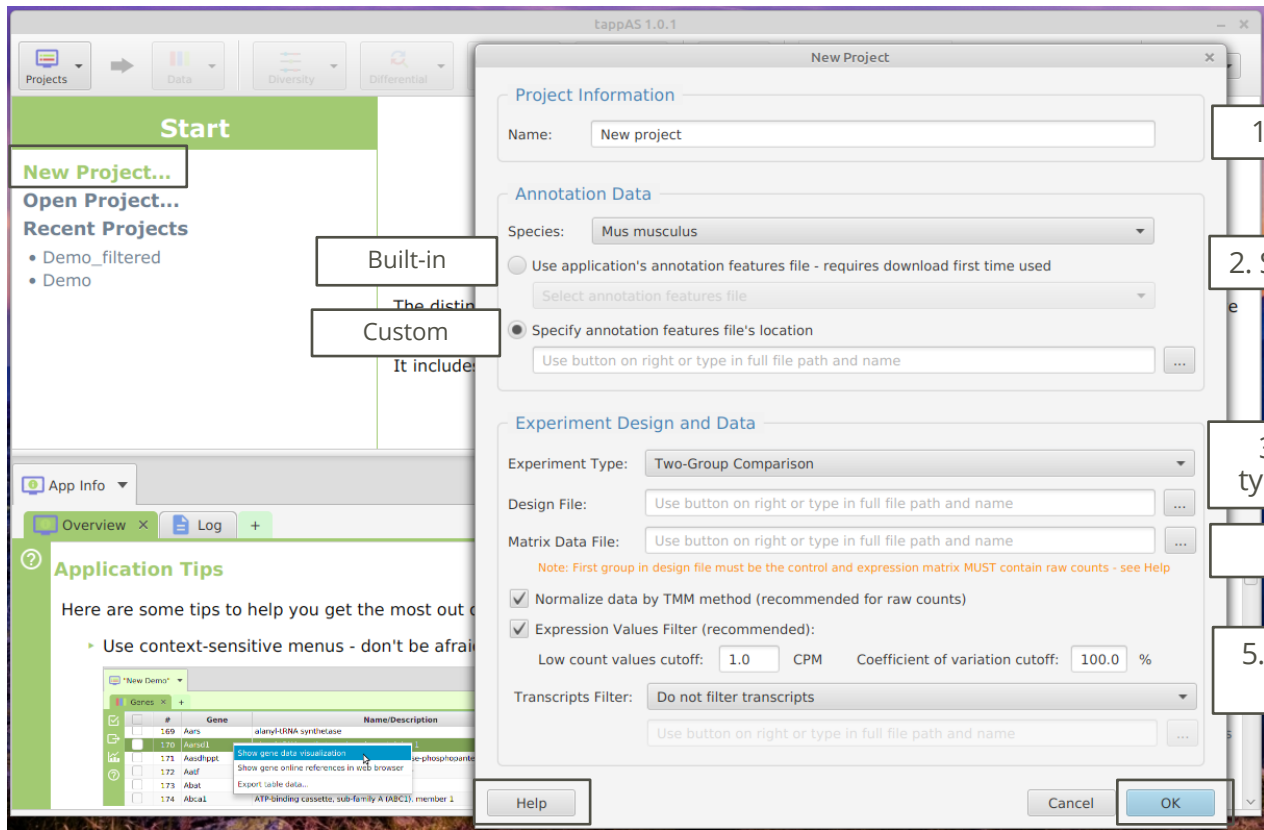


A. Quantification of RNA-Seq short reads via mapping to IsoSeq3 transcriptome

# How do I obtain an isoform-level expression matrix?



# Project creation in tappAS



1. Supply a name for the project

2. Select functional annotation

3. Design type and file

- Two-group comparison
- Single time course
- Multiple time course

4. Expression matrix file

5. Normalization and filtering options

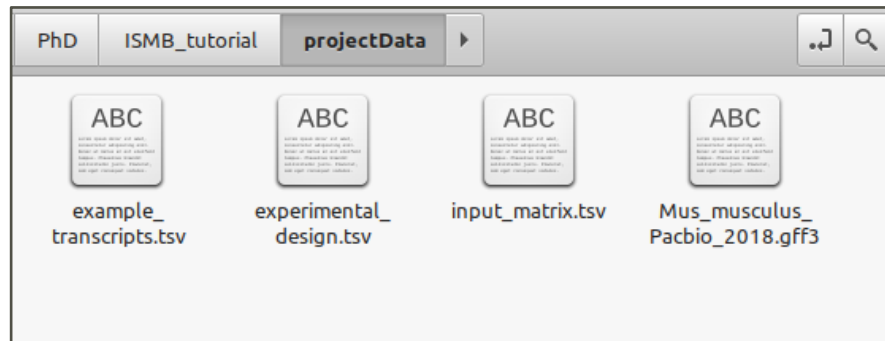


# Project creation in tappAS

- Execute the tappas.jar file with Java, or open a terminal window and run:

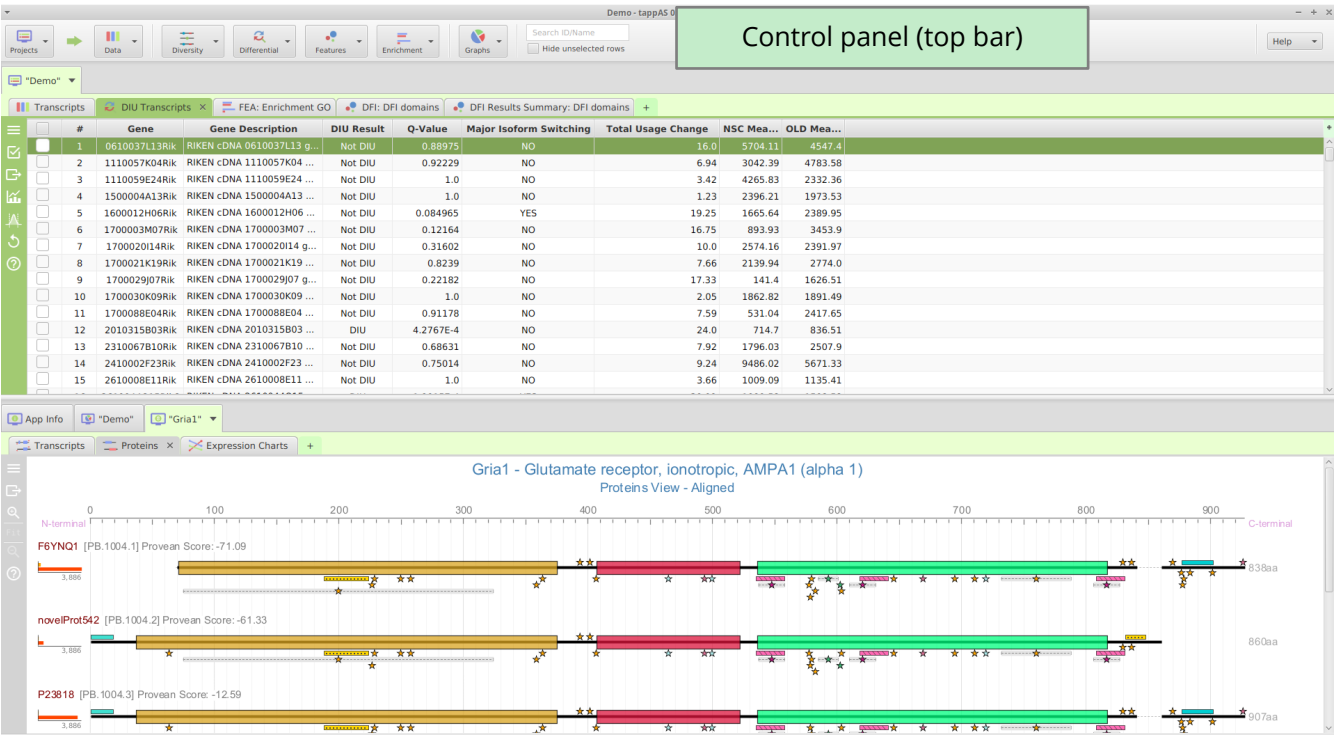
```
java -jar tappas.jar
```

- Locate the required input files:



# Guide to tappAS' interface

# tappAS interface



Control panel (top bar)

Data and result tables panel (top panel)

Visualization panel (bottom panel)

