

# The Metadata Tab

The Partek Flow *Metadata* Tab has an option to import data, and is where sample/cell attributes are managed. This is also where users can modify the location of the project output folder.

- [Import data](#)
  - [Automatically create samples from files](#)
  - [Create a new blank sample](#)
  - [Importing count matrix data](#)
- [Project output directory](#)
- [Sample Annotation](#)
  - [Sample attributes](#)
  - [Adding a categorical attribute](#)
  - [Adding a numeric attribute](#)
  - [Adding a system-wide attribute](#)
  - [Assigning categories or values to attributes](#)
  - [Assigning attributes using a Sample Annotation Text File](#)
  - [Guidelines for preparing the sample annotation text file](#)
  - [Use of attributes as Optional columns in task report tables](#)
- [Deleting or Renaming samples within a Project](#)

## Import data

The *Metadata* tab can be used to import data. To add samples to the project, click **Add data** under *Import*, different import options are displayed using the cascading menu (Figure 1).

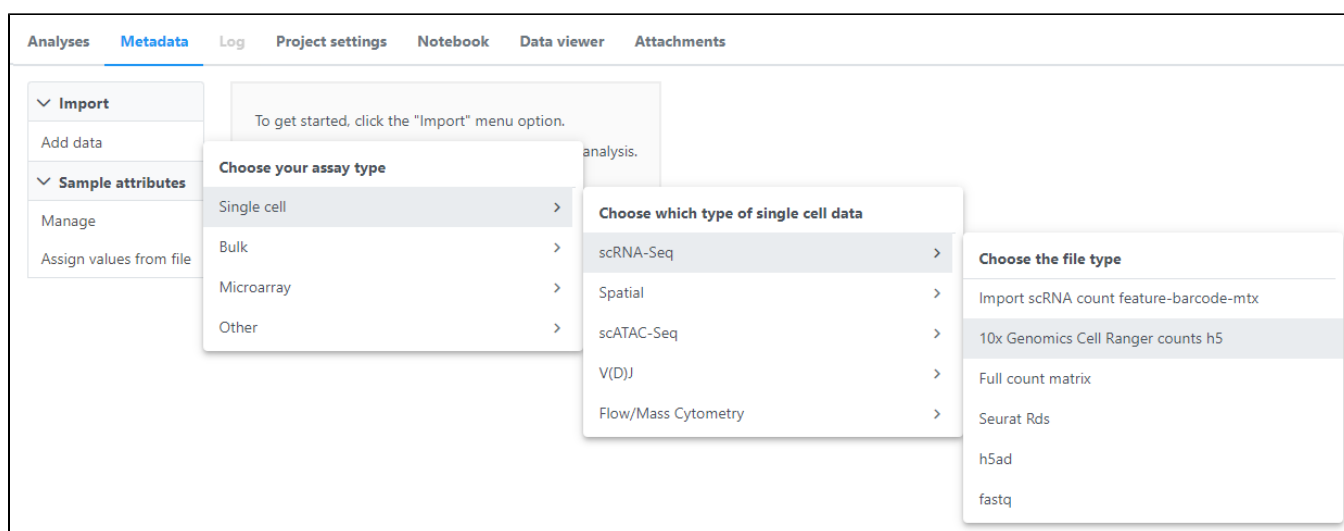


Figure 19. The Partek Flow Metadata tab and selecting options for adding samples

## Automatically create samples from files

This method adds samples by creating them simultaneously as the data gets imported into a project. The sample names are assigned automatically based on filenames.

*Before proceeding, it is ideal that you have already [transfer the data](#) you wish to analyze in a folder (with appropriate permissions) within the Partek Flow server. Please seek assistance from your system administrator in uploading your data directly.*

Select the **Automatically create samples from files** button. The next screen will feature a file browser that will show any folders you have access to in the Partek Flow server (Figure 2). Select a folder by clicking the folder name. Files in the selected folder that have file formats that can be imported by Partek Flow will be displayed and tick-marked on the right panel. You can exclude some files from the folder by unselecting the check mark on the left side of the filename. When you have made your selections, click the **Create sample** button.

