Adding sample attributes

Now that the data has been imported, we need to make a few changes to the data annotation before analysis.

Modifying sample attributes

Notice that the *Sample ID* names in column 1 are gray (Figure 1). This indicates that *Sample ID* is a text factor. Text factors cannot be used as a variable in downstream analysis so we need to change *Sample ID* to a categorical factor.

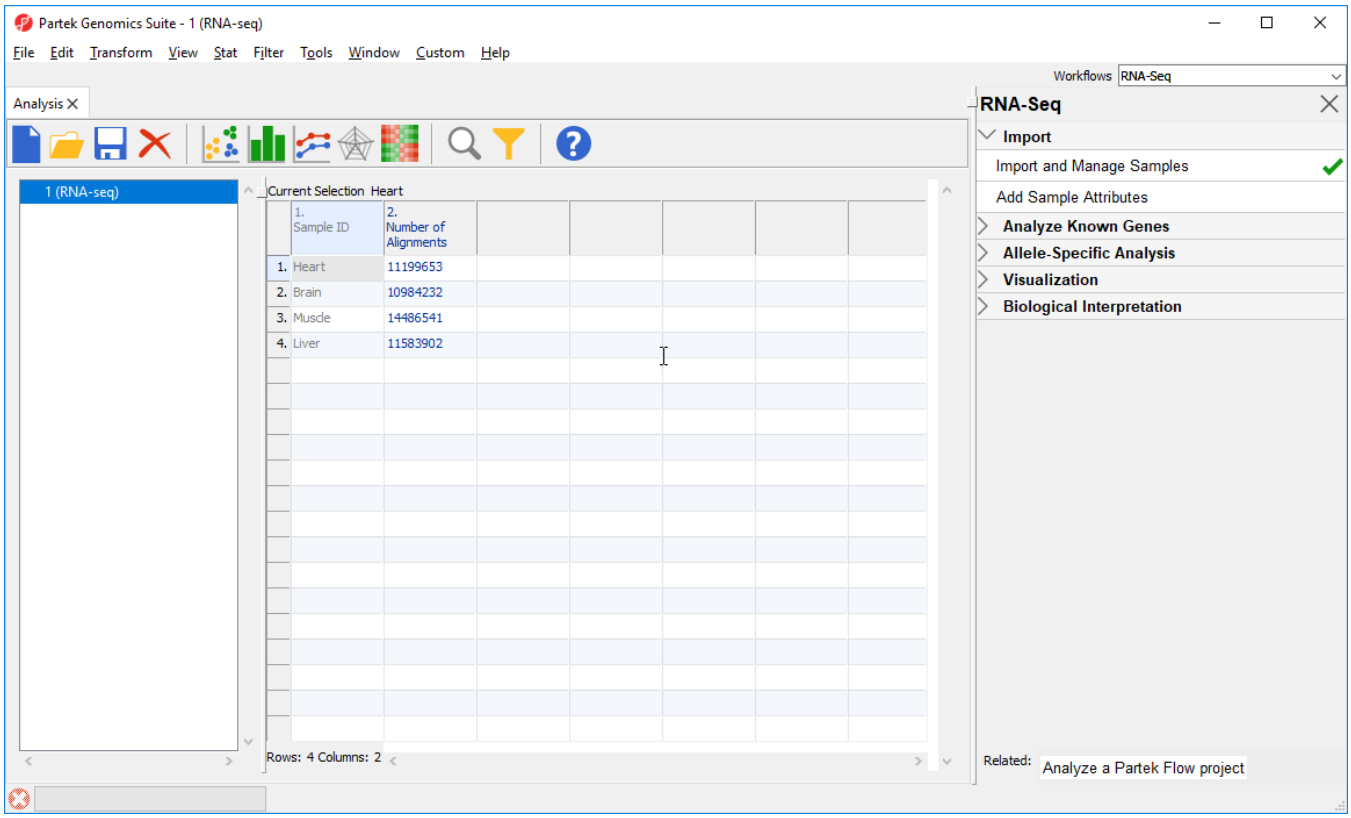

Figure 7. Viewing the imported data in a spreadsheet

- Right-click on the column header to invoke the pop-up menu
- Select **Properties** (Figure 2)

