

# A user guide of ASpedia-R

## 1. A brief introduction

*ASpedia-R* is an R package, and has been developed in R 4.2. *ASpedia-R* composed of multi-omics profile involved in spliced exons and its flanking regions. To explore profile for each AS event, we generated AS event ID rule. One or multiple AS event IDs could query to interpolate multi-omics profile. Our tool is easily executable by sequential workflow: conversion, annotation, and visualization.

## 2. Installation

To install the latest version *ASpedia-R*, you will need to use the latest version of R. We provide *ASpedia-R* from github, so you could easily install to follow below guide line.

To install *ASpedia-R* from github:

```
> install.packages("devtools")
> devtools::install_github("ncc-bioinfo/ASpedia-R")
> library(ASpediaR)
```

## 3. To prepare input file: conversion

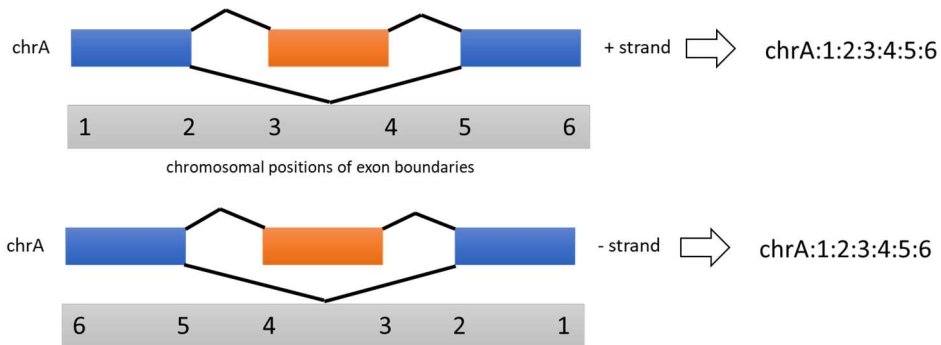
`asr_converter` formats the input file from DAS analysis tool result. The result files of differential AS test applications rMATS, SUPPA, and spliceR could be converted by our system. So, DAS analysis tools name is required to run `asr_converter`. Follow below example to generate AS event IDs.

```
## rMATS
> das.file.name <- system.file("extdata", "rMATS_test.txt", package="ASpediaR")
> rmats.convert.result <- asr_converter(das.file.name, program="rMATS",
as.type="SE")

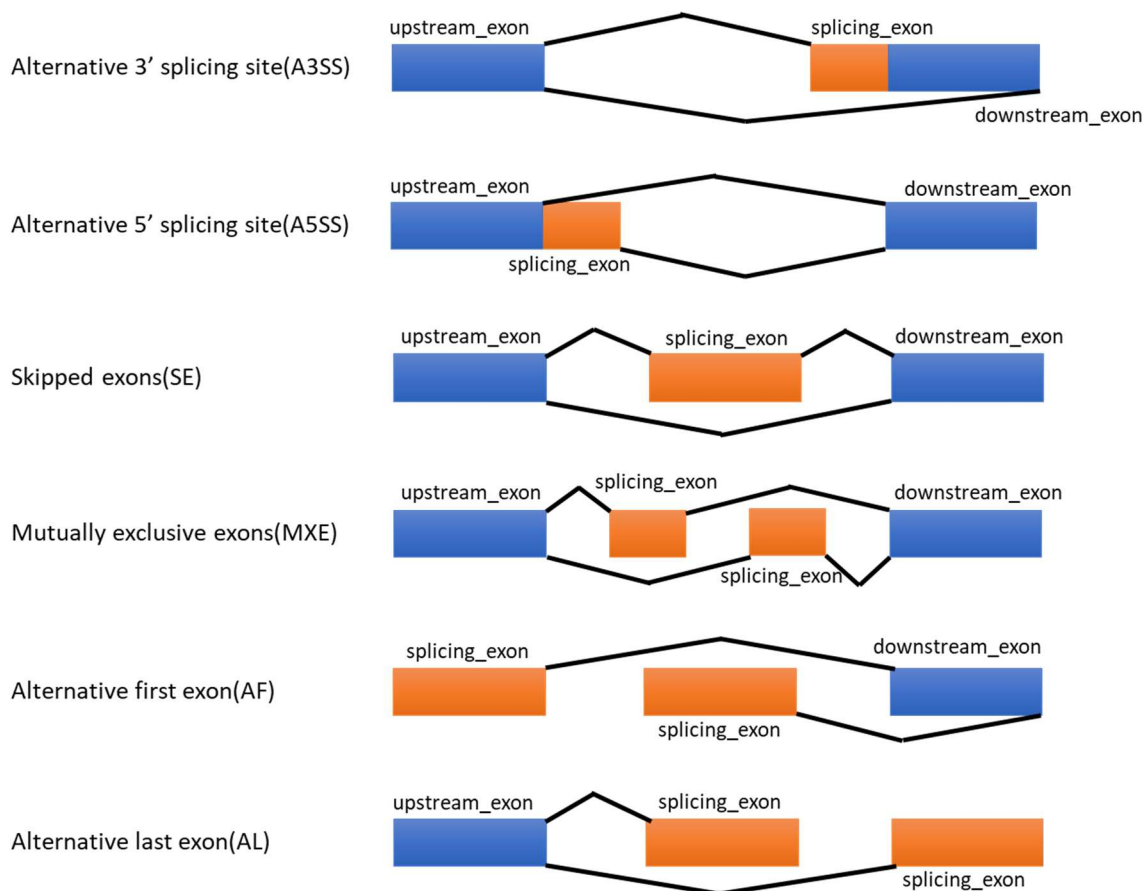
##SUPPA
> das.file.name <- system.file("extdata", "SUPPA_test.txt", package="ASpediaR")
> gtf.file.name <- system.file("extdata", "test_gtf.gtf", package="ASpediaR")
> ioe.file.name <- system.file("extdata", "SUPPA_test.ioe", package="ASpediaR")
> suppa.convert.result <- asr_converter(das.file.name, program="SUPPA",
gtf.file.name=gtf.file.name, ioe.file.name=ioe.file.name)

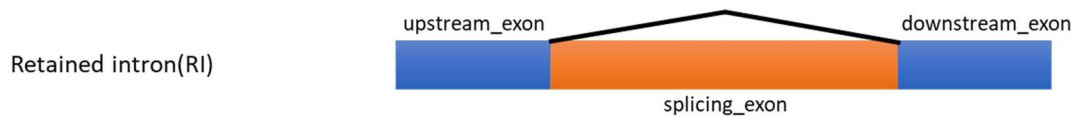
##spliceR
> das.file.name <- system.file("extdata", "spliceR_test.txt",
package="ASpediaR")
> splicer.convert.result <- asr_converter(das.file.name, program="spliceR",
gene.model="Ensembl", genome.version="GRCh38")
```