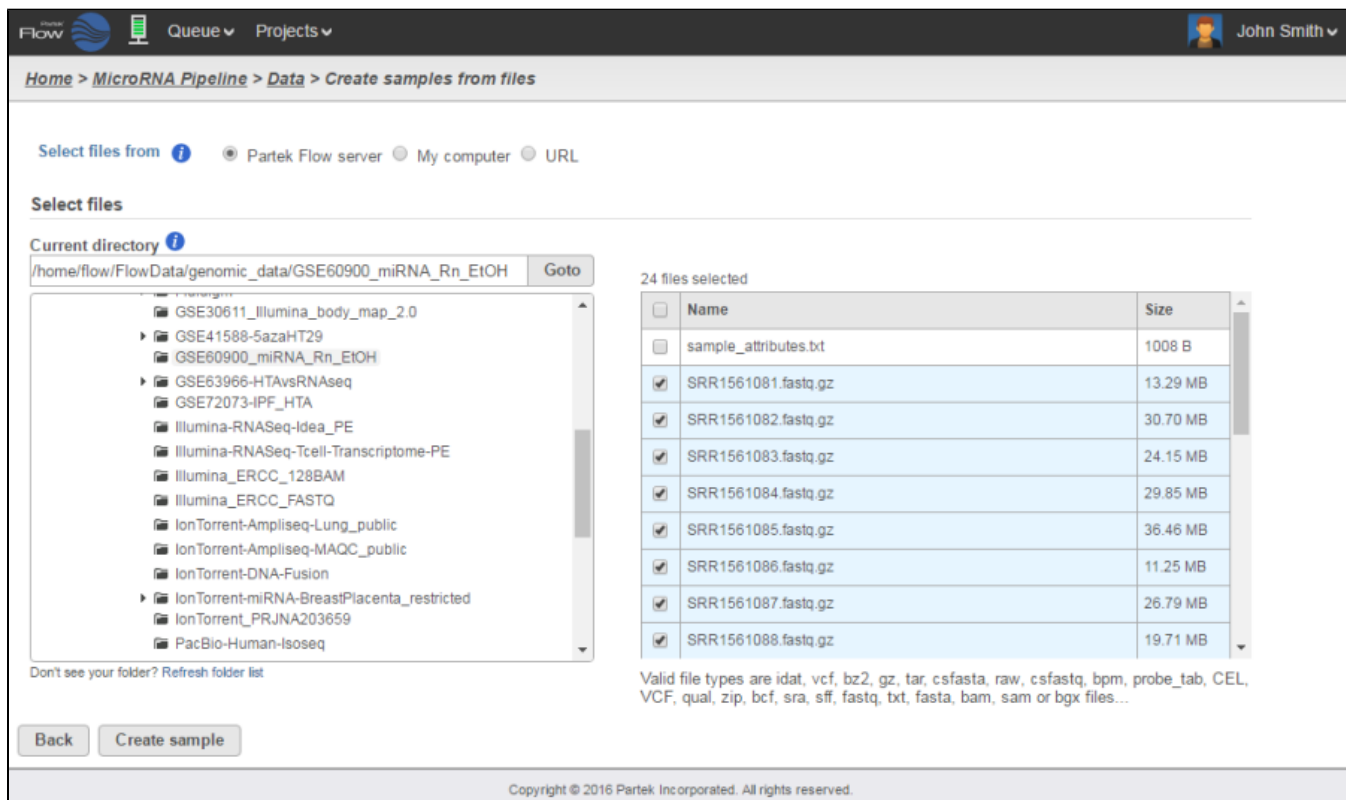


Figure 20. Selecting files in the Partek Flow server to be imported in a project

Alternatively, files can also be uploaded and imported into the project from the user's local computer -only use this option if your file size is less than 500MB. Select the **My computer** radio button (Figure 3) and the options of selecting the local file and the upload (destination) directory will appear. Only one file at a time can be imported to a project using this method.

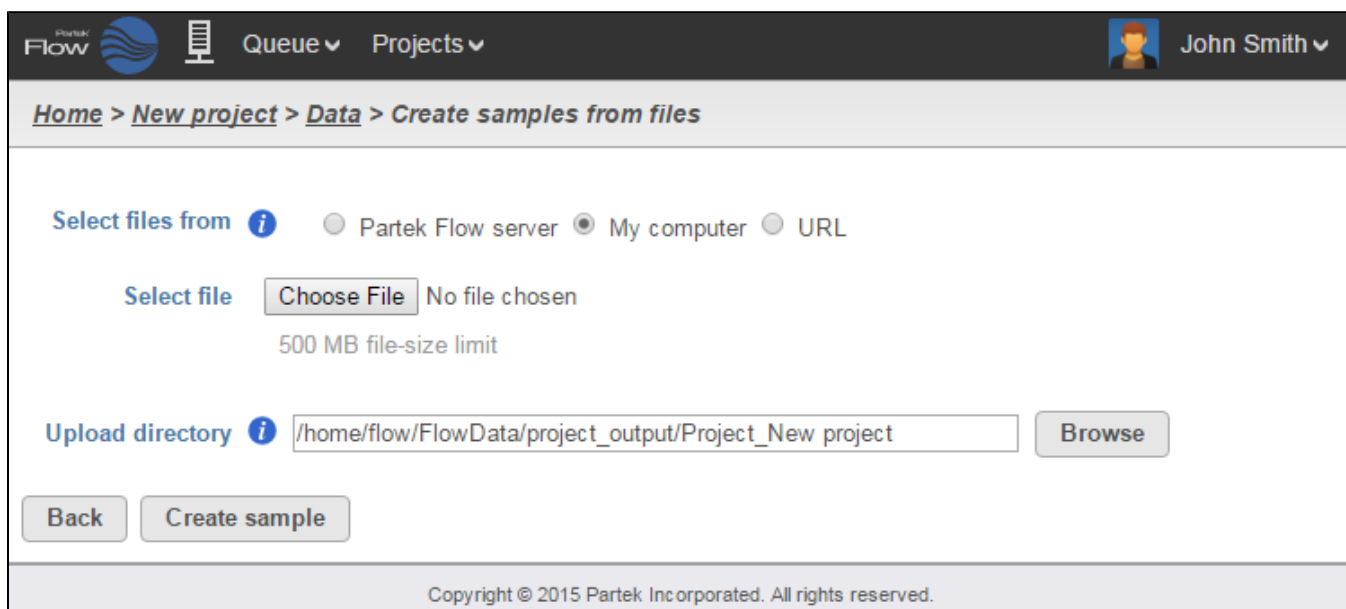


Figure 21. Selecting files from the user's local computer for upload and import

Multiple data files can be compressed a single .zip file before uploading. Partek Flow will automatically unzip the files and put them in the upload directory.

*Please be aware that the use of the method illustrated in Figure 3 highly depends on the speed and latency of the Internet connection between the user's computer and the server where Partek Flow is installed. Given the large size of most genomics data sets, is not recommended in most cases.*

After successful creation of samples from files, the *Data* tab now contains a *Sample management table* (Figure 4). The *Sample name* column in the table is automatically generated based on the filenames and the table is sorted in alphabetical order.

Clicking the on the **Show data files** link on the lower right side of the sample management table will expand the table and reveal the filenames of the files associated with each sample. Conversely, clicking on **Hide data files** will hide the file information.