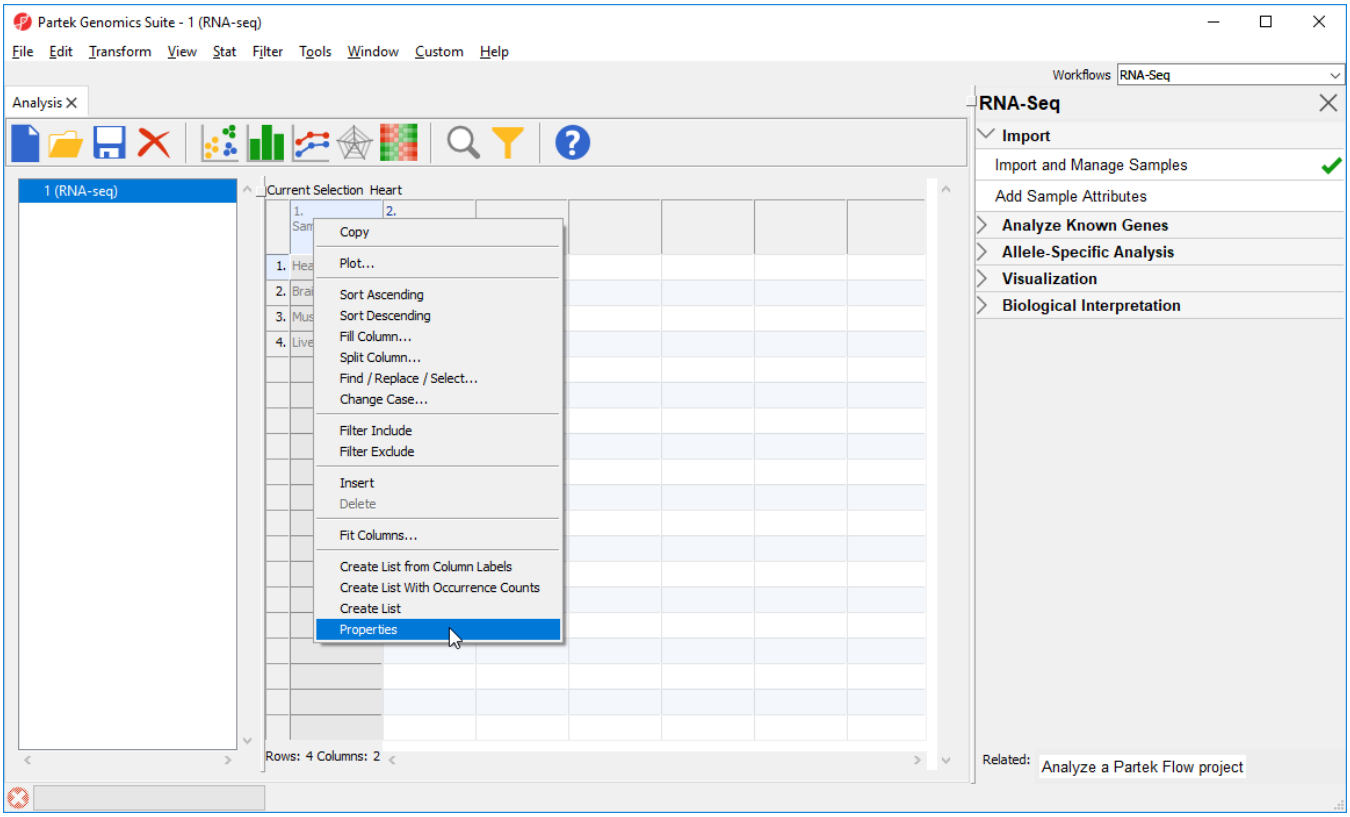

Figure 8. Changing column properties

- Configure the *Properties of Column 1 in Spreadsheet 1* dialog as shown (Figure 3) with *Type* set to **categorical** and *Attribute* set to **factor**

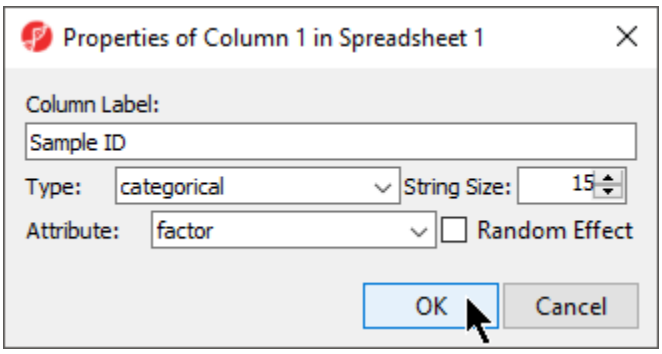

Figure 9. Changing column 1 properties

- Select **OK**

The samples names in column 1 are now black, indicating that they have been changed to a categorical variable. Next, we will add attributes for grouping the data.

