Detecting differential expression in RNA-Seq data

During import, you created a categorical attribute called *Tissue* and assigned the 4 samples to either the *muscle* or *not muscle* groups. This step was to create replicates within a group, albeit this grouping is somewhat artificial and is only used in this tutorial because we want to illustrate ANOVA with a small data set. Replicates are a prerequisite for differential expression analysis using ANOVA.

- Select **Differential Expression Analysis** from the *Analyze Known Genes* section of the *RNA-Seq* workflow

The *Differential Expression Analysis* dialog offers the choice of analyzing at Gene-,Transcript-, or Exon-level.

- Select **Gene-level**
- Specify the **1/gene_rpkm (RNA-Seq_results.gene.rpkm)** spreadsheet from the *Spreadsheet* drop-down menu (Figure 1)

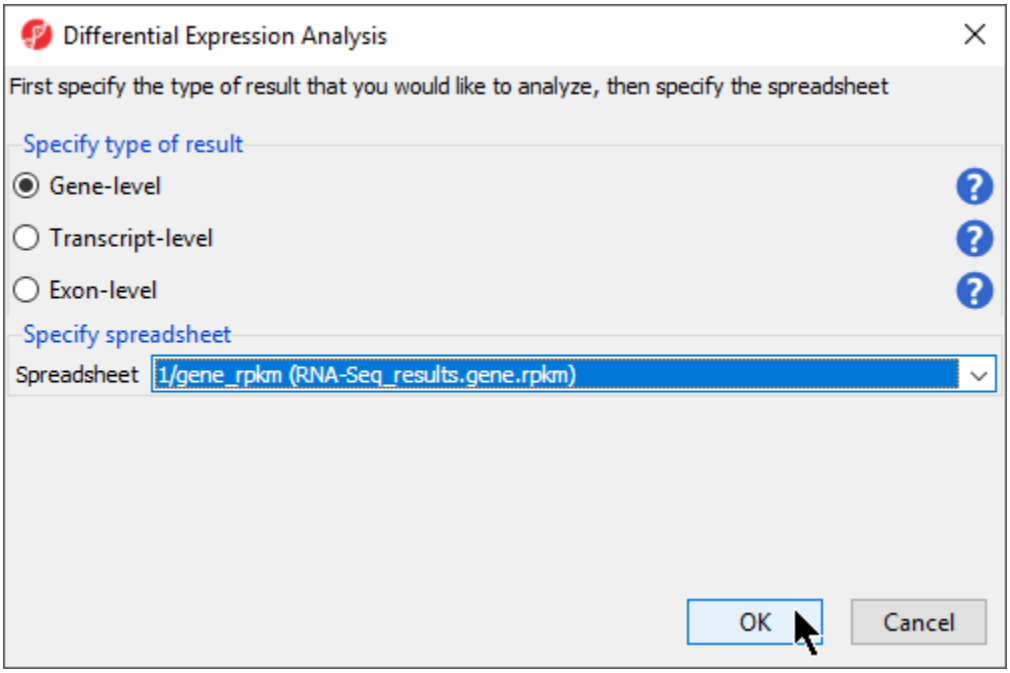


Figure 6. Choosing the type of differential expression analysis

- Select **OK** to open the *ANOVA* dialog

Available factors are listed in the *Experimental Factor(s)* panel on the left-hand side of the dialog.

- Select **Tissue**, then select **Add Factor >** to move **Tissue** to the *ANOVA Factor(s)* panel on the right-hand side of the dialog (Figure 2)

