IMAS : Integrative analysis of Multi-omics data for Alternative Splicing

Seonggyun Han and Younghee Lee

May 2, 2019

Contents

IMAS : Integrative analysis of Multi-omics data for Alternative Splicing

1   Introduction

A balance of alternatively spliced (AS) transcript isoforms is critical for a normal phenotype, and dysregulated expression might result in phenotypic and clinical alterations in certain individuals and groups [1]. For example, the expression ratio (denoted as PSI: Percent Spliced In) of AS transcript isoforms or AS exons is modulated in genes that are associated with many complex diseases, such as cancer and diabetes [2]. Moreover, molecular factors that affect the expression ratio, such as genetic variations (sQTLs) and methylation (defined as me-sQTLs), give additional insight into the molecular mechanisms that explain the regulatory elements in splicing machinery [2]. The Bioconductor R package *IMAS* offers two components. First, *IMAS* provides an integrative analysis tool to detect differential AS events (e.g., exon skipping, intron retention, and 3- and 5-prime alternative splicing sites) using paired and junction reads that are created from high-throughput RNA-seq data. Second, *IMAS* links AS exons that are differentially expressed across groups, defined as PSI, to molecular factors (eg, SNPs and DNA methylation). IMAS can identify sQTLs and me-sQTLs to interpret their molecular functions. In addition, *IMAS* can incorporate the clinical information of a patient and determine whether exon skipping or inclusion affects the clinical outcomes of disease groups Statistical methods (linear regression model [4] or generalized linear mixed model [3] in sQTLs) are used to identify significant differences in PSI values across groups, sQTLs, me-sQTLs, and clinical outcome associations. *IMAS* can accept mapped bam files, which may be created from several mapping tools, such as bwa [5] and bowtie [6]. *IMAS* analyzes all paired-end and junction reads that are extracted using *Rsamtools*. Moreover, all functions of this package accept a simple FPKM matrix dataset of transcripts (output file format of assemble tools, such as cufflinks) through *IVAS*, which is an R/bioconductor package for identifying genetic variants that affect alternative splicing patterns with the FPKM matrix dataset. *IMAS* also provides a function for visualizing all identified AS transcript isoforms, sQTLs, me-sQTLs, and clinical outcomes.

2   The input data set

This subsection introduces the input data. To run the *IMAS* package, a GTF file is required as a reference transcript model. Based on this AS transcript model, AS exons are tabulated. Moreover, *IMAS* accepts the mapped bam file to estimate PSI values in a given AS exon. If the PSI is calculated with simple expression matrix data that consist of FPKM values of isoforms through *IVAS*, the PSI dataset can be analyzed by all functions of *IMAS*. To discover AS exons with differential PSI values between

groups, the specific groups for each individual are required. To identify sQTLs, a genotype data frame for each polymorphism and its position are required. The sQTL analysis is carried out using `sQTLsFinder` in the *IVAS* package. To identify me-sQTLs, the expression and position of methylation loci are needed. Clinical information, such as survival status and survival time, is inputted for survival analysis.

2.1   The transcript model data

The GTF file including gene structures such as genomic coordinate (position) of all exons and transcripts are required. The GTF file should be loaded with `TxDb` object in the *GenomicFeatures* package.

2.2   The expression data

To calculate PSI values of AS exons, *IMAS* uses a mapped bam file as experimental expression data from RNA-seq. The paths of all bam files can be assigned a variable. An example is described in Section 4, Loading data. Also, the PSI values can be calculated from the FPKM matrix dataset through `RatioFromFPKM` (examples of the method with FPKM are described in the *IVAS* package).

	path	names
1	./extdata/bamfiles/sample1.bam	ind1
2	./extdata/bamfiles/sample2.bam	ind2
3	./extdata/bamfiles/sample3.bam	ind3
4	./extdata/bamfiles/sample4.bam	ind4

2.3   The difference in splicing ratio of a given exon between Groups

Once the PSI values are calculated for each AS exon from the mapped bam files or FPKM dataset, then with the PSI, one can identify differentially expressed exons between groups For this analysis, groups for each sample are required as the list type of R. The names of the list data must be labeled "GroupA" and "GroupB."

```
$GroupA
   ind1     ind4
$GroupB
   ind2     ind3
```

2.4   The genotype and genomic position dataset of SNPs

To identify polymorphisms that are associated with the splicing ratio (PSI) of a given exon, the genotype data should be prepared as simple matrix data. *IMAS* carries out this analysis using sqtlfinder in the *IVAS* package. Details and examples are described in the *IVAS* package.