Adding sample attributes

- From the *RNA-seq* workflow panel, select **Add Sample Attributes** to bring up the Add Sample Attributes dialog (Figure 4)

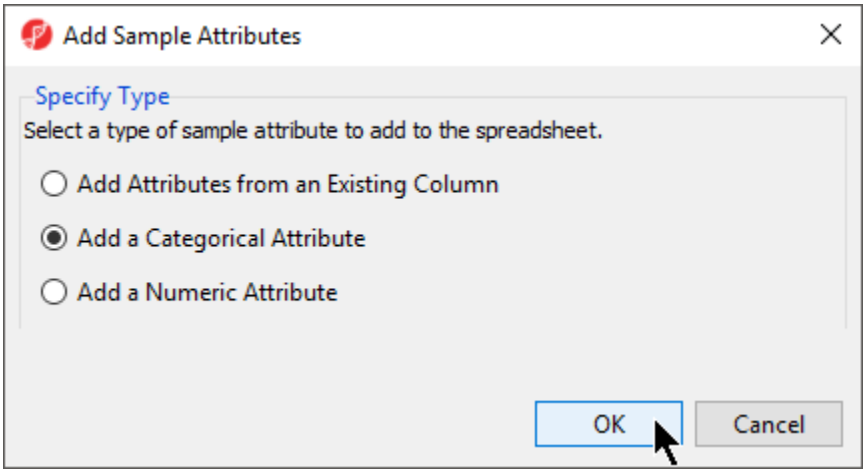

Figure 10. Add Sample Attributes dialog

- Select **Add a Categorical Attribute**
- Select **OK** to bring up the *Create categorical attribute* dialog

Creating a categorical sample attribute allows us to group samples. This is useful for designating samples as replicates, as members of an experimental group, or as sharing a phenotype of interest. In this tutorial, we have four different samples from different tissues and different donors, but to illustrate the available statistical analysis options, we need to divide the samples into two groups: *muscle* (Heart and Muscle) and *not muscle* (Brain and Liver).

- Set *Attribute name:* as *Tissue*
- Rename *Group 1* to *muscle* and *Group 2* to *not muscle*
- Select and drag the samples from the *Unassigned* panel to the correct group panel (Figure 5)