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

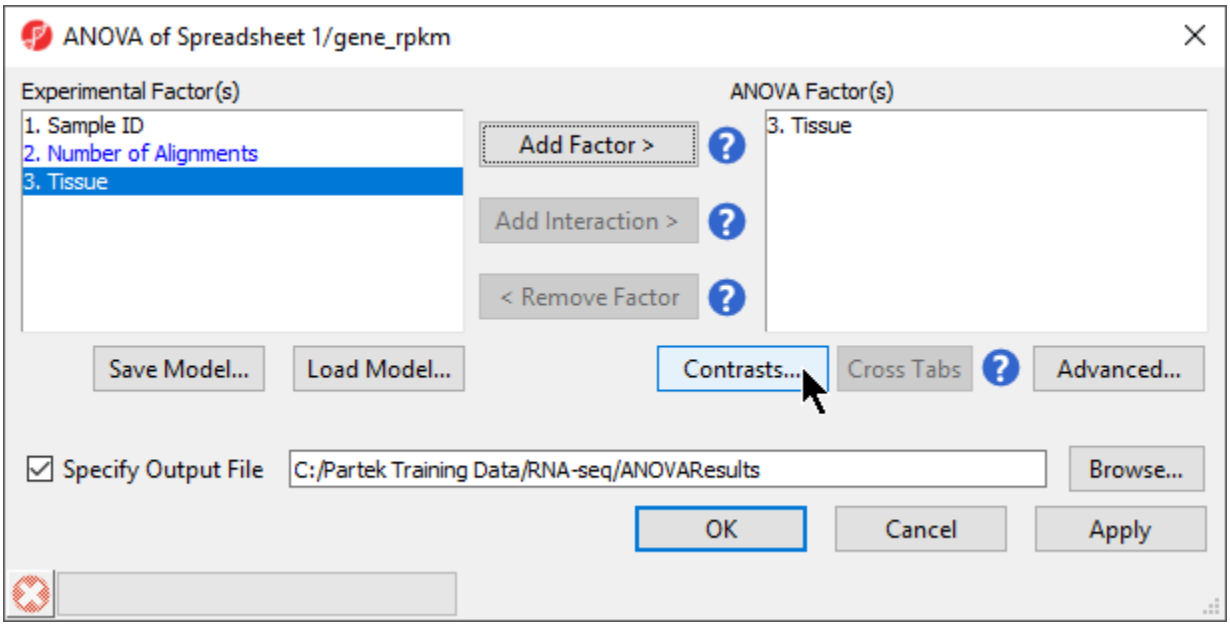


Figure 7. The ANOVA dialog

If the ANOVA were now performed (without contrasts), a p-value for differential expression would be calculated, but it would only indicate if there are differences within the factor *Tissue*; it would not inform you which groups are different or give any information on the magnitude of the difference between groups (fold-change or ratio). To get this more specific information, you need to define linear contrasts.

- Select **Contrasts...** to open the *Configure* dialog
- For *Select Factor/Interaction*, **Tissue** will be the only factor available as it was the only factor included in the ANOVA model in the previous step; if multiple factors were included, they could be selected in the *Select Factor/Interaction:* drop-down menu. The levels in this factor are listed on the *C andidate Level(s)* panel on the left side of the dialog
- For this data set, verify that **No** is selected for *Data is already log transformed?*
- Left click to select **muscle** from the *Candidate Level(s)* panel and move it to the *Group 1* panel (renamed *muscle*) by selecting **Add Contrast Level >** in the top half of the dialog. *Label 1* will be changed to the subgroup name automatically, but you can also manually specify the label name
- Select **not muscle** from the *Candidate Level(s)* panel and move it to the *Group 2* panel (renamed *not muscle)*
- The **Add Contrast** button can now be selected (Figure 3)
- Select **OK** to return to the *ANOVA* dialog