The columns in the expanded view show the files associated with each sample. Files are organized by file type. Any filename extensions that indicate compression (such as .gz) are not shown.

|                 | Sample name | Attributes |                |           | Files                    |            |            |                                                                                                                                                                             |
|-----------------|-------------|------------|----------------|-----------|--------------------------|------------|------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                 |             | Aza dose   | Platform       | Sample ID | fastq                    | bam        | CEL        | +/-                                                                                                                                                                         |
| 1               | ARR-0uM-1   | 0uM        | HG-U133_Plus_2 | 1         |                          | ARR-0uM-1  | GSM1015541 |       |
| 2               | ARR-0uM-2   | 0uM        | HG-U133_Plus_2 | 2         |                          | ARR-0uM-2  | GSM1015542 |       |
| 3               | ARR-0uM-3   | 0uM        | HG-U133_Plus_2 | 3         |                          | ARR-0uM-3  | GSM1015543 |       |
| 4               | ARR-5uM-1   | 5uM        | HG-U133_Plus_2 | 4         |                          | ARR-5uM-1  | GSM1015544 |       |
| 5               | ARR-5uM-2   | 5uM        | HG-U133_Plus_2 | 5         |                          | ARR-5uM-2  | GSM1015545 |       |
| 6               | ARR-5uM-3   | 5uM        | HG-U133_Plus_2 | 6         |                          |            | GSM1015546 |       |
| 7               | ARR-10uM-1  | 10uM       | HG-U133_Plus_2 | 7         |                          | ARR-10uM-1 | GSM1015547 |       |
| 8               | ARR-10uM-2  | 10uM       | HG-U133_Plus_2 | 8         |                          | ARR-10uM-2 | GSM1015548 |       |
| 9               | ARR-10uM-3  | 10uM       | HG-U133_Plus_2 | 9         |                          | ARR-10uM-3 | GSM1015549 |       |
| 10              | NGS-0uM-1   | 0uM        | HiSeq 2000     | 1         | SRR592573_1, SRR592573_2 | NGS-0uM-1  |            |       |
| 11              | NGS-0uM-2   | 0uM        | HiSeq 2000     | 2         | SRR592574_1, SRR592574_2 | NGS-0uM-2  |            |     |
| 12              | NGS-0uM-3   | 0uM        | HiSeq 2000     | 3         | SRR592575_1, SRR592575_2 | NGS-0uM-3  |            |   |
| 13              | NGS-5uM-1   | 5uM        | HiSeq 2000     | 4         | SRR592576_1, SRR592576_2 | NGS-5uM-1  |            |   |
| 14              | NGS-5uM-2   | 5uM        | HiSeq 2000     | 5         | SRR592577_1, SRR592577_2 |            |            |   |
| 15              | NGS-5uM-3   | 5uM        | HiSeq 2000     | 6         | SRR592578_1, SRR592578_2 | NGS-5uM-3  |            |   |
| 16              | NGS-10uM-1  | 10uM       | HiSeq 2000     | 7         | SRR592579_1, SRR592579_2 | NGS-10uM-1 |            |   |
| 17              | NGS-10uM-2  | 10uM       | HiSeq 2000     | 8         | SRR592580_1, SRR592580_2 |            |            |   |
| 18              | NGS-10uM-3  | 10uM       | HiSeq 2000     | 9         | SRR592581_1, SRR592581_2 | NGS-10uM-3 |            |   |
| Hide data files |             |            |                |           |                          |            |            | Download                                                                                                                                                                    |

Figure 22. The sample management table with data files shown

Once a sample is created in a project, the files associated with it can be modified. In the expanded view, mouse over the +/- column of a sample. The highlighted icons will correspond to the options for the sample on that row.

Click the **green icon** () to associate additional files or the **red icon** () to dissociate a file from a sample. You can manually associate multiple files with one sample. Dissociating a file from a sample does not delete the file from the Partek Flow server.

## Create a new blank sample

Samples can be added one at a time by selecting the **Create a new blank sample** option (Figure 5). In the following dialog box, type a sample name and click **Create**. This process creates a sample entry in the sample management table but there is no associated file with it, hence it is a "blank sample."

Expanding the *Sample management table* by clicking **Show data files** on the lower left corner of the table will reveal the option to associate files to the blank sample.

Mouse over the +/- column and click the **green icon** () to associate a file(s) to the sample. Perform the process for every sample in your project.

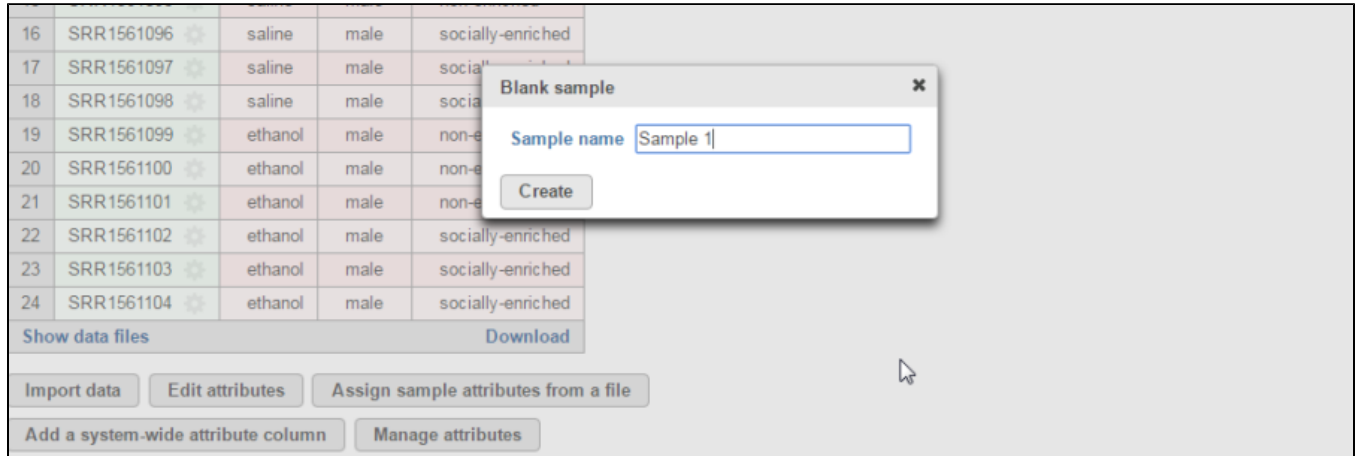


Figure 23. Adding a blank sample

### Importing count matrix data

Alternatively, if you have a matrix of data, such as raw read count data in text format, select **Import count matrix**. The requirements of this text file are listed below:

- The file contains numeric values in a tab-delimited format, samples can be on rows while features (e.g.gene names) are in columns, or vice versa
- The file contains unique sample IDs and feature IDs
- If the data contains sample attribute information, all these attributes have to be either
  - The leftmost columns when samples are on rows (Figure 6)
  - The first few rows when samples are on columns (Figure 7)

| Sample ID | Treatment | Gender | A4GALT | ACO2    | ACR | ADM2    | ADORA2A  | ADORA2A | ADRBK2 | ADSL | AIFM3 | ALG12 | ANKRD54 | ANKRD62 | AP1B1   |
|-----------|-----------|--------|--------|---------|-----|---------|----------|---------|--------|------|-------|-------|---------|---------|---------|
| Sample1   | Treated   | F      | 74     | 104.014 | 0   | 399     | 4.89522  | 4       | 273    | 99   | 36    | 57    | 61      | 0       | 674.999 |
| Sample2   | Treated   | M      | 110    | 145.227 | 0   | 626.996 | 9.20691  | 5       | 377    | 169  | 39    | 76    | 99      | 0       | 939.002 |
| Sample3   | Treated   | F      | 142    | 186.623 | 0   | 856.001 | 9.40078  | 3       | 496    | 184  | 68    | 132   | 110     | 0       | 1154    |
| Sample4   | Treated   | M      | 243    | 249.214 | 1   | 541.999 | 23.3351  | 5       | 153    | 156  | 55    | 193   | 162     | 0       | 1170    |
| Sample5   | Control   | F      | 195    | 206.951 | 1   | 474     | 16.9033  | 10      | 171    | 185  | 36    | 167   | 149     | 0       | 1072.01 |
| Sample6   | Control   | M      | 277    | 213.86  | 1   | 561.999 | 6.44823  | 12      | 239    | 191  | 44    | 196   | 184     | 1       | 1378.01 |
| Sample7   | Control   | F      | 121    | 71.9272 | 0   | 294     | 0.454614 | 4       | 104    | 66   | 32    | 74    | 53      | 0       | 371     |
| Sample8   | Control   | M      | 287    | 232.984 | 0   | 549     | 0.738325 | 4       | 208    | 168  | 41    | 170   | 96.0001 | 0       | 1057    |

Figure 24. Example of sample on row, the first column is sample ID, the 2nd and 3rd columns contain sample attribute information, feature count starts from column 4

| Sample ID      | Sample1  | Sample2 | Sample3  | Sample4  | Sample5  | Sample6  | Sample7  | Sample8  |
|----------------|----------|---------|----------|----------|----------|----------|----------|----------|
| Treatment      | Treated  | Treated | Treated  | Treated  | Control  | Control  | Control  | Control  |
| Gender         | F        | M       | F        | M        | F        | M        | F        | M        |
| A4GALT         | 74       | 110     | 142      | 243      | 195      | 277      | 121      | 287      |
| ACO2           | 104.014  | 145.227 | 186.623  | 249.214  | 206.951  | 213.86   | 71.9272  | 232.984  |
| ACR            | 0        | 0       | 0        | 1        | 1        | 1        | 0        | 0        |
| ADM2           | 399      | 626.996 | 856.001  | 541.999  | 474      | 561.999  | 294      | 549      |
| ADORA2A        | 4.89522  | 9.20691 | 9.40078  | 23.3351  | 16.9033  | 6.44823  | 0.454614 | 0.738325 |
| ADORA2A-AS1    | 4        | 5       | 3        | 5        | 10       | 12       | 4        | 4        |
| ADRBK2         | 273      | 377     | 496      | 153      | 171      | 239      | 104      | 208      |
| ADSL           | 99       | 169     | 184      | 156      | 185      | 191      | 66       | 168      |
| AIFM3          | 36       | 39      | 68       | 55       | 36       | 44       | 32       | 41       |
| ALG12          | 57       | 76      | 132      | 193      | 167      | 196      | 74       | 170      |
| ANKRD54        | 61       | 99      | 110      | 162      | 149      | 184      | 53       | 96.0001  |
| ANKRD62P1-PARF | 0        | 0       | 0        | 0        | 0        | 1        | 0        | 0        |
| AP1B1          | 674.999  | 939.002 | 1154     | 1170     | 1072.01  | 1378.01  | 371      | 1057     |
| AP1B1P1        | 0        | 0       | 0        | 0        | 0        | 0        | 0        | 0        |
| APOBEC3A       | 0        | 0       | 0        | 4.92E-10 | 0        | 0        | 0        | 0        |
| APOBEC3A_B     | 2        | 0       | 1.90E-24 | 1.19678  | 1.93E-27 | 1.31E-34 | 0        | 1.88E-23 |
| APOBEC3B       | 2.12E-06 | 4       | 4        | 31.6882  | 24.9496  | 34.7395  | 4        | 8        |
| APOBEC3B-AS1   | 0        | 0       | 0        | 1.06538  | 1        | 1.06281  | 0        | 0        |
| APOBEC3C       | 0        | 0       | 0        | 4        | 3        | 4        | 0        | 3        |
| APOBEC3D       | 0        | 0       | 0        | 0        | 1.05044  | 1.05495  | 0        | 0        |
| APOBEC3F       | 0        | 0       | 0        | 7.04965  | 4        | 10.1427  | 0        | 2        |
| APOBEC3G       | 1        | 0       | 0        | 1        | 2        | 4        | 0        | 2        |
| APOBEC3H       | 0        | 1       | 0        | 0        | 0        | 0        | 0        | 0        |
| APOL1          | 161      | 234     | 262      | 761.998  | 836.005  | 791      | 262      | 636.001  |

Figure 25. Example of sample on column, the first row is sample ID, the 2nd and 3rd rows contain sample attribute information, feature count starts from column 4

Like all other input files, you can upload the file from the *Partek Flow server*, *My Computer* or via a *URL*. Uploading the file brings up a file preview window (Figure 8). The preview of the first few rows and columns of the text file should help you determine on which rows/columns the relevant counts are located (the preview will display up to 100 rows and 100 columns). Inspect the text preview and indicate the orientation of the text file under **File format>Input format**.

