

# 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 Llerma<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>4,6,9</sup>

The screenshot shows the GitHub repository page for 'ConesaLab / tappAS'. The 'Releases' tab is selected, showing the 'Latest release' is 'tappAS v1.0.1' by 'aarzalluz' (released 27 days ago). The release notes mention a fixed bug in DFI analysis for Multiple Time Course designs, updated contact page, GO annotation update, and miscellaneous changes. Assets listed include 'tappas.jar' (17.4 MB), 'Source code (zip)', and 'Source code (tar.gz)'.

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

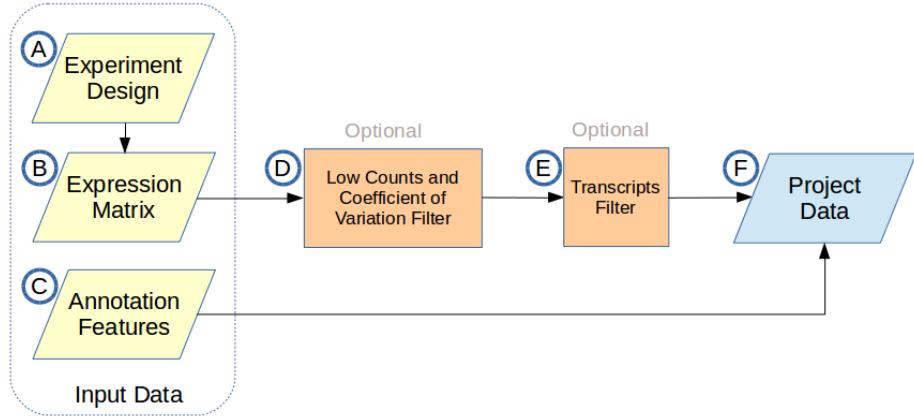
- Inputs and project creation.
- Brief guide to tappAS' interface.
- Analysis options.

# 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.

sample	group
NSC1	NSC
NSC2	NSC
OLD1	OLD
OLD2	OLD

NSC1	NSC2	OLD1	OLD2
PB.1668.1	13090.25	8635.61	5911.69
PB.972.1	2900.91	2220.34	2890.57
PB.5017.1	3833.06	2861.24	2889.58
PB.5017.2	691.37	353.12	81.23
PB.5017.3	3755.89	755.2	1005.15
PB.5017.4	416.69	250.44	323.03
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)\*

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.

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)\*

```

PB.3189.4 tappAS transcript 1 1399 . + . ID=XM_006524897; primary_class=full-splice_match; PosType=T
PB.3189.4 tappAS gene 1 1399 . + . ID=Qpcf; Name=Qpcf; Desc=glutaminyl-peptide cyclotransferase (glutaminyl cyclase); PosType=T
PB.3189.4 tappAS CDS 10 948 . + . ID=NP_001303658.1; PosType=T
PB.3189.4 scanForMotifs PAS 1369 1398 . + . ID=MammalianPolyadenylationElement; Name=MammalianPolyadenylationElement; Desc=Mammalian Polyadenylation Element; PosType=T
PB.3189.4 scanForMotifs 3UTRmotif 1148 1222 . + . ID=tRNAlikestructure; Name=tRNAlikestructure; Desc=tRNA like structure; PosType=T
PB.3189.4 miRWalk miRNA_Binding 1090 1097 . + . ID=mmu-miR-139-5p; Name=mmu-miR-139-5p; Desc=3UTR; PosType=T
PB.3189.4 miRWalk miRNA_Binding 1100 1107 . + . ID=mmu-miR-376b-3p; Name=mmu-miR-376b-3p; Desc=3UTR; PosType=T
PB.3189.4 miRWalk miRNA_Binding 1294 1301 . + . ID=mmu-miR-374b-5p; Name=mmu-miR-374b-5p; Desc=3UTR; PosType=T
PB.3189.4 tappAS genomic 1 1 . + . Chr=chr17; PosType=G
PB.3189.4 tappAS exon 79052257 79052388 . + . Chr=chr17; PosType=G
PB.3189.4 tappAS exon 79070673 79070951 . + . Chr=chr17; PosType=G
PB.3189.4 tappAS exon 79077482 79077658 . + . Chr=chr17; PosType=G
PB.3189.4 tappAS exon 79079467 79079566 . + . Chr=chr17; PosType=G
PB.3189.4 tappAS exon 79081747 79081863 . + . Chr=chr17; PosType=G
PB.3189.4 tappAS exon 79089623 79090216 . + . Chr=chr17; PosType=G
PB.3189.4 tappAS splice_junction 79052389 79070672 . + . ID=known_canonical; Chr=chr17; PosType=G
PB.3189.4 tappAS splice_junction 79070952 79077481 . + . ID=known_canonical; Chr=chr17; PosType=G
PB.3189.4 tappAS splice_junction 79077659 79079466 . + . ID=known_canonical; Chr=chr17; PosType=G
PB.3189.4 tappAS splice_junction 79079567 79081746 . + . ID=known_canonical; Chr=chr17; PosType=G
PB.3189.4 tappAS splice_junction 79081864 79089622 . + . ID=known_canonical; Chr=chr17; PosType=G
PB.3189.4 tappAS protein 1 313 . + . ID=NP_001303658.1; PosType=P
PB.3189.4 Provean FunctionalImpact . . . . ID=lesser; Score=-111.14; PosType=N
PB.3189.4 GeneOntology C . . . . ID=G0:0070062; Name=extracellular exosome; PosType=N
PB.3189.4 GeneOntology F . . . . ID=G0:0008270; Name=zinc ion binding; PosType=N
PB.3189.4 GeneOntology F . . . . ID=G0:0016603; Name=glutaminyl-peptide cyclotransferase activity; PosType=N
PB.3189.4 GeneOntology P . . . . ID=G0:0017186; Name=peptidyl-pyroglutamic acid biosynthetic process, using glutaminyl-peptide cyclotransferase; PosType=N