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

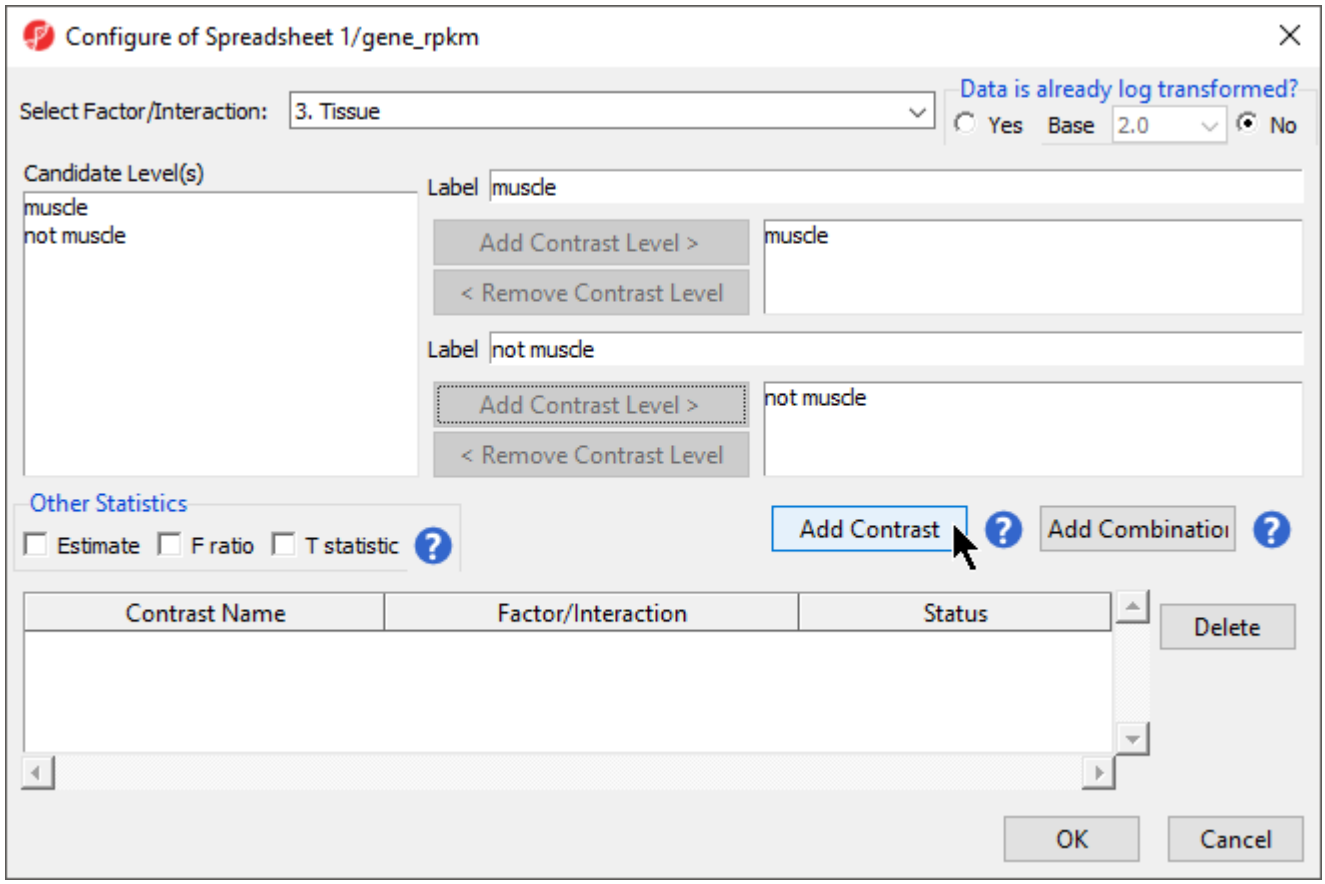
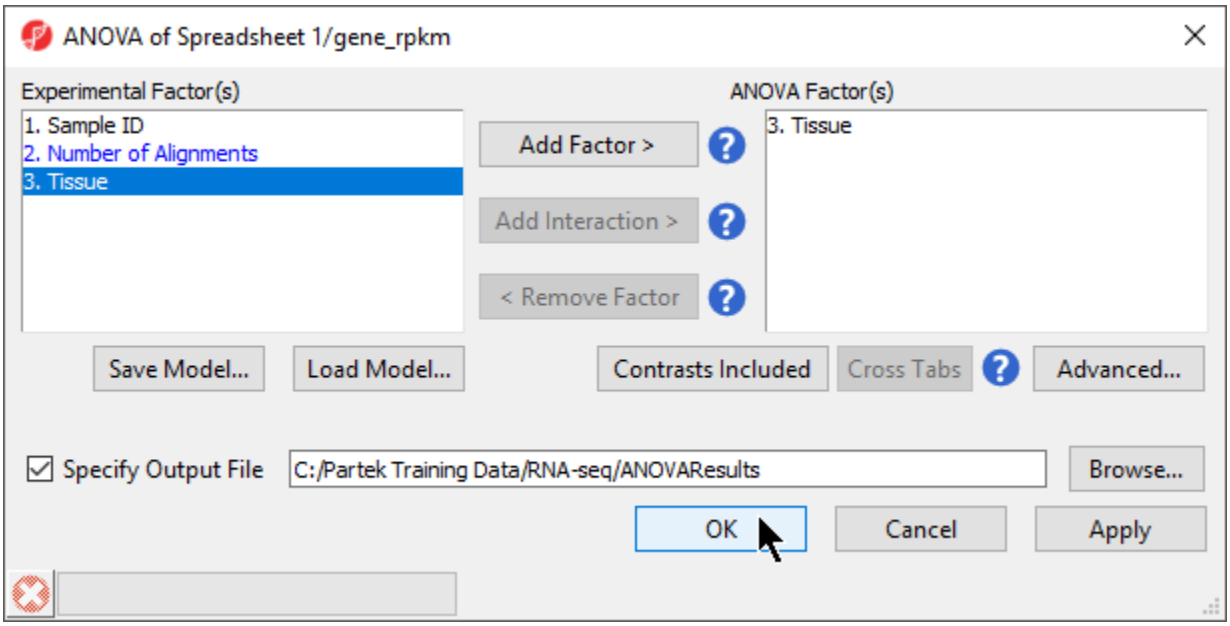


