alternativeSplicingEvents.hg19

April 11, 2018

`alternativeSplicingEvents.hg19.rda`
Alternative splicing event annotation for Human (hg19)

Description

Data frame containing alternative splicing events per event type. Each splicing event is characterised by its chromosome, strand, splice junction coordinates, associated gene and event identifier from the source program.

The following event types are available:

- Skipped Exon (SE)
- Mutually Exclusive Exon (MXE)
- Alternative First Exon (AFE)
- Alternative Last Exon (ALE)
- Alternative 3' Splice Site (A3SS)
- Alternative 5' Splice Site (A5SS)
- Retained Intron (RI)
- Tandem UTR

The splicing events were compiled from the annotation files used by the alternative splicing quantification tools MISO, VAST-TOOLS, SUPPA and rMATS (see details).

Details

The annotation files from MISO and VAST-TOLLS were retrieved from the links in "Sources". SUPPA and rMATS identify alternative splicing events and generate annotation files from a transcripts annotation file in GTF format.

The transcripts annotation file used was retrieved from UCSC Table Browser by selecting the GRCh37/hg19 assembly, "Genes and Gene Predictions" group, "UCSC Genes" track, "knownGene"" table for all genome in the GTF format. Unfortunately, this file has a bug where the "transcript_id" is the same as the "gene_id". To fix the transcripts' identifier from the GTF file, the exact same information was retrieved from UCSC Table Browser in TXT format and the following R script was run: https://gist.github.com/nuno-agostinho/24a5e4926b6e08e988c0

After obtaining the resulting annotation files, all files were parsed to obtain the identifier, chromosome, strand and coordinates of each splicing event per event type. Then, a full outer join was performed on the annotation of the different programs per event type using `dplyr::full_join`.

Note that before the outer join, some rMATS coordinates should increment one unit to be compared to the annotation of other files and the "chr" prefix present in the chromosome field of some annotation files was also removed. The genes were based on an annotation file containing gene symbols retrieved from the UCSC table browser.

For more details, check the `make-data.R` file.

Source

- **MISO:** https://miso.readthedocs.io/en/fastmiso/annotation.html
- **rMATS:** http://rnaseq-mats.sourceforge.net/user_guide.htm
- **SUPPA:** https://bitbucket.org/regulatorygenomicsupf/suppa
- **VAST-TOOLS:** http://vastdb.crg.eu/libs/

Transcripts annotation files used to generate events from SUPPA and rMATS and annotation with gene symbols were retrieved from UCSC Table Browser (https://genome.ucsc.edu/cgi-bin/hgTables).

References

- **MISO:** Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature methods. 2010;7(12):1009-1015. doi:10.1038/nmeth.1528.
- **rMATS:** Shen S, Park JW, Lu Z, et al. rMATS: Robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. Proceedings of the National Academy of Sciences of the United States of America. 2014;111(51):E5593-E5601. doi:10.1073/pnas.1419161111.
- **SUPPA:** Alamancos GP, Pagès A, Trincado JL, Bellora N, Eyras E. Leveraging transcript quantification for fast computation of alternative splicing profiles. RNA. 2015;21(9):1521-1531. doi:10.1261/rna.051557.115.
- **VAST-TOOLS:** Irimia M, Weatheritt RJ, Ellis J, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. Cell. 2014;159(7):1511-1523. doi:10.1016/j.cell.2014.11.035.

Examples

```
library(AnnotationHub)
hub <- AnnotationHub()

## Load the alternative splicing events for Human (hg19)
events <- query(hub, "alternativeSplicingEvents.hg19")[[1]]
```

Index