2.5   The expression and position dataset of methylation

IMAS identifies methylation loci that are associated with the splicing ratio (PSI) of a given exon. A dataset needs to include the expression of methylation (i.e., beta value) and the genomic position of each methylation. In this file, each row and column label the methylation IDs and individual names, respectively.

The expressions of methylations

	ind1	ind2	ind3	ind4
methyl1	14.4	20.2	21.2	20.1
methyl2	7.4	7.2	6.2	10.1
methyl3	13.2	15.8	11.2	15.9

The positions of methylations

Methyl	CHR	locus
methyl1	1	4834015
methyl2	11	61523247
methyl3	11	2382791

2.6   Clinical information

The clinical information of matched samples that have expression data is used to test the association of a splicing ratio (PSI) of a given exon with clinical outcomes. Clinical information must include the survival status and survival time, which are the first and second columns in the file, respectively. In addition, each row represents an individual, labeled with an ID.

	status	sur_time
ind1	0	100
ind2	1	302
ind3	1	402
ind4	0	81

3 Example dataset

We generated our own simulation data for 50 breast cancer patients as an example data set in order to demonstrate how to use function in the Manual of the *IMAS* package. The simulation dataset comprises clinical data, random group assignment to PR-positive or negative, mapped reads, methylation status for five loci, and genotypes for five SNPs. As a demonstration case, we define an exon located in chr11:100962491-100962607 in the PRGA gene. The exon is assumed to have differential PSI values between two groups, PR-positive and PR-negative. The differential exon expression is also assumed to be associated with SNP and methylation level. These simulation data can be loaded from the *IMAS* package.

4 Loading data

For this analysis, load the following data and an object from *IMAS*, the group information for each individual, the genotypes and genomic coordinates of SNPs, the expression and genomic coordinates of methylation, clinical information, and the `TxDb` object from GTF.
Installation of *IMAS* package :

```
> if (!requireNamespace("BiocManager", quietly=TRUE))
+     install.packages("BiocManager")
> BiocManager::install("IMAS")
```

Loading *IMAS* package :

```
> library(IMAS)
```

Loading expression data :

```
> data(bamfilestest)
> ext.dir <- system.file("extdata", package="IMAS")
> samplebamfiles[,"path"] <-
+     paste(ext.dir,"/samplebam/",samplebamfiles[,"path"],".bam",sep="")
```

Loading group information :

```
> data(sampleGroups)
```

Loading SNP data :

```
> data(samplesnp)
```

Loading SNP position data :

```
> data(samplesnplocus)
```

Loading methylation expression data :

```
> data(samplemethyl)
```

Loading methylation position data :

```
> data(samplemethyllocus)
```

Loading clinical information data :

```
> data(sampleclinical)
```

Loading `TxDb` object :

```
> sampledb <- list.files(ext.dir,pattern="DB",full.names=TRUE)
> transdb <- loadDb(sampledb)
> transdb

TxDb object:
# Db type: TxDb
# Supporting package: GenomicFeatures
# Data source: ./test.ref
# Organism: NA
# Taxonomy ID: NA
# miRBase build ID: NA
# Genome: NA
# transcript_nrow: 8
# exon_nrow: 23
# cds_nrow: 11
# Db created by: GenomicFeatures package from Bioconductor
# Creation time: 2017-02-06 07:49:33 -0700 (Mon, 06 Feb 2017)
# GenomicFeatures version at creation time: 1.24.5
# RSQLite version at creation time: 1.1-2
# DBSCHEMAVERSION: 1.1
```

`TxDb` can be created from the GTF reference file using the `makeTxDbFromGFF` function in *GenomicFeatures* in R/Bioconductor.

5 Running *IMAS* for integrative alternative splicing analysis

All results from each function of the *IMAS* package are saved by adding slots in the `ASdb` object, which is a s4 class type object. The last row of the `ASdb` object indicates the names of currently saved slots. All functions of *IMAS* allow one to use a multi-thread through a `foreach` function.

