TEC_splicing_2021

Annotations for Alternative Splicing Events (ASEs) were generated from the hg38 or mm10 GFT files (igenome_hg38.gtf, igenome_mm10.gtf) using MATS 3.0.8.
We considered the MATS-generated files "fromGTF.XXX.txt" for A3SS, A5SS, RI and SE.
Human ASE annotations are in the folder: "Human hg38-based ASEs"
Mouse ASE annotations are in the folder: "Mouse mm10-based ASEs"
Alternatively, custom files based on the same file structure as "fromGTF.XXX.txt" can be considered for downstream analysis.

"ASE_label.R" parses the hg39 or mm10 GFT files and the MATS-generated ASE annotation files to tag each transcript isoform with splice-in or -out ASEs.
A single transcript isoform can feature multiple ASEs:
1 is for a spliced-in SE
2 for a spliced-in A3SS
3 for a spliced-in A5SS
4 for a spliced-in RI
9 for a spliced-out ASE
0 na
The output file is named "ASE_label_output.csv". An example is shown in the "Sample" folder.
"ASE_label_V2.R" is used by default. A more recent V3 version fixing some "warning issues" is provided.

############## example: Steap2

NM_001285471	Steap2	99
NM_001103157	Steap2	320
NM_001285470	Steap2	29
NM_028734	Steap2	20
NM_001103156	Steap2	921
NM_001285469	Steap2	20

- A "A5SS" is spliced-in in "NM_001103157" and spliced-out in "NM_001103156". The other four transcript isoforms are not related to this ASE.
- A "A3SS" is spliced-in in "NM_001103157", "NM_001285470", "NM_028734", "NM_001103156" and "NM_001285469", and spliced-out in "NM_001285471".
- A "SE" is spliced-in in "NM_001103156" and spliced-out in "NM_001285471" and "NM_001285470". The other three transcript isoforms are not related to this ASE.
############

"PSI_ASE.R" parses the "ASE_label_output.csv" file and the frequency files (for genes and isoforms) generated by Tophat2, as exemplified for a mTEChi sample in the "Sample" folder.
"PSI_ASE.R" calculates the percent splicing inclusion (PSI) of each ASE listed in the "ASE_label_output.csv" file.
The output file is named "PSI_ASE_output.csv". The output PSI file is provided in the "Sample" folder.

############ example: Steap2

```
Steap2            0.00587022
Steap2            1
Steap2            0.99918201
```

PSI values for each ASE identified in Steap2 for the mTEChi sample (provided in the "Sample" folder)
###########