

Sample Table from a Text File

- [Select the Text File](#)
- [Create the Sample Table](#)
- [Include Data Import](#)

For new projects, a sample table can be automatically created from a tab-delimited text file. There are several advantages of creating a sample table in this manner:

- You can define multiple samples and attributes even before data has been imported, allowing you to:
 - customize the name of your samples, thus avoiding the automatic sample names generated based on file names
 - import sample sheets as defined by the instrument that generated your data
- You can simultaneously create the sample table *and* import data, allowing you to:
 - combine several files into one sample
 - import data located in multiple subdirectories

This process of generating a sample table based on a text file can only be done once per project. Additional samples or attributes can still be added using the **Import data** or **Manage attributes** buttons under the *Data Tab*.

Select the Text File

The text file must be created outside of Partek® Flow® (you can use software such as Partek® Genomics Suite®, Microsoft® Excel® or any text editor). A valid text file is a tab-delimited text file that contains one sample per row and columns containing sample information. At least one column must have unique entries and will be suggested as Sample IDs. Additional columns may contain numeric or categorical attributes and (optional) filenames. Examples of text files are shown in [Figures 3](#) and [6](#).

To select the text file, create a new project and in the blank *Data Tab* (no samples have been imported yet), click the **Assign values from a file** button (Figure 1).

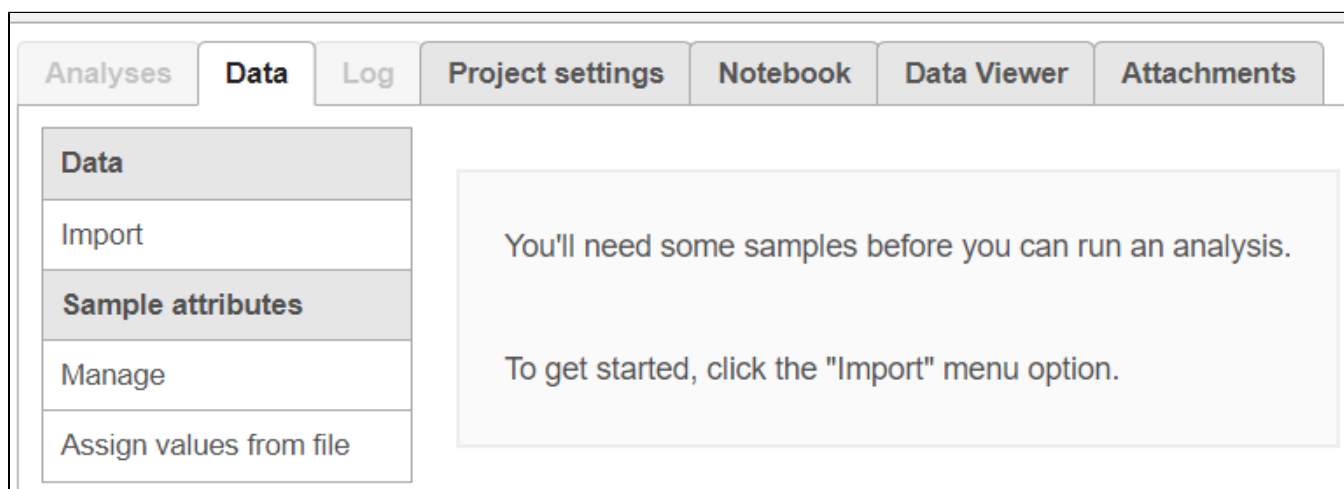


Figure 14. A new project with no samples in the Data Tab

Navigate to the file using the browser as shown in Figure 2. The text file may be located in either the **Partek Flow server**, **My computer**, or from a **URL**.

Check the box next to the text file that you want to use and click **Next**.