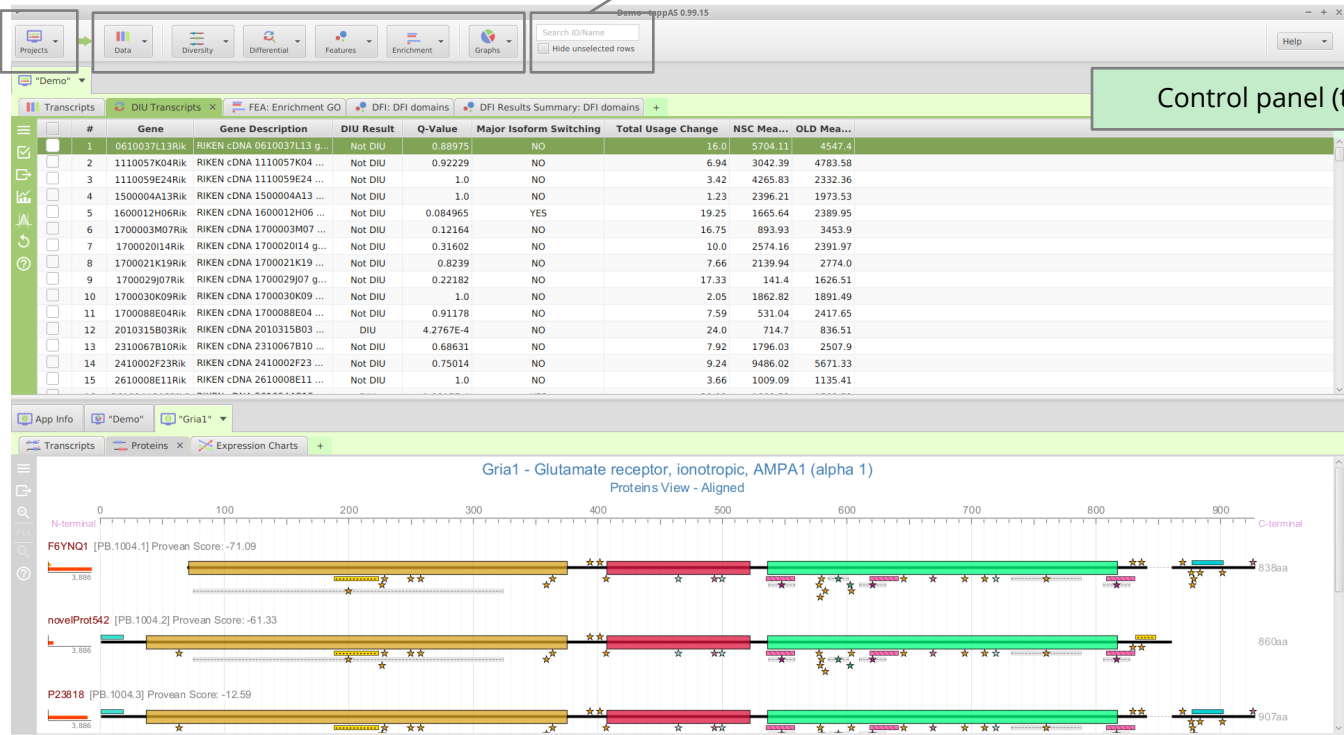
# tappAS interface

Create new and view extant projects

Run analyses and view results

Search bar for data and results panel

Control panel (top bar)



# tappAS interface

Run analyses and view results

Control panel (top bar)

The screenshot displays the tappAS interface with a table of gene data, a protein alignment chart, and an analysis parameter dialog box.

#	Gene	Gene Description	DIU Result	Q-Value	Major Isoform Switching	Total Usage Change	NSC
1	0610037L13Rik	RIKEN cDNA 0610037L13 g...	Not DIU	0.88975	NO	16.0	5
2	1110057K04Rik	RIKEN cDNA 1110057K04 ...	Not DIU	0.92229	NO	6.94	3
3	1110059E24Rik	RIKEN cDNA 1110059E24 ...	Not DIU	1.0	NO	3.42	4
4	1500004A13Rik	RIKEN cDNA 1500004A13 ...	Not DIU	1.0	NO	1.23	2
5	1600012H06Rik	RIKEN cDNA 1600012H06 ...	Not DIU	0.084965	YES	19.25	1
6	1700003M07Rik	RIKEN cDNA 1700003M07 ...	Not DIU	0.12164	NO	16.75	1
7	1700020I14Rik	RIKEN cDNA 1700020I14 g...	Not DIU	0.31602	NO	10.0	2
8	1700021K19Rik	RIKEN cDNA 1700021K19 ...	Not DIU	0.8239	NO	7.66	2
9	1700029J07Rik	RIKEN cDNA 1700029J07 g...	Not DIU	0.22182	NO	17.33	1
10	1700030K09Rik	RIKEN cDNA 1700030K09 ...	Not DIU	1.0	NO	2.05	1
11	1700088E04Rik	RIKEN cDNA 1700088E04 ...	Not DIU	0.91178	NO	7.59	1
12	2010315B03Rik	RIKEN cDNA 2010315B03 ...	DIU	4.2767E-4	NO	24.0	1
13	2310067B10Rik	RIKEN cDNA 2310067B10 ...	Not DIU	0.68631	NO	7.92	1
14	2410002F23Rik	RIKEN cDNA 2410002F23 ...	Not DIU	0.75014	NO	9.24	9
15	2610008E11Rik	RIKEN cDNA 2610008E11 ...	Not DIU	1.0	NO	3.66	1

The protein alignment chart shows three protein sequences: F6YNQ1 [PB.1004.1] Provean Score: -71.09, novelProt542 [PB.1004.2] Provean Score: -61.33, and P23818 [PB.1004.3] Provean Score: -12.59. The chart displays N-terminal, C-terminal, and internal domain alignments with star markers indicating specific features.

The 'Differential Feature Inclusion Analysis Parameters' dialog box is open, showing the following settings:

- Project: "Demo"
- Annotation Feature(s): ☒ SIGNALP\_EUK, ☒ PAR-CLIP, ☒ PFAM, ☒ TMHMM
- Check Using: Feature genomic position
- Analysis Method: DEXSeq
- Analysis Parameters:
  - Name: Enter Differential Feature Inclusion Analysis name
  - Significance Level: 0.05
  - Filtering: ☐ Filter minor isoforms, Fold filtering (default: 2)
  - Expression difference: 2.0 (default: 2)

Analysis pre-run window to select parameters

Analysis info and help (built-in documentation)

# tappAS interface

Multi-option side bar

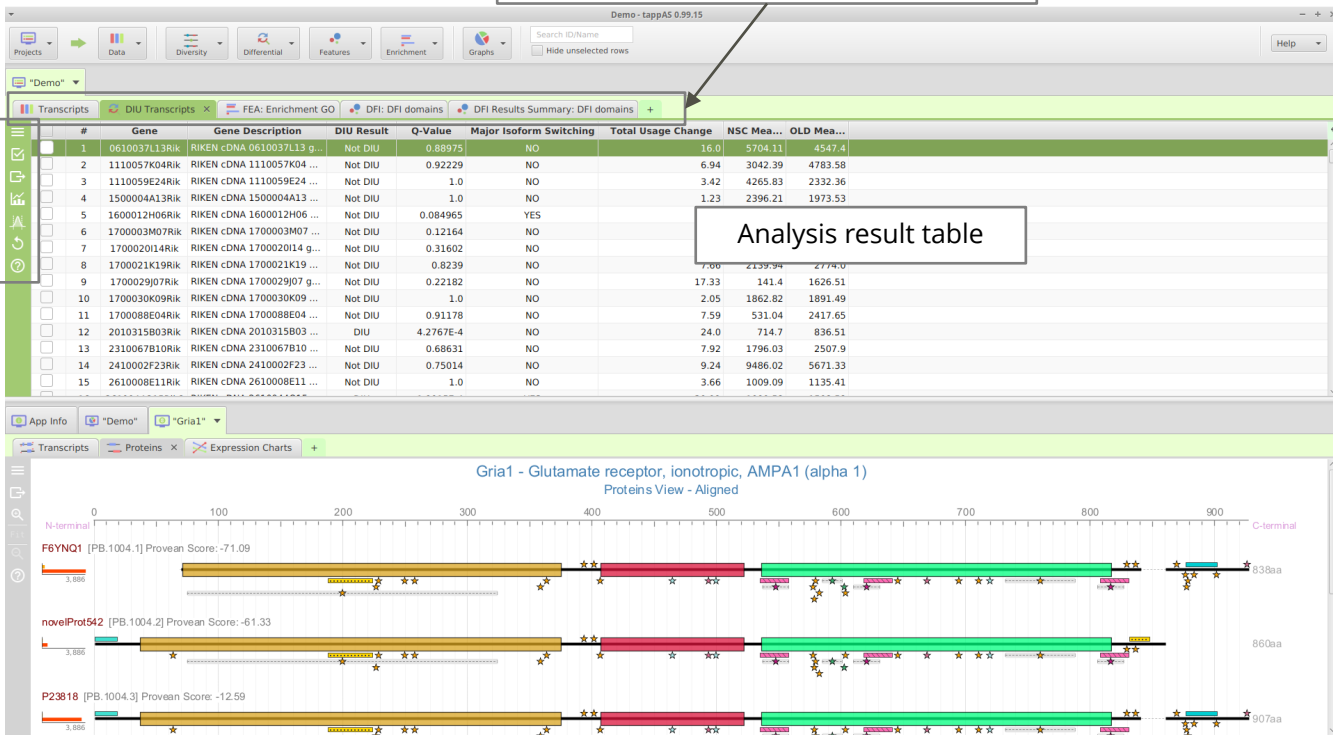
- Analysis log
- Filter table
- Export table
- Visualization options
- Change significance level
- Re-run analysis
- Help

Result subtabs for multiple analyses

Analysis result table

Data and result tables panel

Contains expression matrices for genes, transcripts and CDS, analysis result tables and result summary tables for up to 3 projects simultaneously.