5.1 Tabulating alternatively spliced exons

IMAS tabulates AS exons and categorizes them into four types of AS patterns: exon skipping, alternative 3-prime splice site, alternative 5-prime splice site, and intron retention. The `Splicingfinder` function in the *IVAS* package is used in this step. If a single gene is inputted, AS exons are defined only as a single gene. The result will be saved in the slot, named "SplicingModel" of the `ASdb` object.
To use this function :

```
> ASdb <- Splicingfinder(GTFdb=transdb,calGene="ENSG00000082175",Ncor=1)

[1] "------------------Processing : chr 11 ------------------"

> ASdb <- ExonsCluster(ASdb,transdb)
> ASdb

Splicing Models : ES = 5 Rows & ASS = 2 Rows & IR = 0 Rows
#ASdb object with SplicingModel

> head(slot(ASdb,"SplicingModel")$"ES")

  Index EnsID            Nchr Strand 1stEX
1 "ES1" "ENSG00000082175" "11" "-"    "100920660-100920790"
2 "ES2" "ENSG00000082175" "11" "-"    "100933178-100933483"
3 "ES3" "ENSG00000082175" "11" "-"    "100962491-100962607"
4 "ES4" "ENSG00000082175" "11" "-"    "100933178-100933483"
5 "ES5" "ENSG00000082175" "11" "-"    "100933178-100933483"
  2ndEX                 DownEX                UpEX
1 "NA"                  "100912392-100912833" "100922155-100922299"
2 "NA"                  "100922155-100922299" "100962491-100996889"
3 "NA"                  "100922155-100933483" "100996738-100996889"
4 "100962491-100962607" "100922155-100922299" "100996738-100996889"
5 "100962491-100962607" "100922155-100922299" "100996738-100996889"
  1st_des                                  2nd_des
1 "100920660-100920738,100920660-100920790" "NA"
2 "100933178-100933483"                     "NA"
3 "100962491-100962607"                     "NA"
4 "100933178-100933483"                     "100962491-100962607"
```

```
5 "100933178-100933483"                     "100962491-100962607"
  Do_des
1 "100912392-100912833,100912676-100912833"
2 "100922155-100922299"
3 "100922155-100922299,100933178-100933483"
4 "100922155-100922299"
5 "100922155-100922299"
  Up_des                                     Types
1 "100922155-100922201,100922155-100922299" "ES"
2 "100962491-100962607,100996738-100996889" "ES"
3 "100996738-100996889"                      "ES"
4 "100996738-100996889"                      "ES"
5 "100996738-100996889"                      "MXE"
```

As an example, search for sQTLs in the ENSG00000082175 gene. The first column, labeled "Index," is a unique name for each splicing model, and the index number of the first column is a generally used as an identifier, commonly used in other *IMAS* functions. One can use the argument for "out.dir" to save the results in a specific directory.

5.2 Estimating the expression ratio (Percent Spliced In) of AS exons with bam files of RNA-seq

The `RatioFromReads` function calculates the PSI values of AS exons, based on splicing models in the `ASdb`. The result will be saved in the "Ratio" slot of `ASdb`. This needs matrix data, including the paths of bam files and the sample names, which will be labeled in the columns of the results of this function in the first and second columns, respectively.

```
> ASdb <- RatioFromReads(ASdb=ASdb,samplebamfiles,readsInfo="paired",
+     readLen=50,inserSize=40,minr=3,CalIndex="ES3",Ncor=1)
> ASdb

Splicing Models : ES = 5 Rows & ASS = 2 Rows & IR = 0 Rows
Ratio : ES = 1 Rows by 50 samples & ASS = 0 Rows by 0 samples & IR = 0 Rows by 0 samples
#ASdb object with SplicingModel & Ratio

> head(rbind(slot(ASdb,"Ratio")$"ES"[,1:20]))

    Index EnsID            Nchr 1stEX                 2ndEX
[1,] "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
    DownEX                UpEX                Types BreastPatient_1
[1,] "100922155-100933483" "100996738-100996889" "ES"  "0.948307921160283"
    BreastPatient_10    BreastPatient_11    BreastPatient_12
[1,] "0.954270674836534" "0.954270674836534" "0.948307921160283"
```

```
        BreastPatient_13       BreastPatient_14       BreastPatient_15
[1,] "0.960176810051979" "0.942287730952875" "0.960176810051979"
        BreastPatient_16       BreastPatient_17       BreastPatient_18
[1,] "0.948307921160283" "0.942287730952875" "0.954270674836534"
        BreastPatient_19       BreastPatient_2
[1,] "0.954270674836534" "0.971822420433813"
```

As an example, one single index number, ES3, returns the PSI value of a splicing pattern, which is an AS exon located in chr11:100962491-100962607. One can use the argument for "out.dir" to save the results in a specific directory.

