Detect alt-splicing (ANOVA)

Alternative splicing results in a single gene coding for multiple protein isoforms, so this task can only be invoked from transcript level data. The algorithm is is based on ANOVA to detect genes with multiple transcripts showing expression changes differently in different biology groups, e.g. a gene has two transcripts: A and B, transcript A is showing up-regulation in the treated group comparing to the control group, while B is showing down regulation in treated group.

Alt-splicing ANOVA dialog

The alt-splicing dialog is very similar to ANOVA dialog, since the analysis is based on the ANOVA model specified. To setup an ANOVA model, first chose factors from the available sample attributes. The factors can be categorical or numeric attribute(s). Click on a **check box** to select and click **Add factors** button to add a factor to the model (Figure 1).

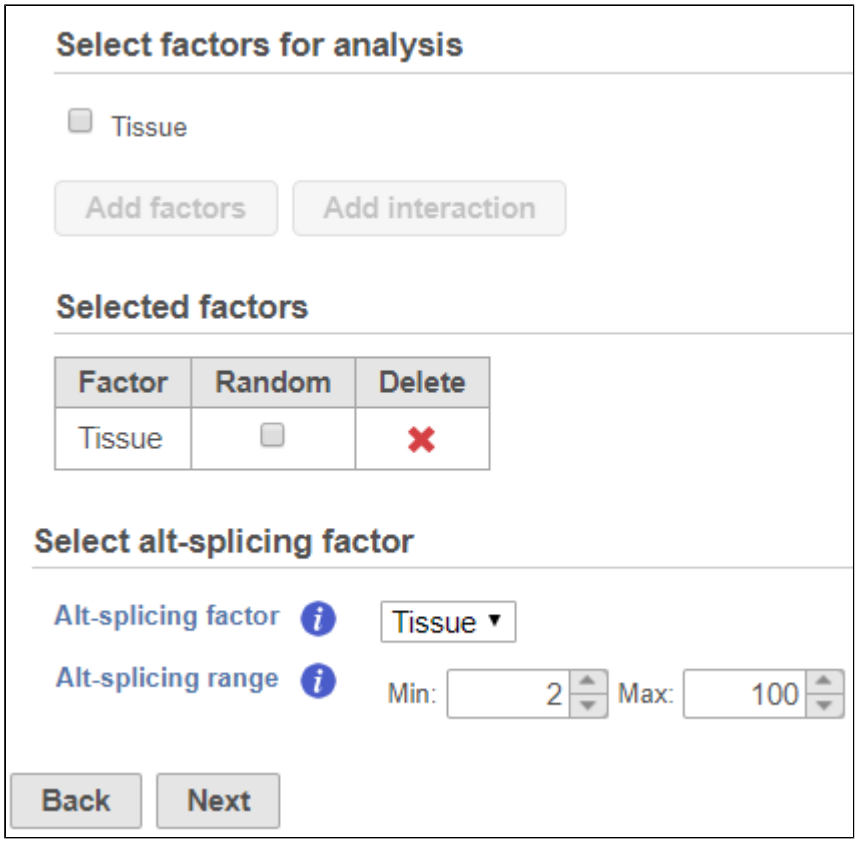

Figure 6. Alt-splicing dialog: selecting factors and alt-splicing factors

Only one alt-splicing factor needs to be selected from the ANOVA factors. The ANOVA model performed is based on the factors specified in the dialog, while the transcript ID and transcript ID interaction with alt-splicing factor effects are added into the model automatically.

- Transcript ID effect: not all transcripts of a gene are expressed at the same level, so transcript ID is added to the model to account for transcript-to. transcript differences.
- Interaction of transcript ID with alt-splicing factor: that effect is used to estimate whether different transcripts have different expression among the levels of the same factor.

Suppose there is an experiment designed to detect transcripts showing differential expression in two tissue groups: liver vs muscle. The alt-splicing ANOVA dialog allows you to specify the ANOVA model that in this analysis is the *Tissue*. The alt-splicing factor is chosen from the ANOVA factor(s), so the alt-splicing factor is also *Tissue* (Figure 1).

The alt-splicing range will limit analysis to genes possessing the number of transcripts in the specified range. Lowering the maximum number of transcripts will increase the speed of analysis.