```

*Arabidopsis thaliana*  
Ensembl



*Drosophila melanogaster*  
Ensembl



*Homo sapiens*  
Ensembl, RefSeq



*Zea Mays*  
Ensembl

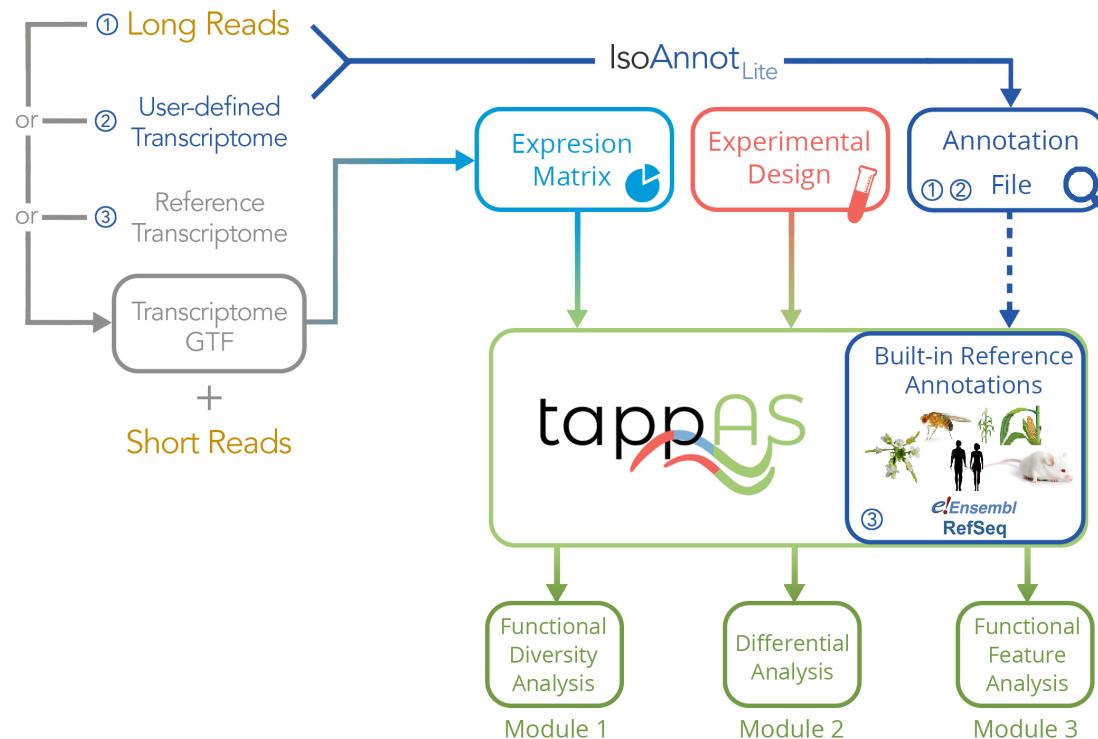


*Mus musculus*  
Demo, Ensembl, RefSeq

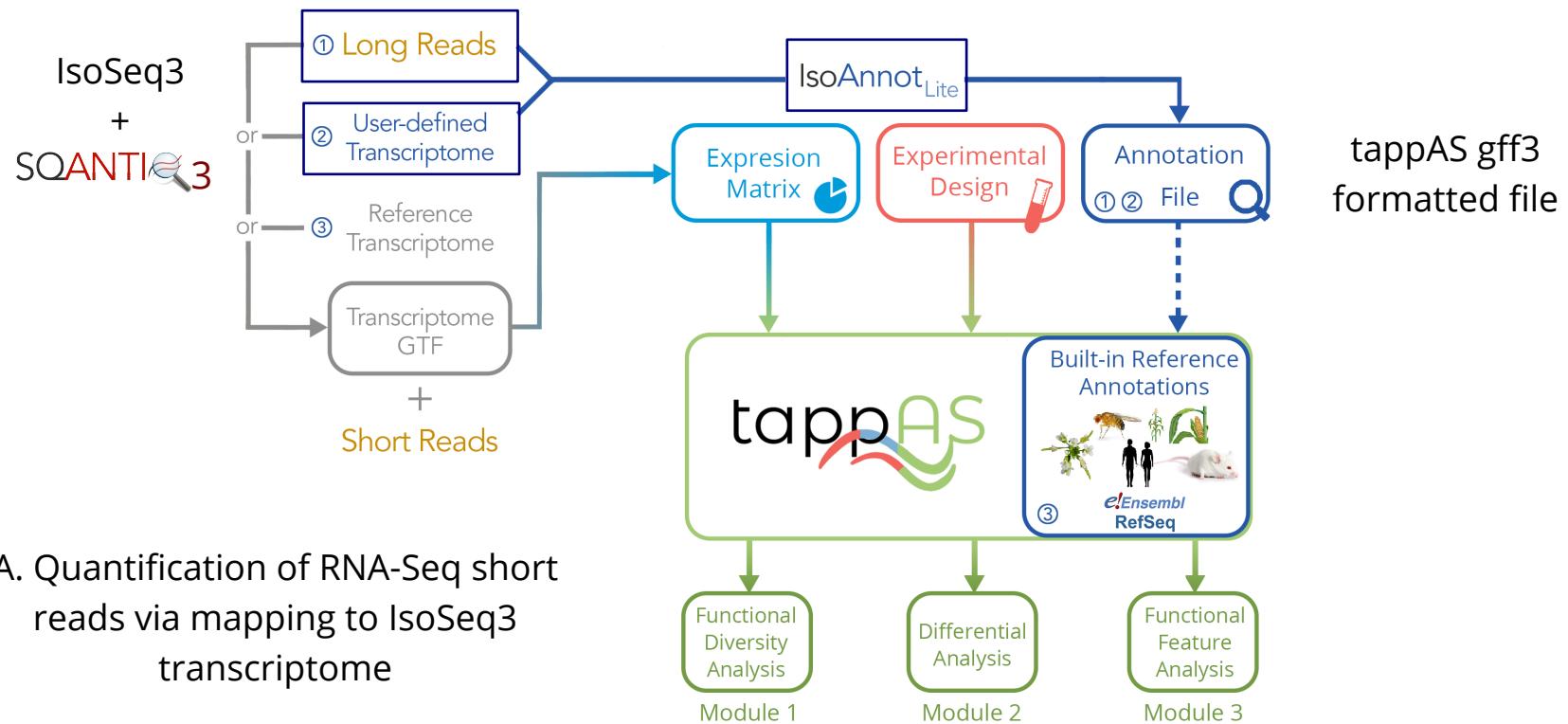


\*Provided by tappAS for supported species/transcriptomes, can also be provided by user.

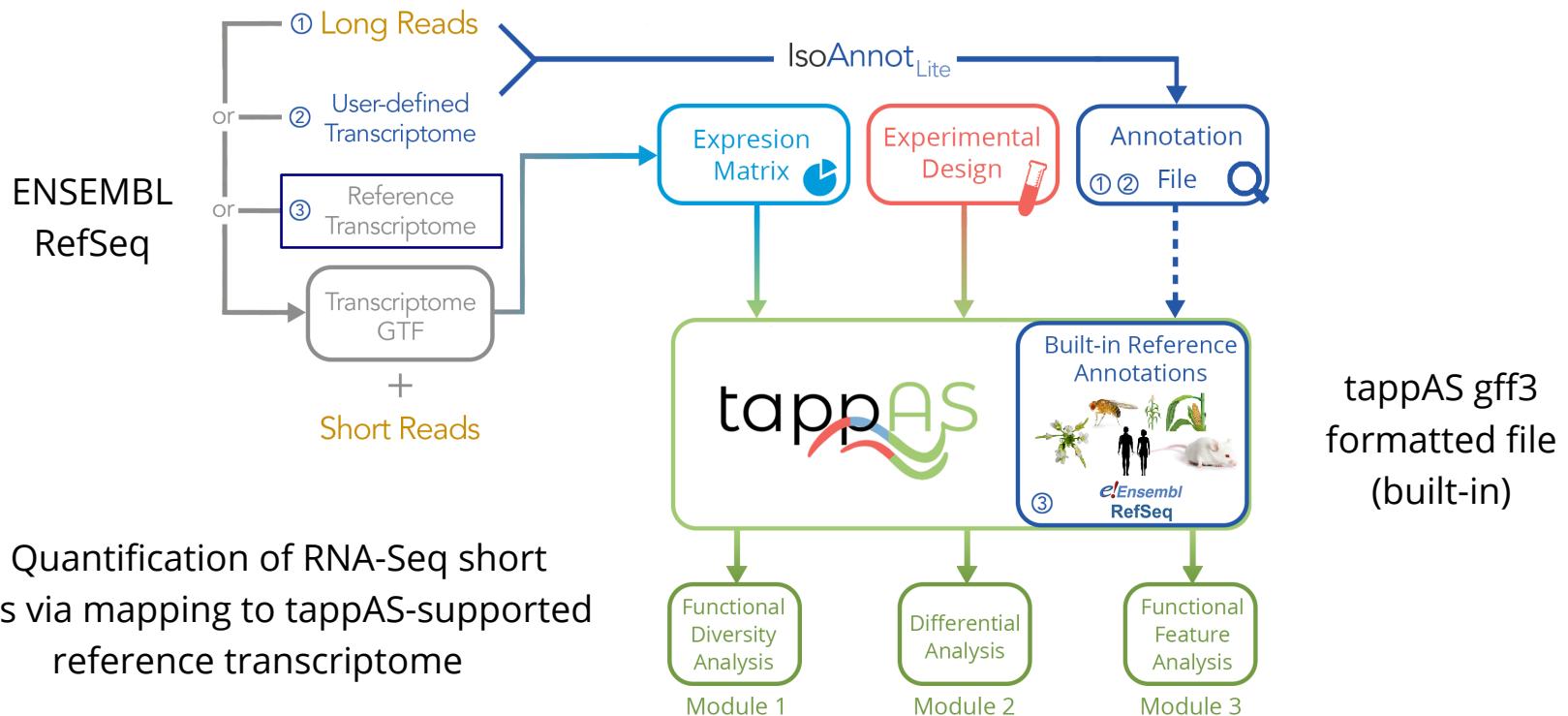
# How do I obtain an isoform-level expression matrix?



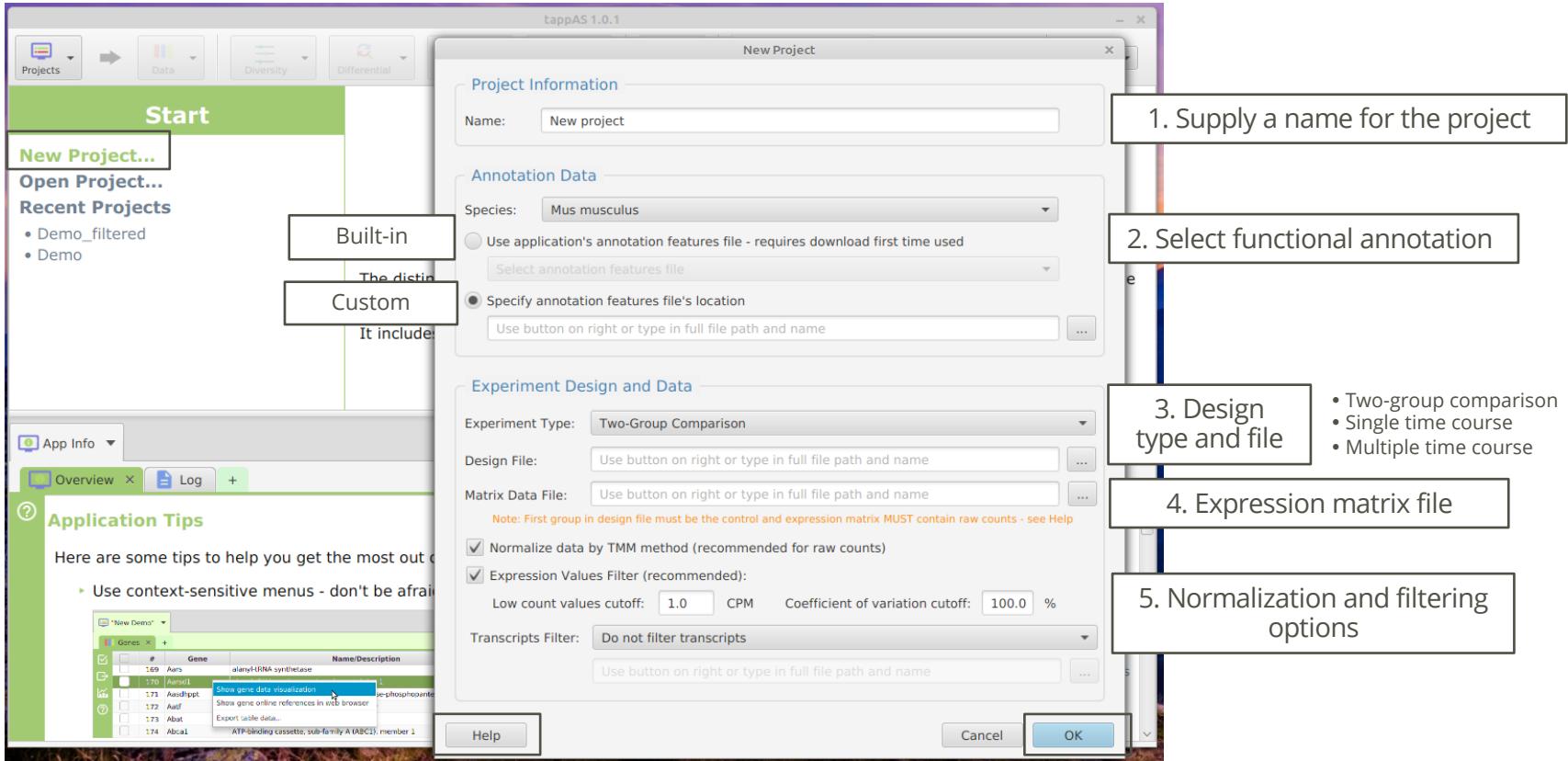
# How do I obtain an isoform-level expression matrix?



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



# Project creation in tappAS



The screenshot shows the tappAS 1.0.1 software interface. On the left, there's a sidebar with 'Start' (highlighted in green), 'New Project...', 'Open Project...', 'Recent Projects' (with items 'Demo\_filtered' and 'Demo'), 'App Info', and 'Overview' (selected). Below 'Overview' is an 'Application Tips' section with a tip about context-sensitive menus. The main workspace shows a 'Genes' table with several rows of gene information.

