NanoString-Splice

Table of Contents

1. What is NanoString-Splice?

NanoString-Splice is a web service to analyze the NanoString data for pre-mRNA splicing. It considers the potential impact of different hybridization efficiency of the two probes used in two splice isoforms and uses an optimization method to adjust the percent-splice-in (***PSI***) values. It also provides functions for drug dose-response analysis.

2. Download the source code

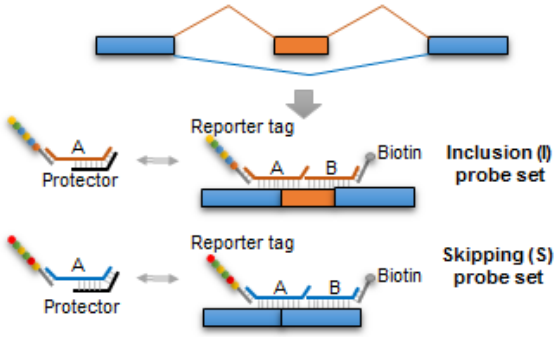
NanoString-Splice is written in R. it can be downloaded from http://www.bioinformatics.org/nanostr-splice .

3. How to fill out the web-form?

There are two forms available: one for large amount of input data (batch analysis). It was designed for "single compound, multiple targets" analysis. It deals with normalized NanoString counts data for I- and S- probes and calculates *Ei*-adjusted *PSI* values and then do dose-response analysis. Dose-response curves will be generated for each individual target in seperated plot. Another form (Custom

dose-response curve analysis) is designed for "multiple compounds, single target" analysis. Only one plot (figure) will be generated with all compound-responses shown in different colors. The webforms are self-explanatory and click the "example" links to see examples.

4. What are I- and S-probes?



I- and S-probes are two NanoString probe sets targeting the inclusion (I) and skipping (S) isoforms of an alternative exon.

5. What is Ei-adjusted-PSI?

"Ei" represents the relative hybridization efficiency of I-probe over S-probe. NanoString-Splice uses the R function “optimize” to find the *Ei* which minimizes the coefficient of variation (*CV*) of the total adjusted counts *T=I/Ei+S* (*I* and *S* are the normalized counts for I- and S- probe respectively) for multiple samples.

6. Can I use NanoString-Splice to analyze other types of data?

Yes! The tool is optimized for NanoString data to study RNA splicing. However, it is designed to be flexible. Functions like *Ei*-adjustion, *PSI* calculation can be "turned off". Users can provide their own data (including counts data or percentage data) for dose-response analysis. Many parameters can be changed before running the tool (eg. *Ec5x*, *Ec_PSI30* values can be calculated). It's also possible to study non-dose-response data. For example, one can calculate the *Ei-adjusted-PSIs* to study splicing changes in cells with certain factor knockdown.

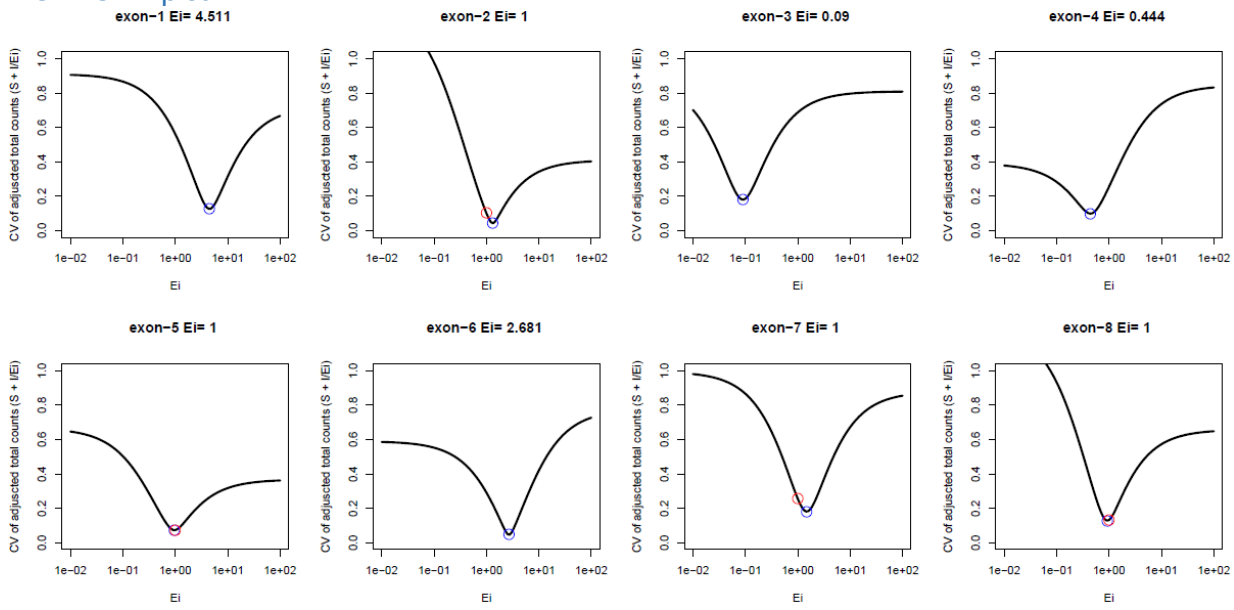
7. How to interpret the output data?