5.3 Detecting alternatively spliced exons with differential PSIs between groups

The `CompGroupAlt` function calculates significant differences in PSIs between groups using linear regression. The result will be saved in the "GroupDiff" slot of `ASdb`. The list object includes sample names for each group and is labeled "GroupA" and "GroupB." The sample names should be matched with the column names of the "Ratio" slot in the `ASdb` object, described above (5.2).

```
> ASdb <- CompGroupAlt(ASdb=ASdb,GroupSam=GroupSam,Ncor=1,CalIndex="ES3")
> ASdb

Splicing Models : ES = 5 Rows & ASS = 2 Rows & IR = 0 Rows
Ratio : ES = 1 Rows by 50 samples & ASS = 0 Rows by 0 samples & IR = 0 Rows by 0 samples
Differential Ratio : ES = 1 Rows & ASS = 0 Rows & IR = 0 Rows
#ASdb object with SplicingModel & Ratio & GroupDiff

> head(slot(ASdb,"GroupDiff")$"ES")
        Index EnsID              Nchr 1stEX                 2ndEX
Diff.P "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
        DownEX                 UpEX                    Types Diff.P
Diff.P "100922155-100933483" "100996738-100996889" "ES"   "9.41906778666016e-05"
        Fdr.p
Diff.P "9.41906778666016e-05"
```

You can use the argument for "out.dir" to save your results in a specific directory.

5.4 sQTLs analysis

`sQTLsFinder` in the *IVAS* package is used to identify significant SNPs that are associated with AS ratio (PSI). The result will be saved in the slot labeled "sQTLs" of `ASdb`.

```
> ASdb <- sQTLsFinder(ASdb=ASdb,Total.snpdata=samplesnp,Total.snplocus=samplesnplocus,
+     GroupSam=NULL,Ncor=1,CalIndex="ES3",method="lm")
> ASdb

Splicing Models : ES = 5 Rows & ASS = 2 Rows & IR = 0 Rows
Ratio : ES = 1 Rows by 50 samples & ASS = 0 Rows by 0 samples & IR = 0 Rows by 0 samples
Differential Ratio : ES = 1 Rows & ASS = 0 Rows & IR = 0 Rows
sQTLs : ES = 5 Rows & ASS = 0 Rows & IR = 0 Rows
#ASdb object with SplicingModel & Ratio & GroupDiff & sQTLs

> head(slot(ASdb,"sQTLs")$"ES")

  SNP     Index EnsID              Nchr 1stEX               2ndEX
1 "SNP1" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
2 "SNP2" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
3 "SNP3" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
4 "SNP4" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
5 "SNP5" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
  DownEX                UpEX                  Types pByGeno
1 "100922155-100933483" "100996738-100996889" "ES"  "0.00800343068370079"
2 "100922155-100933483" "100996738-100996889" "ES"  "0.00347996675111845"
3 "100922155-100933483" "100996738-100996889" "ES"  "0.704647458467744"
4 "100922155-100933483" "100996738-100996889" "ES"  "0.738751326919983"
5 "100922155-100933483" "100996738-100996889" "ES"  "0.00307967379995783"
  FdrByGeno           diff   met
1 "0.0133390511395013"  "diff" "lm"
2 "0.00869991687779613" "diff" "lm"
3 "0.738751326919983"   "diff" "lm"
4 "0.738751326919983"   "diff" "lm"
5 "0.00869991687779613" "diff" "lm"
```

One can use the argument for "out.dir" to save the results in a specific directory.