# tappAS interface

Multi-option side bar

- Analysis log
- Filter table
- Export table
- Visualization options
- Change significance level
- Re-run analysis
- Help

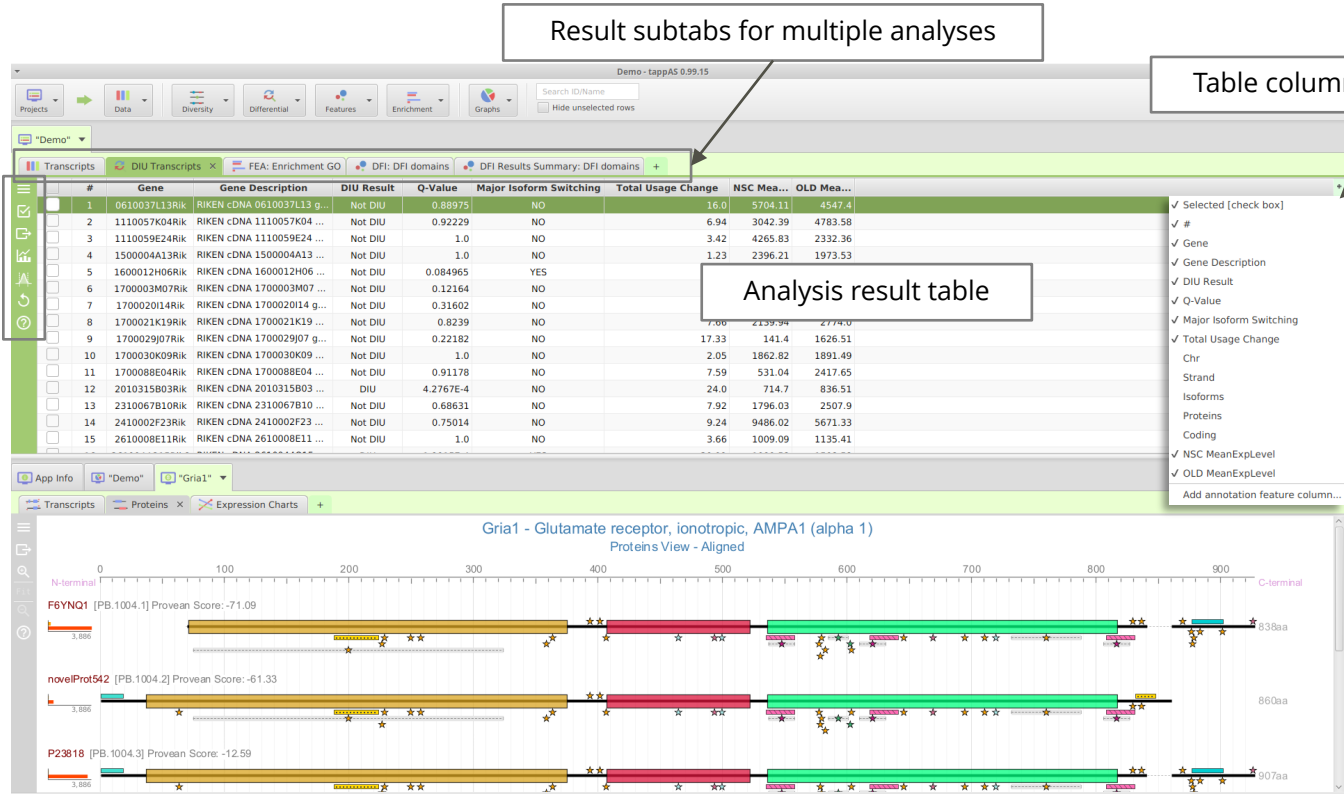
Result subtabs for multiple analyses

Table column customization

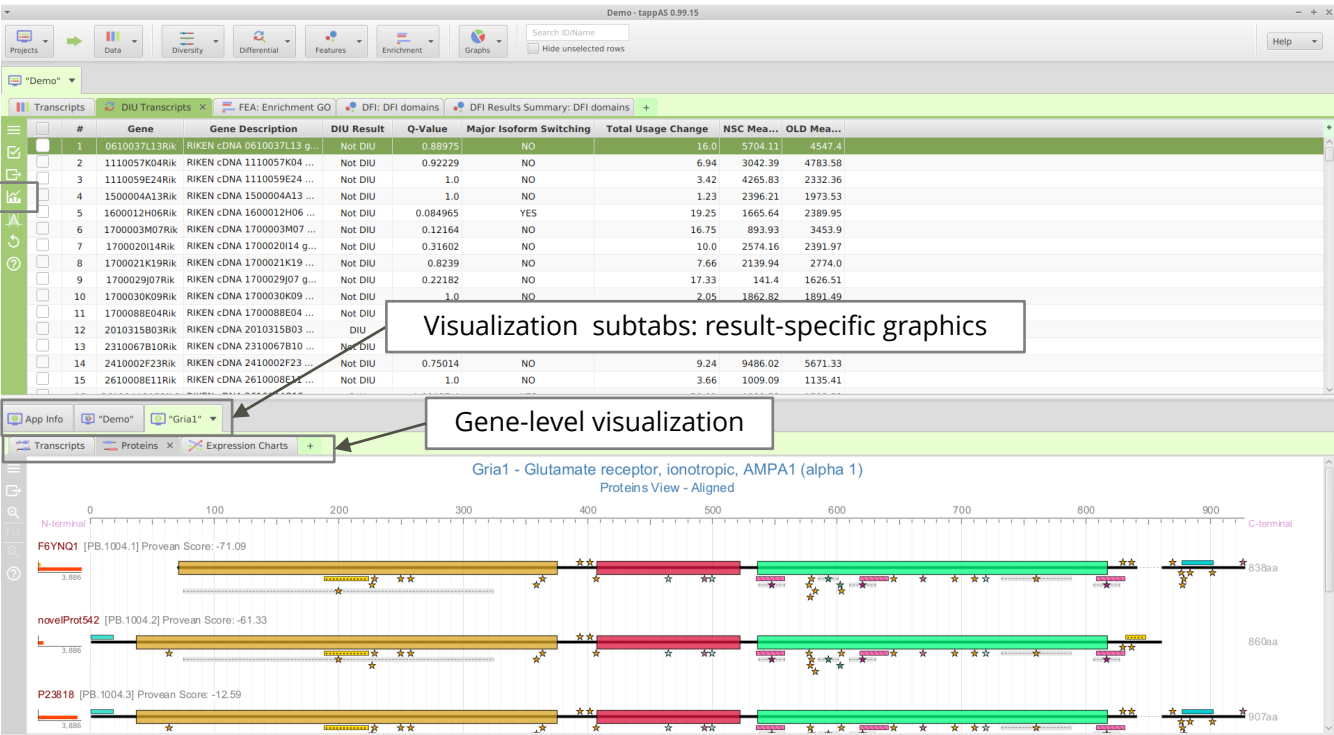
Analysis result table

Data and result tables panel

Contains expression matrices for genes, transcripts and CDS, analysis result tables and result summary tables for up to 3 projects simultaneously.



# tappAS interface



Analysis-specific visualization options

Multi-option side bar

- Analysis log
- Export image
- Zoom/adjust to screen.
- Help.