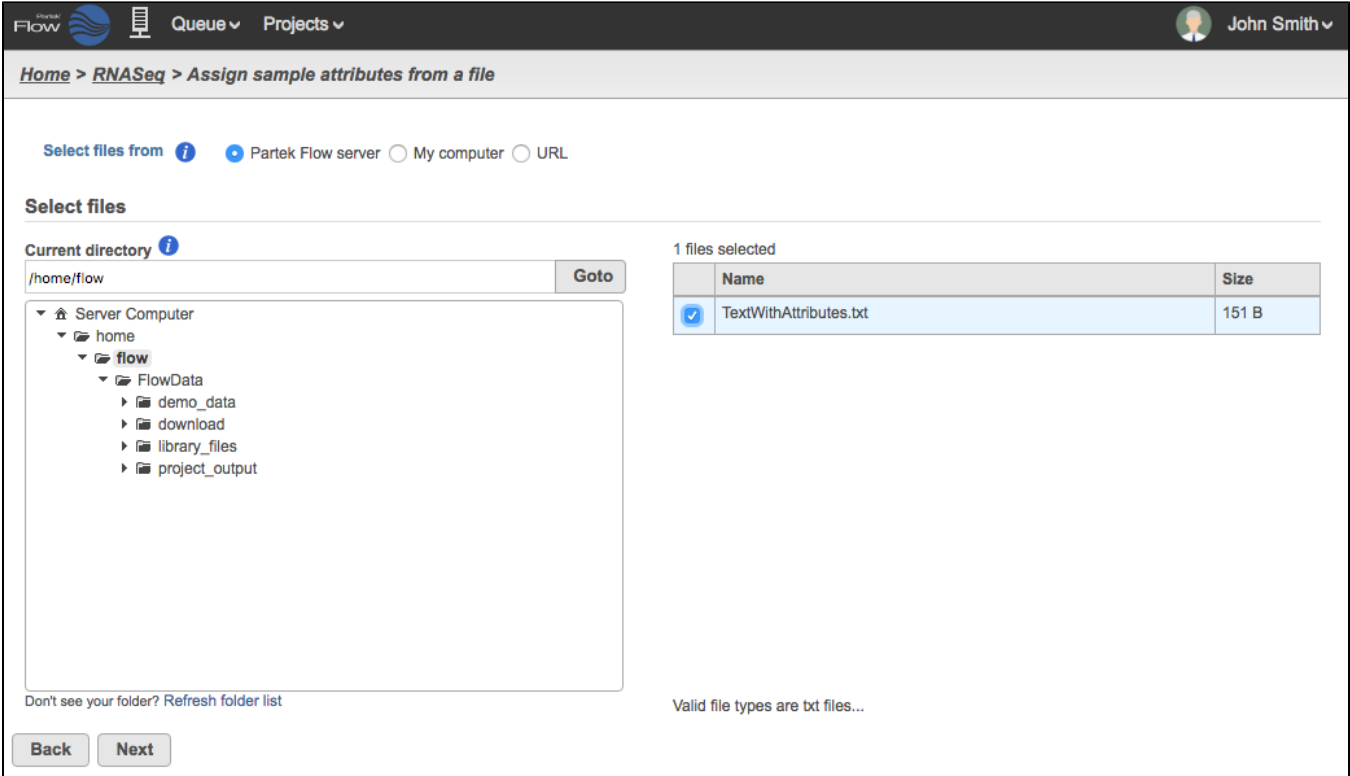


Figure 15. Locating a text file that contains sample information

Create the Sample Table

Text files that contain only sample IDs and attributes such as the one shown in Figure 3 can be imported to create a sample table with no associated files. (If you wish to couple the table creation with data import, you must include the filenames as described in the section below on how to [Include Data Import](#)).

Sample name	Freezer Location	Treatment	Genotype	Age
NA1031	A8	Case	wt	15
NA1032	A9	Case	mutant	18
NA1033	A10	Control	wt	17
NA1034	A11	Control	mutant	17

Figure 16. Text with attributes

The text file will be summarized as in Figure 4. The first two columns show the headers and example terms parsed from the text file in Figure 3. The suggested attribute names can be renamed before import. Columns that contain unique entries are recognized as possible *Sample IDs* and can be selected using the radio button. You can choose which attributes to include and, if applicable, whether they are numeric or categorical. The **Show/hide file preview** link allows you to preview the text of the tab delimited text file you are using.

In the example in Figure 4, the columns for "Sample name" and "Freezer Location" are both unique and the former is selected as the *Sample ID*. The "Freezer location" has been deselected and it will not be included in the resulting Sample table. Since "Age" has all number terms, the *Attribute type* column for it is a drop down menu to choose between *Numeric* and *Categorical*. There are no filenames in the text file so the *Files* column is empty.