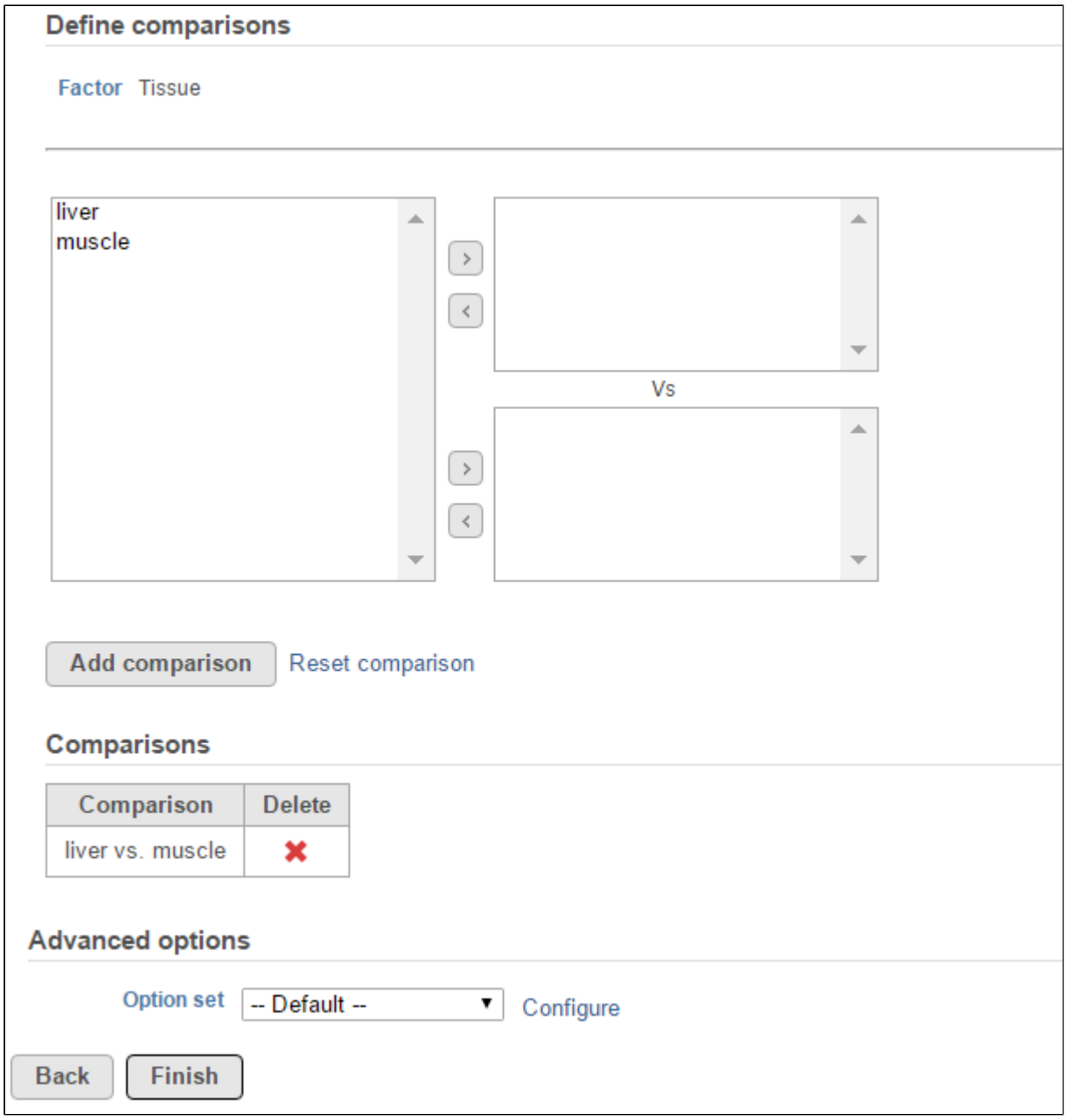

Figure 7. Alt-splicing ANOVA comparisons setup dialog

Click **Next** to setup the comparisons (Figure 2). The levels (i.e. subgroups) of the alt-splicing factor will be displayed in the left panel; click to select a level name and move it to one of the panels on the right. The fold change calculation on the comparison will use the group in the top panel as the numerator, and the group in the bottom panel as the denominator. Click on **Add comparison** button to add a comparison to the comparisons table.