**New Project Dialog (Top Right):**

- 1. Supply a name for the project**: A text input field contains 'New project'.
- 2. Select functional annotation**: A dropdown 'Species' menu is set to 'Mus musculus'. Below it, two radio button options are shown: 'Use application's annotation features file - requires download first time used' (unchecked) and 'Specify annotation features file's location' (checked), with a text input field for the file path.
- 3. Design type and file**: Under 'Experiment Design and Data', the 'Experiment Type' is set to 'Two-Group Comparison'. Below it, a note says 'Note: First group in design file must be the control and expression matrix MUST contain raw counts - see Help'.
- 4. Expression matrix file**: Fields for 'Design File' and 'Matrix Data File' are present, each with a text input field and a browse button.
- 5. Normalization and filtering options**: A checkbox 'Normalize data by TMM method (recommended for raw counts)' is checked. Another checkbox 'Expression Values Filter (recommended)' is also checked. Below these are 'Low count values cutoff' (set to 1.0), 'CPM', 'Coefficient of variation cutoff' (set to 100.0), and '%'. A 'Transcripts Filter' dropdown is set to 'Do not filter transcripts'.

**Buttons at the bottom of the dialog:**

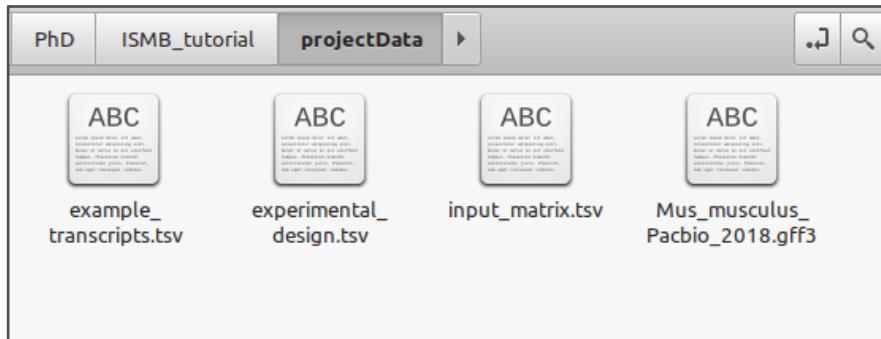
- Help
- Cancel
- OK

# 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)**

Demo - tappAS 0

Projects Data Diversity Differential Features Enrichment Graphs Search ID/Name Hide unselected rows

\* Demo Transcripts DIU Transcripts FEA: Enrichment GO DFI: DFI domains DFI Results Summary: DFI domains +

#	Gene	Gene Description	DIU Result	Q-Value	Major Isoform Switching	Total Usage Change	NSC Mea...	OLD Mea...
1	0610037L13Rik	RIKEN cDNA 0610037L13 g...	Not DIU	0.88975	NO	16.0	5704.11	4547.4
2	1110057K04Rik	RIKEN cDNA 1110057K04 ...	Not DIU	0.92229	NO	6.94	3042.39	4783.58
3	1110059E24Rik	RIKEN cDNA 1110059E24 ...	Not DIU	1.0	NO	3.42	4265.83	2332.36
4	1500004A13Rik	RIKEN cDNA 1500004A13 ...	Not DIU	1.0	NO	1.23	2396.21	1973.53
5	1600012H06Rik	RIKEN cDNA 1600012H06 ...	Not DIU	0.084965	YES	19.25	1665.64	2389.95
6	1700003M07Rik	RIKEN cDNA 1700003M07 ...	Not DIU	0.12164	NO	16.75	893.93	3453.9
7	170002011Rik	RIKEN cDNA 1700020114 g...	Not DIU	0.31602	NO	10.0	2574.16	2391.97
8	1700021K19Rik	RIKEN cDNA 1700021K19 ...	Not DIU	0.8239	NO	7.66	2139.94	2774.0
9	1700029J07Rik	RIKEN cDNA 1700029J07 g...	Not DIU	0.22182	NO	17.33	141.4	1626.51
10	1700030K09Rik	RIKEN cDNA 1700030K09 ...	Not DIU	1.0	NO	2.05	1862.82	1891.49
11	170008E04Rik	RIKEN cDNA 170008E04 ...	Not DIU	0.91178	NO	7.59	531.04	2417.65
12	2010315B03Rik	RIKEN cDNA 2010315B03 ...	DIU	4.2767E-4	NO	24.0	714.7	836.51
13	2310067B10Rik	RIKEN cDNA 2310067B10 ...	Not DIU	0.68631	NO	7.92	1796.03	2507.9
14	2410002F23Rik	RIKEN cDNA 2410002F23 ...	Not DIU	0.75014	NO	9.24	9486.02	5671.33
15	2610008E11Rik	RIKEN cDNA 2610008E11 ...	Not DIU	1.0	NO	3.66	1009.09	1135.41

App Info Demo Gria1 \*

Transcripts Proteins Expression Charts +

Gria1 - Glutamate receptor, ionotropic, AMPA1 (alpha 1)  
Proteins View - Aligned

N-terminal 0 100 200 300 400 500 600 700 800 900 C-terminal

F6YNQ1 [PB.1004.1] Provean Score: -71.09  
novelProt542 [PB.1004.2] Provean Score: -61.33  
P23818 [PB.1004.3] Provean Score: -12.59

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

Demo - tappAS 0.99.15

Help

Control panel (top bar)

The interface consists of two main windows. The top window, titled 'Demo - tappAS 0.99.15', is a 'Data' analysis panel. It features a top navigation bar with tabs for Data, Diversity, Differential, Features, Enrichment, and Graphs. Below this is a search bar labeled 'Search ID/Name' with a 'Hide unselected rows' checkbox. The main area displays a table of gene expression data with columns for Gene, DIU Result, Q-Value, Major Isoform Switching, Total Usage Change, NSC Mea..., and OLD Mea... . The bottom window, titled 'Gria1 - Glutamate receptor, ionotropic, AMPA1 (alpha 1)', is a 'Proteins' analysis panel. It shows three protein sequence alignments for F6YNQ1, novelProt542, and P23818, with Provean Scores of -71.09, -61.33, and -12.59 respectively. The alignments show the N-terminal and C-terminal ends of the proteins with various domains and motifs highlighted.

#	Gene	Gene Description	DIU Result	Q-Value	Major Isoform Switching	Total Usage Change	NSC Mea...	OLD Mea...
1	0610037L13Rik	RIKEN cDNA 0610037L13 g...	Not DIU	0.88975	NO	16.0	5704.11	4547.4
2	1110057K04Rik	RIKEN cDNA 1110057K04 ...	Not DIU	0.92229	NO	6.94	3042.39	4783.58
3	1110059E24Rik	RIKEN cDNA 1110059E24 ...	Not DIU	1.0	NO	3.42	4265.83	2332.36
4	1500004A13Rik	RIKEN cDNA 1500004A13 ...	Not DIU	1.0	NO	1.23	2396.21	1973.53
5	1600012H06Rik	RIKEN cDNA 1600012H06 ...	Not DIU	0.084965	YES	19.25	1665.64	2389.95
6	1700003M07Rik	RIKEN cDNA 1700003M07 ...	Not DIU	0.12164	NO	16.75	893.93	3453.9
7	170002011R4Rik	RIKEN cDNA 1700020114 g...	Not DIU	0.31602	NO	10.0	2574.16	2391.97
8	1700021K19Rik	RIKEN cDNA 1700021K19 ...	Not DIU	0.8239	NO	7.66	2139.94	2774.0
9	1700029J07Rik	RIKEN cDNA 1700029J07 g...	Not DIU	0.22182	NO	17.33	141.4	1626.51
10	1700030K09Rik	RIKEN cDNA 1700030K09 ...	Not DIU	1.0	NO	2.05	1862.82	1891.49
11	170008E04Rik	RIKEN cDNA 170008E04 ...	Not DIU	0.91178	NO	7.59	531.04	2417.65
12	2010315B03Rik	RIKEN cDNA 2010315B03 ...	DIU	4.276764	NO	24.0	714.7	836.51
13	2310067B10Rik	RIKEN cDNA 2310067B10 ...	Not DIU	0.68631	NO	7.92	1796.03	2507.9
14	2410002F23Rik	RIKEN cDNA 2410002F23 ...	Not DIU	0.75014	NO	9.24	9486.02	5671.33
15	2610008E11Rik	RIKEN cDNA 2610008E11 ...	Not DIU	1.0	NO	3.66	1009.09	1135.41

App Info Demo "Gria1"

Gria1 - Glutamate receptor, ionotropic, AMPA1 (alpha 1)  
Proteins View - Aligned

N-terminal 0 100 200 300 400 500 600 700 800 900 C-terminal

F6YNQ1 [PB.1004.1] Provean Score: -71.09  
novelProt542 [PB.1004.2] Provean Score: -61.33  
P23818 [PB.1004.3] Provean Score: -12.59

3,886 438aa 860aa 907aa

# tappAS interface

Run analyses and view results

Projects
 Data
 Diversity
 Differential
 Features
 Enrichment
 Graphs

\*Demo\*
Transcripts
DIU Transcripts
FEA: Enrichment GO
DFI: DFI domains
DFI Results Summary: DFI domains
+ [New]

#	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	3
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	170002011Rik	RIKEN cDNA 1700020114 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	2
10	1700030K09Rik	RIKEN cDNA 1700030K09 ...	Not DIU	1.0	NO	2.05	1
11	170008E04Rik	RIKEN cDNA 170008E04 ...	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	1
15	2610008E11Rik	RIKEN cDNA 2610008E11 ...	Not DIU	1.0	NO	3.66	1

App Info
\*Demo\*
Gria1
+ [New]

Transcripts
Proteins
Expression Charts
+ [New]

**Gria1 - Glutamate receptor, ionotropic, /**  
Proteins View - Aligned

F6YNQ1 [PB.1004.1] Provean Score: -71.09

novelProt542 [PB.1004.2] Provean Score: -61.33

P23818 [PB.1004.3] Provean Score: -12.59

Control panel (top bar)

Project "Demo"

**Annotation Feature(s)**

- SIGNALP\_EUK
- PAR-CLIP
- PFAM
- TMHMM

You may select annotation features from multiple sources.

**Check Using:** Feature genomic position

**Analysis Method**

Method: DEXSeq

**Analysis Parameters**

Name:

Significance Level:

Filtering:  Filter minor isoforms

Fold filtering:  (default: 2)

Help Cancel OK

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

The screenshot displays the tappAS software interface. At the top, there is a navigation bar with tabs: Projects, Data, Diversity, Differential, Features, Enrichment, and Graphs. A search bar labeled 'Search ID/Name' and a checkbox for 'Hide unselected rows' are also present. Below the navigation bar, a dropdown menu shows 'Demo'. The main area contains two panels. The top panel is titled 'Analysis result table' and shows a table with columns: #, Gene, Gene Description, DIU Result, Q-Value, Major Isoform Switching, Total Usage Change, NSC Mea..., and OLD Mea... (partially visible). The table lists 15 entries, each corresponding to a RIKEN cDNA entry. The bottom panel is titled 'Gria1 - Glutamate receptor, ionotropic, AMPA1 (alpha 1) Proteins View - Aligned' and displays three protein alignment charts for F6YNQ1, novelProt542, and P23818. Each chart shows the protein sequence with various domains and scores.