Partek Flow Queue Projects John Smith

Home > RNASeq > Assign sample attributes from a file

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

Attribute name	Terms	Import	Files	Sample ID	Attribute type
Sample name	NA1031, NA1032, NA1033, NA...	<input type="checkbox"/>		<input checked="" type="radio"/>	Categorical
Freezer Location	A10, A11, A8, A9	<input type="checkbox"/>		<input type="radio"/>	Categorical
Treatment	Case, Control	<input checked="" type="checkbox"/>			Categorical
Genotype	mutant, wt	<input checked="" type="checkbox"/>			Categorical
Age	15, 17, 18	<input checked="" type="checkbox"/>			Numeric

Show/hide file preview

TextWithAttributes.txt (151 B)

	Sample name	Freezer Location	Treatment	Genotype	Age
1	NA1031	A8	Case	wt	15
2	NA1032	A9	Case	mutant	18
3	NA1033	A10	Control	wt	17
4	NA1034	A11	Control	mutant	17

<< 1-5 of 5 rows >>

Back Import

Figure 17. Selecting the attributes and Sample ID

Click **Import** and the table will be created (Figure 5). Click **Show data files** to expand the table and associate files (see the [Adding samples](#) section).

Partek Flow Queue Projects John Smith

Home > RNASeq

Analyses Data Log Project settings

	Sample name	Attributes		
		Genotype	Treatment	Age
1	NA1031	wt	Case	15
2	NA1032	mutant	Case	18
3	NA1033	wt	Control	17
4	NA1034	mutant	Control	17

Show data files Download

You'll need to associate some files with the samples before you can run an analysis.

Click "Add samples" or "Show data files" and "Associate files with this sample".

Import data Edit attributes Assign sample attributes from a file

Add a system-wide attribute column Manage attributes

Project output directory /home/flow/FlowData/project_output/Project_RNASeq (11.04 TB free)

Figure 18. Sample table created from a text file with attributes

Include Data Import

If you have a text file that contains sample IDs and attributes as well as the filenames of your data, you can create the sample table and start the data import at the same time. This is particularly useful for projects where multiple files are associated with the same sample (e.g., a sample ran in multiple lanes in the sequencer).

For this type of import, the text file may be located in either the **Partek Flow server**, **My computer**, or from a **URL**. However, the data files should already be uploaded to the **Partek Flow server** (see section on [file transfers](#)).

In the text file, each filename to be associated with the sample must be separated by a tab. That means, if you are using a spreadsheet software to generate the text file, there is a maximum of one file per column. You also need to add headers such as file1, file2, etc., to define the columns. There is no limit on the number of columns on the text file. It is also possible to have a different number of files associated with different samples.

The filenames must show the proper extensions of data types compatible with Partek Flow (see [Types of Data](#)).

Text file and data are both in server and in same directory

If the text file and the data files are both in the *same* directory of the **Partek Flow server**, simply include the filenames in the text file as shown in Figure 6. You do not have to specify the file path.

Sample name	Treatment	Genotype	Age	file1	file2	file3	file4
NA1031	Case	wt	15	NA1031_S25_L007_R1_001.fastq.gz	NA1031_S25_L007_R2_001.fastq.gz	NA1031_S45_L008_R1_001.fastq.gz	NA1031_S45_L008_R2_001.fastq.gz
NA1032	Case	mutant	18	NA1032_S46_L008_R1_001.fastq.gz	NA1032_S46_L008_R2_001.fastq.gz	NA1032_S26_L007_R1_001.fastq.gz	NA1032_S26_L007_R2_001.fastq.gz
NA1033	Control	wt	17	NA1033_S47_L008_R1_001.fastq.gz	NA1033_S47_L008_R2_001.fastq.gz	NA1033_S27_L007_R1_001.fastq.gz	NA1033_S27_L007_R2_001.fastq.gz
NA1034	Control	mutant	17	NA1034_S28_L007_R1_001.fastq.gz	NA1034_S28_L007_R2_001.fastq.gz	NA1034_S48_L008_R1_001.fastq.gz	NA1034_S48_L008_R2_001.fastq.gz

Figure 19. Text with attributes and filenames

Text file and data are in different directories

If the text file and the data files are in different directories, you must modify Figure 6 to include the file paths. For text files located in the **Partek Flow server**, you can use *relative* or *absolute* file paths. For text files located in **My Computer** or **URL** you must use an *absolute* file path. (Remember that in either case, the data files must have already been uploaded in the **Partek Flow server**.)