Other various formats of AS profiles could be also applied for our database retrieval system by following AS event ID rule. Our AS event ID composed of chromosome, and exon boundaries to reflect transcript direction. An input file should include `as_id` column to be IDs corresponding to each AS event. `as_id` column follows below rules.



upstream\_exon, splicing\_exon, and downstream\_exon in `asr_converter` result has different form by `as_type`.





### 3. Annotation

Annotation of `asr_converter` result using ASpedia database. ASpedia database consist of alternative splicing related DNA and RNA features, that separated by gene model and genome version. AS ID is a key role for ASpedia database search.

```
> annotation.result.dir <- paste0(system.file("extdata", package="ASpediaR"),
"/annotation_result")
> annotation.result <- asr_annotation(rmats.convert.result,
gene.model="Ensembl", genome.version="GRCh38", result.dir=annotation.result.dir)
> head(annotation.result, c(5, 10))
```

For example, annotation result is generated to `data.frame` object.

gene_symbol	chr	as_id
1 PHACTR4	chr1	chr1:28407410:28407463:28457832:28457855:28459085:28459258
2 MACF1	chr1	chr1:39479798:39480009:39480920:39481030:39484601:39484730
3 PTPRF	chr1	chr1:43598720:43598913:43602071:43602097:43603416:43603533
4 ADGRL2	chr1	chr1:81951957:81952142:81952987:81953025:81955877:81956060
5 ADAM15	chr1	chr1:155061415:155061489:155061904:155061975:155062245:155062369
6 ADAM15	chr1	chr1:155061415:155061489:155061904:155062117:155062245:155062369

gene_id	as_description_id	as_type
1 ENSG00000204138	SE:Exon2,Exon3(-)	SE
2 ENSG00000127603	SE:Exon99(+)	SE
3 ENSG00000142949	SE:Exon14(+)	SE
4 ENSG00000117114	SE:Exon12(+)	SE
5 ENSG00000143537	SE:Exon21(+)	SE
6 ENSG00000143537	SE:ENST00000526491 Exon21-ENST00000355956 Exon20,Exon21	SE

strand	gene_name	locus_group	location
1 +	phosphatase and actin regulator 4	protein-coding gene	1p35.3
2 +	microtubule actin crosslinking factor 1	protein-coding gene	1p34.3
3 +	protein tyrosine phosphatase receptor type F	protein-coding gene	1p34.2
4 +	adhesion G protein-coupled receptor L2	protein-coding gene	1p31.1
5 +	ADAM metalloproteinase domain 15	protein-coding gene	1q21.3
6 +	ADAM metalloproteinase domain 15	protein-coding gene	1q21.3

