

Microarray Toolkit

With the Partek® Flow® Microarray Toolkit, you can import and analyze microarray data with the same ease as any sequencing analysis pipeline. This document covers the following parts.

- [Supported Platforms](#)
- [Data Import](#)
- [Conversion to Aligned Reads](#)
- [Visualizing Microarray Probes](#)
- [Downstream Tasks](#)
- [References](#)
- [Additional Assistance](#)

Supported Platforms

Partek Flow supports popular gene expression microarray platforms including Affymetrix GeneChips (.CEL) and Illumina BeadChips (.idat). You can also import text output from Illumina Genome Studio in the form of tab-delimited text files with probe IDs and AVGSIGNAL values. For the latter, each sample should correspond to one text file.

For custom-made Affymetrix .CEL and Illumina .idat or AVGSIGNAL data generated from Genome Studio files, refer to the documentation on [Importing Custom Microarrays](#) for additional instructions regarding the probe files.

Data Import

To import microarray files into a project, go to the *Data* tab of the project and select the **Import Data** button. If the data is not already in an existing Partek Flow project, select the **Automatically create samples from files** button. You will be presented with three different ways to upload your data files:

- Directly from your Partek Flow server
- From My Computer
- From a URL

The most efficient way of importing array data would be to import them directly from your Partek Flow server. When you select this option, navigate to the folder containing your data files. Valid file types will be selectable for upload. Click on the files you would like to import into the project and select the **Create sample** button (Figure 1).

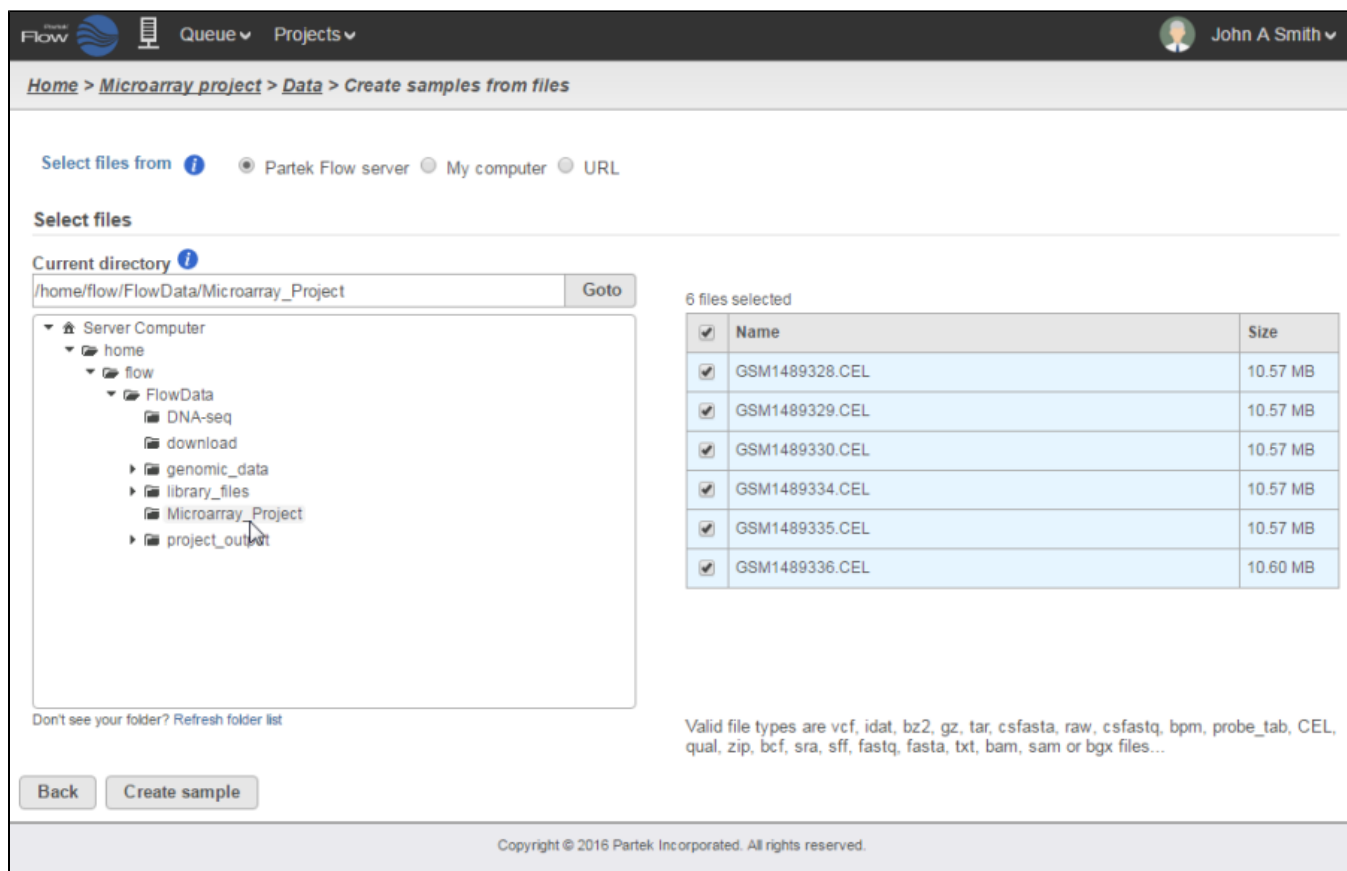


Figure 10. Importing microarray files into Partek Flow

During the upload process, Partek Flow will determine the specific platform of your dataset. If they are commonly used chips, such as Affymetrix HTA arrays, the gene and transcript annotations associated with the platform will automatically be downloaded in the *Library File Manager*.