7.1. List of output files

Input file:
input.txt
input.txt.exclusionProbe.counts.txt (input S-probe counts)
input.txt.inclusionProbe.counts.txt (input I-probe counts)
Output Hybridization efficiency index (Ei) and Percent-spliced-in (PSI) data:
input.txt.PSI.out.txt (calculated PSI values)
input.txt.PSI.Ei_plot.pdf (CV of total counts (I/Ei + S) vs. Ei plot)
input.txt.PSI.Ei.out.txt (Ei output table)
Output of Dose-response (DR) analysis results:
input.txt.exclusionProbe.counts.txt.drc_fitting.pdf (fitting curves of S-probe)
input.txt.exclusionProbe.counts.txt.drc_fitting.out.txt (DR results for S-probe)
input.txt.inclusionProbe.counts.txt.drc_fitting.pdf (fitting curves of I-probe)
input.txt.inclusionProbe.counts.txt.drc_fitting.out.txt (DR results for I-probe)
input.txt.PSI.drc_fitting.pdf (fitting curves of PSI data)
input.txt.PSI.out.txt.drc_fitting.out.txt (DR results for PSI)
All input and output data combined:
input.txt.output.combine.txt
All results zipped:
output.zip

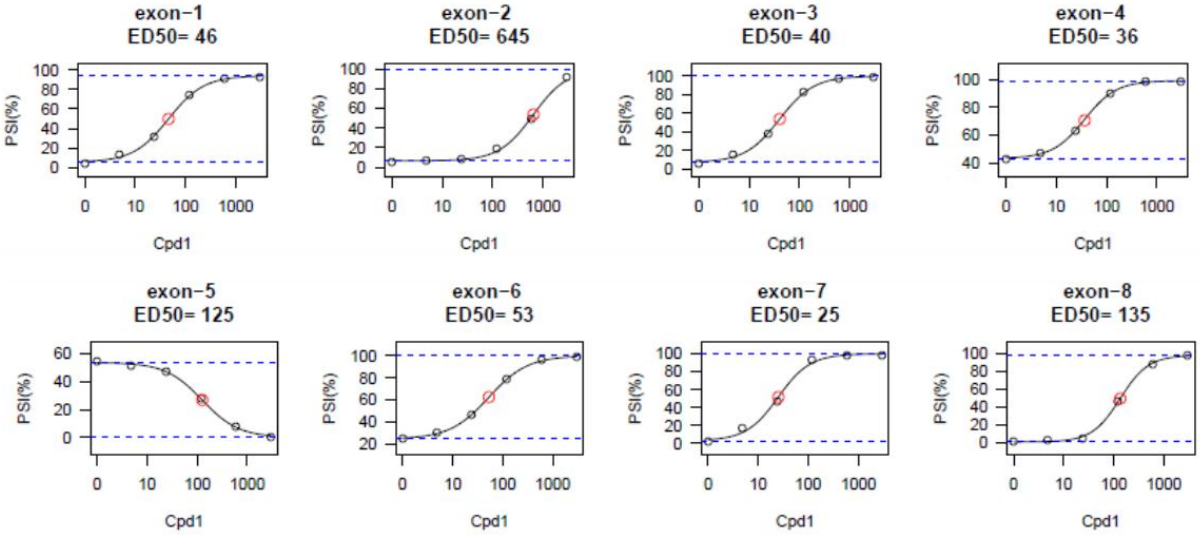
Above example shows the output files for batch analysis using default I- and S- probes counts as input. The description of the file is listed in the parentheses.