If the read counts are based on a compatible annotation file in Partek Flow, you can specify that annotation file under **Gene/feature annotation**. Select the appropriate genome build and annotation model for your count data. Select the **Contain sample attributes** checkbox if your data includes additional sample information.

## Start row

Data (including labels) starts on row

sample on row.txt (21.63 KB)

|   | 1         | 2         | 3      | 4      | 5       | 6   | 7       | 8        | 9           | 10     | 11   | 12    | 13    | 14      | 15                | 16      | 17      | 18       | 19         | 20       |
|---|-----------|-----------|--------|--------|---------|-----|---------|----------|-------------|--------|------|-------|-------|---------|-------------------|---------|---------|----------|------------|----------|
| 1 | Sample ID | Treatment | Gender | A4GALT | AC02    | ACR | ADM2    | ADORA2A  | ADORA2A-AS1 | ADRBK2 | ADSL | AIFM3 | ALG12 | ANKRD54 | ANKRD62P1-PARP4P3 | AP1B1   | AP1B1P1 | APOBEC3A | APOBEC3A_B | APOBEC3B |
| 2 | Sample1   | Treated   | F      | 74     | 104.014 | 0   | 399     | 4.89522  | 4           | 273    | 99   | 36    | 57    | 61      | 0                 | 674.999 | 0       | 0        | 2          | 2.12E-06 |
| 3 | Sample2   | Treated   | M      | 110    | 145.227 | 0   | 626.996 | 9.20691  | 5           | 377    | 169  | 39    | 76    | 99      | 0                 | 939.002 | 0       | 0        | 0          | 4        |
| 4 | Sample3   | Treated   | F      | 142    | 186.623 | 0   | 856.001 | 9.40078  | 3           | 496    | 184  | 68    | 132   | 110     | 0                 | 1154    | 0       | 0        | 1.90E-24   | 4        |
| 5 | Sample4   | Treated   | M      | 243    | 249.214 | 1   | 541.999 | 23.3351  | 5           | 153    | 156  | 55    | 193   | 162     | 0                 | 1170    | 0       | 4.92E-10 | 1.19678    | 31.6882  |
| 6 | Sample5   | Control   | F      | 195    | 206.951 | 1   | 474     | 16.9033  | 10          | 171    | 185  | 36    | 167   | 149     | 0                 | 1072.01 | 0       | 0        | 1.93E-27   | 24.9496  |
| 7 | Sample6   | Control   | M      | 277    | 213.86  | 1   | 561.999 | 6.44823  | 12          | 239    | 191  | 44    | 196   | 184     | 1                 | 1378.01 | 0       | 0        | 1.31E-34   | 34.7395  |
| 8 | Sample7   | Control   | F      | 121    | 71.9272 | 0   | 294     | 0.454614 | 4           | 104    | 66   | 32    | 74    | 53      | 0                 | 371     | 0       | 0        | 0          | 4        |
| 9 | Sample8   | Control   | M      | 287    | 232.984 | 0   | 549     | 0.738325 | 4           | 208    | 168  | 41    | 170   | 96.0001 | 0                 | 1057    | 0       | 0        | 1.88E-23   | 8        |

1-9 of 9 rows

## File format

Input format

Use annotation file ☒

Assembly

Gene/feature annotation

Contain sample attributes ☒

## Column information

Sample ID on column

Sample attributes start on column

Feature IDs start on column

## Row information

Feature ID on row

Sample ID start on row

## Counts format

Counts in log base

[Back](#) [Next](#)

Figure 26. Importing count matrix data preview. This example is showing a count matrix text file has samples on rows with sample attributes columns

The example above is showing an example text file with samples listed on rows. The gene ID is compatible with the *hg19 RefSeq Transcripts - 2016-08-01* annotation model. Under the **Column information** and **Row information** sections, indicate the location of the Sample ID, which in this case is on *Column 1*. Indicate the sample attribute location by marking where it starts, which in the example is at *Column 2*. Mark the Feature ID, which in this case are gene IDs and starts at *Column 4*.

If the data has been log transformed, specify the base under **Counts format**.

## Project output directory

The project output directory is the folder within the Partek Flow server where all output files produced during analysis will be stored.

*The default directory is configured by the Partek Flow Administrator under the Settings menu (under **System Preferences** > **Default project output directory**).*

If the user does not override the default, the task output will go to a subdirectory with the name of the Project.

The user has the option of specifying an existing folder or creating a new one as the project output directory. To do so, click the  icon next to the directory and specify or create a new folder in the dialog box.

## Sample Annotation

After samples have been added in the project, additional information about the samples can be added. Information such as disease type, age, treatment, or sex can be annotated to the data by assigning the *Attributes* for each sample. Certain tasks in Partek Flow, such as Gene-Specific Analysis, require that samples be assigned attributes in order to do statistical comparisons between groups or samples. As attributes are added to the project, additional columns in the sample management table will be created.

### Sample attributes

Attributes can be managed or created within a project. Under the *Data* tab, click the button to open the *Manage attributes* page (Figure 9).