ANOVA advanced options

Click on the *Configure* to customize Advanced options (Figure 3).

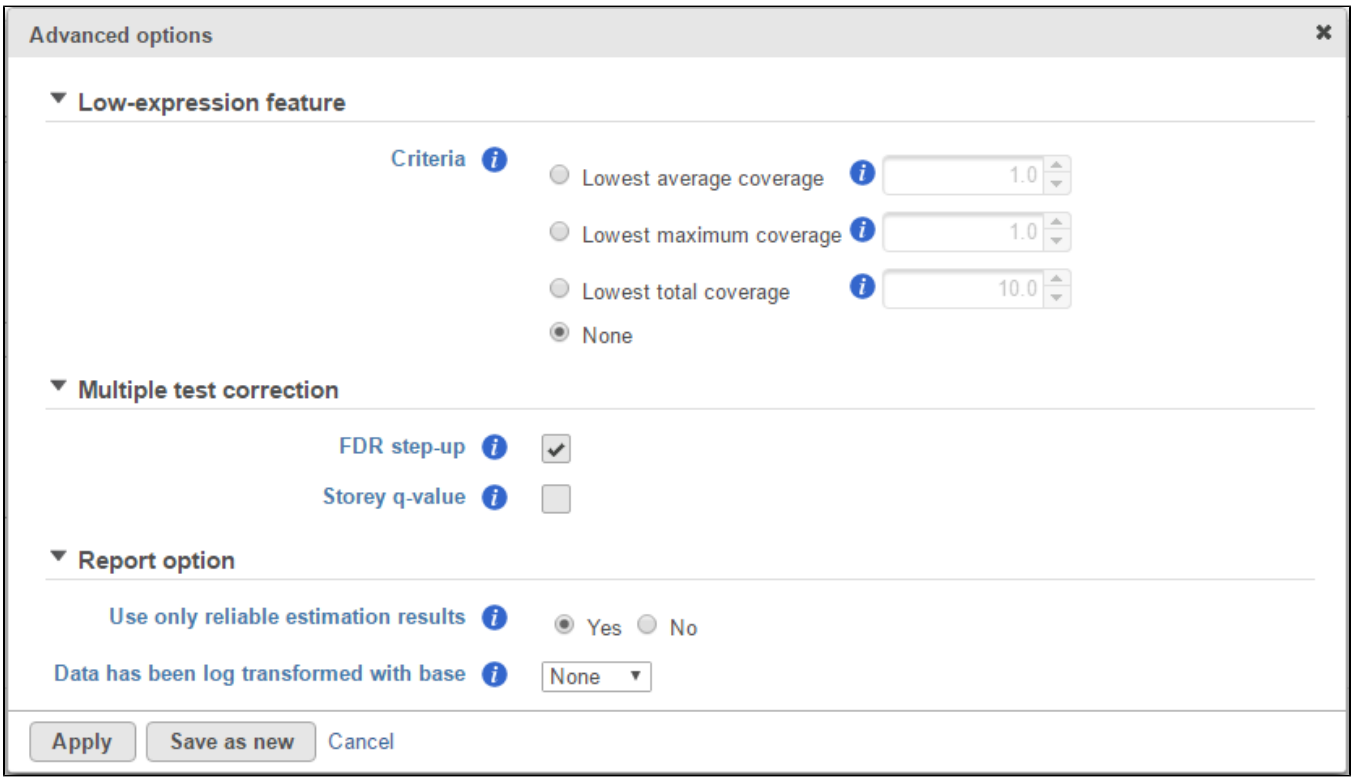

Figure 8. Configuring advanced options when running alt-splicing ANOVA

Low-expression feature and *Multiple test correction* sections are the same as the matching GSA advanced option, so see GSA advanced options discussio n.

Report option

- *User only reliable estimation results:* There are situations when a model estimation procedure does not fail outright, but still encounters some difficulties. In this case, it can even generate p-value and fold change on the comparisons, but they are not reliable, i.e. they can be misleading. Therefore, the default is set to *Yes*.
- *Data has been log transformed with base*: showing the current scale of the input data on this task.