Visualization subtabs: result-specific graphics

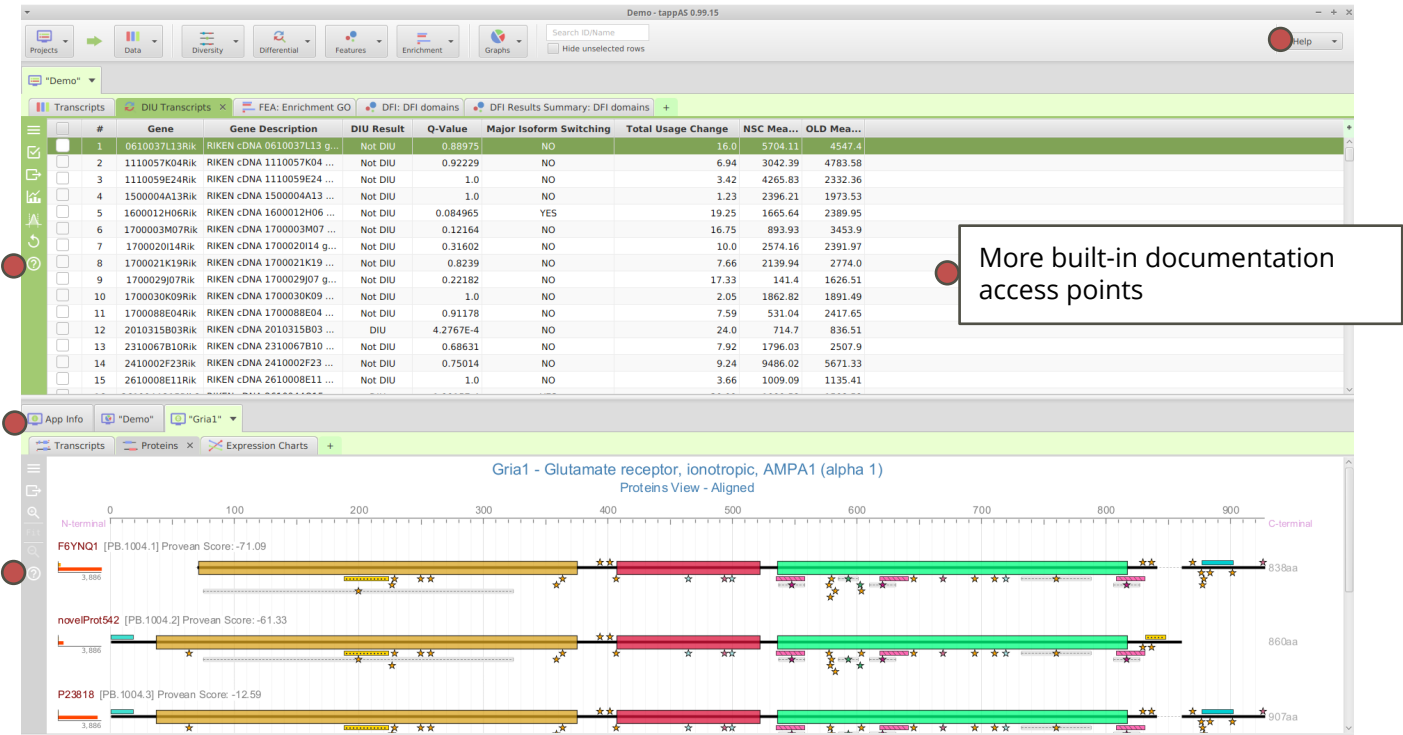
Gene-level visualization

Visualization panel

Contains gene-level visualization and result summary visualizations for the project under the project name subtab (e.g. Demo).



# tappAS interface

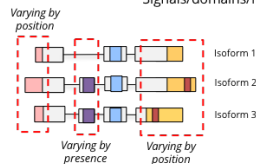


# tappAS analysis modules

## Module 1: Functional Diversity Analysis

Positional approach  
APA/UTR/CDS...

Presence/absence approach  
Signals/domains/NMD...



## Module 2: Differential Analysis

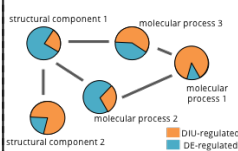
Global analysis : gene expression levels

Intra-gene analysis: AltTP

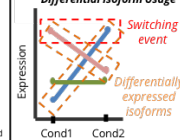
Differential Expression  
Transcript/Gene/CDS level

Differential Isoform Usage  
Transcript/CDS level

Functional Enrichment of DE vs DIU results

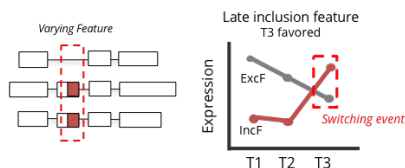


Differential Isoform Usage



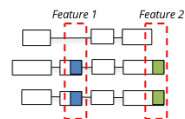
## Module 3: Feature Analysis

Differential Feature Inclusion

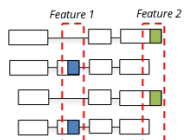


Co-DFI

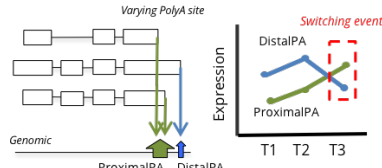
• Co-inclusion



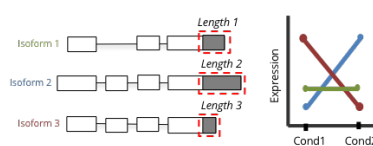
• Mutual exclusion



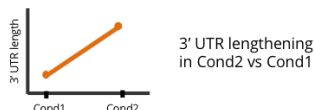
Differential Polydenylation



UTR shortening/lengthening



$$3UTRLength_{Cond1} = Expr_1 \cdot Length_1 + Expr_2 \cdot Length_2 + Expr_3 \cdot Length_3$$

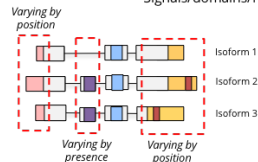


3' UTR lengthening  
in Cond2 vs Cond1

## Module 1: Functional Diversity Analysis

Positional approach  
APA/UTR/CDS...

Presence/absence approach  
Signals/domains/NMD...



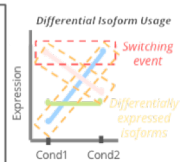
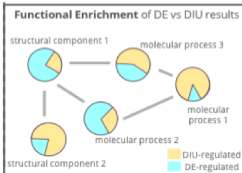
## Module 2: Differential Analysis

Global analysis : gene expression levels

Intra-gene analysis: AltTP

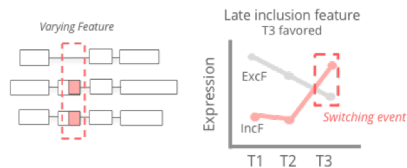
Differential Expression  
Transcript/Gene/CDS level

Differential Isoform Usage  
Transcript/CDS level



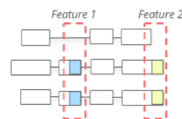
## Module 3: Feature Analysis

Differential Feature Inclusion

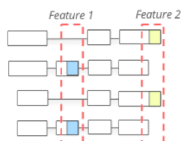


Co-DFI

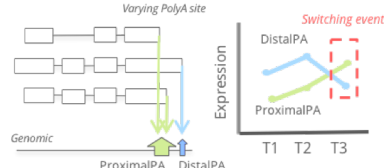
· Co-inclusion



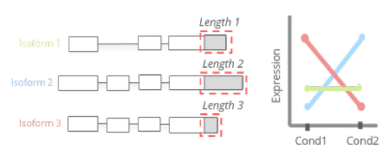
· Mutual exclusion



Differential Polydenylation



UTR shortening/lengthening

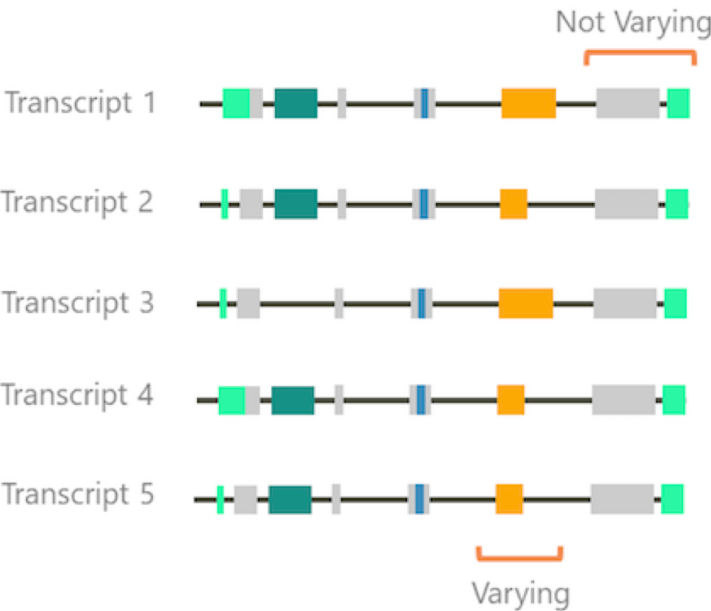


$$3UTRLength_{Cond1} = Expr_1 \cdot Length_1 + Expr_2 \cdot Length_2 + Expr_3 \cdot Length_3$$