Relative path

You can use a path relative to the location of the text file you are using to create the sample table. For example in Figure 7, the text file is located in a directory called "download" while the files are in a subdirectory called MyData, then the filenames must include the path /MyData/. An example would be: `/MyData/NA1031_S25_L007_R1_001.fastq.gz`

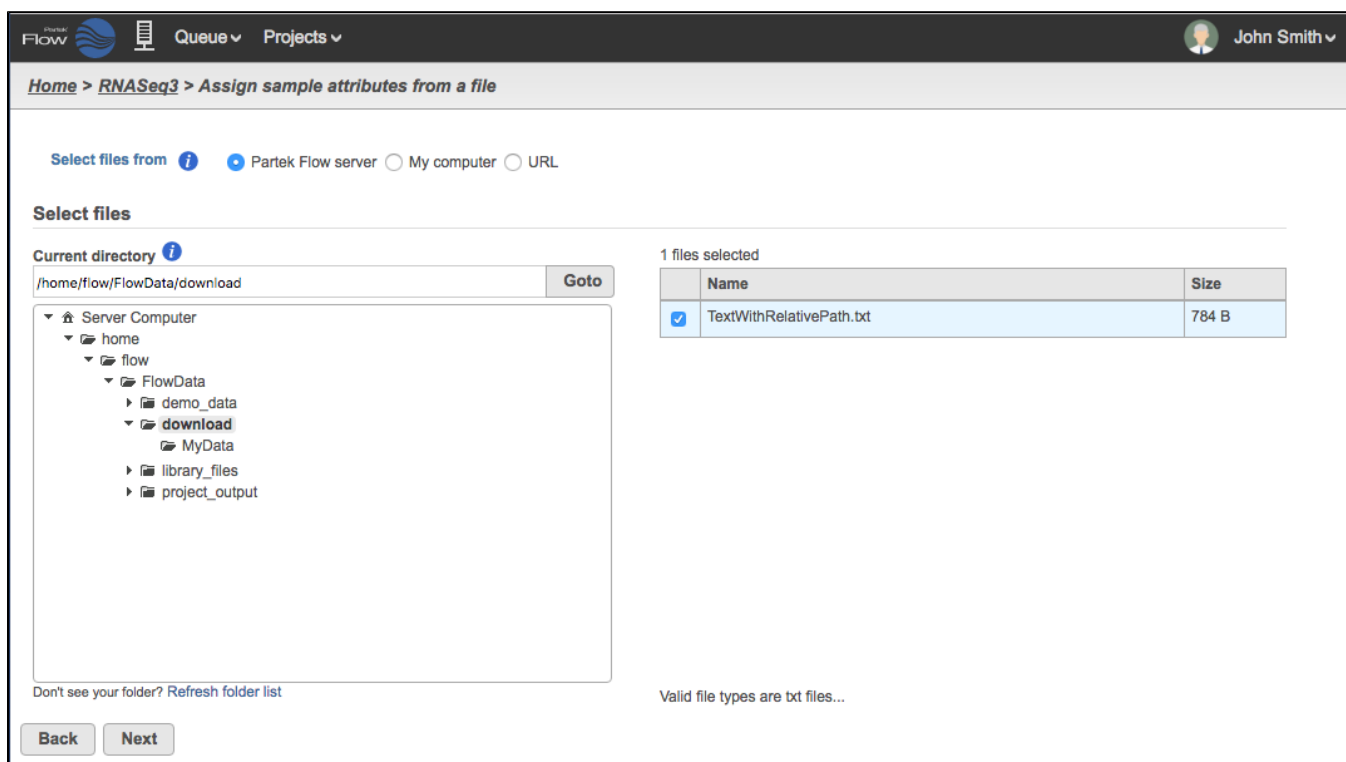


Figure 20. The text file is located in a directory called "download" while the files are located in a directory one level below called "MyData"

Absolute path

You can use the full path to the file based on your Partek Flow server. You must include the paths starting from the Partek Flow home directory (Figure 8, red box). For typical installations, the path begins with /home/flow/FlowData/ and so the filenames to include in the text file may look like this: /home/flow/FlowData/download/MyData/NA1031_S25_L007_R1_001.fastq.gz

Partek
Flow

Queue

Projects

John Smith

Home > RNASeq3 > Assign sample attributes from a file

Select files from

Partek Flow server

My computer

URL

Select files

Current directory

/home/flow/FlowData

Goto

Server Computer

home

flow

FlowData

demo_data

download

MyData

library_files

project_output

1 files selected

Don't see your folder? [Refresh folder list](#)

Valid file types are txt files...