Alt-splicing ANOVA report

For this analysis, only genes with more than one transcript will be included in the calculation. The report format is the same as ANOVA report, each row represent a transcript, and besides statistics information on the specified comparisons, there is also alt-splicing information at the right end of the table. That information is represented by the p-value of interaction of transcript ID with alt-splicing factor. Note that the transcripts of the same gene should have the same p-value. Small p-value indicates significant alt-splicing event, hence the table is sorted based on that p-value by default (Figure 4).

				liver vs muscle						Alt-splicing Info (Tissue)	
View	Gene ID	Transcript ID	Total counts	P-value	FDR step up	Ratio	Fold change	LSMean(liver)	LSMean(muscle)	P-value	FDR step up
	SLC25A3	NM_213611	116.87	9.24E-12	5.98E-11	0.17	-5.83	4.28	24.94	3.92E-20	6.87E-18
	SLC25A3	NM_005888	1,777.10	1.49E-21	7.05E-20	1.8E-3	-556.74	0.80	443.48	3.92E-20	6.87E-18
	SLC25A3	NM_002635	781.17	3.94E-12	2.81E-11	6.31	6.31	168.56	26.73	3.92E-20	6.87E-18
	TACC2	NM_001291877	137.33	5.19E-24	3.35E-22	2.51E-3	-399.01	0.09	34.25	6.92E-20	1.15E-17
	TACC2	NM_001291879	209.81	7.31E-3	9.72E-3	0.42	-2.40	15.44	37.02	6.92E-20	1.15E-17
	TACC2	NM_206862	879.46	2.17E-21	1.01E-19	4.11E-3	-243.16	0.90	218.96	6.92E-20	1.15E-17
	TACC2	NM_206861	321.71	5.91E-25	4.58E-23	1.27E-3	-790.42	0.10	80.33	6.92E-20	1.15E-17
	TACC2	NM_001291876	248.22	1.13E-27	1.31E-25	3.28E-4	-3,049.83	0.02	62.04	6.92E-20	1.15E-17
	PRG4	NM_005807	295.84	1.48E-4	2.44E-4	1.28	1.28	41.48	32.48	1.28E-19	2.02E-17
	PRG4	NM_001303232	675.16	5.79E-17	1.21E-15	2.62	2.62	122.12	46.67	1.28E-19	2.02E-17
	PRG4	NM_001127710	1,366.96	6.81E-13	5.83E-12	1.96	1.96	226.34	115.40	1.28E-19	2.02E-17
	PRG4	NM_001127709	1,806.24	2.81E-21	1.29E-19	3.95	3.95	360.32	91.24	1.28E-19	2.02E-17
	PRG4	NM_001127708	258.20	2.77E-9	1.05E-8	0.62	-1.60	24.82	39.73	1.28E-19	2.02E-17
	CSDE1	NM_001007553	421.42	3.57E-5	6.4E-5	0.47	-2.14	33.54	71.81	6.22E-19	9.35E-17
	CSDE1	NM_007158	1,337.13	3.85E-4	5.99E-4	0.54	-1.87	116.54	217.74	6.22E-19	9.35E-17
	CSDE1	NM_001242893	659.32	5.79E-11	3.16E-10	0.22	-4.48	30.07	134.76	6.22E-19	9.35E-17
	CSDE1	NM_001130523	2,475.96	1.21E-23	7.23E-22	0.02	-44.74	13.53	605.46	6.22E-19	9.35E-17
	CSDE1	NM_001242891	844.98	4.01E-24	2.69E-22	0.02	-48.49	4.27	206.98	6.22E-19	9.35E-17
	CSDE1	NM_001242892	189.57	5.02E-12	3.49E-11	0.20	-5.07	7.81	39.58	6.22E-19	9.35E-17
	LIMCH1	NM_001330672	386.63	1.37E-20	5.58E-19	4.86E-3	-205.89	0.47	96.19	7.41E-19	1.06E-16
	CD99	NM_002414	164.02	0.43	0.46	1.21	1.21	22.42	18.58	9.41E-19	1.29E-16
	CD99	NM_002414.1	164.02	0.43	0.46	1.21	1.21	22.42	18.58	9.41E-19	1.29E-16
	ASPH	NM_004318	155.30	1E-3	1.48E-3	2.16	2.16	26.54	12.29	1.48E-18	1.95E-16
	SCP2	NM_002979	342.91	4.31E-11	2.4E-10	9.74	9.74	77.75	7.98	1.64E-18	2.06E-16
	PKM	NM_001206799	3,400.86	4.64E-20	1.65E-18	0.03	-33.30	24.79	825.42	1.97E-18	2.39E-16