#	Gene	Gene Description	DIU Result	Q-Value	Major Isoform Switching	Total Usage Change	NSC Mea...	OLD Mea...	
1	0610037L13Rik	RIKEN cDNA 0610037L13 g...	Not DIU	0.88975	NO		16.0	5704.11	4547.4
2	1110057K04Rik	RIKEN cDNA 1110057K04 g...	Not DIU	0.92229	NO		6.94	3042.39	4783.58
3	1110059E24Rik	RIKEN cDNA 1110059E24 g...	Not DIU	1.0	NO		3.42	4265.83	2332.36
4	1500004A13Rik	RIKEN cDNA 1500004A13 g...	Not DIU	1.0	NO		1.23	2396.21	1973.53
5	1600012H06Rik	RIKEN cDNA 1600012H06 g...	Not DIU	0.084965	YES				
6	1700003M07Rik	RIKEN cDNA 1700003M07 g...	Not DIU	0.12164	NO				
7	2010020114Rik	RIKEN cDNA 2010020114 g...	Not DIU	0.31602	NO				
8	1700021K19Rik	RIKEN cDNA 1700021K19 g...	Not DIU	0.8239	NO		17.33	141.4	1626.51
9	1700029J7Rik	RIKEN cDNA 1700029J07 g...	Not DIU	0.22182	NO				
10	1700030K09Rik	RIKEN cDNA 1700030K09 g...	Not DIU	1.0	NO		2.05	1862.82	1891.49
11	1700088E04Rik	RIKEN cDNA 1700088E04 g...	Not DIU	0.91178	NO		7.59	531.04	2417.65
12	2010315B03Rik	RIKEN cDNA 2010315B03 g...	DIU	4.2767E-4	NO		24.0	714.7	836.51
13	2310067B10Rik	RIKEN cDNA 2310067B10 g...	Not DIU	0.68631	NO		7.92	1796.03	2507.9
14	2410002F23Rik	RIKEN cDNA 2410002F23 g...	Not DIU	0.75014	NO		9.24	9486.02	5671.33
15	2610008E11Rik	RIKEN cDNA 2610008E11 g...	Not DIU	1.0	NO		3.66	1009.09	1135.41

Gria1 - Glutamate receptor, ionotropic, AMPA1 (alpha 1)  
Proteins View - Aligned

N-terminal 0 100 200 300 400 500 600 700 800 900 C-terminal

F6YNQ1 [PB.1004.1] Provean Score: -71.09  
novelProt542 [PB.1004.2] Provean Score: -61.33  
P23818 [PB.1004.3] Provean Score: -12.59

3,886 438aa 860aa 907aa

## 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**

**Data and result tables panel**

**Analysis result table**

#	Gene	Gene Description	DIU Result	Q-Value	Major Isoform Switching	Total Usage Change	NSC Mea...	OLD Mea...
1	0610037L13Rik	RIKEN cDNA 0610037L13 g...	Not DIU	0.88975	NO	16.0	5704.11	4547.4
2	1110057K04Rik	RIKEN cDNA 1110057K04 ...	Not DIU	0.92229	NO	6.94	3042.39	4783.58
3	1110059E24Rik	RIKEN cDNA 1110059E24 ...	Not DIU	1.0	NO	3.42	4265.83	2332.36
4	1500004A13Rik	RIKEN cDNA 1500004A13 ...	Not DIU	1.0	NO	1.23	2396.21	1973.53
5	1600012H06Rik	RIKEN cDNA 1600012H06 ...	Not DIU	0.084965	YES			
6	1700003M07Rik	RIKEN cDNA 1700003M07 ...	Not DIU	0.12164	NO			
7	2010020114Rik	RIKEN cDNA 2010020114 ...	Not DIU	0.31602	NO			
8	1700021K19Rik	RIKEN cDNA 1700021K19 ...	Not DIU	0.8239	NO	17.33	141.4	1626.51
9	1700029J7Rik	RIKEN cDNA 1700029J07 ...	Not DIU	0.22182	NO			
10	1700030K09Rik	RIKEN cDNA 1700030K09 ...	Not DIU	1.0	NO	2.05	1862.82	1891.49
11	1700088E04Rik	RIKEN cDNA 1700088E04 ...	Not DIU	0.91178	NO	7.59	531.04	2417.65
12	2010315B03Rik	RIKEN cDNA 2010315B03 ...	DIU	4.2767E-4	NO	24.0	714.7	836.51
13	2310067B10Rik	RIKEN cDNA 2310067B10 ...	Not DIU	0.68631	NO	7.92	1796.03	2507.9
14	2410002F23Rik	RIKEN cDNA 2410002F23 ...	Not DIU	0.75014	NO	9.24	9486.02	5671.33
15	2610008E11Rik	RIKEN cDNA 2610008E11 ...	Not DIU	1.0	NO	3.66	1009.09	1135.41

**Gria1 - Glutamate receptor, ionotropic, AMPA1 (alpha 1)**  
**Proteins View - Aligned**

F6YNQ1 [PB.1004.1] Provean Score: -71.09  
novelProt542 [PB.1004.2] Provean Score: -61.33  
P23818 [PB.1004.3] Provean Score: -12.59

Annotations include: 3.886, 438aa, 860aa, 907aa.

# tappAS interface

Analysis-specific visualization options

The screenshot shows the tappAS software interface with several panels and subtabs:

- Top Navigation Bar:** Projects, Data, Diversity, Differential, Features, Enrichment, Graphs.
- Search Bar:** Search ID/Name, Hide unselected rows.
- Subtab Selection:** Transcripts, DIU Transcripts (highlighted), FEA: Enrichment GO, DFI: DF1 domains, DFI Results Summary: DF1 domains.
- Data Table:** Shows a list of genes with various metrics. A callout box points to the "Visualization subtabs: result-specific graphics" subtab.

#	Gene	Gene Description	DIU Result	Q-Value	Major Isoform Switching	Total Usage Change	NSC Mea...	OLD Mea...
1	0610037L13Rik	RIKEN cDNA 0610037L13 g...	Not DIU	0.88975	NO	16.0	5704.11	4547.4
2	1110057K04Rik	RIKEN cDNA 1110057K04 ...	Not DIU	0.92229	NO	6.94	3042.39	4783.58
3	1110059E24Rik	RIKEN cDNA 1110059E24 ...	Not DIU	1.0	NO	3.42	4265.83	2332.36
4	1500004A13Rik	RIKEN cDNA 1500004A13 ...	Not DIU	1.0	NO	1.23	2396.21	1973.53
5	1600012H06Rik	RIKEN cDNA 1600012H06 ...	Not DIU	0.084965	YES	19.25	1665.64	2389.95
6	1700003M07Rik	RIKEN cDNA 1700003M07 ...	Not DIU	0.12164	NO	16.75	893.93	3453.9
7	1700020I14Rik	RIKEN cDNA 1700020I14 ...	Not DIU	0.31602	NO	10.0	2574.16	2391.97
8	1700021K19Rik	RIKEN cDNA 1700021K19 ...	Not DIU	0.8239	NO	7.66	2139.94	2774.0
9	1700029J07Rik	RIKEN cDNA 1700029J07 ...	Not DIU	0.22182	NO	17.33	141.4	1626.51
10	1700030K09Rik	RIKEN cDNA 1700030K09 ...	Not DIU	1.0	NO	2.05	1862.82	1891.49
11	1700088E04Rik	RIKEN cDNA 1700088E04 ...	Not DIU					
12	2010315B03Rik	RIKEN cDNA 2010315B03 ...	DIU					
13	2310067B10Rik	RIKEN cDNA 2310067B10 ...	Not DIU					
14	2410002F23Rik	RIKEN cDNA 2410002F23 ...	Not DIU	0.75014	NO	9.24	9486.02	5671.33
15	2610008E11Rik	RIKEN cDNA 2610008E11 ...	Not DIU	1.0	NO	3.66	1009.09	1135.41

- Result-Specific Graphics Subtab:** Shows a callout box labeled "Visualization subtabs: result-specific graphics".
- Gene-Level Visualization Subtab:** Shows a callout box labeled "Gene-level visualization".
- Protein View - Aligned:** Displays protein sequences for Gria1, F6YNQ1, novelProt542, and P23818, showing their alignment and Provean scores.

Multi-option side bar

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

Visualization panel

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

# tappAS interface

Demo - tappAS 0.99.15

Transcripts DIU Transcripts FEA: Enrichment GO DFI: DF1 domains DFI Results Summary: DF1 domains +

#	Gene	Gene Description	DIU Result	Q-Value	Major Isoform Switching	Total Usage Change	NSC Me...	OLD Me...
1	0610037L13Rik	RIKEN cDNA 0610037L13 g...	Not DIU	0.88975	NO	16.0	5704.11	4547.4
2	1110057K04Rik	RIKEN cDNA 1110057K04 ...	Not DIU	0.92229	NO	6.94	3042.39	4783.58
3	1110059E24Rik	RIKEN cDNA 1110059E24 ...	Not DIU	1.0	NO	3.42	4265.83	2332.36
4	1500004A13Rik	RIKEN cDNA 1500004A13 ...	Not DIU	1.0	NO	1.23	2396.21	1973.53
5	1600012H06Rik	RIKEN cDNA 1600012H06 ...	Not DIU	0.084965	YES	19.25	1665.64	2389.95
6	1700003M07Rik	RIKEN cDNA 1700003M07 ...	Not DIU	0.12164	NO	16.75	893.93	3453.9
7	170002011Rik	RIKEN cDNA 1700020114 g...	Not DIU	0.31602	NO	10.0	2574.16	2391.97
8	1700021K19Rik	RIKEN cDNA 1700021K19 ...	Not DIU	0.8239	NO	7.66	2139.94	2774.0
9	1700029J07Rik	RIKEN cDNA 1700029J07 g...	Not DIU	0.22182	NO	17.33	141.4	1626.51
10	1700030K09Rik	RIKEN cDNA 1700030K09 ...	Not DIU	1.0	NO	2.05	1862.82	1891.49
11	170008E04Rik	RIKEN cDNA 170008E04 ...	Not DIU	0.91178	NO	7.59	531.04	2417.65
12	2010315B03Rik	RIKEN cDNA 2010315B03 ...	DIU	4.276764	NO	24.0	714.7	836.51
13	2310067B10Rik	RIKEN cDNA 2310067B10 ...	Not DIU	0.68631	NO	7.92	1796.03	2507.9
14	2410002F23Rik	RIKEN cDNA 2410002F23 ...	Not DIU	0.75014	NO	9.24	9486.02	5671.33
15	2610008E11Rik	RIKEN cDNA 2610008E11 ...	Not DIU	1.0	NO	3.66	1009.09	1135.41