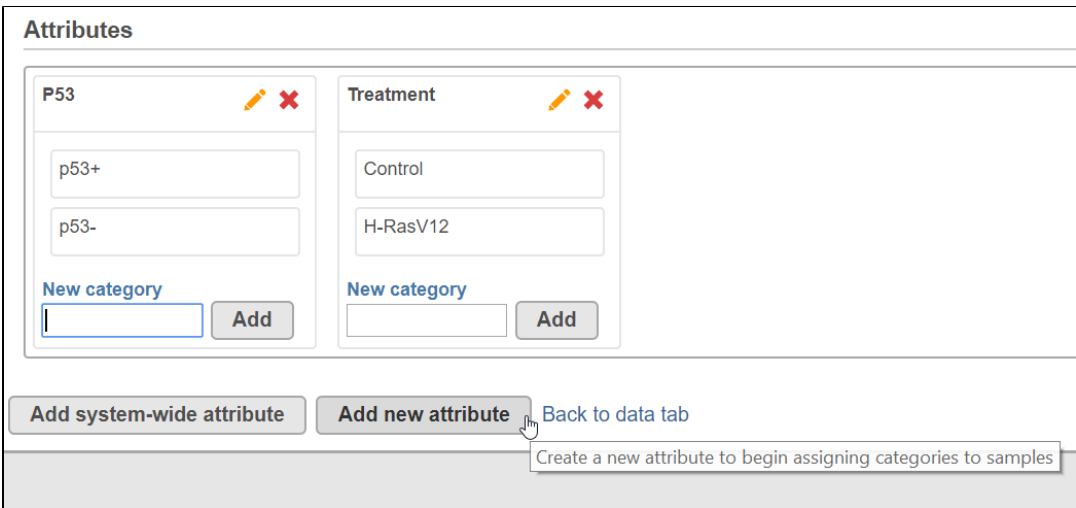


Figure 27. Managing attributes

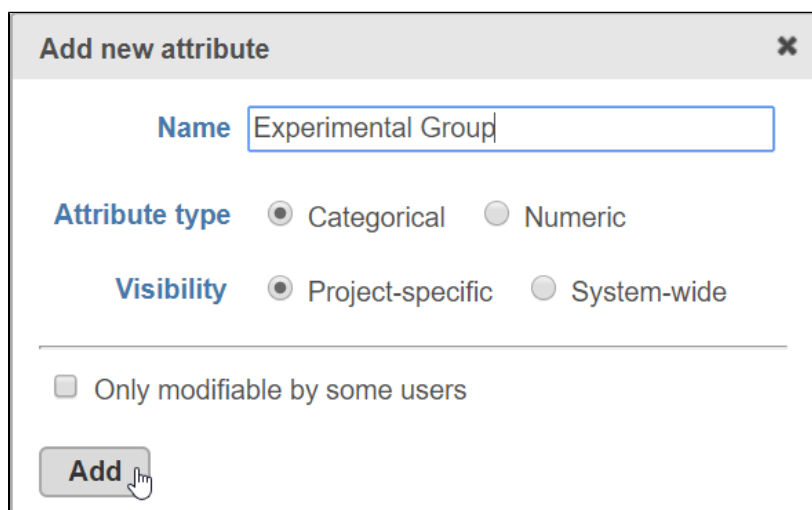
To prepare for later data analysis using statistical tools, attributes can either be categorical or numeric (i.e., continuous).

### Adding a categorical attribute

For categorical attributes, there are two levels of visibility. *Project-specific categorical attributes* are visible only within the current project. *System-wide categorical attributes* are visible across all the projects within the Partek Flow server, and are useful for maintaining uniformity of terms. Importing samples in a new project will retain the system-wide attributes, but not the project-specific attributes.

A feature of Partek Flow is the use of controlled vocabulary for *categorical attributes*, allowing samples to be assigned only within pre-defined categories. It was designed to effectively manage content and data and allow teams to share common definitions. The use of standard terms minimizes confusion.

To add a categorical attribute in the *Manage attributes* page, click the **Add new attribute** (Figure 10). In the dialog box, type a **Name** for the attribute, select the **Categorical** radio button next to **Attribute type**, select the visibility of the attribute and then click the **Add** button.



**Add new attribute** [X]

**Name**

**Attribute type** ☒ Categorical ☐ Numeric

**Visibility** ☒ Project-specific ☐ System-wide

---

☐ Only modifiable by some users

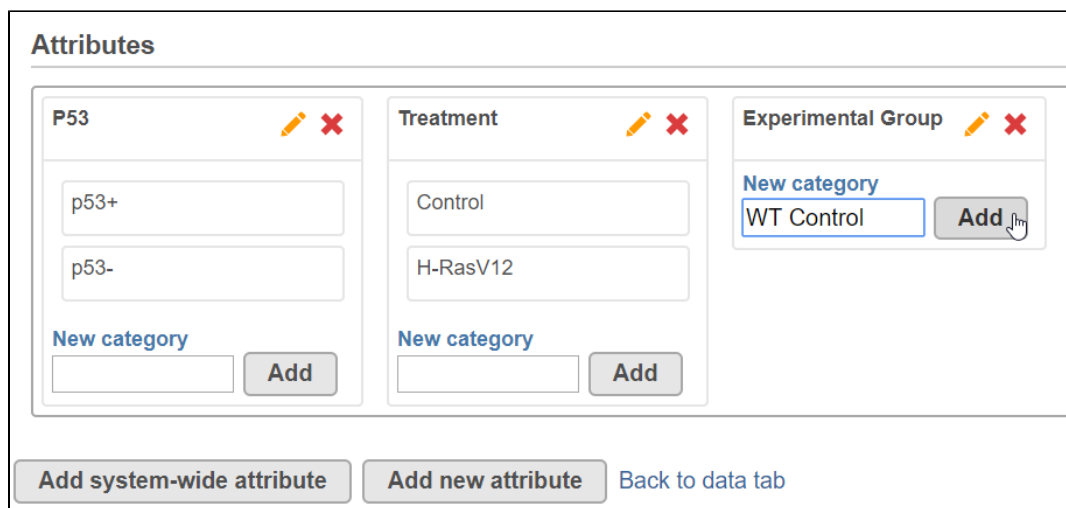
**Add** [Hand cursor]

Figure 28. Adding a categorical attribute and defining the categories

Individual categories for the attribute must then be entered. Enter a name of the **New category** in the *New category* text box and click **Add** (Figure 11). The *Name* of the new category will show up in the table. The category can also be edited by clicking  or deleted by clicking  (visible on mouse-over). Repeat to add additional categories within the attribute.

Repeat the process for additional attributes of the samples in your study. When done, click **Back to sample management table**. Categorical attributes will default to *Project-specific* visibility.

Click an attribute name to drag and drop can change the order of the attributes displayed on the data tab. Click on group name to drag and drop vertically can change the order of the group name, which can be reflected on visualization.