Figure 8. Defining linear contrasts

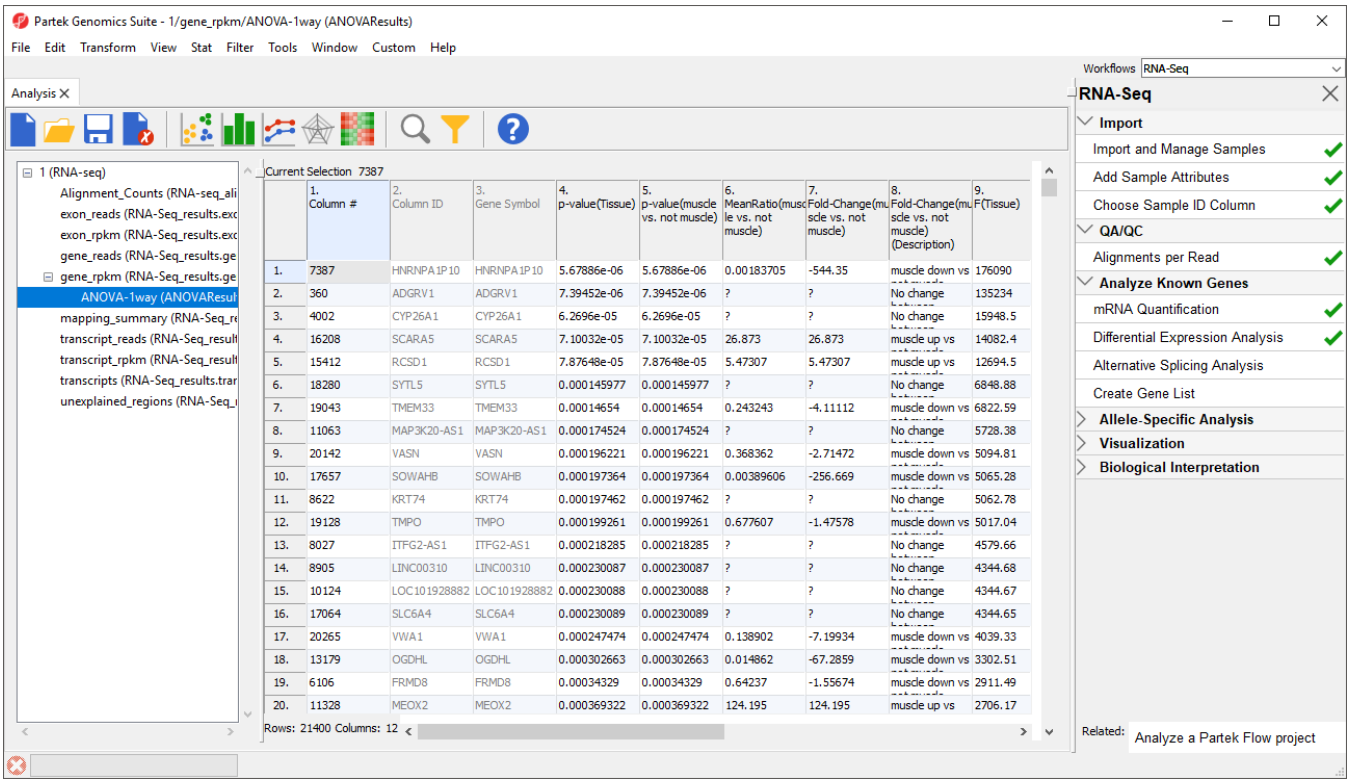
- Select **OK** to perform the ANOVA as configured (Figure 4)



Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

Figure 9. Fully configured ANOVA

Once the ANOVA has been performed on each gene in the data set, an ANOVA child spreadsheet *ANOVA-1way (ANOVAResults)* will appear under the *gene_rpkm* spreadsheet (Figure 5). The format of the ANOVA spreadsheet is similar for all workflows. Mouse over each column title for a description of the column contents.