Total annotation result includes several columns for AS-related multi-omics features. Description of annotation result columns as follows:

Column	Description	Example
<b>gene_symbol</b>	HGNC approved gene symbol.	TP73

<b>chr</b>	Chromosome.	chr1
<b>as_id</b>	Alternative splicing ID.	chr1:3645891:3646012:3647491:3647629:3648027:3648120
<b>as_description_id</b>	Renamed alternative splicing ID.	SE: Involving multiple isoforms
<b>as_type</b>	Alternative splicing type.	SE
<b>strand</b>	Strand orientation of genomic coordinates.	+
<b>gene_name</b>	HGNC approved name for the gene.	tumor protein p73
<b>locus_group</b>	A group name for a set of related locus types as defined by the HGNC.	protein-coding gene
<b>location</b>	Cytogenetic location of the gene.	1p36.32
<b>gene_id</b>	REFSEQ or ENSEMBL gene ID.	ENSG00000078900
<b>transcript_id</b>	REFSEQ or ENSEMBL transcript ID.	ENST00000346387,ENST00000604479,ENST00000378280,ENST00000604566
<b>exon_inclusion_transcript_id</b>	Transcript id included by this alternative splicing.	ENST00000346387,ENST00000604479
<b>exon_exclusion_transcript_id</b>	Transcript id excluded by this alternative splicing.	ENST00000378280,ENST00000604566
<b>GO_BP</b>	Gene Ontology terms describing pathways and processes of given gene symbol.	APOPTOSIS GO;POSITIVE REGULATION OF TRANSCRIPTION
<b>GO_CC</b>	Gene Ontology terms describing localization of given gene symbol.	MITOCHONDRION;NUCLEUS
<b>GO_MF</b>	Gene Ontology terms describing molecular activity of given gene symbol.	DNA BINDING;TRANSCRIPTION FACTOR ACTIVITY
<b>conservation_score</b>	Average conservation scores of exonic and intronic regions for this AS region. Conservation scores are reported by each database, exon and intron information is in parentheses.	hg19.100way.phastCons:(E)0.304/(I)0.041; phastCons46way.placentalMammals:(E)0.234/(I)0.045; phastCons46way.primates:(E)0.253/(I)0.072; phastCons46way.vertebrate:(E)0.326/(I)0.041
<b>dbSNP_variant</b>	Variants (point mutations) of splicing site in dbSNP for this AS regions. dbSNP IDs, position of point mutations, refs, and alts are reported.	rs368114063,chr17:76212746,G>A; rs67347443,chr17:76221717,C>CT
<b>COSMIC_variant</b>	Variants (point mutations) of splicing site in COSMIC for this AS regions. Position of point mutations, refs, and alts are reported.	chr6:3646012,CG>C;chr6:3646013,G>A
<b>SPIDEX_variant</b>	Variants (point mutations) of splicing site in SPIDEX for this AS region. Position of point mutations, refs, and alts are reported.	chr4:3645891,G>A/C/T;chr4:3645892,T>A/C/G
<b>miRNA_binding_site</b>	miRNA binding sites of 3' UTR region predicted by TargetScan for this AS region. miRNA binding regions and miRNA IDs are reported.	chr10:2038702-2038709,miR-125/351; chr5:2038405-2038413,miR-153
<b>repeat</b>	Overlapping repeats regions with this AS region. Repeat database names and repeat regions are	RepeatMasker,SINE,chr17:304973-305104; Simple Repeats,chr3:305848-305880