Once the samples have been uploaded, the *Data tab* will list each sample along with the *Platform* associated with each sample. If Affymetrix .CEL files were uploaded, it will also include the date that the array was scanned (Figure 2). This information may be helpful in assessing possible batch effects.

The screenshot shows the Partek Flow web interface for a 'Microarray project'. The top navigation bar includes 'Queue' and 'Projects' menus, and a user profile for 'John A Smith'. The breadcrumb trail is 'Home > Microarray project'. Below this, there are tabs for 'Analyses', 'Data' (selected), 'Log', and 'Project settings'. The 'Data' tab displays a table with the following data:

	Sample name	Attributes	
		Platform	Scan date
1	GSM1489328	RaGene-1_0-st-v1	04/28/2011
2	GSM1489329	RaGene-1_0-st-v1	04/28/2011
3	GSM1489330	RaGene-1_0-st-v1	04/28/2011
4	GSM1489334	RaGene-1_0-st-v1	04/28/2011
5	GSM1489335	RaGene-1_0-st-v1	04/28/2011
6	GSM1489336	RaGene-1_0-st-v1	04/28/2011

Below the table are buttons for 'Show data files' and 'Download'. Further down, there are buttons for 'Import data', 'Edit attributes', 'Assign sample attributes from a file', 'Add a system-wide attribute column', and 'Manage attributes'. At the bottom, it shows the 'Project output directory' as '/home/flow/FlowData/project_output/Project_Microarray project (16.62 TB free)'. The footer contains the copyright notice: 'Copyright © 2016 Partek Incorporated. All rights reserved.'

Figure 11. Imported microarrays will automatically include Platform and, in some cases, Scan date in the Attributes page

You can now add additional sample attributes to your microarray project. For more information on how to set up your project, please refer to the documentation on *Creating a project*.

Custom annotations

Partek Flow also supports custom-made Affymetrix and Illumina arrays. Once the data files have been uploaded, you will also need to supply a custom annotation file which contains information on the probe sequences associated with the array. Please refer to the document on [Importing Custom Microarrays](#) for additional information on how to create and upload these annotations.

Conversion to Aligned Reads

Once the microarray data has been uploaded, the *Microarray intensity* data node will appear in the *Analyses* tab. To convert this data to aligned reads, select the **Microarray intensity** data node and select the aligner you would like to use for conversion (Figure 3).