App Info Demo "Gria1"

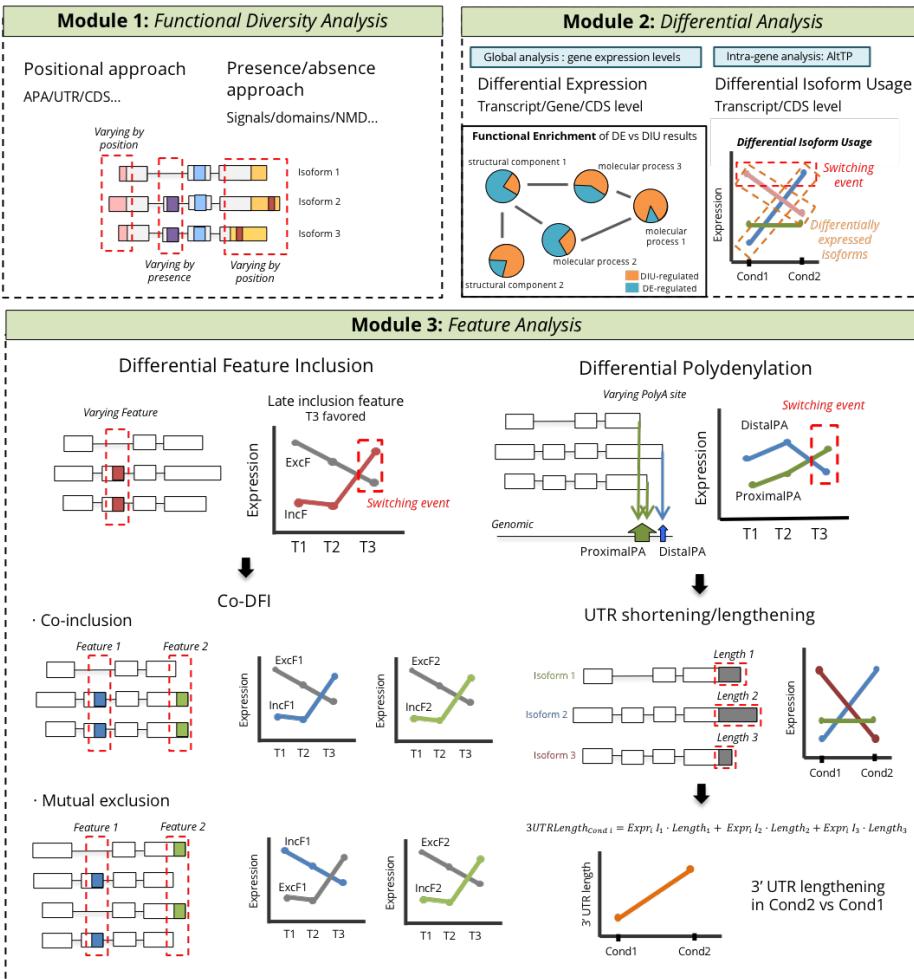
Gria1 - Glutamate receptor, ionotropic, AMPA1 (alpha 1)  
Proteins View - Aligned

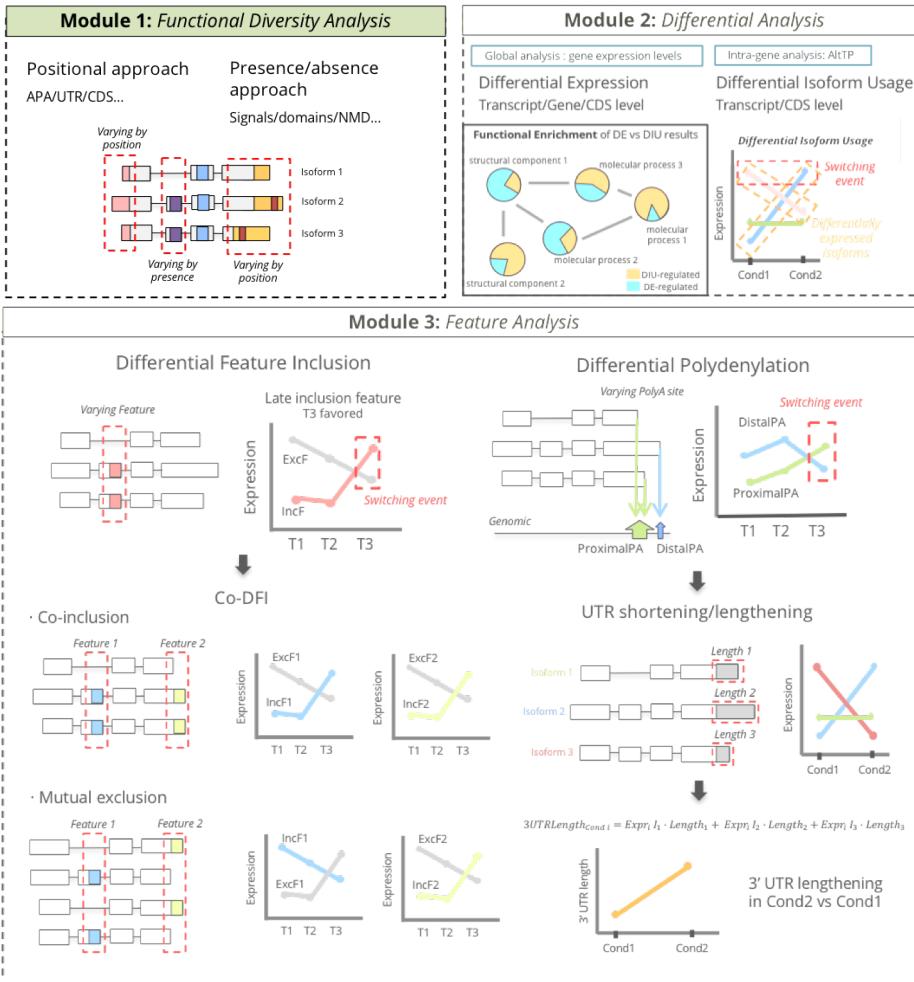
N-terminal 0 100 200 300 400 500 600 700 800 900 C-terminal

F6YNQ1 [PB.1004.1] Provean Score: -71.09  
novelProt542 [PB.1004.2] Provean Score: -61.33  
P23818 [PB.1004.3] Provean Score: -12.59

# tappAS analysis modules

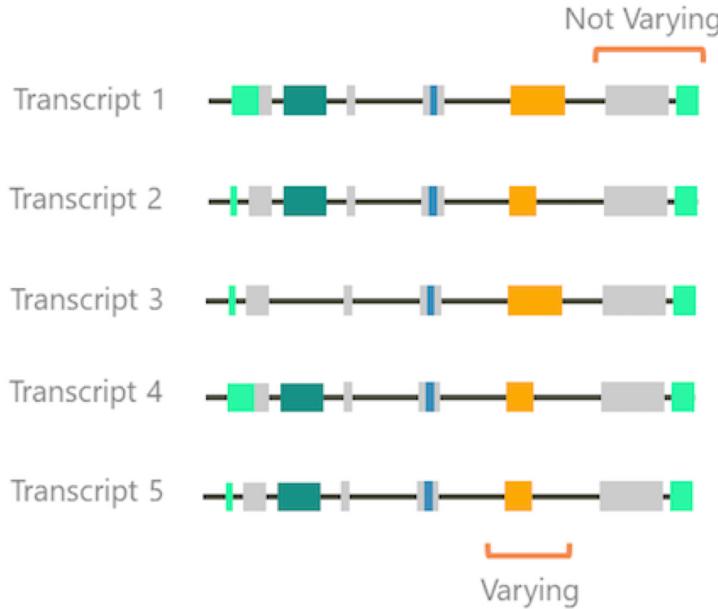
---



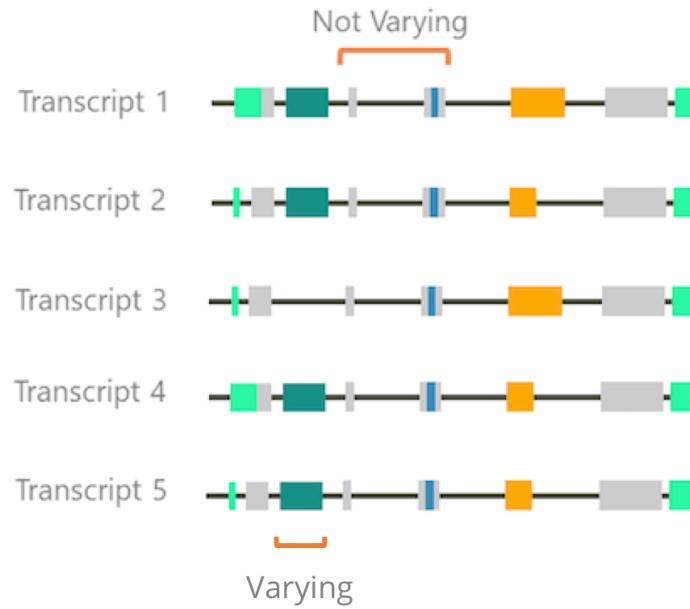


# Functional Diversity Analysis

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)**

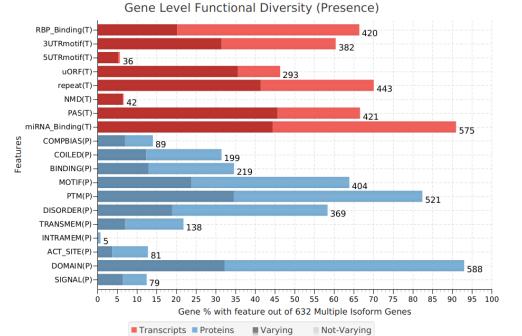
- SIGNALP\_EUK
- PAR-CLIP
- PFAM
- TMHMM
- scanForMotifs
- RepeatMasker
- UniProtKB/Swiss-Prot\_Phosphosite
- COILS

[Clear All](#) [Check All](#)