	1. Column #	2. Column ID	3. Gene Symbol	4. p-value(Tissue)	5. p-value(muscle vs. not muscle)	6. MeanRatio(musc le vs. not muscle)	7. Fold-Change(mu scle vs. not muscle)	8. Fold-Change(mu scle vs. not muscle) (Description)	9. F(Tissue)
1.	7387	HNRNPA1P10	HNRNPA1P10	5.67886e-06	5.67886e-06	0.00183705	-544.35	muscle down vs	176090
2.	360	ADGRV1	ADGRV1	7.39452e-06	7.39452e-06	?	?	No change	135234
3.	4002	CYP26A1	CYP26A1	6.2696e-05	6.2696e-05	?	?	No change	15948.5
4.	16208	SCARA5	SCARA5	7.10032e-05	7.10032e-05	26.873	26.873	muscle up vs	14082.4
5.	15412	RCSD1	RCSD1	7.87648e-05	7.87648e-05	5.47307	5.47307	muscle up vs	12694.5
6.	18280	SYTL5	SYTL5	0.000145977	0.000145977	?	?	No change	6848.88
7.	19043	TMEM33	TMEM33	0.00014654	0.00014654	0.243243	-4.11112	muscle down vs	6822.59
8.	11063	MAP3K20-AS1	MAP3K20-AS1	0.000174524	0.000174524	?	?	No change	5728.38
9.	20142	VASN	VASN	0.000196221	0.000196221	0.368362	-2.71472	muscle down vs	5094.81
10.	17657	SOWAHB	SOWAHB	0.000197364	0.000197364	0.00389606	-256.669	muscle down vs	5065.28
11.	8622	KRT74	KRT74	0.000197462	0.000197462	?	?	No change	5062.78
12.	19128	TMPO	TMPO	0.000199261	0.000199261	0.677607	-1.47578	muscle down vs	5017.04
13.	8027	ITFG2-AS1	ITFG2-AS1	0.000218285	0.000218285	?	?	No change	4579.66
14.	8905	LINC00310	LINC00310	0.000230087	0.000230087	?	?	No change	4344.68
15.	10124	LOC101928882	LOC101928882	0.000230088	0.000230088	?	?	No change	4344.67
16.	17064	SLC6A4	SLC6A4	0.000230089	0.000230089	?	?	No change	4344.65
17.	20265	VWA1	VWA1	0.000247474	0.000247474	0.138902	-7.19934	muscle down vs	4039.33
18.	13179	OGDHL	OGDHL	0.000302663	0.000302663	0.014862	-67.2859	muscle down vs	3302.51
19.	6106	FRMD8	FRMD8	0.00034329	0.00034329	0.64237	-1.55674	muscle down vs	2911.49
20.	11328	MEOX2	MEOX2	0.000369322	0.000369322	124.195	124.195	muscle up vs	2706.17

Figure 10. Viewing ANOVA results

In this tutorial, the overall p-value for the factor (column 4) is the same as the p-value for the linear contrast (column 5) as there are only two levels within *Tissue*. If we had more than two groups, the overall p-value and the linear contrast p-values would most likely differ. You can also see the ? symbol in the ratio/fold-change columns (6 and 7) for several genes that also have a low p-value because there are zero reads in one of the groups, thus making it impossible to calculate ratios and fold-changes between groups.

For using ANOVA with more complicated experimental designs, including multiple factors and linear contrasts, please refer to Identifying differentially expressed genes using ANOVA in the Gene Expression Analysis tutorial.

« RNA-Seq mRNA quantification Creating a gene list with advanced options »

Additional Assistance

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

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

Your Rating: Results: 34 rates

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.