5.5   me-sQTLs analysis

The `MEsQTLFinder` function identifies methylation locus that may affect AS events. This `MEsQTLFinder` function uses a linear regression model with the expression of each methylation as variables. The result will be saved in the "Me.sQTLs" slot of the `ASdb` object.

```
> ASdb <- MEsQTLFinder(ASdb=ASdb,Total.Medata=sampleMedata,Total.Melocus=sampleMelocus,
+     GroupSam=GroupSam,Ncor=1,CalIndex="ES3")
> ASdb

Splicing Models : ES = 5 Rows & ASS = 2 Rows & IR = 0 Rows
Ratio : ES = 1 Rows by 50 samples & ASS = 0 Rows by 0 samples & IR = 0 Rows by 0 samples
```

```
    Differential Ratio : ES = 1 Rows & ASS = 0 Rows & IR = 0 Rows
    sQTLs : ES = 5 Rows & ASS = 0 Rows & IR = 0 Rows
    Me-sQTLs : ES = 5 Rows & ASS = 0 Rows & IR = 0 Rows
    #ASdb object with SplicingModel & Ratio & GroupDiff & sQTLs & Me.sQTLs

    > head(slot(ASdb,"Me.sQTLs")$"ES")

      MeID        Index EnsID           Nchr 1stEX             2ndEX
    1 "Methyl_1" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
    2 "Methyl_2" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
    3 "Methyl_3" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
    4 "Methyl_4" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
    5 "Methyl_5" "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
      DownEX               UpEX                 Types pByMet
    1 "100922155-100933483" "100996738-100996889" "ES"  "7.06767598610055e-05"
    2 "100922155-100933483" "100996738-100996889" "ES"  "0.522194583543479"
    3 "100922155-100933483" "100996738-100996889" "ES"  "0.182660148910586"
    4 "100922155-100933483" "100996738-100996889" "ES"  "0.000536141211354368"
    5 "100922155-100933483" "100996738-100996889" "ES"  "0.000106433265859588"
      fdrByMet              pByGroups             fdrByGroups
    1 "0.00026608316464897" "3.99615688872986e-18" "9.99039222182465e-18"
    2 "0.522194583543479"   "0.876097599018987"    "0.876097599018987"
    3 "0.228325186138232"   "0.332355282013401"    "0.415444102516751"
    4 "0.000893568685590613" "1.70066122503774e-18" "8.5033061251887e-18"
    5 "0.00026608316464897" "2.28504465545911e-14" "3.80840775909852e-14"
```

One can use the argument for "out.dir" to save the results in a specific directory.

5.6 Global effect of SNP and methylation for alternative splicing among individuals

The GroupSam parameter specifies a group for each individual. The default option for this parameter is that there is no specified group for grouping individuals. With this default option, the `sQTLsFinder` and `MEsQTLFinder` functions identify sQTLs and me-sQTLs for a set of all individuals. The option of the parameters for assigning specific groups is set for each individual, which forms groups: GroupA and GroupB, as described in Section 5.3. This option identifies sQLs and me-sQTLs that are associated with GroupA and GroupB.

5.7 Clinical analysis

The `ClinicAnalysis` function identifies which exons ,that are differentially expressed in groups, are associated with clinical outcomes. The result of this analysis will be saved in the "Clinical" slot of `ASdb`.

```
> ASdb <- ClinicAnalysis(ASdb=ASdb,ClinicalInfo=Clinical.data,Ncor=1,CalIndex="ES3")
> ASdb

Splicing Models : ES = 5 Rows & ASS = 2 Rows & IR = 0 Rows
Ratio : ES = 1 Rows by 50 samples & ASS = 0 Rows by 0 samples & IR = 0 Rows by 0 samples
Differential Ratio : ES = 1 Rows & ASS = 0 Rows & IR = 0 Rows
sQTLs : ES = 5 Rows & ASS = 0 Rows & IR = 0 Rows
Me-sQTLs : ES = 5 Rows & ASS = 0 Rows & IR = 0 Rows
Clinical Analysis : ES = 1 Rows & ASS = 0 Rows & IR = 0 Rows
#ASdb object with SplicingModel & Ratio & GroupDiff & sQTLs & Me.sQTLs & Clinical

> head(slot(ASdb,"Clinical")$"ES")

  Index EnsID             Nchr 1stEX              2ndEX
1 "ES3" "ENSG00000082175" "11" "100962491-100962607" "NA"
  DownEX                UpEX                  Types Pvalue
1 "100922155-100933483" "100996738-100996889" "ES"  "0.000875323192621613"
  Fdr.p
1 "0.000875323192621613"
```

