

# 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\_rpk (RNA-Seq\_results.gene.rpk)** spreadsheet from the *Spreadsheet* drop-down menu (Figure 1)

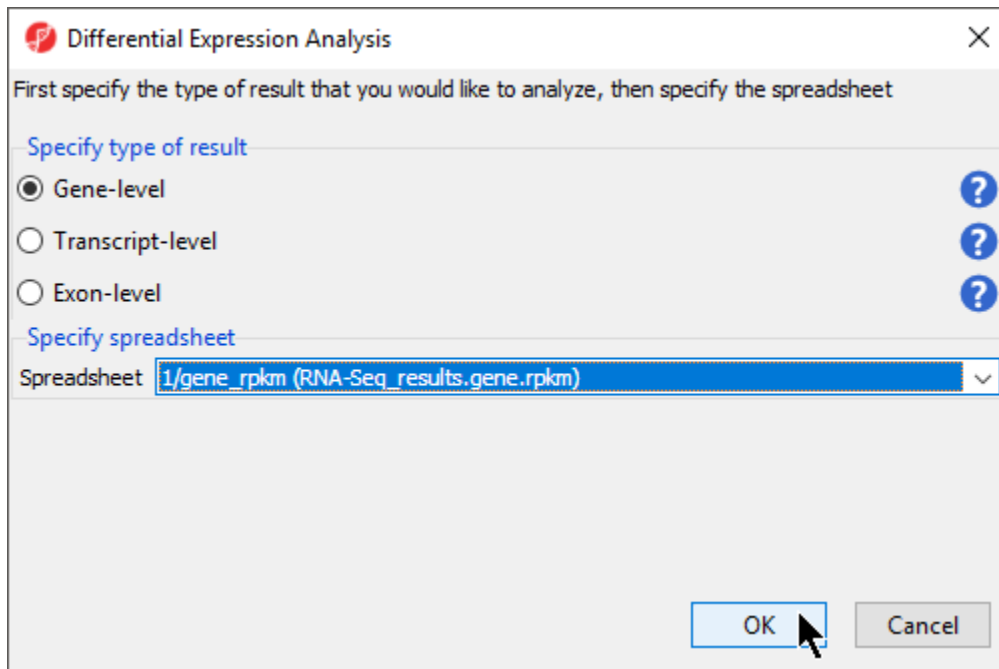


Figure 6. Choosing the type of differential expression analysis

- 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)

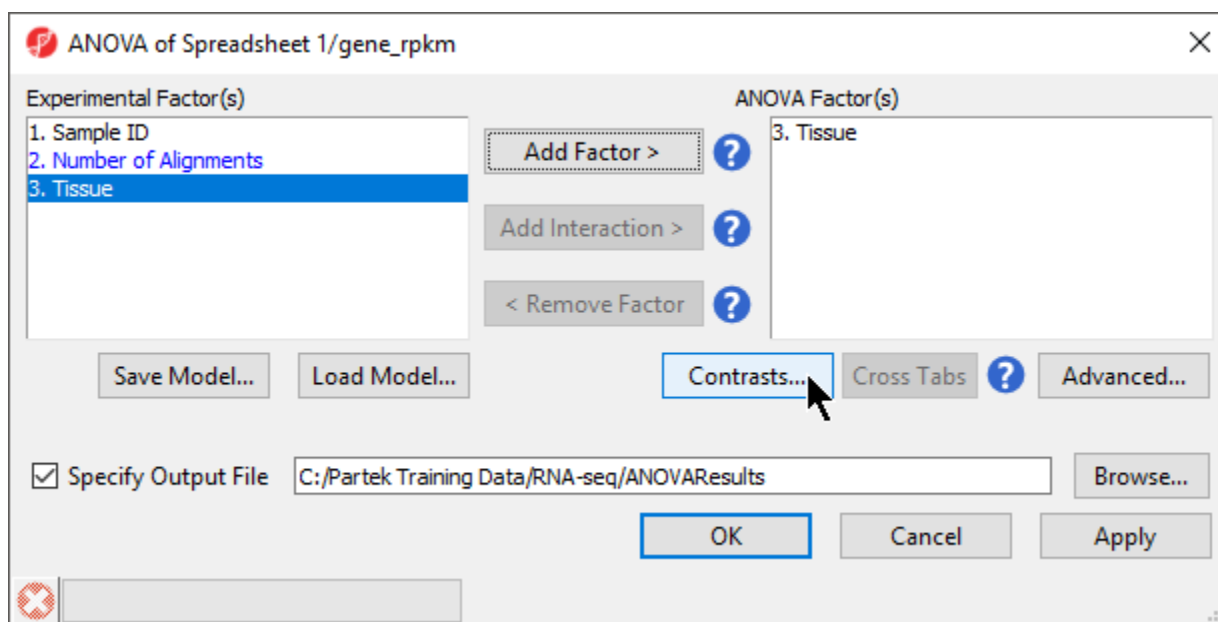


Figure 7. The ANOVA dialog

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 *Candidate 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

**Configure of Spreadsheet 1/gene\_rpkm**

Select Factor/Interaction: 3. Tissue Data is already log transformed?  
☐ Yes Base 2.0 ☒ No

Candidate Level(s)  
 muscle  
 not muscle

Label muscle  
 Add Contrast Level >  
 < Remove Contrast Level

Label not muscle  
 Add Contrast Level >  
 < Remove Contrast Level

Other Statistics  
☐ Estimate ☐ F ratio ☐ T statistic ?