	reported. And only RepeatMasker is reported repeat class information.	
<b>NMD</b>	NMD sites in known stop codons for this AS region. Chromosomes and NMD sites are reported.	chr12:120636530;chr12:120636541
<b>COSMIC_NMD</b>	NMD sites in novel variant stop codons inferring from COSMIC for this AS region. Chromosomes and NMD sites are reported.	chr10:103344469;chr10:103344504
<b>dbSNP_NMD</b>	NMD sites in novel variant stop codons inferring from dbSNP for this AS region. Chromosomes and NMD sites are reported.	chr10:70644615;chr10:70645026
<b>protein_domain</b>	Protein domain using Pfam for this AS region. Pfam domain ID, Pfam domain name, and genomic region are reported. And if the information of proteomic region is available, also reported.	PF07647,SAM domain (Sterile alpha motif), 415-478,chr8:3647601-3649379
<b>protein_translational_modification</b>	PTM sites collected from PhosphositePlus for this AS region. PTM types, proteomic regions, and genomic regions are reported.	Chain,p310-1400,chr2:49924743-49940115; Phosphorylation site,AA1056-1056,chr11:49932703-49932705
<b>RBP_splicing_factor</b>	Summary of RBP around alternative splicing regions, and peak detection p-values. Target proteins, p-values, peak genomic region, and peak length are reported.	HNRNPU,6.533499884e-07,chr15:3649426-3649497,72; SF3B1,1.022357417e-08,chr1:3647540-3647574,35
<b>isoform_PPI_a</b>	Transcript IDs in this isoform.	ENST00000378288;ENST00000378295
<b>isoform_PPI_b</b>	Protein interaction partners with isoform_PPI_a.	ITCH/TP73/NEDD4/UBC; WWOX/HMGB1/PIN1/MAGEA2B/E6/TP73/YAP1
<b>isoform_subcellular_localization_id</b>	Transcript IDs in this isoform.	Q9HBH9-1;Q9HBH9-2
<b>isoform_subcellular_localization</b>	Isoform specific Subcellular localizations.	Cytoplasm;Nucleus > PML body

asr\_annotation additionally provide gene enrichment test result between annotation result genes and reference genes.

### Example

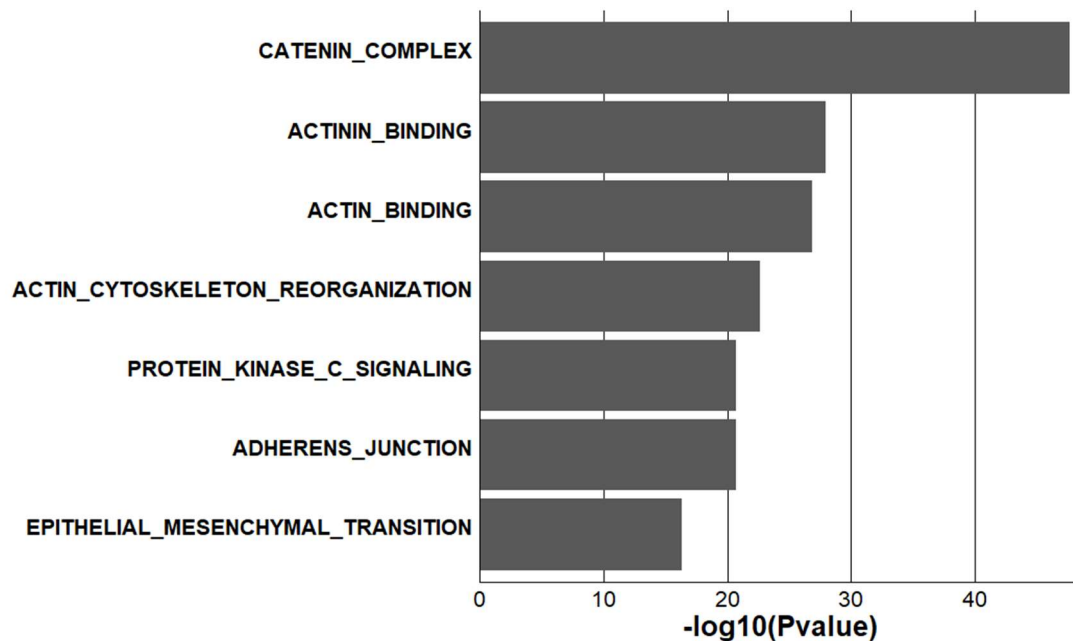
```
## gene model and genome version
> annotation.gene.list <- unique(annotation.result$gene_symbol)
> gsea.result.dir <- paste0(system.file("extdata", package="ASpediarR"),
"/gsea_result")
> mining_gsea(annotation.gene.list, gene.model="Ensembl",
genome.version="GRCh38", result.dir=gsea.result.dir)
```

```
## reference gene list from user input
> test.gene.list.file.name <- system.file("extdata", "test_whole_gene.txt",
package="ASpediarR")
```

```

> test.gene.list <- unique(read.table(test.gene.list.file.name, header=FALSE,
stringsAsFactors=FALSE)$V1)
> gsea.result.dir <- paste0(system.file("extdata", package="ASpediaR"),
"/gsea_result")
> mining_gsea(annotation.gene.list, gsea.gene.list=test.gene.list,
result.dir=gsea.result.dir)