7.2. *CV* vs. *Ei* plot



Above is an example of *CV* vs. *Ei* plot output from NanoString-Splice batch analysis. The initial estimated *Ei* (with minimal *CV*) was marked by a blue circle. If the difference of $CV_{(minimal)}$ and $CV_{(Ei=1)}$ (red circle) is less than 0.1, *Ei*=1 was used because the *Ei* does not decrease the *CV* significantly.

7.3. Dose-response curves



Above is an example of output fitting curves of adjusted-PSI data from NanoString-Splice batch analysis. *ED50* values is marked by a red circle. The lower and upper limit are shown as blue dashed lines.

7.4. Output table

The file “input.txt.output.combine.txt” has all the calculated values combined. The example of column names and their meanings are shown below:

Column name	**Meaning**
exon_id	ID of the targets from user input
I_DMSO.1	I-probe counts from user input
I_DMSO.2	
I_Cpd_4.8nM	
I_Cpd_24nM	
I_Cpd_120nM	
I_Cpd_600nM	
I_Cpd_3uM	
S_DMSO.1	S-probe counts from user input
S_DMSO.2	
S_Cpd_4.8nM	
S_Cpd_24nM	
S_Cpd_120nM	
S_Cpd_600nM	
S_Cpd_3uM	
Ei_estimated	Estimated Ei value
Ei_used	Finally used Ei value. This may be different compared to Ei_estimated. Eg. user may forbid Ei-adjustion, in this case, all Ei_used will be 1. user can also provide their own Ei list. In this case, the Ei_used will be the values from the user input Ei file.
PSI_DMSO.1	Calculated adjusted-PSI values using Ei_used for adjustion (when Ei_used=1, adjusted-PSI will be the same as raw-PSI)
PSI_DMSO.2	

PSI_Cpd_4.8nM	
PSI_Cpd_24nM	
PSI_Cpd_120nM	
PSI_Cpd_600nM	
PSI_Cpd_3uM	
Pval.Cpd1	*P*-value for the dose-response curve (DRC) fitting using PSI data using drc R-package.
steepness.Cpd1	steepness in drc fitting using PSI data
lowerLimit.Cpd1	lower limit in drc fitting using PSI data
upperLimit.Cpd1	upper limit in drc fitting using PSI data
ED50_PSI.Cpd1	ED50 in drc fitting using PSI data
EC_dPSI10.Cpd1	effective concentration when PSI change by 10%
EC_dPSI30.Cpd1	effective concentration when PSI change by 30%
Pval.I.Cpd1	*P*-value for the dose-response curve fitting using I-probe counts data
steepness.I.Cpd1	steepness in drc fitting using I-probe counts data
lowerLimit.I.Cpd1	lower limit in drc fitting using I-probe counts data
upperLimit.I.Cpd1	upper limit in drc fitting using I-probe counts data
ED50_Exp.I.Cpd1	ED50 in drc fitting using I-probe counts data
EC_5x.I.Cpd1	effective concentration when I-probe counts change by 5 fold
EC_10x.I.Cpd1	effective concentration when I-probe counts change by 10 fold
Pval.S.Cpd1	*P*-value for the dose-response curve fitting using S-probe counts data
steepness.S.Cpd1	steepness in drc fitting using S-probe counts data
lowerLimit.S.Cpd1	lower limit in drc fitting using S-probe counts data
upperLimit.S.Cpd1	upper limit in drc fitting using S-probe counts data
ED50_Exp.S.Cpd1	ED50 in drc fitting using S-probe counts data
EC_5x.S.Cpd1	effective concentration when S-probe counts change by 5 fold
EC_10x.S.Cpd1	effective concentration when S-probe counts change by 10 fold

8. Prerequisite of using the R source code

Install R. “Rscript” command should be available in terminal. The “drc” (Analysis of Dose-Response Curves) R package is required for the dose-response analysis.

File “NanoString-Splice.inc.R” has all the functions required to run NanoString-Splice. Put “NanoString-Splice.inc.R” in the same directory with the other two .R files (batch_analysis_wrapper.R and draw_custom_DRC.R).

9. Use Rscript batch_analysis_wrapper.R

9.1. Description

“batch_analysis_wrapper.R” is the main R script which deals with the batch analysis. Use it to analyze PSI and dose-response (optional) for a list of targets.

9.2. Usage

```
Rscript batch_analysis_wrapper.R \
  -compound "Cpd1" \
```

-sample_names "s1 s2 s3 s4 s5 s6" \
-doses "dose1 dose2 dose3 dose4 dose5 dose6" \
-input_data_f "path\to\input\file" \
[options...]