Back

Next

Figure 21. The Current directory (red box) shows the file path to be included in a text file containing absolute paths

Once you have selected the text file with filenames and/or file paths, click **Next**. The text file will be summarized as in Figure 9. If the file types are valid, Partek Flow will recognize the filenames and locate the files within the directory. The *Files* column represent files to be imported.

Partek Flow

Queue▼Projects▼

John Smith▼

Home > RNASeq2 > Assign sample attributes from a file

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

Attribute name	Terms	Import	Files	Sample ID	Attribute type
Sample name	NA1031, NA1032, NA1033, NA...	<input type="checkbox"/>		<input checked="" type="radio"/>	Categorical
Treatment	Case, Control	<input checked="" type="checkbox"/>			Categorical
Genotype	mutant, wt	<input checked="" type="checkbox"/>			Categorical
Age	15, 17, 18	<input checked="" type="checkbox"/>			Numeric
file1	NA1031_S25_L007_R1_001.fa...	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	Categorical
file2	NA1031_S25_L007_R2_001.fa...	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	Categorical
file3	NA1031_S45_L008_R1_001.fa...	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	Categorical
file4	NA1031_S45_L008_R2_001.fa...	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="radio"/>	Categorical

Show/hide file preview

TextWithFilenames.txt (656 B)

1	Sample name	Treatment	Genotype	Age	file1	file2	file3	file4
2	NA1031	Case	wt	15	NA1031_S25_L007_R1_001.fastq.gz	NA1031_S25_L007_R2_001.fastq.gz	NA1031_S45_L008_R1_001.fastq.gz	NA1031_S45_L008_R2_001.fastq.gz
3	NA1032	Case	mutant	18	NA1032_S46_L008_R1_001.fastq.gz	NA1032_S46_L008_R2_001.fastq.gz	NA1032_S26_L007_R1_001.fastq.gz	NA1032_S26_L007_R2_001.fastq.gz
4	NA1033	Control	wt	17	NA1033_S47_L008_R1_001.fastq.gz	NA1033_S47_L008_R2_001.fastq.gz	NA1033_S27_L007_R1_001.fastq.gz	NA1033_S27_L007_R2_001.fastq.gz
5	NA1034	Control	mutant	17	NA1034_S28_L007_R1_001.fastq.gz	NA1034_S28_L007_R2_001.fastq.gz	NA1034_S48_L008_R1_001.fastq.gz	NA1034_S48_L008_R2_001.fastq.gz

<<1 - 5 of 5 rows>>

BackImport

Figure 22. Selecting the attributes, Sample ID and filenames

Click **Import** and the table will be created. After a few seconds, the server icon will be animated (Figure 10). This means that the process of importing data files have started. Click the **Queue** dropdown and select **View queued tasks...** to see the queue of import tasks (Figure 11).

Partek Flow

Queue▼Projects▼

John Smith▼

Home > RNASeq2

AnalysesDataQueue▼Settings

Project estimated end: Unknown

	Sample name	Treatment	Age
1	NA1031	wt	Case
2	NA1032	mutant	Case
3	NA1033	wt	Control
4	NA1034	mutant	Control

Show data filesDownload

Import data

Edit attributes

Assign sample attributes from a file

Add a system-wide attribute column

Manage attributes

Project output directory /home/flow/FlowData/project_output/Project_RNASeq2 (10.27 TB free)

Figure 23. Sample table created from a text file and data is queued for import

Queue ▾

Projects ▾

John Smith ▾

Home > Queued tasks

There are 4 tasks in the queue and Partek Flow is set to run up to 100 project(s) simultaneously.

Status	Task	Project	User	Submitted	End	Cancel
	[Import] Import unaligned reads	RNASeq2	John Smith	28 Sep 2016, 12:34 PM CDT	28 Sep 2016, 12:37 PM CDT	
Waiting	[Import] Import unaligned reads	RNASeq2	John Smith	28 Sep 2016, 12:34 PM CDT	Unknown	
Waiting	[Import] Import unaligned reads	RNASeq2	John Smith	28 Sep 2016, 12:34 PM CDT	Unknown	
Waiting	[Import] Import unaligned reads	RNASeq2	John Smith	28 Sep 2016, 12:34 PM CDT	Unknown	

View recent activity

- Waiting for upstream tasks to complete
 - Waiting for system resources
 - Limited by number of running projects
 - Cannot run with current system configuration

Time estimates are being continuously updated and will become more accurate.