```



## 4. Visualization

Result plot consist of DNA and RNA features plot and PPI plot.

```

> gtf.file.name <- system.file("extdata", "test_gtf.gtf", package="ASpediaR")
> plot.result.dir <- paste0(system.file("extdata", package="ASpediaR"),
"/plot_result")

##using gene name
> asr_plot(annotation.result, gtf.file.name, gene.model="Ensembl",
genome.version="GRCh38", gene.name="FGFR2", result.dir=plot.result.dir)

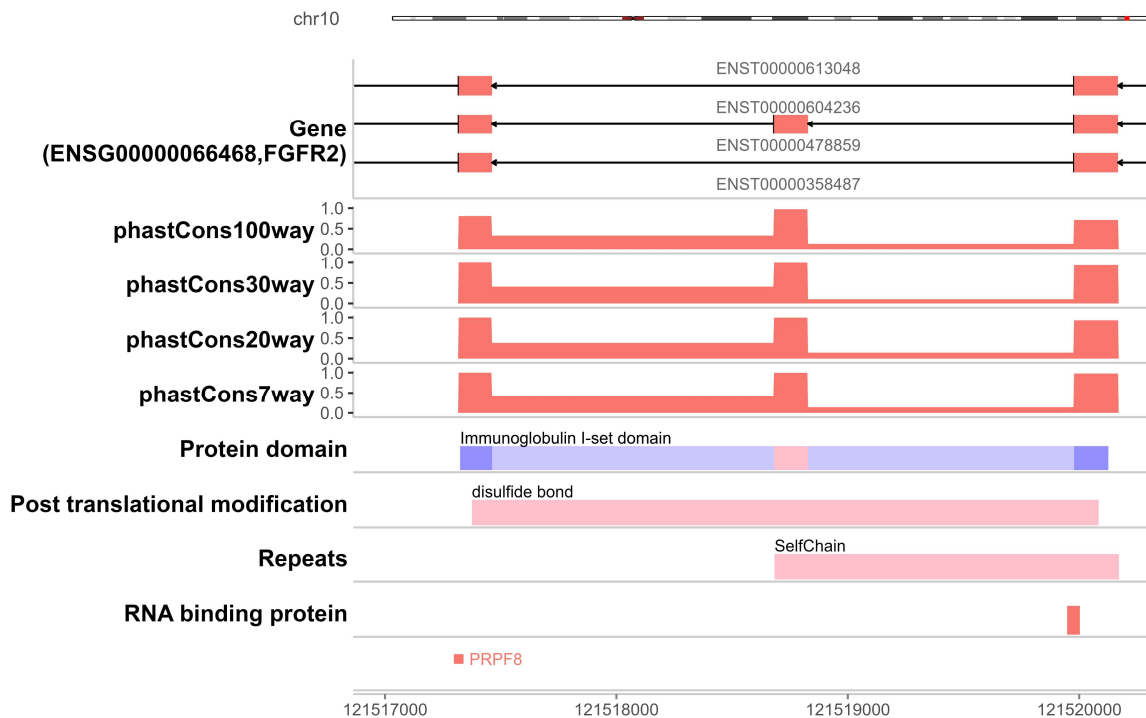
##using AS ID
> asr_plot(annotation.result, gtf.file.name, gene.model="Ensembl",
genome.version="GRCh38",
as.id="chr10:121520169:121519979:121518829:121518682:121517463:121517319",
result.dir=plot.result.dir)

##track lists and track heights are change
> asr_plot(annotation.result, gtf.file.name, gene.model="Ensembl",
genome.version="GRCh38",
as.id="chr10:121520169:121519979:121518829:121518682:121517463:121517319",
heights.list=c(1, 2, 1, 1, 1, 1, 1, 1, 1, 1), list.of.plot=c("conservation",

```

```
"domain", "PTM", "repeats", "RBP"), result.dir=plot.result.dir)
```

## - DNA and RNA feature plot



## - PPI plot

```
> ppi.result.dir <- paste0(system.file("extdata", package="ASpediaR"),
"/ppi_result")

##using gene name
> asr_plot_ppi(annotation.result, gene.name="FGFR2", result.dir=ppi.result.dir)

##using AS ID
> asr_plot_ppi(annotation.result,
as.id="chr10:121520169:121519979:121518829:121518682:121517463:121517319",
result.dir=ppi.result.dir)
```