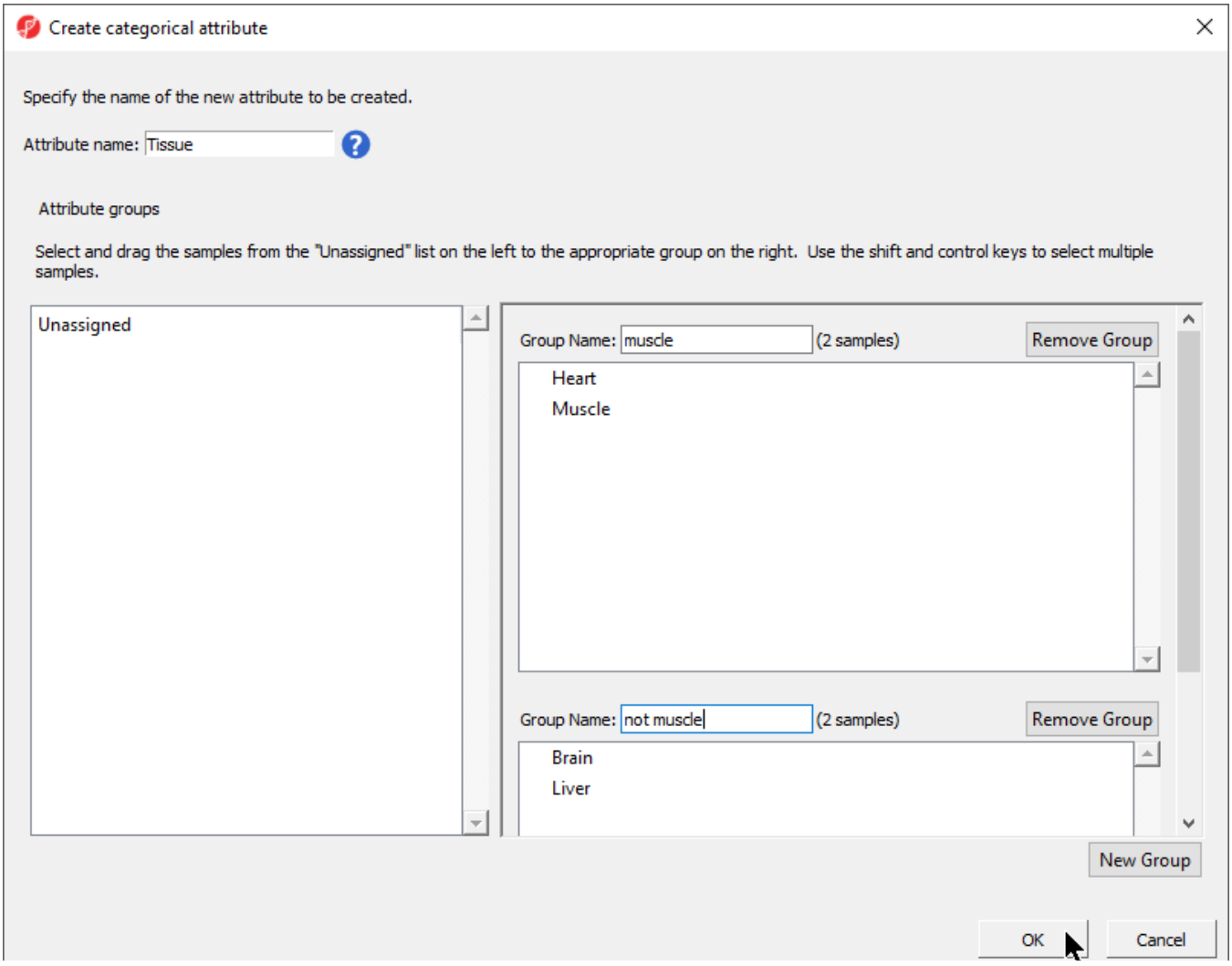

Figure 11. Creating a categorical attribute

- Select **OK**
- Select **No** from the *Add another attribute?* dialog
- Select **Yes** from the *Save spreadsheet 1* dialog

The attribute will now appear as a new column in the *RNA-seq* spreadsheet with the heading *Tissue* and the groups *muscle* and *not muscle*.

Choosing Sample ID column

The next available step in the *Import* panel of the *RNA-seq* workflow is **Choose Sample ID Column.** Verifying the correct column is designated the Sample ID becomes particularly important when data from multiple experiments is being combined.

- Select **Choose Sample ID Column** from the *Import* panel of the *RNA-Seq* workflow
- Select **OK** (Figure 6)

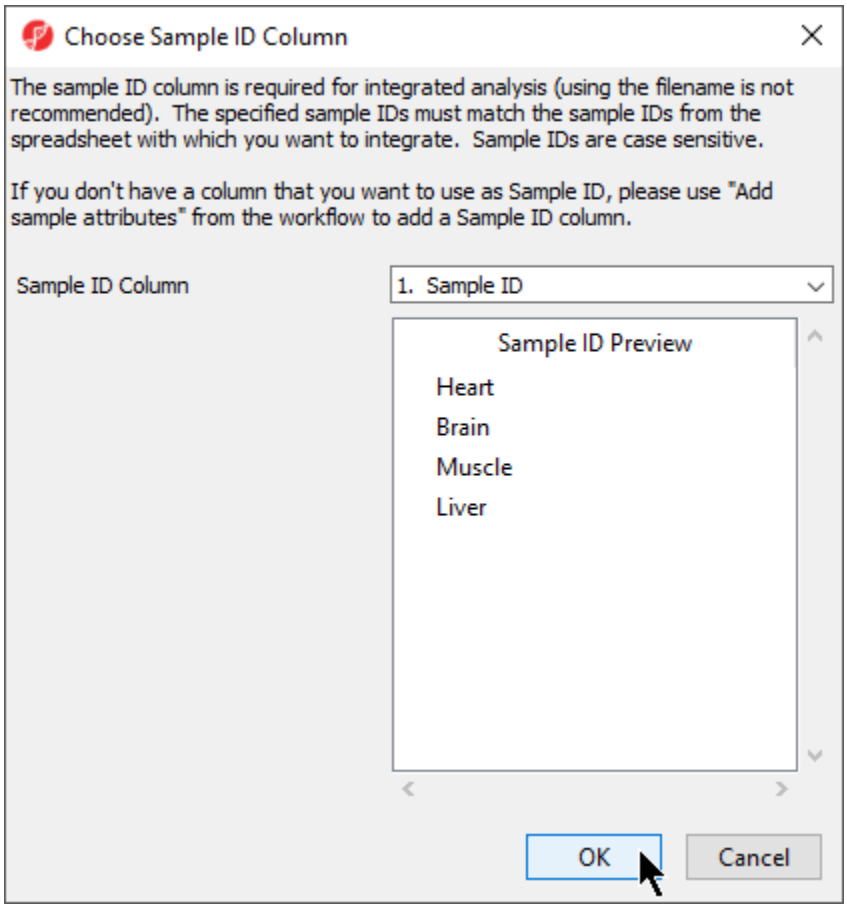

Figure 12. Choosing the correct column as Sample ID

« Importing aligned reads RNA-Seq mRNA quantification »

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