**Attributes**

| P53   | Treatment   | Experimental Group   |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <input type="text" value="p53+"/><br><input type="text" value="p53-"/><br><b>New category</b><br><input type="text"/><br><b>Add</b>                                         | <input type="text" value="Control"/><br><input type="text" value="H-RasV12"/><br><b>New category</b><br><input type="text"/><br><b>Add</b>                                        | <b>New category</b><br><input type="text" value="WT Control"/><br><b>Add</b> [Hand cursor]                                                                                                     |

**Add system-wide attribute**
**Add new attribute**
[Back to data tab](#)

Figure 29. Adding categories within an attribute

## Adding a numeric attribute

To add a numeric attribute in the *Manage attributes* page, click the **Add new attribute**. In the dialog box (Figure 13), type a **Name** for the attribute, select the **Numeric** radio button next to **Attribute type**, and then click the **Add** button. Some optional parameters for numeric attributes include the *Minimum value*, *Maximum value*, and *Units*. When done, click **Add** to return to the *Manage attributes* page. Repeat the process add more numeric attributes. When done, click **Back to sample management table**.

**Add new attribute** [X]

Name:

Attribute type: ☐ Categorical ☒ Numeric

Visibility: ☒ Project-specific ☐ System-wide

---

☒ Minimum value:

☒ Maximum value:

☒ Specify a unit type:

Unit:

**Add**

Figure 30. Adding a numeric attribute and specifying the units

## Adding a system-wide attribute

Since system-wide attributes do not have to be created by the current user, they only need to be added to the sample management table in a project.

In the *Data* tab, click **Add a system-wide attribute** button. In the dialog box that follows (Figure 14), a drop down menu is located next to **Add attribute** where you can select the *System-wide attribute* you would like to add to the project. Once selected, it will be recognized automatically as either *Categorical*, *system-wide* or the *Numeric attribute*.

For an *System-wide categorical attribute*, the different categories are listed and you have the option of pre-filling the columns with N/A (or any other category within the attribute). Click **Add column** and you will return to the *Data* Tab.

|      |              |
|------|--------------|
| male | non-enriched |
| male | n            |
| male | n            |
| male | s            |
| male | s            |
| male | s            |
| male | n            |
| male | n            |
| male | n            |
| male | s            |
| male | s            |
| male | s            |

**Add a system-wide attribute column** [X]

Add attribute:

Categories within Condition include: Control, Disease

Automatically fill with:

**Add column**

Don't see what you need here? [Manage attributes](#)

Figure 31. Adding a system-wide categorical attribute column

## Assigning categories or values to attributes

After adding all the desired attributes to a project, the sample management table will show a new column for each attribute (Figure 15). The columns will initially as "N/A", as the samples have not yet been categorized or assigned a value. To edit the table, click **Edit attributes**. Assign the sample attributes by using a drop down for categorical attributes (controlled vocabulary) or typing with a keyboard for numeric attributes.

|                                 | Sample name | Attributes               |           |                    |
|---------------------------------|-------------|--------------------------|-----------|--------------------|
|                                 |             | P53                      | Treatment | Experimental Group |
| 1                               | SRR3541289  | p53-                     | Control   | N/A                |
| 2                               | SRR3541290  | p53-                     | Control   | N/A                |
| 3                               | SRR3541291  | p53-                     | Control   | N/A                |
| 4                               | SRR3541292  | p53-                     | H-RasV12  | N/A                |
| 5                               | SRR3541293  | p53-                     | H-RasV12  | N/A                |
| 6                               | SRR3541294  | p53-                     | H-RasV12  | N/A                |
| 7                               | SRR3541295  | p53+                     | Control   | N/A                |
| 8                               | SRR3541296  | p53+                     | Control   | N/A                |
| 9                               | SRR3541297  | p53+                     | Control   | N/A                |
| 10                              | SRR3541298  | p53+                     | H-RasV12  | N/A                |
| 11                              | SRR3541299  | p53+                     | H-RasV12  | N/A                |
| 12                              | SRR3541300  | p53+                     | H-RasV12  | N/A                |
| <a href="#">Show data files</a> |             | <a href="#">Download</a> |           |                    |

[Edit attributes](#)

[Assign sample attributes from a file](#)

[Add a system-wide attribute column](#)

[Apply attributes to the samples shown here](#)

[Manage attributes](#)

Figure 32. A sample management table prior to assigning attribute columns for a new attribute

When all the attributes have been entered, click **Apply changes** and the sample management table will be updated. After editing the sample table, make sure there are no fields with blank or N/A values before proceeding. To rename or delete attributes, click **Manage attributes** from the *Data* tab to access the *Manage attributes* page.

## Assigning attributes using a Sample Annotation Text File

Another way to assign attributes to samples in the *Data* tab is to use a text file that contains the table of attributes and categories/values. This table is prepared outside of Partek Flow using any text editing software capable of saving tab-separated text files.

Using a text editor, prepare a table containing the attributes. An example is shown in Figure 16. There should only be one tab between columns with no extra tabs after the last column. In this particular example, the first column contains the filename and the text file is saved as *Sampleinfo.txt*.

```

|FileName → Type → Time in storage → Sample source
HBR.WT.DID-143.fastq.gz → HBR → 1 → Human Brain Reference RNA
HBR.WT.GOG-139.fastq.gz → HBR → 2 → Human Brain Reference RNA
HBR.WT.LUC-140.fastq.gz → HBR → 3 → Human Brain Reference RNA
HBR.WT.LUC-141.fastq.gz → HBR → 4 → Human Brain Reference RNA
HBR.WT.POZ-124.fastq.gz → HBR → 5 → Human Brain Reference RNA
UHR.WT.DID-144.fastq.gz → UHR → 1 → Universal Human Reference RNA
UHR.WT.GOG-140.fastq.gz → UHR → 2 → Universal Human Reference RNA
UHR.WT.POZ-125.fastq.gz → UHR → 3 → Universal Human Reference RNA
UHR.WT.POZ-126.fastq.gz → UHR → 4 → Universal Human Reference RNA
UHR.WT.POZ-127.fastq.gz → UHR → 5 → Universal Human Reference RNA

```

Figure 33. A sample annotation text file. This view shows tab stops

The first row of the table in the text file contains the attributes (as headers). The first *column* of the table in the text file, regardless of the header of the first column, should contain either the sample names or the file names of the samples already added in Partek Flow. The first column is the unique identifier that will match the samples to the correct values or categories.