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

## Category result:

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

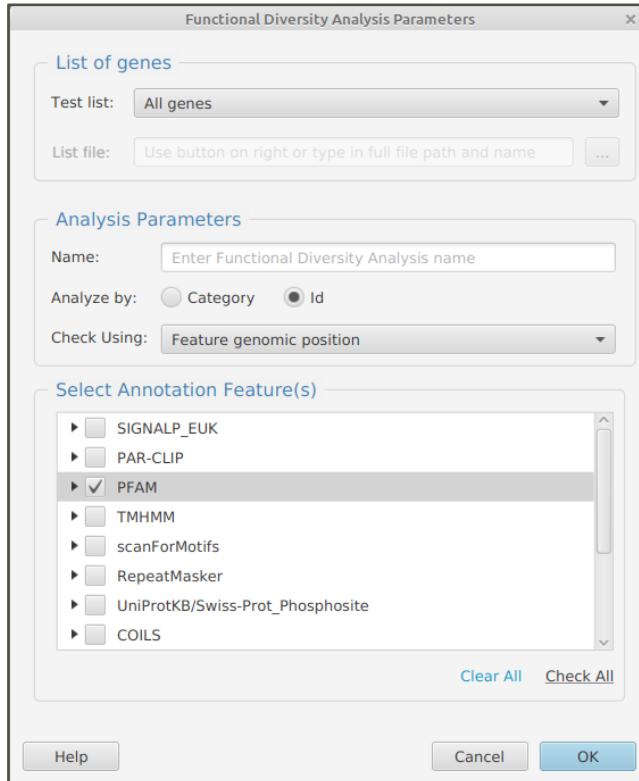


FDA: FDA\_pres X +

**Transcript Annotation**

#	Gene	NMD	RBP_Binding	repeat	SUTRmotif	uORF	miRNA_Binding	3UTRmotif	PAS
1	Cyth4						Not Varying	Not Varying	Not Varying
2	Rnh1				Varying	Varying	Not Varying	Not Varying	Not Varying
3	Pop4			Not Varying	Not Varying	Varying	Not Varying	Not Varying	Not Varying
4	Smc6			Varying	Varying	Varying	Varying	Varying	Varying
5	Ganc				Varying	Not Varying	Varying	Varying	Varying
6	Smc2					Not Varying	Varying	Varying	Varying
7	Ipo11			Varying	Not Varying		Not Varying	Not Varying	Varying
8	Tex264				Not Varying	Not Varying		Not Varying	
9	Gm10825					Not Varying	Not Varying	Not Varying	Varying
10	Scp2			Varying			Not Varying	Not Varying	Not Varying
11	Myc			Not Varying	Varying		Not Varying	Not Varying	Not Varying
12	Shmt1				Varying		Varying	Varying	Varying
13	Akt2			Varying	Varying		Not Varying	Varying	Not Varying
14	Cyb5d2				Not Varying		Varying	Not Varying	Not Varying

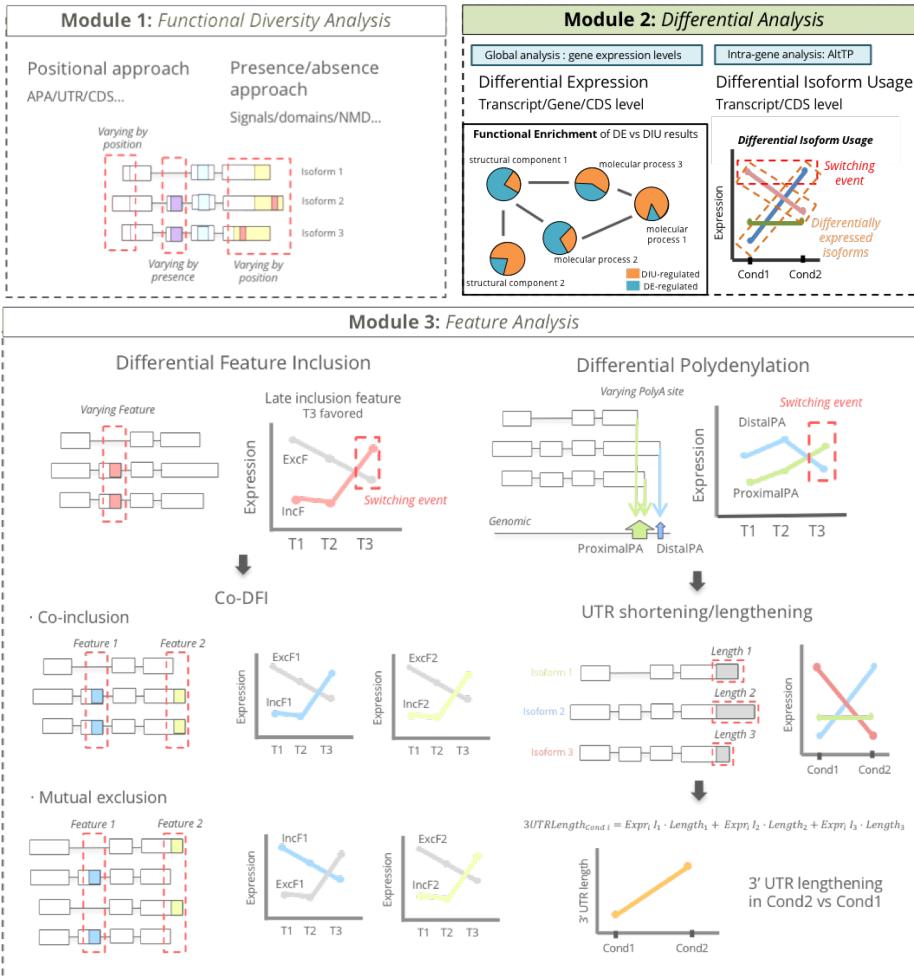
# Functional Diversity Analysis (FDA)



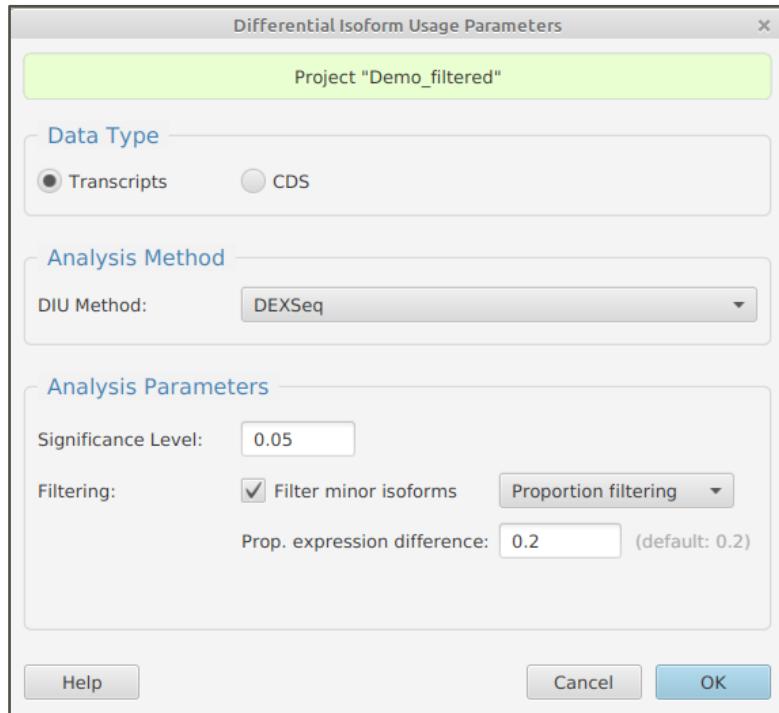
ID result:

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

#	Feature ID	Feature Description	P-Value	AdjP-Value	Protein Annotation	
					VARYING	NOT VARYING
1	PFAM/DOMAIN/PF00169	PH domain	0.005	1.0	8	3
2	PFAM/DOMAIN/PF00069	Protein kinase domain	0.4258	1.0	7	13
3	PFAM/DOMAIN/PF00096	Zinc finger, C2H2 type	0.241	1.0	6	8
4	PFAM/DOMAIN/PF00076	RNA recognition motif. (a.k.a. RRM,...)	0.7906	1.0	5	15
5	PFAM/DOMAIN/PF00041	Fibronectin type III domain	0.0771	1.0	4	2
6	PFAM/DOMAIN/PF12796	Ankyrin repeats (3 copies)	0.2902	1.0	4	5
7	PFAM/DOMAIN/PF13927	Immunoglobulin domain	0.1745	1.0	3	2
8	PFAM/DOMAIN/PF01352	KRAB box	0.3729	1.0	3	4
9	PFAM/DOMAIN/PF13432	Tetratricopeptide repeat	0.0952	1.0	2	0
10	PFAM/DOMAIN/PF00307	Calponin homology (CH) domain	0.0952	1.0	2	0
11	PFAM/DOMAIN/PF01756	Acyl-CoA oxidase	0.0952	1.0	2	0
12	PFAM/DOMAIN/PF00433	Protein kinase C terminal domain	0.0952	1.0	2	0
13	PFAM/DOMAIN/PF009379	FERM N-terminal domain	0.0952	1.0	2	0
14	PFAM/DOMAIN/PF07719	Tetratricopeptide repeat	0.0952	1.0	2	0