9.3. Options

-compound	Compound name. The default is "Cpd1".
-sample_names	Sample names separated by spaces. This is a required option.
-doses	Doses match with sample_names. Numeric numbers separated by spaces. Eg. "0 4.8 24 120 600 3000".
-input_data_f	Path to the input file. This is a required option.
-DataType	Data type. Could be one of these values: "countsIS" (I- and S-probe counts), "PSI" (PSI data only), "counts" (counts data only). The default is "countsIS".
-Ei_option	Ei option. Could be one of these values: "DoAdjustion", "userProvide" (user will provide Ei table), "noAdjustion". The default is "DoAdjustion".
-input_Ei_f	Path to input Ei file. Provide file path when "-Ei_option" is "userProvide".
-DRC_option	Dose-response curve (DRC) option. Choose "yes" (do DRC analysis) or "no" (do not do DRC analysis).
-PSI_cal_rnumMin	Minimum total counts (I+S) required to calculate a PSI value. (default is 20)
-No_response_dose_plateau	The value to return when there is no dose-response. (default is 10000)
-NoResponse_P_cutoff	The P-value cutoff required to call a dose-response. (default is 0.05)
-min_PSI_response	The minimum percentage change of PSI required to call a dose-response. (default is 10)
-min_fold_response	The minimum fold change of counts data required to call a dose-response. (default is 1.5)
-dose_at_delta_PSIs	Calculate the compound dose which trigger the change of absolute PSI value by certain levels. Default is "10 30", meaning "EC_dPSI10" and "EC_dPSI30" will be calculated.
-dose_at_FoldChanges	Calculate the compound dose which trigger the fold change of counts by certain levels. Default is "5 10", meaning "EC_5x" and "EC_10x" will be calculated.

9.4. Output

Output files will be in the same directory as input file (input_data_f). See "List of output files" for details.

10. Use Rscript draw_custom_DRC.R

10.1. Description

"draw_custom_DRC.R" is the R script to do custom dose-response analysis. Use it to generate dose-response curves for several compounds and/or targets.

10.2. Usage

```
Rscript draw_custom_DRC.R \
        -project_name "my project" \
        -IDs "Cpd1 Cpd2"  \
        -doses_str "xx xx xx xx xx xx|xx xx xx xx xx xx" \
        -data_str "xx xx xx xx xx xx|xx xx xx xx xx xx" \
        -mycolors "black red" \
        [options...]
```

10.3. Options

-IDs	IDs of data. Could be compound IDs. (default is "Cpd1 Cpd2")
-doses_str	Doses string. Should be formatted like "xx xx xx xx xx xx|xx xx xx xx xx xx". Eg. "0 4.8 24 120 600 3000|0 4.8 24 120 600 3000". If all data use the same set of doses, it also accept just one dose series (eg. "0 4.8 24 120 600 3000").
-data_str	Data string. Should be formatted like "xx xx xx xx xx xx|xx xx xx xx xx xx". Eg. "3 25 66 94 97 98|3 3 4 10 54 84".
-mycolors	Colors used to represent each ID in the same order. Eg. "black red blue green cyan pink". A full list of R color names can be found in NanoString-splice website (Link).
-DataType	Data type. Could be one of these values: "PSI" (PSI data), "counts" (counts data). The default is "PSI".
-output_root	Path to the output folder.
-LLsearch_range	Lower limit parameter search range. For PSI data only. The default is "0 100", meaning the lower limit can be any value between 0 to 100%.
-ULsearch_range	Upper limit parameter search range. For PSI data only. The default is "0 100", meaning the upper limit can be any value between 0 to 100%.
-No_response_dose_plateau	The value to return when there is no dose-response. (default is 10000)
-NoResponse_P_cutoff	The P-value cutoff required to call a dose-response. (default is 0.05)
-min_PSI_response	The minimum percentage change of PSI required to call a dose-response. (default is 10)
-min_fold_response	The minimum fold change of counts data required to call a dose-response. (default is 1.5)
-dose_at_delta_PSIs	Calculate the compound dose which trigger the change of absolute PSI value by certain levels. Default is "10 30", meaning "EC_dPSI10" and "EC_dPSI30" will be calculated.
-dose_at_FoldChanges	Calculate the compound dose which trigger the fold change of counts by certain levels. Default is "5 10", meaning "EC_5x" and "EC_10x" will be calculated.
-file_format	Output image file format. Can choose "png", "bmp", "jpeg", "tiff" or "pdf". Default is "png".

10.4. Output

One image file and one txt file will be generated in directory specified by “output_root”.