To upload sample attributes, click **Assign sample attributes from a file** in the *Data tab*. Then indicate where the attribute text file is stored and navigate to it. Partek Flow will parse the text file and present attributes that will be available for import (Figure 17).

Select the attributes you want to import by clicking the *Import* check box. Imported attributes that do not currently exist in the project will create new *project-specific attributes*.

Home > *Illumina body map (paired end)* > Assign sample attributes from a file

Imported attributes that do not currently exist will create new Project-specific attributes.

| Attribute name                              | Terms                           | Import                              | Attribute type |
|---------------------------------------------|---------------------------------|-------------------------------------|----------------|
| <input type="text" value="Tissue"/>         | 16 Tissues mixture, adipose...  | <input checked="" type="checkbox"/> | Categorical    |
| <input type="text" value="Age"/>            | 19 years, 29 years, 37 years... | <input checked="" type="checkbox"/> | Categorical    |
| <input type="text" value="Sex"/>            | <not provided>, female, male    | <input checked="" type="checkbox"/> | Categorical    |
| <input type="text" value="Library_layout"/> | PAIRED, SINGLE                  | <input checked="" type="checkbox"/> | Categorical    |

Show/hide file preview

Back

Import

Copyright © 2016 Partek Incorporated. All rights reserved.

Figure 34. Assigning attributes of samples using a sample annotation text file

You can change the name of a specific attribute by editing the *Attribute name* text box. Columns containing letter characters are automatically selected as *categorical attributes*. Columns containing numbers are suggested to be *numeric attributes* and can be changed to categorical using the drop down menu under *Attribute type*.

## Guidelines for preparing the sample annotation text file

- The first column is always the unique identifier and can refer only to *File names* or *Sample names*.
- If using *Sample names* in the first column, they must match the entries of the *Sample name* column in the *Sample management table*.
- If using *File names* in the first column, use the filenames shown in the *fastq* column of the expanded sample management table (see Figure 4) then add the extension *.gz*. All filenames must include the complete file extension (e.g., *Samplename.fastq.gz*).
- The header name of the first column of the table (top left cell of our text table) is irrelevant but should not be left blank. Whether the first column contains *File names* or *Sample names* will be chosen during the process.
- The last column cannot have empty values
- Missing data (blank cells) can only be handled if the attribute is numeric. If it is categorical, please put a character in it.

It is advisable to use *Sample name* as the first column identifier when:

- Samples are associated with more than one file (for instance, paired-end reads and/or technical replicates).
- The files were imported in the SRA format (from the Sequence Read Archive database). In Partek Flow, they are automatically converted to the FASTQ format. Consequently, their filenames would change once they are imported. The new file names can be seen by expanding the sample management table, the new extension would be *.fastq.gz*.

If attributes are assigned from two different text files, the following will happen:

- If the previous attributes have the same header and type (both are either categorical or numeric), the values are overwritten.
- If there are different/additional headers on the "second round" of assignment, these new attributes will be appended to the table.
- For *numeric attributes*, a "blank" value will not override a previous value.

## Use of attributes as Optional columns in task report tables

The attributes assigned to the samples within the Data Tab will be associated with the samples throughout the project. During the course of analysis, Partek Flow tasks generate various tables and any attributes associated with a sample can be included in the table as optional columns. An example is shown in Figure 18 for a pre-alignment QA/QC report where the Optional columns link on the top left of the table reveal the different sample attributes.

|            |        | Treatment ▾       | Total reads ▾ | Avg. read length ▾ | Avg. read quality ▾ | % N ▾  | % GC ▾ |
|------------|--------|-------------------|---------------|--------------------|---------------------|--------|--------|
| e          |        | non-enriched      | 1,179,519     | 23.06              | 28.86               | 19.64% | 48.33% |
| e          |        | non-enriched      | 2,204,453     | 23.66              | 31.19               | 12.39% | 47.89% |
| e          |        | non-enriched      | 1,883,429     | 23.16              | 29.19               | 17.90% | 48.09% |
| e          |        | socially-enriched | 2,271,689     | 23.21              | 29.91               | 15.99% | 47.97% |
| SRR1561085 | saline | socially-enriched | 2,612,831     | 23.60              | 30.95               | 12.96% | 48.23% |

Figure 35. Optional columns include sample attributes

## Deleting or Renaming samples within a Project

In the *Data* tab, each sample can be renamed or deleted from the project by clicking the gear icon next to the sample name. The gear icon is readily visible upon mouse over (Figure 19). Sample can only be deleted if no analysis has been performed on the data yet. If any analysis has been performed on the data node, then delete sample operation is invisible. You can perform filter samples in downstream analysis if you want to exclude certain samples in further analysis. Deleting a sample from a project does not delete the associated files, which will remain on the disk.

|   | Sample name | Attributes |
|---|-------------|------------|
|   |             | Aza dose   |
| 1 | SRR592573   | 0uM        |
| 2 | SRR592574   |            |
| 3 | SRR592575   |            |
| 4 | SRR592576   | 5uM        |
| 5 | SRR592577   | 5uM        |
| 6 | SRR592578   | 5uM        |
| 7 | SRR592579   | 10uM       |
| 8 | SRR592580   | 10uM       |
| 9 | SRR592581   | 10uM       |

[Show data files](#)
[Download](#)

[Add samples](#)
[Assign sample attributes from a file](#)

[Add a system-wide attribute column](#)
[Manage attributes](#)

Project output directory ⓘ /home/flow/FlowData/project\_output/Project\_Project 1 (18.06 TB free) ✎

Figure 36. Renaming or deleting a sample

You can download your completed *Sample management table* by clicking the **Download** link at the lower corner of the table. This will export a tab-delimited text file with contents of the table.

# Additional Assistance

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

[« Creating a New Project](#) [The Analyses Tab](#) »



Your Rating: Results: 40 rates