You can use the argument for "out.dir" to save your results in a specific directory.
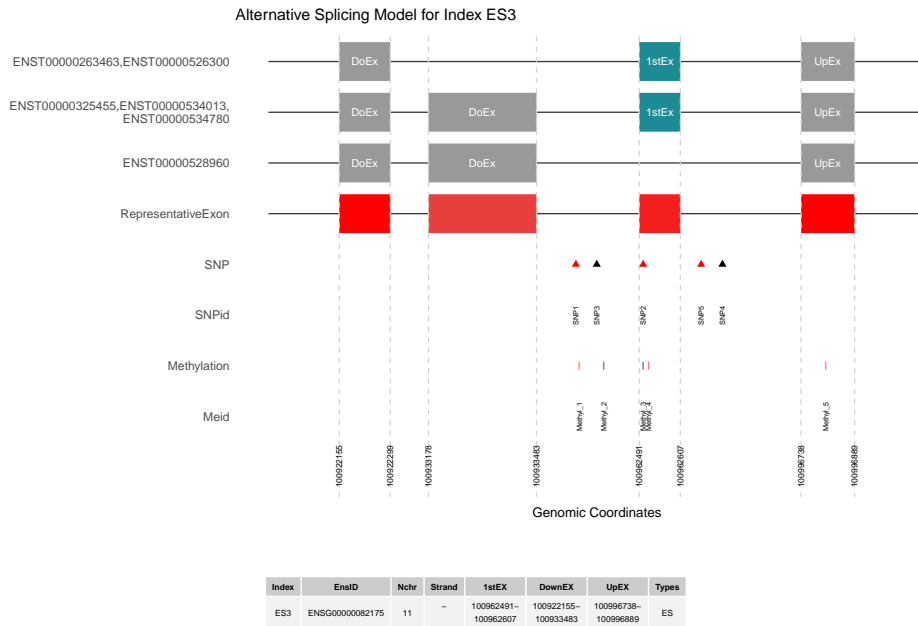
6   Visualizing the result

The `ASvisualization` function visualizes the information that is saved in `ASdb` slots, such as AS models, PSI values of AS exons, sQTLs, me-sQTLs, and clinical results. With an index number as an input of this function, the splicing model, SNP, methylation, clinical information is plotted and saved to one single pdf file that is named with the index number.

```
> exon.range <- exonsBy(transdb,by="tx")
> select.cns <- c("TXCHROM","TXNAME","GENEID","TXSTART","TXEND","TXSTRAND")
> txTable <- select(transdb, keys=names(exon.range),columns=select.cns,keytype="TXID")
> ASvisualization(ASdb,CalIndex="ES3",txTable,exon.range,samplesnp,samplesnplocus,
+               sampleMedata,sampleMelocus,GroupSam,Clinical.data,out.dir=tempdir())
```
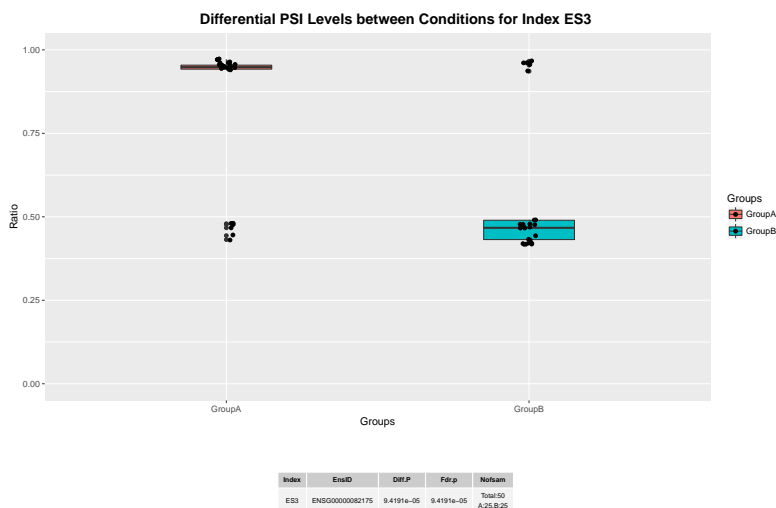
The plots of results from the example analysis are as follows:

Alternative Splicing Model for Index ES3

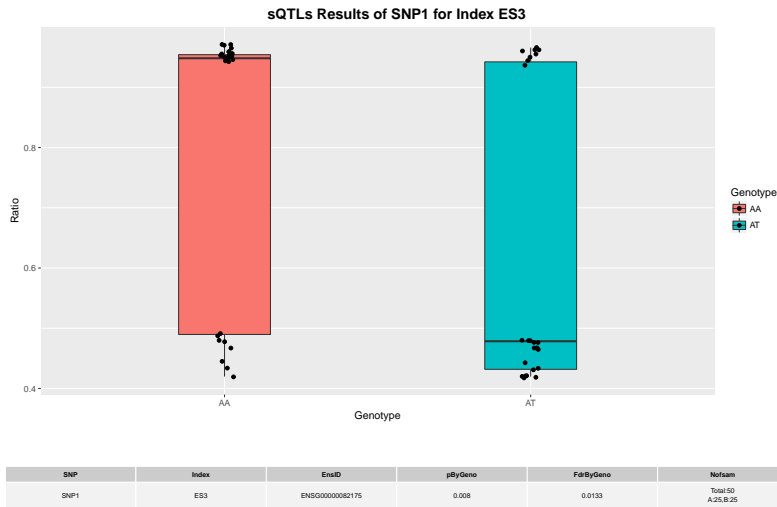| Index	EnsID	Nchr	Strand	1stEX	DownEX	UpEX	Types
ES3	ENSG00000082175	11	–	100962491–100962607	100922155–100933483	100996738–100996889	ES