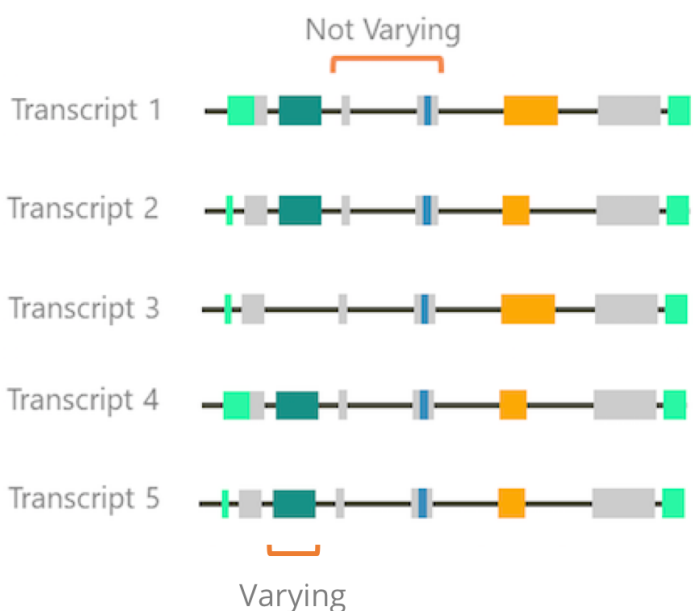


# Functional Diversity Analysis

Genomic position variance



Presence variance



# Functional Diversity Analysis

Functional Diversity Analysis Parameters

List of genes

Test list: All genes

List file: Use button on right or type in full file path and name

Analysis Parameters

Name: Enter Functional Diversity Analysis name

Analyze by: Category Id

Check Using: Feature genomic position

Select Annotation Feature(s)

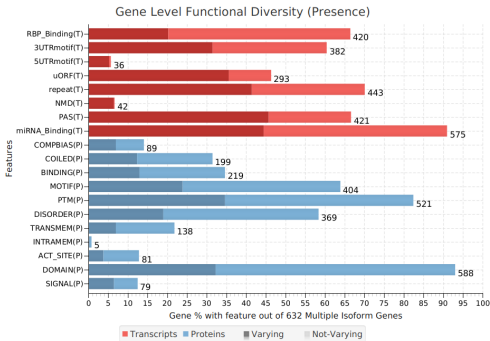
☒ SIGNALP\_EUK  
☒ PAR-CLIP  
☒ PFAM  
☒ TMHMM  
☒ scanForMotifs  
☒ RepeatMasker  
☒ UniProtKB/Swiss-Prot\_Phosphosite  
☒ COILS

Clear All Check All

Help Cancel OK

Category result:

- Gene-level, for each feature.
- Varying/not varying.



Transcript Annotation											
	#	Gene	NMD	RBP_Binding	repeat	5UTRmotif	uORF	miRNA_Binding	3UTRmotif	PAS	
<input type="checkbox"/>	1	Cyth4						Not Varying	Not Varying	Not Varying	
<input type="checkbox"/>	2	Rnh1			Varying		Varying	Not Varying			
<input type="checkbox"/>	3	Pop4		Not Varying	Not Varying		Varying	Not Varying	Not Varying		
<input type="checkbox"/>	4	Smc6		Varying	Varying		Varying	Varying	Varying	Varying	
<input type="checkbox"/>	5	Ganc			Varying		Not Varying	Varying	Varying		
<input type="checkbox"/>	6	Smc2			Not Varying			Varying	Varying	Varying	
<input type="checkbox"/>	7	Ipo11		Varying	Not Varying			Not Varying	Not Varying	Varying	
<input type="checkbox"/>	8	Tex264		Not Varying	Not Varying			Not Varying			
<input type="checkbox"/>	9	Gm10825			Not Varying		Not Varying		Not Varying	Varying	
<input type="checkbox"/>	10	Scp2		Varying				Not Varying	Not Varying	Not Varying	
<input type="checkbox"/>	11	Myc		Not Varying	Varying			Not Varying	Not Varying	Not Varying	
<input type="checkbox"/>	12	Shmt1			Varying			Varying	Varying	Varying	
<input type="checkbox"/>	13	Akt2		Varying	Varying		Not Varying	Varying	Varying	Not Varying	
<input type="checkbox"/>	14	Cyb5d2			Not Varying		Varying	Not Varying		Not Varying	

# Functional Diversity Analysis (FDA)

Functional Diversity Analysis Parameters

**List of genes**

Test list: All genes

List file: Use button on right or type in full file path and name

**Analysis Parameters**

Name: Enter Functional Diversity Analysis name

Analyze by: ☐ Category ☒ Id

Check Using: Feature genomic position

**Select Annotation Feature(s)**

- ☐ SIGNALP\_EUK
- ☐ PAR-CLIP
- ☒ PFAM
- ☐ TMHMM
- ☐ scanForMotifs
- ☐ RepeatMasker
- ☐ UniProtKB/Swiss-Prot\_Phosphosite
- ☐ COILS

Clear All Check All

Help Cancel OK

ID result:

- Feature ID-level within a category.
- No. of varying/not varying occurrences across genes.

FDA: PFAM_pres							
	#	Feature ID	Feature Description	P-Value	AdjP-Value	Protein Annotation	
						VARY...▼	NOT VARYING
<input type="checkbox"/>	1	PFAM/DOMAIN/PF00169	PH domain	0.005	1.0	8	3
<input type="checkbox"/>	2	PFAM/DOMAIN/PF00069	Protein kinase domain	0.4258	1.0	7	13
<input type="checkbox"/>	3	PFAM/DOMAIN/PF00096	Zinc finger, C2H2 type	0.241	1.0	6	8
<input type="checkbox"/>	4	PFAM/DOMAIN/PF00076	RNA recognition motif. (a.k.a. RRM,...	0.7906	1.0	5	15
<input type="checkbox"/>	5	PFAM/DOMAIN/PF00041	Fibronectin type III domain	0.0771	1.0	4	2
<input type="checkbox"/>	6	PFAM/DOMAIN/PF12796	Ankyrin repeats (3 copies)	0.2902	1.0	4	5
<input type="checkbox"/>	7	PFAM/DOMAIN/PF13927	Immunoglobulin domain	0.1745	1.0	3	2
<input type="checkbox"/>	8	PFAM/DOMAIN/PF01352	KRAB box	0.3729	1.0	3	4
<input type="checkbox"/>	9	PFAM/DOMAIN/PF13432	Tetratricopeptide repeat	0.0952	1.0	2	0
<input type="checkbox"/>	10	PFAM/DOMAIN/PF00307	Calponin homology (CH) domain	0.0952	1.0	2	0
<input type="checkbox"/>	11	PFAM/DOMAIN/PF01756	Acyl-CoA oxidase	0.0952	1.0	2	0
<input type="checkbox"/>	12	PFAM/DOMAIN/PF00433	Protein kinase C terminal domain	0.0952	1.0	2	0
<input type="checkbox"/>	13	PFAM/DOMAIN/PF09379	FERM N-terminal domain	0.0952	1.0	2	0
<input type="checkbox"/>	14	PFAM/DOMAIN/PF07719	Tetratricopeptide repeat	0.0952	1.0	2	0

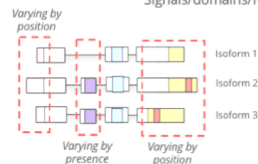
## Module 1: Functional Diversity Analysis

### Positional approach

APA/UTR/CDS...

### Presence/absence approach

Signals/domains/NMD...



## Module 2: Differential Analysis

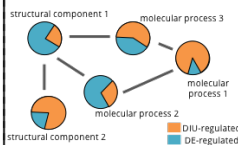
Global analysis : gene expression levels

Intra-gene analysis: AltTP

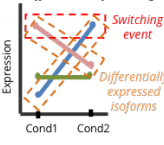
Differential Expression  
Transcript/Gene/CDS level

Differential Isoform Usage  
Transcript/CDS level

### Functional Enrichment of DE vs DIU results

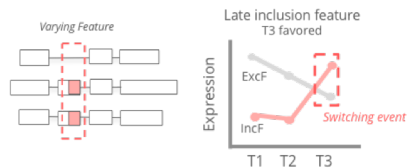


### Differential Isoform Usage



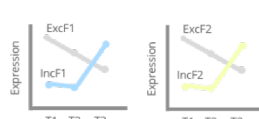
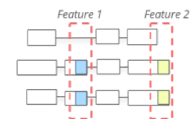
## Module 3: Feature Analysis

### Differential Feature Inclusion

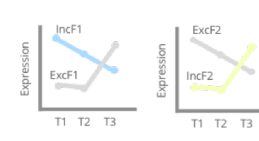
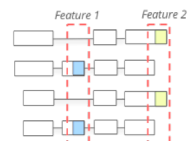


Co-DFI

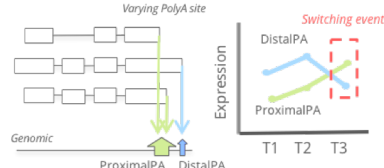
· Co-inclusion



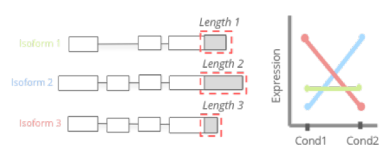
· Mutual exclusion



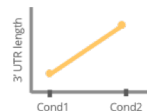
### Differential Polydenylation



### UTR shortening/lengthening



$$3UTRLength_{Cond1} = Expr_1 \cdot Length_1 + Expr_2 \cdot Length_2 + Expr_3 \cdot Length_3$$



3' UTR lengthening  
in Cond2 vs Cond1



# Differential Isoform Usage Analysis (DIU)

Differential Isoform Usage Parameters

Project "Demo\_filtered"