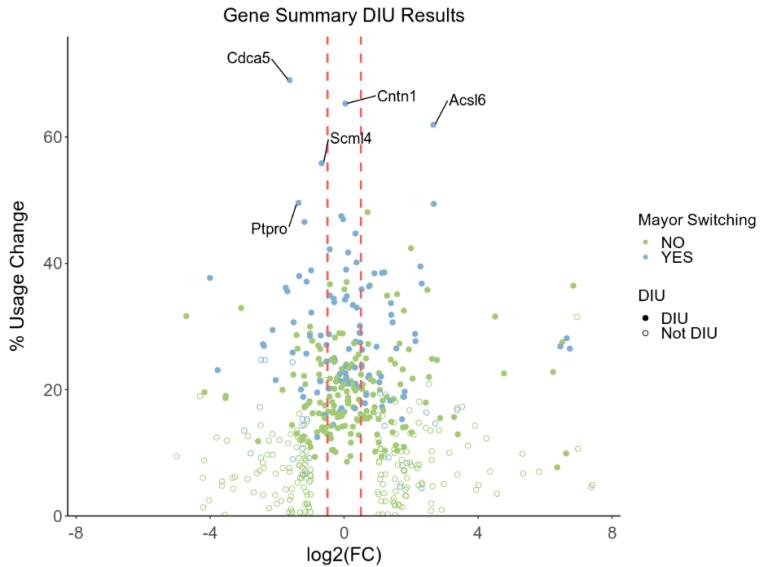


# Differential Isoform Usage Analysis (DIU)

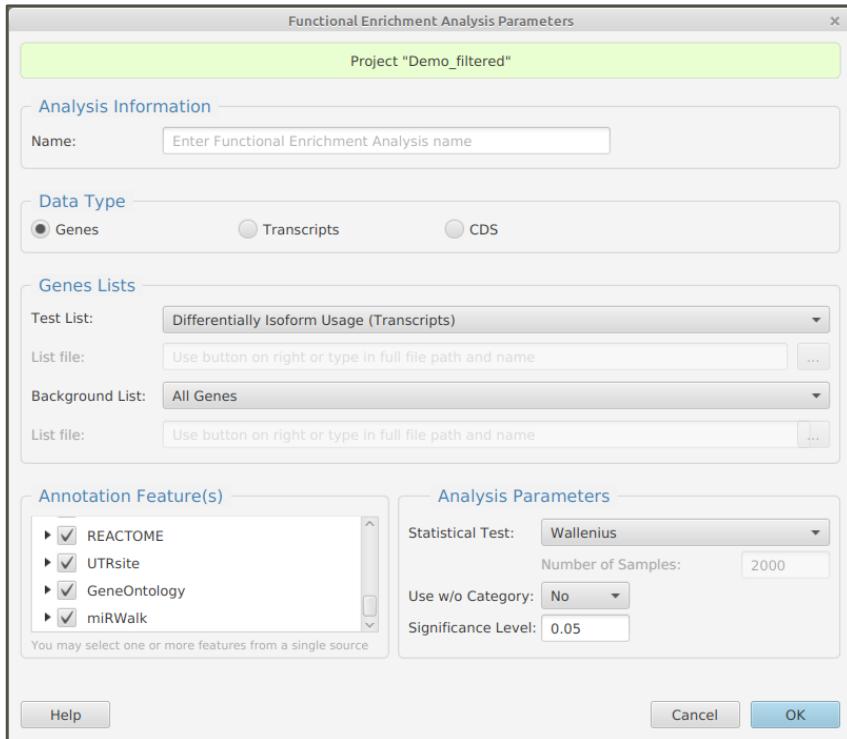


Result:

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

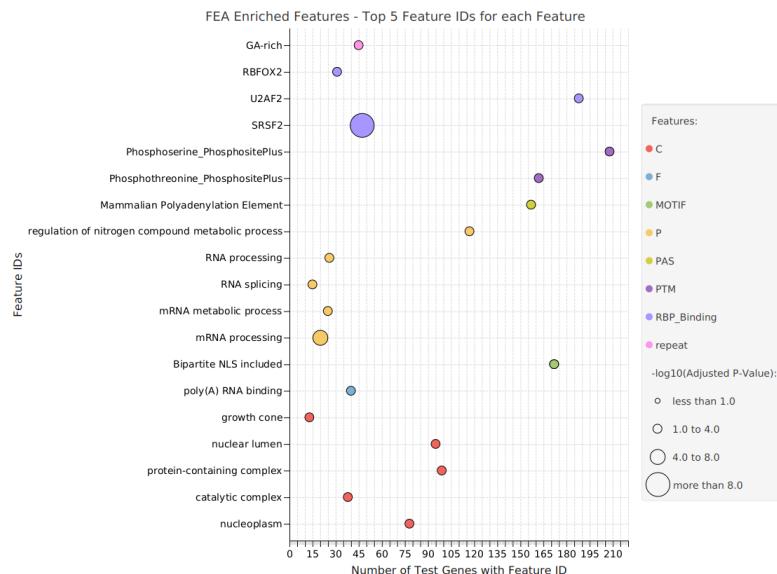


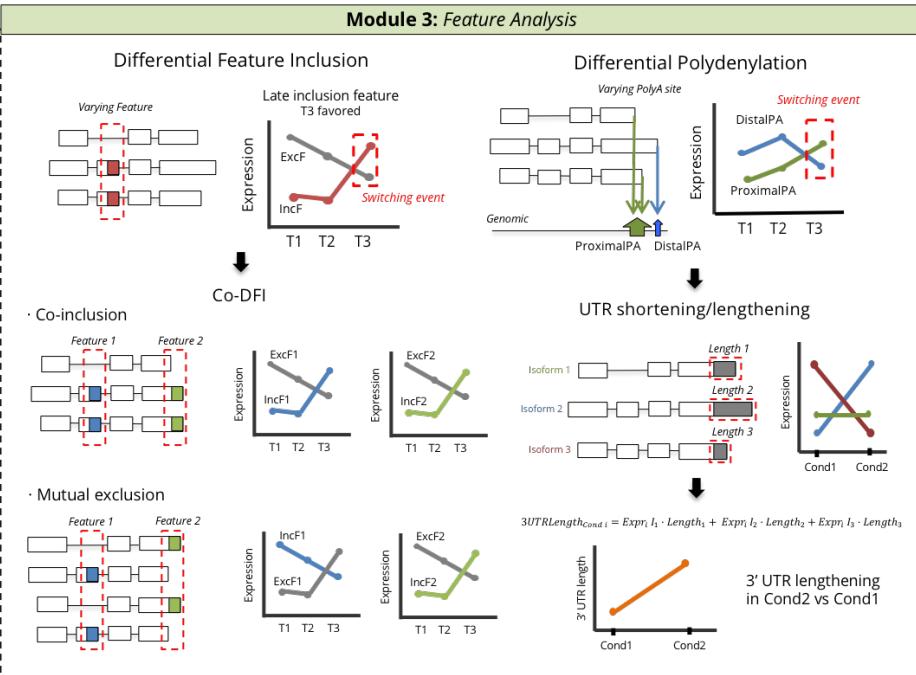
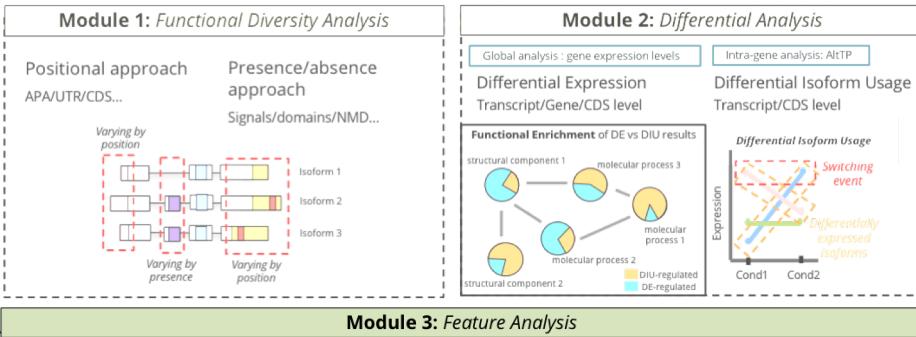
# Enrichment analyses (FEA and GSEA)



Result:

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





# 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:

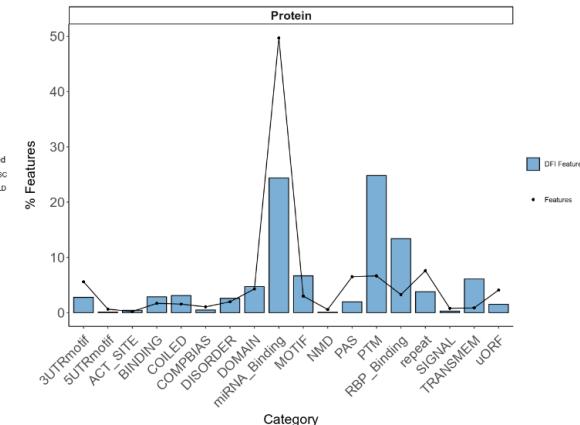
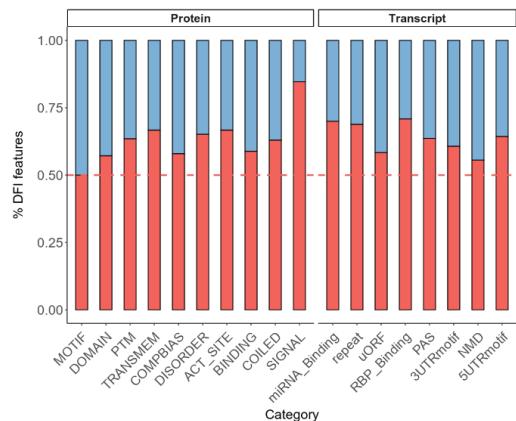
- Filter minor isoforms
- [Proportion filtering](#)

Prop. expression difference:  (default: 0.2)

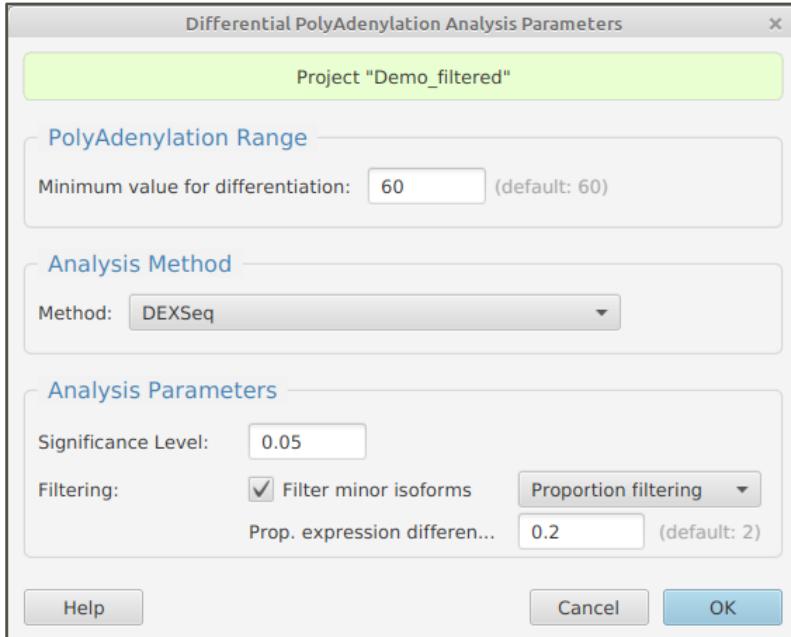
[Help](#) [Cancel](#) [OK](#)