The first page of the pdf result file shows the splicing model of the inputted index. The x-axis and y-axis indicate the genomic coordinates and transcripts with each splicing pattern, respectively. The exon in cyan indicates the alternatively spliced target, and the grey represents neighboring exons. The representative exons are shown in red, and the brightness represents the number of transcripts that have the respective exons. The SNP and methylation information is shown below the exons. The significant SNPs and methylations have the red color.

Differential PSI Levels between Conditions for Index ES3

Index	EnsID	Diff.P	Fdr.p	Nofsam
ES3	ENSG00000082175	9.4191e–05	9.4191e–05	Total:50 A:25,B:25

This plot shows the distribution of PSI values between two groups.

sQTLs Results of SNP1 for Index ES3

| SNP	Index	EnsID	pByGeno	FdrByGeno	Nofsam
SNP1	ES3	ENSG00000082175	0.008	0.0133	Total:50 A:25,B:25

This plot shows the sQTL results. The distribution of PSI values across genotypes is shown as a boxplot.

Me−sQTLs Results of Methyl_1 for Index ES3

MeID	Index	EnsID	pByMet	fdrByMet	pByGroups	fdrByGroups	Nofsam
Methyl_1	ES3	ENSG00000082175	7.0677e−05	3e−04	3.9982e−18	9.9904e−18	Total:50 A:25,B:25
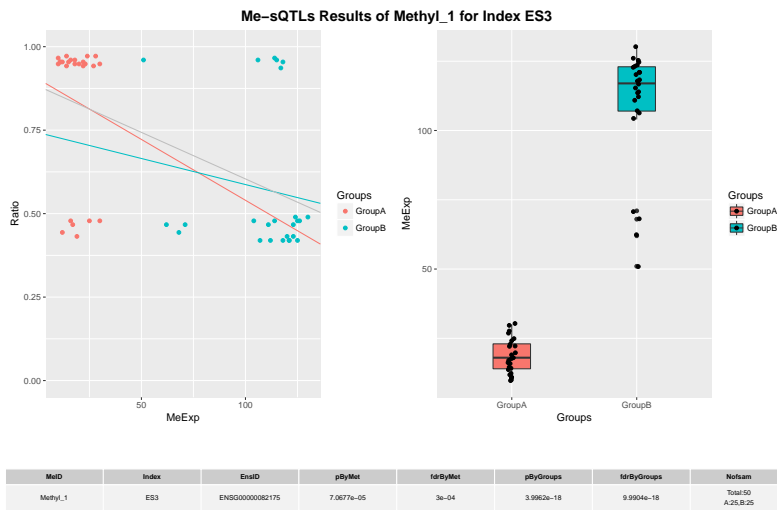
The result of the significant Me-sQTLs is shown in two plots: the left panel is the distribution of PSI values versus methylation expression, and the right panel indicates the methylation levels in each group.
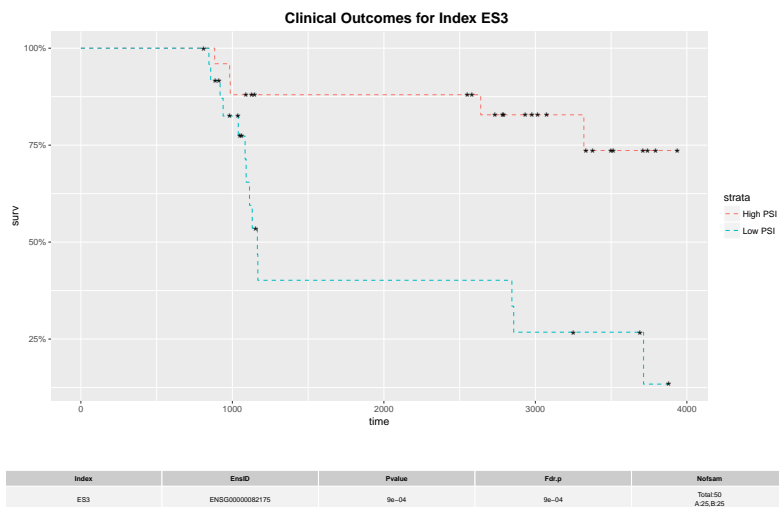
Expected survival outcomes of two groups, expressing high and low PSIs, are plotted.