Data Type

☒ Transcripts ☐ CDS

Analysis Method

DIU Method: DEXSeq

Analysis Parameters

Significance Level: 0.05

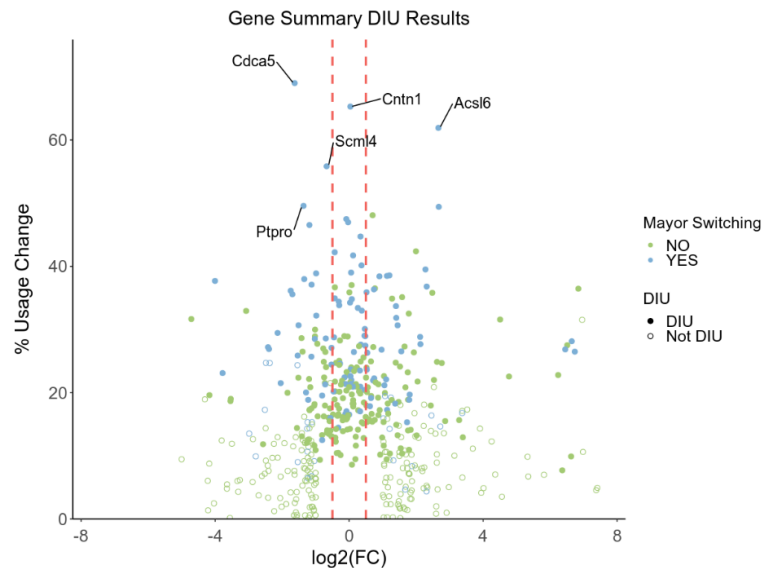
Filtering: ☒ Filter minor isoforms Proportion filtering

Prop. expression difference: 0.2 (default: 0.2)

Help Cancel OK

Result:

- Isoform switching.
- DIU/not DIU.
- Total Usage Change.



# Enrichment analyses (FEA and GSEA)

Functional Enrichment Analysis Parameters

Project "Demo\_filtered"

**Analysis Information**

Name:

**Data Type**

☒ Genes ☐ Transcripts ☐ CDS

**Genes Lists**

Test List:

List file:

Background List:

List file:

**Annotation Feature(s)**

☒ REACTOME  
☒ UTRsite  
☒ GeneOntology  
☒ miRWalk

You may select one or more features from a single source

**Analysis Parameters**

Statistical Test:

Number of Samples:

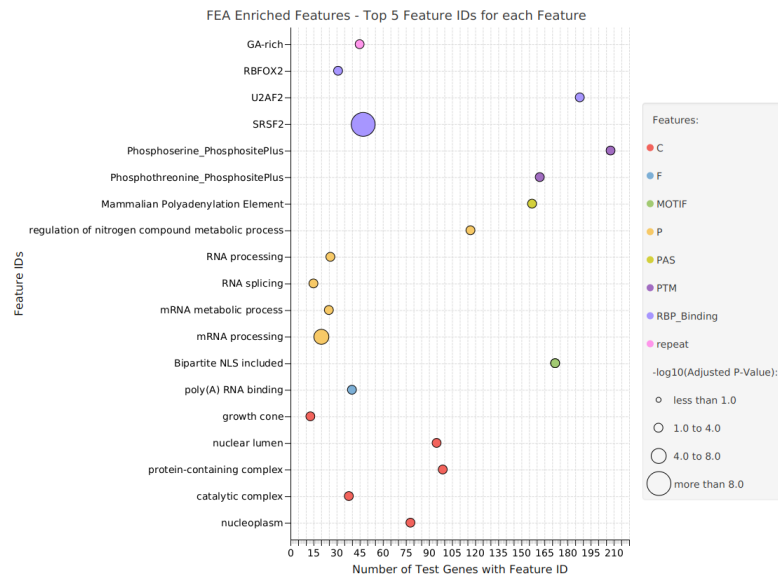
Use w/o Category:

Significance Level:

Help Cancel OK

Result:

- Significantly enriched feature categories.
- No. of genes with feature in test list/background list.



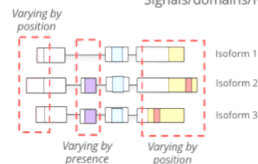
## Module 1: Functional Diversity Analysis

### Positional approach

APA/UTR/CDS...

### Presence/absence approach

Signals/domains/NMD...



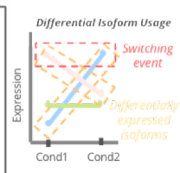
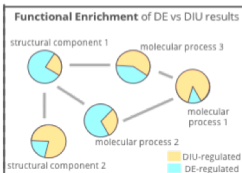
## Module 2: Differential Analysis

Global analysis : gene expression levels

Intra-gene analysis: AltTP

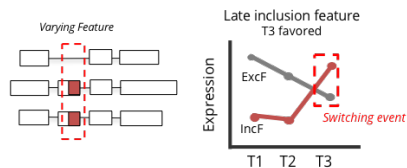
Differential Expression  
Transcript/Gene/CDS level

Differential Isoform Usage  
Transcript/CDS level



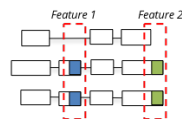
## Module 3: Feature Analysis

### Differential Feature Inclusion

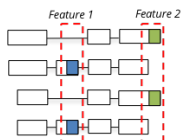


Co-DFI

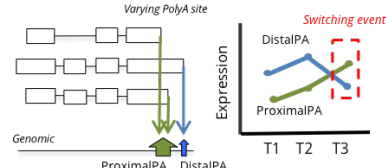
· Co-inclusion



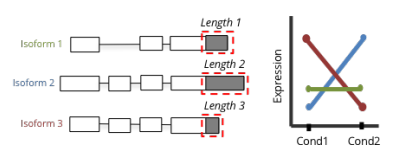
· Mutual exclusion



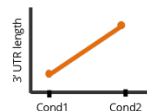
### Differential Polydenylation



### UTR shortening/lengthening



$$3'UTRLength_{Cond1} = Expr_1 \cdot Length_1 + Expr_2 \cdot Length_2 + Expr_3 \cdot Length_3$$



3' UTR lengthening  
in Cond2 vs Cond1

# Differential Feature Inclusion Analysis (DFI)

Differential Feature Inclusion Analysis Parameters

Project "Demo\_filtered"

Annotation Feature(s)

- ☒ SIGNALP\_EUK
- ☒ NMD
- ☒ PAR-CLIP
- ☒ PFAM

You may select annotation features from multiple sources. [Clear All](#) [Check All](#)

Check Using: Feature presence

Analysis Method

Method: DEXSeq

Analysis Parameters

Name:

Significance Level:

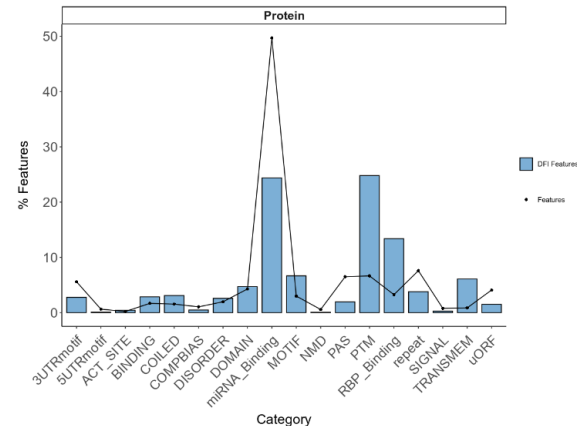
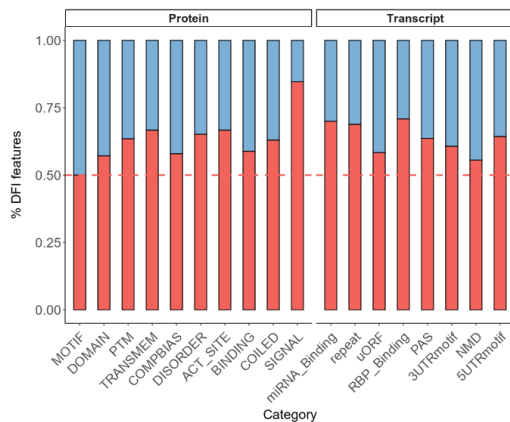
Filtering: ☒ Filter minor isoforms Proportion filtering

Prop. expression difference:  (default: 0.2)

[Help](#) [Cancel](#) [OK](#)

Result:

- DFI status and significance.
- Global statistical testing of each functional category.
- Favored condition/time point.
- Switching for feature usage and  $\Delta FI$ .