## Result:

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

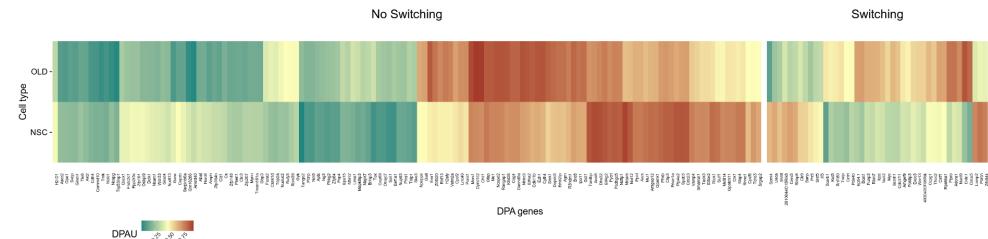


# Differential Polyadenylation Analysis (DPA)



## 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)

UTR Lengthening Analysis Parameters

Project "Demo\_filtered"

Lengthening Range

Minimum value for differentiation:  (default: 100)

Analysis Method

Method:

Analysis Parameters

Significance Level:

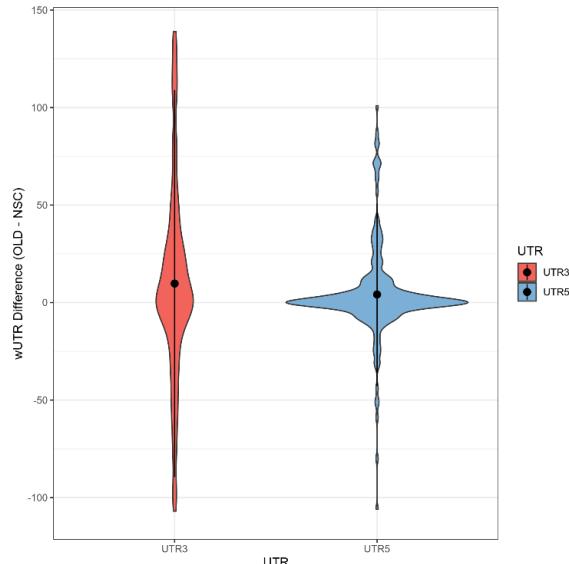
Filtering:

Filter minor isoforms  

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

Result:

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



# Hands on time!

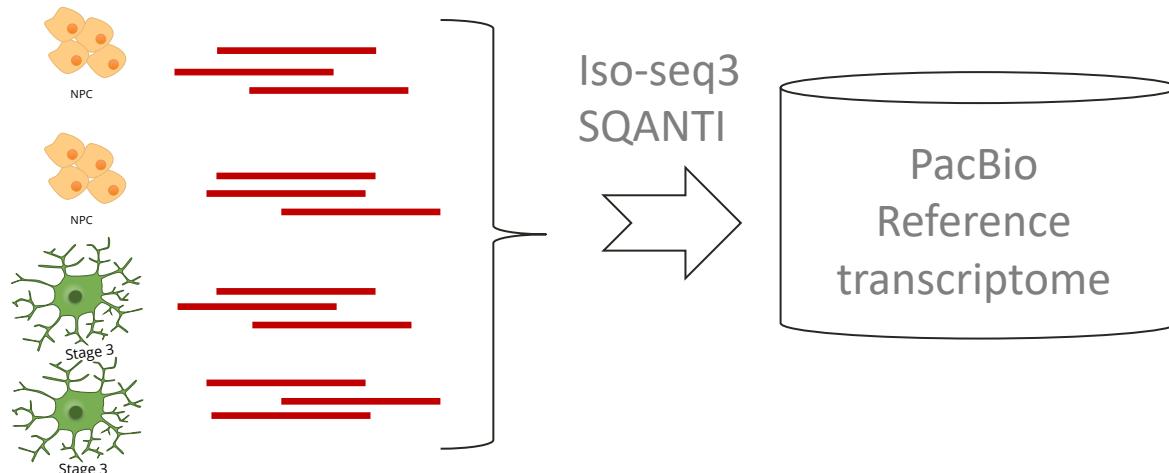
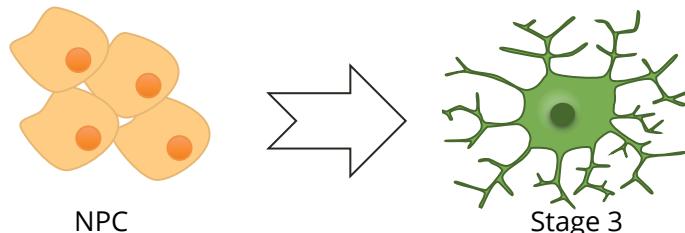
---



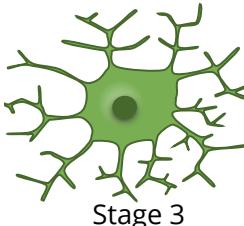
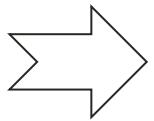
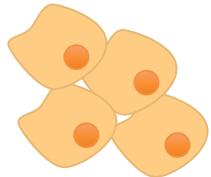
# Demo data: OPC differentiation from NPC



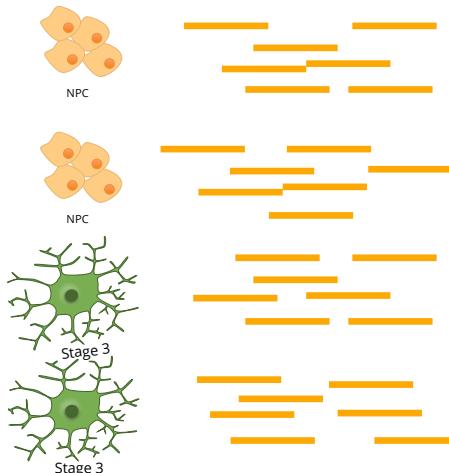
Neural Precursor Cells      Oligodendrocytes



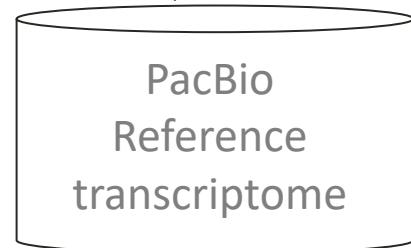
# Demo data: OPC differentiation from NPC



illumina®



Mapping  
STAR



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

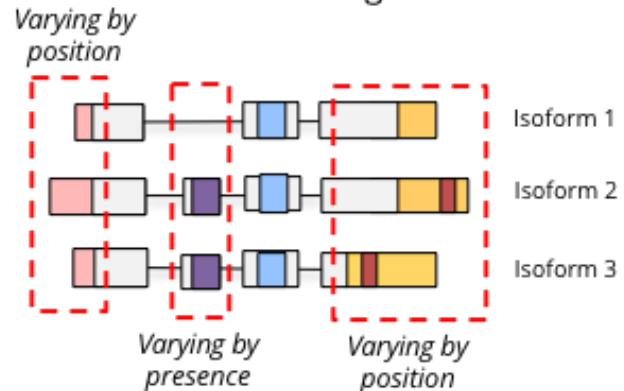
# tappAS hands-on

## Exercise 1: Functional Diversity Analysis (FDA)

### Module 1: Functional Diversity Analysis

Positional approach  
APA/UTR/CDS...

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



# 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?

# tappAS hands-on

## 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!

# tappAS hands-on

## 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.

# tappAS hands-on

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

### Module 2: Differential Analysis

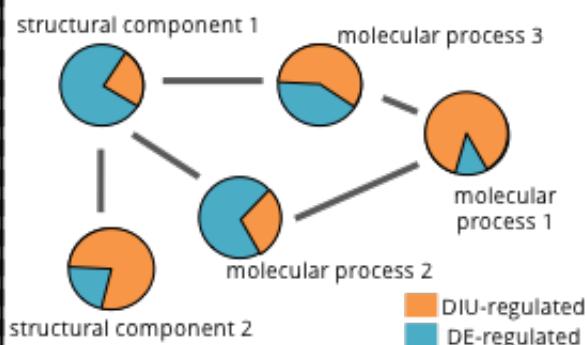
Global analysis : gene expression levels

Differential Expression  
Transcript/Gene/CDS level

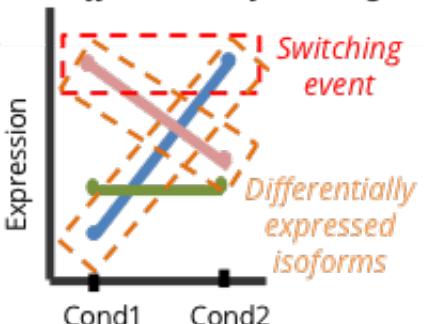
Intra-gene analysis: AltTP

Differential Isoform Usage  
Transcript/CDS level

#### Functional Enrichment of DE vs DIU results



#### Differential Isoform Usage



# tappAS hands-on

## 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?

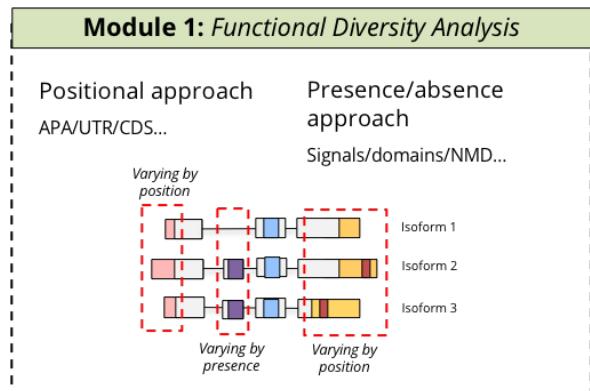
[Tip: you can right click on any functional category, and select «Drill down data» to view the genes that contain a given feature within the test list.]

**1. Extra excercise:** 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).

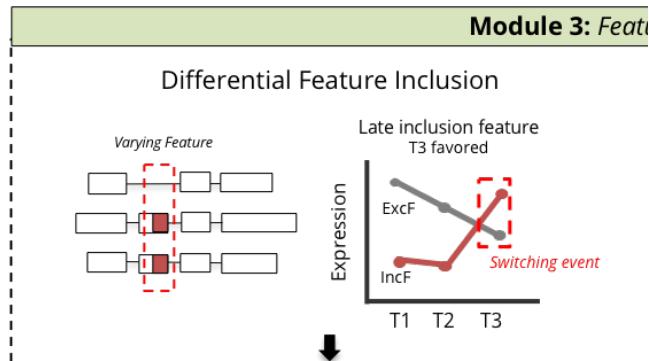
## Exercise 4: Differential Feature Inclusion (DFI) analysis

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



# 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.  
[!] 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?

# tappAS hands-on

## 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?

# tappAS hands-on

## 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.

# tappAS hands-on

## 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?