7   Session Information

```
R version 3.6.0 (2019-04-26)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Ubuntu 18.04.2 LTS

Matrix products: default
BLAS:   /home/biocbuild/bbs-3.9-bioc/R/lib/libRblas.so
LAPACK: /home/biocbuild/bbs-3.9-bioc/R/lib/libRlapack.so

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=C
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats4    parallel  stats     graphics  grDevices utils     datasets
[8] methods   base

other attached packages:
 [1] IMAS_1.8.0               IVAS_2.4.0               ggplot2_3.1.1
```

```
 [4] GenomicFeatures_1.36.0 AnnotationDbi_1.46.0   Biobase_2.44.0
 [7] GenomicRanges_1.36.0   GenomeInfoDb_1.20.0    IRanges_2.18.0
[10] S4Vectors_0.22.0       BiocGenerics_0.30.0

loaded via a namespace (and not attached):
 [1] httr_1.4.0             tidyr_0.8.3
 [3] bit64_0.9-7            splines_3.6.0
 [5] foreach_1.4.4          assertthat_0.2.1
 [7] BiocManager_1.30.4     blob_1.1.1
 [9] GenomeInfoDbData_1.2.1 Rsamtools_2.0.0
[11] yaml_2.2.0             progress_1.2.0
[13] pillar_1.3.1           RSQLite_2.1.1
[15] lattice_0.20-38        glue_1.3.1
[17] digest_0.6.18          XVector_0.24.0
[19] minqa_1.2.4            colorspace_1.4-1
[21] htmltools_0.3.6        Matrix_1.2-17
[23] plyr_1.8.4             XML_3.98-1.19
[25] pkgconfig_2.0.2        biomaRt_2.40.0
[27] zlibbioc_1.30.0        purrr_0.3.2
[29] scales_1.0.0           BiocParallel_1.18.0
[31] lme4_1.1-21            tibble_2.1.1
[33] withr_2.1.2            SummarizedExperiment_1.14.0
[35] lazyeval_0.2.2         survival_2.44-1.1
[37] magrittr_1.5           crayon_1.3.4
[39] memoise_1.1.0          evaluate_0.13
[41] nlme_3.1-139           MASS_7.3-51.4
[43] doParallel_1.0.14      tools_3.6.0
[45] prettyunits_1.0.2      hms_0.4.2
[47] BiocStyle_2.12.0       matrixStats_0.54.0
[49] stringr_1.4.0          munsell_0.5.0
[51] DelayedArray_0.10.0    Biostrings_2.52.0
[53] compiler_3.6.0         rlang_0.3.4
[55] nloptr_1.2.1           grid_3.6.0
[57] RCurl_1.95-4.12        iterators_1.0.10
[59] labeling_0.3           bitops_1.0-6
[61] rmarkdown_1.12         boot_1.3-22
[63] gtable_0.3.0           codetools_0.2-16
[65] DBI_1.0.0              R6_2.4.0
[67] GenomicAlignments_1.20.0 gridExtra_2.3
[69] knitr_1.22             dplyr_0.8.0.1
[71] rtracklayer_1.44.0     bit_1.1-14
[73] ggfortify_0.4.6        stringi_1.4.3
[75] Rcpp_1.0.1             tidyselect_0.2.5
[77] xfun_0.6
```

References

[1] Jamal Tazia, Nadia Bakkoura, Stefan Stamm. Alternative splicing and disease. Biochim Biophys Acta, 2008. doi: 10.1016/j.bbadis.2008.09.017.

[2] Seonggyun Han, Hyeim Jung, Kichan Lee, Hyunho Kim, Sangsoo Kim. 2017. Genome wide discovery of genetic variants affecting alternative splicing patterns in human using bioinformatics method. 2017. Genes Genom. doi:10.1007/s13258-016-0466-7

[3] N.E. Breslow and D.G. Clayton. Approximate Inference in Generalized Linear Mixed Models. Journal of the American Statistical Association. 1993. 88 421: 9-25.

[4] Chambers, J. M. Linear models. Chapter 4 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth, and Brooks Cole. 1992.

[5] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009. doi: 10.1093/bioinformatics/btp324.

[6] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012. doi: 10.1038/nmeth.1923.