Add Contrast ? Add Combination ?

| Contrast Name | Factor/Interaction | Status | Delete |
|---------------|--------------------|--------|--------|
|               |                    |        |        |

OK Cancel

Figure 8. Defining linear contrasts

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

**ANOVA of Spreadsheet 1/gene\_rpkm**

Experimental Factor(s)  
 1. Sample ID  
 2. Number of Alignments  
 3. Tissue

Add Factor > ?  
 Add Interaction > ?  
 < Remove Factor ?

ANOVA Factor(s)  
 3. Tissue

Save Model... Load Model... Contrasts Included Cross Tabs ? Advanced...

☒ Specify Output File C:/Partek Training Data/RNA-seq/ANOVAResults Browse...

OK Cancel Apply

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.

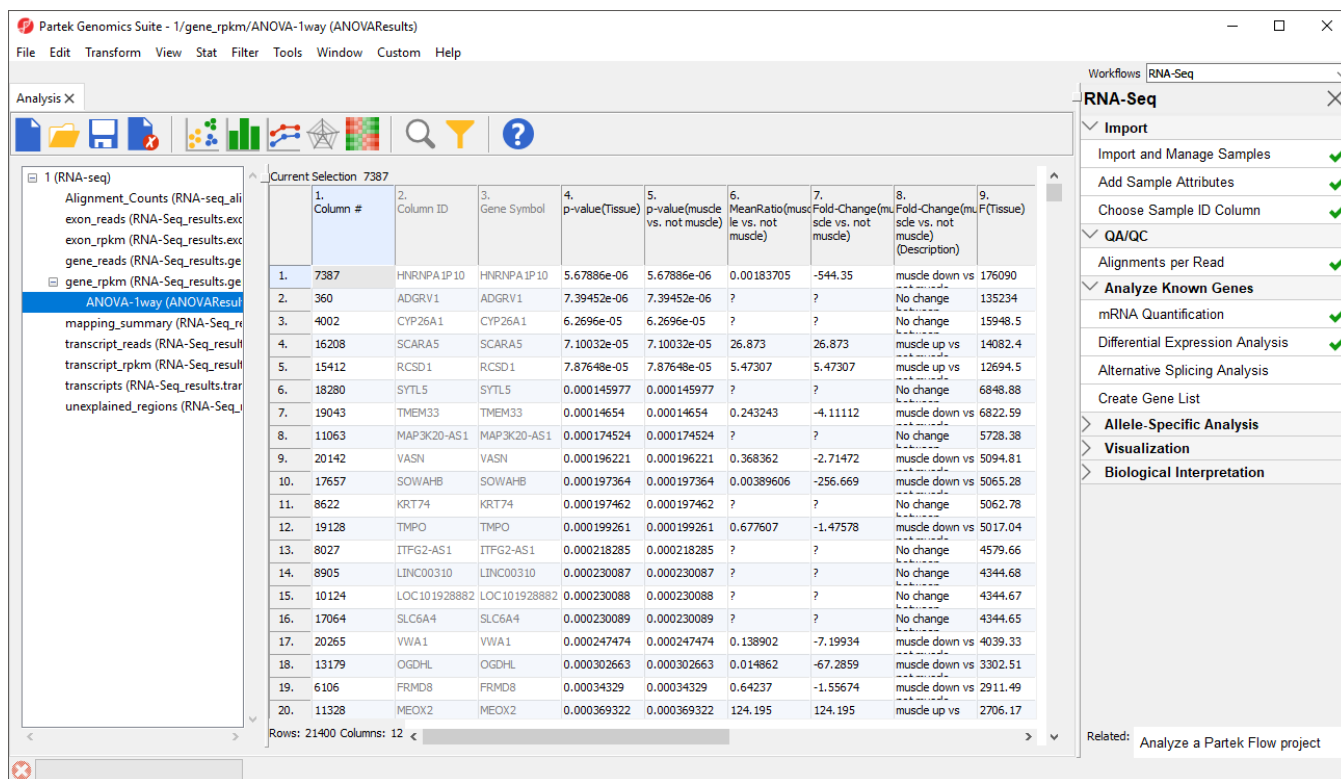


Figure 10. Viewing ANOVA results

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 *77 ssue*. 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.



Your Rating:



Results:



34 rates