Figure 24. Task Queue showing import tasks for each sample

At this stage, you can also go to the *Analyses Tab* of the project and see that the data node has been created but the color is light blue, which indicates the import is not complete (Figure 12, left). Once all the files have been imported from the queue, the data node will appear darker (Figure 12, right).

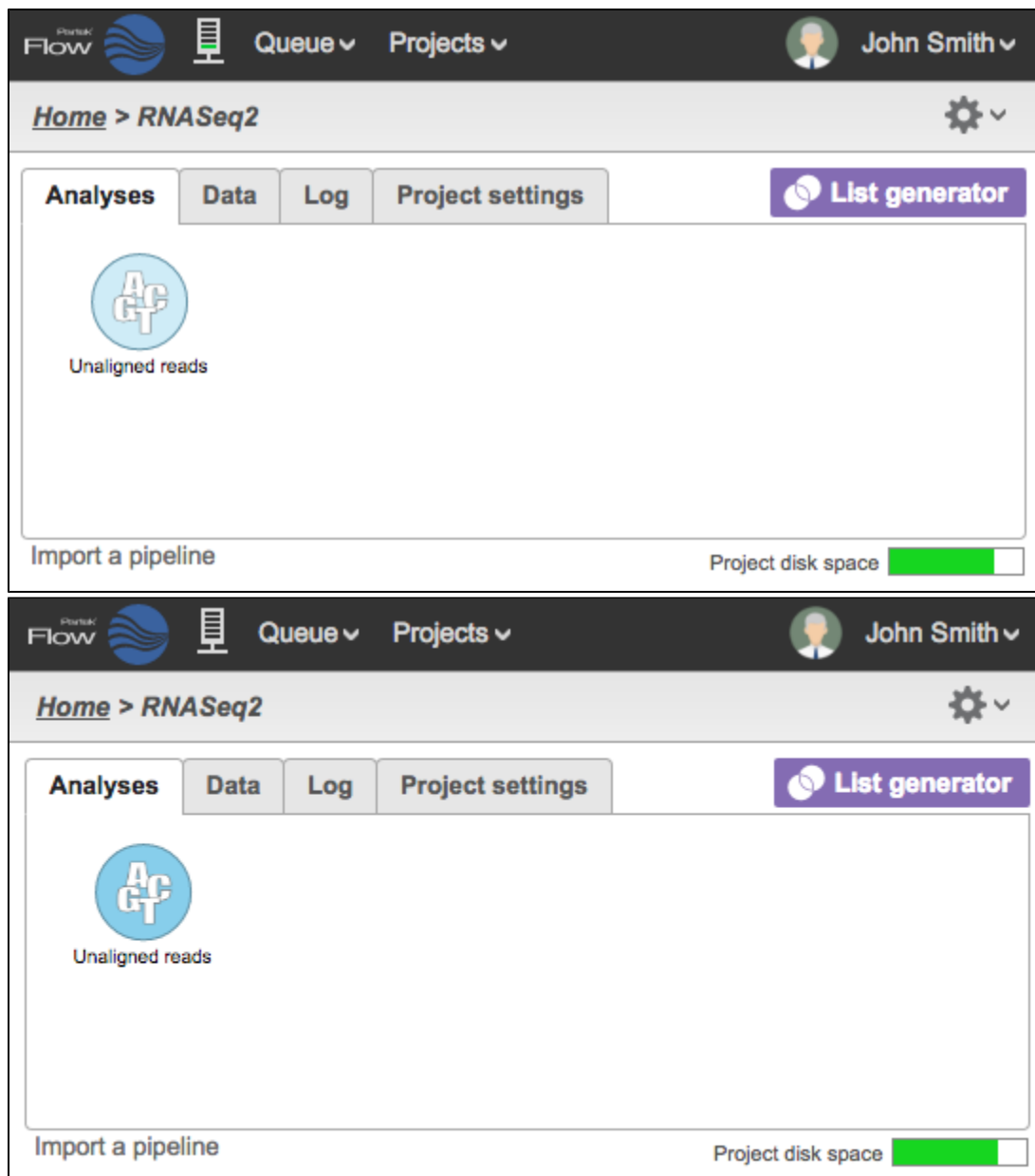


Figure 25. Analyses Tab showing data node during import (left) and once the import is completed (right)

To view the files associated with the data, go to the *Data Tab* and click **Show data files** to expand the table. Figure 13 shows that four files were successfully imported for each sample. You can add or delete samples as described in the [Adding samples](#) section.