# Differential Polyadenylation Analysis (DPA)

Differential PolyAdenylation Analysis Parameters

Project "Demo\_filtered"

**PolyAdenylation Range**

Minimum value for differentiation:  (default: 60)

**Analysis Method**

Method:

**Analysis Parameters**

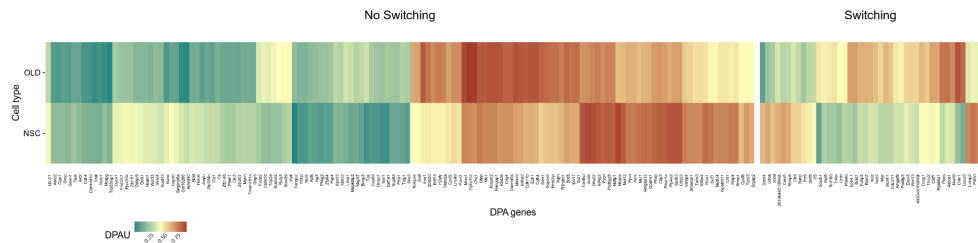
Significance Level:

Filtering: ☒ Filter minor isoforms

Prop. expression differen...  (default: 2)

Result:

- DPA status and significance.
- Proximal and distal polyA site estimated expression.
- Distal condition.
- Switching for polyA site usage and  $\Delta$ DPAU.



# 3' and 5' UTR Lengthening (UL)

Result:

- Expression-weighted UTR length (wUTR) per gene.
- Significance of change in UTR length across conditions for 3' and 5' UTRs.

UTR Lengthening Analysis Parameters

Project "Demo\_filtered"

**Lengthening Range**

Minimum value for differentiation:  (default: 100)

**Analysis Method**

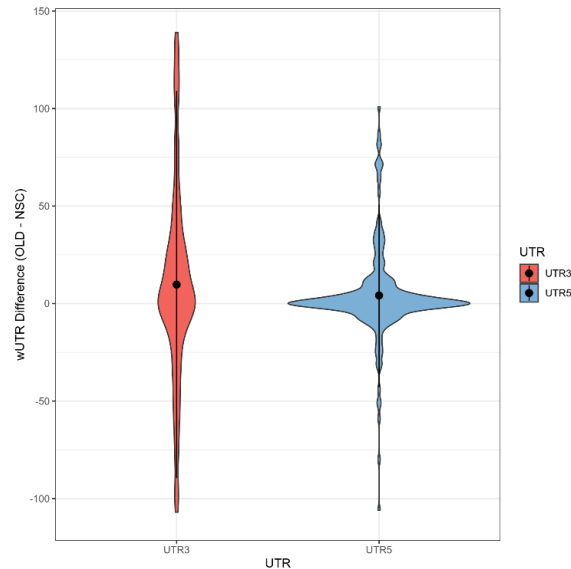
Method:

**Analysis Parameters**

Significance Level:

Filtering: ☒ Filter minor isoforms

Prop. expression differen...  (default: 0.2)



# Hands on time!

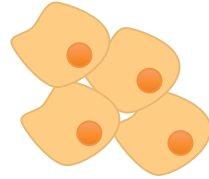
tappAS

# Demo data: OPC differentiation from NPC

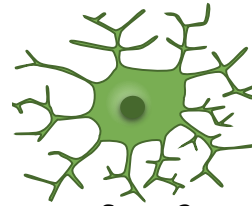
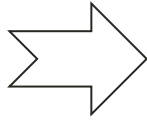


Neural Precursor Cells

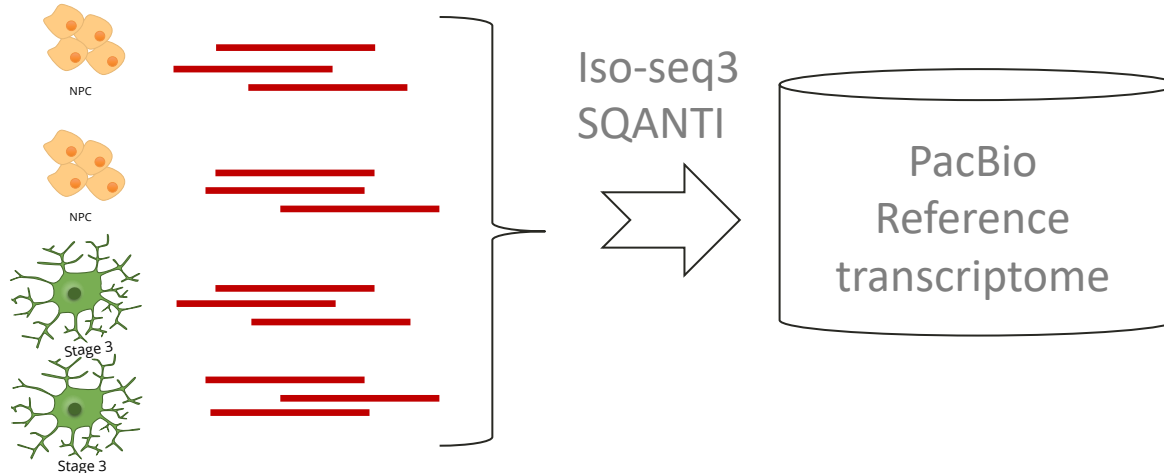
Oligodendrocytes



NPC

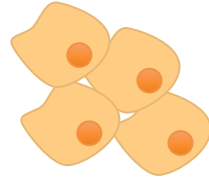


Stage 3

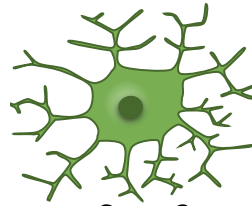
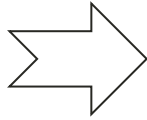




# Demo data: OPC differentiation from NPC

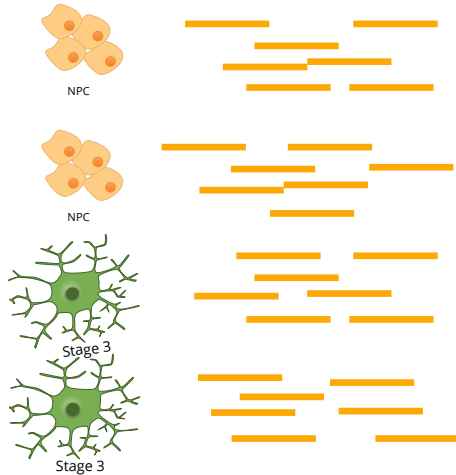


NPC



Stage 3

illumina®



Mapping  
STAR

PacBio  
Reference  
transcriptome



	NPC1	NPC2	OLD1	OLD2
PB1	-	-	-	-
PB2	-	-	-	-
PB3	-	-	-	-
...				
...				

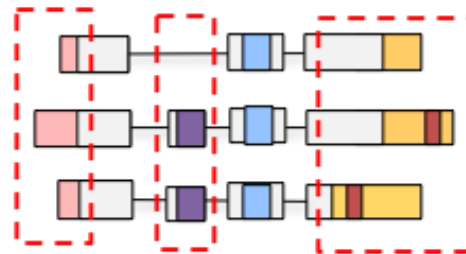
## Exercise 1: Functional Diversity Analysis (FDA)

### Module 1: Functional Diversity Analysis

Positional approach  
APA/UTR/CDS...

Presence/absence  
approach  
Signals/domains/NMD...

*Varying by  
position*



Isoform 1

Isoform 2

Isoform 3

*Varying by  
presence*

*Varying by  
position*

# tappAS hands-on

## Exercise 1: Functional Diversity Analysis (FDA)

1. Run a category FDA analysis using **genomic position** as the *varying* criteria.
2. Run a category FDA analysis using **presence** as the *varying* criteria.
3. **Visualization:** [Tip: use the left-hand side bar, and look for the graphics icon. Results will show in bottom panel.]
4. Look up gene **Rnh1** gene and find the differences in varying status for the Domain category between the presence and genomic position options
5. Run two ID-level FDA for PFAM domains, one using each *varying* criteria. Select the most significant varying ID and find the gene differences in the annotation.
6. **Visualization:** open the combined ID results tab and its results summary graphics. How can the results be interpreted?