Figure 9. Alt-splicing report. Clicking on the column header sorts the table. Panel on the left filters the table

In the example above (Figure 4), the alt-splicing p-value of gene SLC25A3 is very small which indicates that this gene shows preferential transcript expression across tissues. There are 3 splicing variants of the gene: NM_213611, NM_005888 and NM_002635. Fold change clarifies that NM_005888 has higher expression in the muscle relative to the liver (negative fold change, liver as the reference category), while NM_002635 has higher expression in the liver.

To visualize the difference, click on the **Browse to location** icon ( ) (Figure 5). The 3rd exon is differentially expressed between NM_005888 and NM_002635. Muscle primarily expresses NM_005888 while liver primarily uses NM_002635.

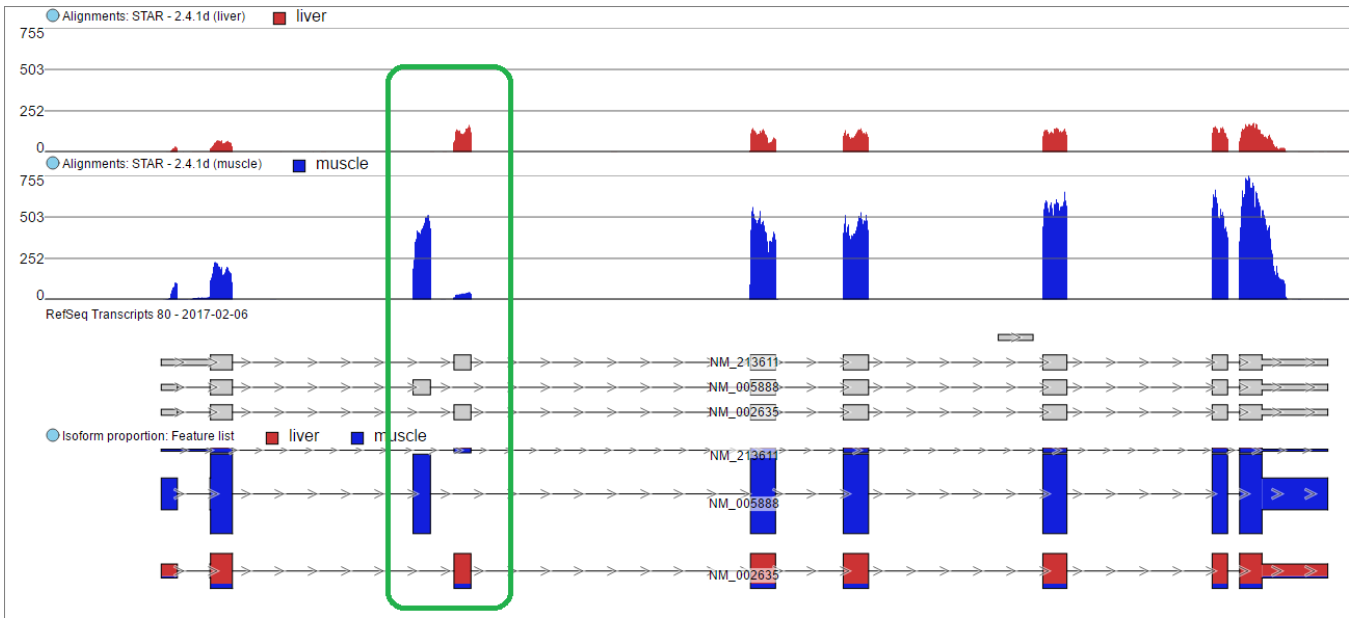

Figure 10. Isoform proportion track in chromosome view visualising alternative splicing. Differential usage of exon 3 is highlighted

Additional Assistance

If you need additional assistance, please visit our support page to submit a help ticket or find phone numbers for regional support.

Your Rating: Results: 40 rates