## Exercise 4: Differential Feature Inclusion (DFI) analysis

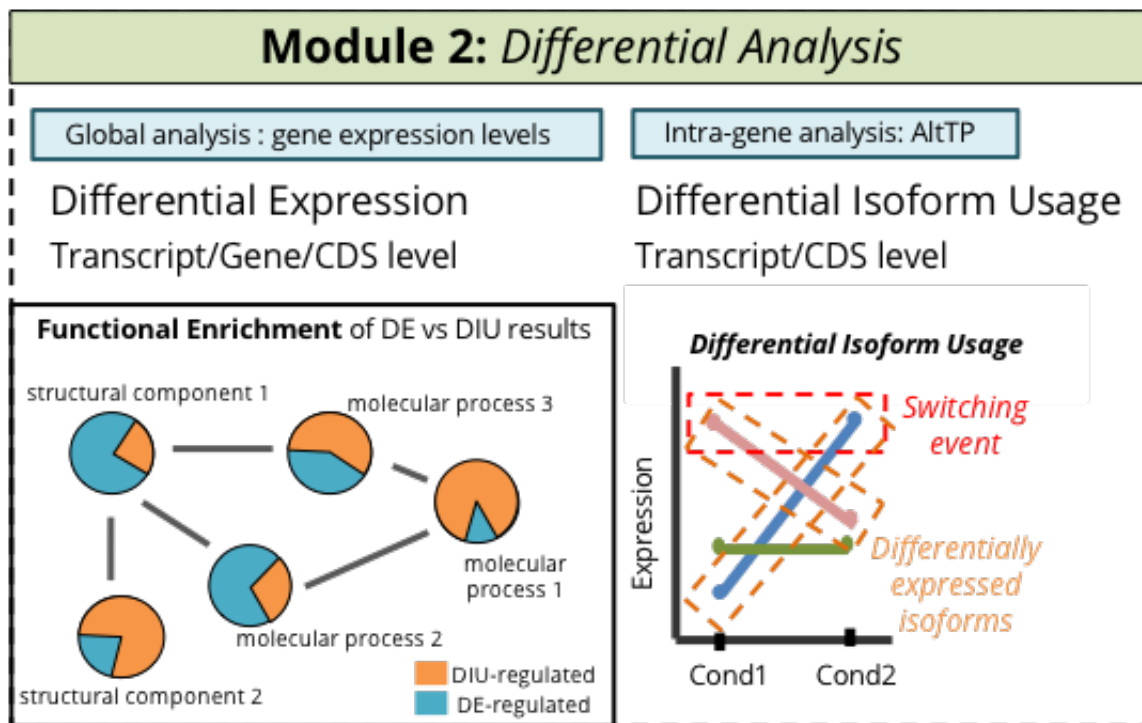
1. Run a **DFI analysis** for selected annotation categories (PAR-CLIP, NLS mapper, mirWalk, PFAM, UTRsite). You may use **genomic position** or **presence** as the differential criteria.

**NOTE:** this analysis will take ~10 min, so we'll leave this running and continue with Exercise 2.  
We'll come back to Exercise 4 and the DFI results later!

## Exercise 2: Differential Expression (DE) and Differential Isoform Usage (DIU)

1. Run a **DE analysis for genes**. You may use any of the available analysis methods.
  2. **Filtering**: filter de DEA genes tab by DE status. How many genes are DE? Export the list of DE genes (gene IDs only).  
[Tip: use the left hand-side bar, and look for the filtering and export icons]
  3. **Filtering**: try adding the «Isoform» column and filter again by isoform no. to detect how many DE genes have more than one isoform.  
[Tip: use the [+] button on the right side of the top panel]
1. **Visualization**: open the Results Summary graphics and try to find the no. of DE genes and the no. of multi-isoform DE genes in this other view.

## Exercise 2: Differential Expression (DE) and Differential Isoform Usage (DIU)



## Exercise 2: Differential Expression (DE) and Differential Isoform Usage (DIU)

5. Run a **DIU analysis for transcripts**. Run a DIU analysis for CDS, i.e. a DCU (Differential Coding region Usage) analysis.

[Tip: do not forget to apply the minor isoform filtering option at the bottom of the window!]

6. **Filtering**: filter the DIU results table by DIU status. Export the DIU gene IDs list. Do the same for the DCU results table (DIU protein).

7. Combine results and search for the ***Ctnnd1* gene**. Obtain gene visualization. Open the Expression Charts, and notice that there is isoform switching for two of its isoforms. Is the gene DIU? And DE?

[Tip: use the «Search» bar at the upper bar of the application, and then right-click on the gene.]

8. **Visualization**: open the DIU results summary. What is the relationship between isoform switching and % Usage Change?

## Exercise 3: Coupling the Differential and Enrichment Modules.

1. Run a **Functional Enrichment Analysis** using **DE genes as test list** and all genes as background list. Select GeneOntology as the functional database for analysis.
2. Run a **Functional Enrichment Analysis for DIU genes**. To study the functional categories that are regulated by splicing, select all functional annotation databases.
3. Run the same analysis for **DCU genes**. How do results change when coding potential is considered?
4. **Visualization**: for the DIU and DCU gene enrichments, open the Enriched Features Chart. Customize the no. of features shown in the chart. Export the images.



## Exercise 3: coupling the Differential and Enrichment Modules.

5. Run a **Functional Enrichment Analysis**, using DIU genes as test list and DE genes as background. Select «Use input file...» and use the exported files in the previous exercise.

[Tip: use the left hand-side bar, and look for the graphics and export icons]

6. For the NLS category, find the list of genes that contains the feature in the test list. How many are there?

[Tip: you can right click on any functional category, and select «Drill down data» to view the genes that contain a given feature within the test list.]

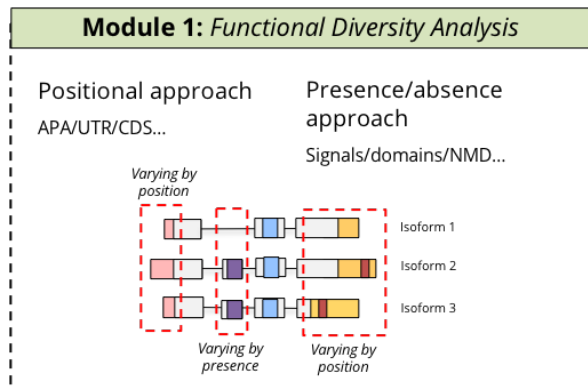
1. **Extra exercise:** go back to the DE and DIU results, filter and export them as **gene ranked lists** (i.e. gene ID and p-value). Perform **unidimensional and multidimensional Gene Set Enrichment Analysis (GSE)**.

# tappAS hands-on

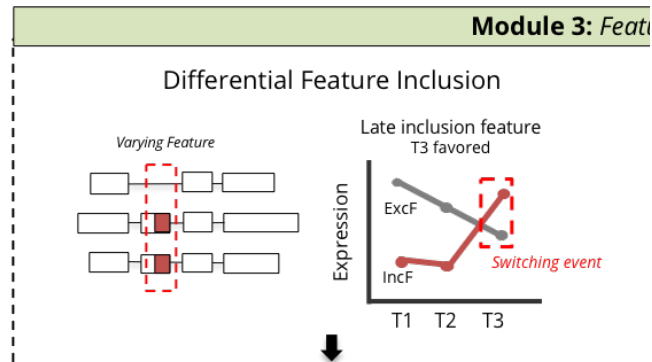
## Exercise 4: Differential Feature Inclusion (DFI) analysis

1. Run a **DFI analysis** for all annotation categories. You may use **genomic position** or **presence** as the differential criteria.

### FDA only uses annotation



### DFI uses gene expression



## Exercise 4: Differential Feature Inclusion (DFI) analysis

1. Run a **DFI analysis** for all annotation categories. You may use **genomic position** or **presence** as the differential criteria.  
[!] The analysis should have finished by now -if not, it may be due to low RAM. tappAS can be run with more RAM by using the `java -XmsAM -XmxBM -jar tappas.jar` command
2. **Visualization:** open the Results Summary graphics. What categories are globally significantly DFI in the transcriptome?
3. Select a DFI gene and open the **feature ID data visualization**. Open the Expression Charts in the gene data visualization for that gene, and compare them. Is there switching for the feature? Is there switching for the isoforms?

## Exercise 4: Differential Feature Inclusion (DFI) analysis

4. For this gene, open the **transcript and protein annotations**. Can you locate the DFI feature in the transcript and/or protein models? Which are the isoforms that include them, and how does their expression change?
2. Open the **DFI Results Summary** tab. What is the most frequently DFI functional annotation category? In which condition is its inclusion favored?
1. Open the **coDFI Results** tab. Which pair of functional features are more frequently co-DFI?

## Exercise 5: other analyses in the Features module.

1. Run a **Differential PolyAdenylation (DPA)** analysis and a  
[Tip: notice the minimum differentiation value threshold. The higher, the more stringent the analysis, i.e. higher bp difference between the polyA sites of all isoforms will be required to consider a gene with two polyA sites, distal and proximal.]
2. **Visualization:** open the Results Summary graphics, and look for the total no. of DPA genes.
3. Select a DPA gene, and open the **feature ID data visualization**. Is there polyA site switching?  
How does this compare to the gene's isoforms expression? Can you locate the different polyA sites in the transcript model annotations?
4. Run a **UTR Lengthening (UL)** analysis.

## Exercise 5: other analyses in the Features module.

### 4. Run a UTR Lengthening (UL) analysis.

[Tip: notice the minimum differentiation value threshold. The higher, the more stringent the analysis, i.e. higher bp difference between the polyA sites of all isoforms will be required to consider a gene with two UTRs, a shorter and a longer.]

1. **Visualization:** open the Results Summary graphics. Is there significant 3' and/or 5' UTR lengthening between the conditions?

2. Add the «Isoforms» column to the UL results table. Find a gene with many isoforms, and open the **gene data visualization**. How many different UTR lengths do you see, and how do you think they were collapsed?