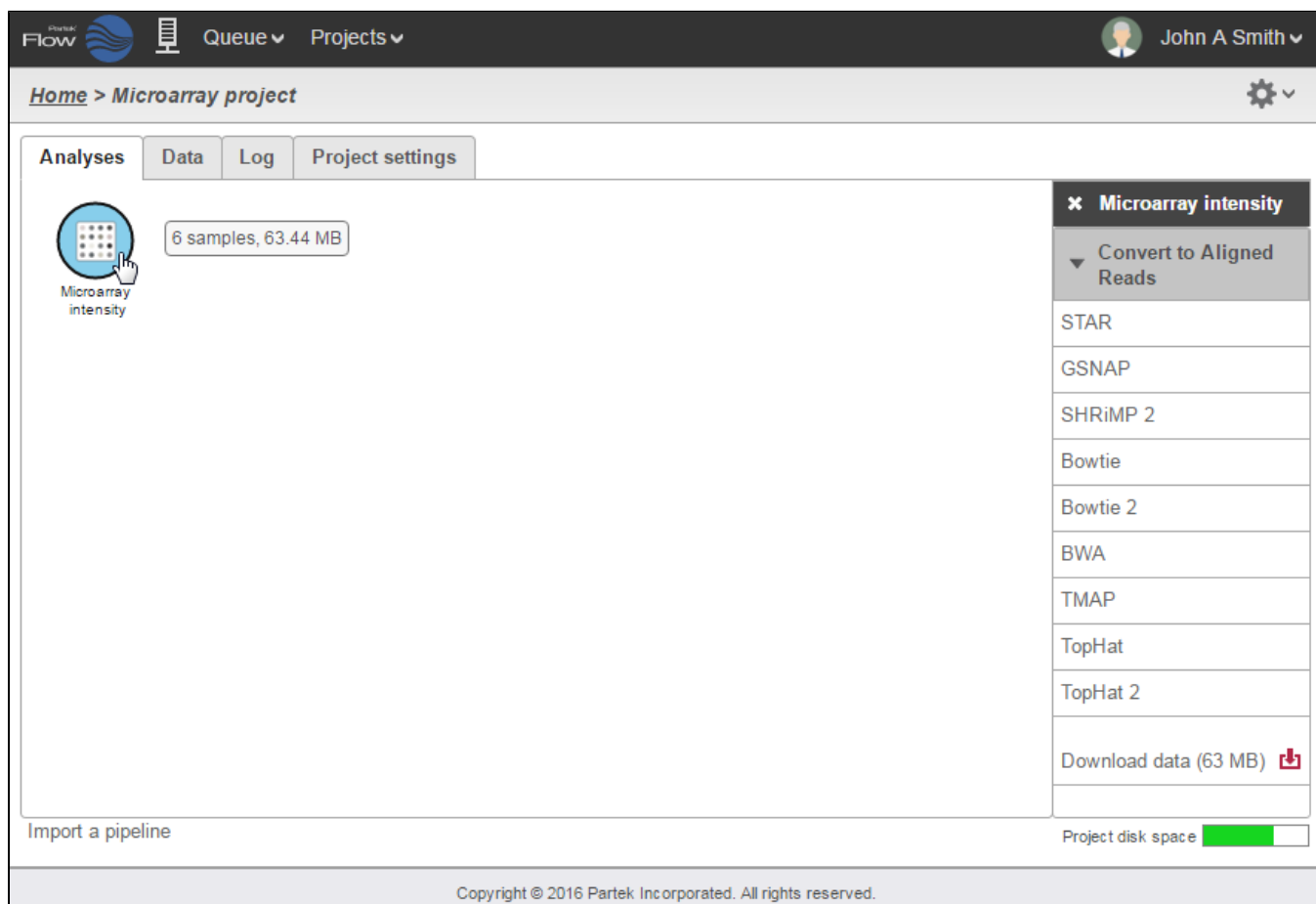


Figure 12. Converting Microarray intensity to Aligned reads

Selecting an aligner will bring up the *Microarray Conversion options* page (Figure 4). Under the *Select probe sequence file* section, make sure that the correct *Chip name* is specified. For non-custom arrays, these should be set automatically to the platform detected upon importing the array.

Select probe sequence file

Chip name
MTA-1_0

[Change selection](#)

Select coverage depth

Nominal read coverage depth (in millions) ⓘ

Advanced options

Option set
-- Default --
[Configure](#)

Back
Next

Figure 13. The Microarray Conversion options page

If for some reason you wish to override that selection, then click the **Change selection** link and select the chip name in the drop-down menu or select **New chip...** if you would like to upload a different chip annotation (Figure 5).

Select probe sequence file

Chip name

RaGene-1_0-st-v1 ▼
AG
RaGene-1_0-st-v1
New chip...

Import options

Figure 14. Selecting a different probe sequence file

In the *Select coverage depth* section, you can set the nominal read coverage depth for your array. This is by default set at 20 million reads. However, you can adjust this number for better accuracy or if you are evaluating low-expressing genes. If you are comparing the microarray data with a corresponding NGS dataset, you can use this option to scale your data to a more comparable order of magnitude. Take note that setting this to a higher number will result in longer computation times overall.

Advanced options

Clicking the *Configure* link under the *Advanced options* section to view the advanced options that can be modified. These are generally transformations applied to microarray data (Figure 6). Note that both interrogating and control (if present) probes are used during fitting and adjustment.

Advanced options

Sequence options

Consolidation of identical probe sequences

mean

Specify annotation sequence format

none

Sequence correction transform

Sequence correction transform

GCCN sequence correction transform

Signal space transformation (SST)

Signal space transformation (SST)

Scale intensities floor

1

Scale intensities ceiling

1000000

Scale intensities low

20

Scale intensities high

50000

Scale intensities detect previous transform

65535

RMA background transform

RMA background transform

Quantile normalization

Quantile normalization

Apply

Save as new

Cancel

Figure 15. Advanced options for Microarray conversion

The *Specify annotation sequence format* drop-down menu allows you to specify the format of the probe sequences associated with your microarrays. Please consult with the manufacturer of your microarrays to determine the exact specification of their probe annotation files. The options available for sequence annotation include:

- None: Assumes that the sequences in the microarray annotation are the probe sequences printed on the array
- Complement: Assumes that the sequence is for the region being interrogated by the array
- Reverse complement: Assumes that the sequence is the reverse complement of the interrogated sequence. This is particularly useful if you have primer sequences as annotations

The *Consolidation of identical probe sequences* option is available to take into account probes that have identical sequences within the same array. This is particularly common in microRNA arrays, where identical mature miRNA sequence can exist for multiple organisms.

By default, this is set to take the *mean* of the intensities the probes and utilizes that to calculate the read count for the probe sequence. Clicking on the drop-down menu allows you to either turn off this feature (*none*) or use the *median* instead (Figure 6).

Import options

Nominal read coverage depth (in millions) i

Consolidation of identical probe sequences i mean none mean median

Advanced options

Option set -- Default -- Configure

Figure 16. Changing the way identical probes are consolidated

Clicking the *Sequence correction transform* checkbox corrects intensity values for sequence-specific effects. It is turned off by default and we only advise that you select this option if there are wide differences in the GC content of your probes.

Both the *GCN sequence correction transform* and *Scale intensity transforms* were developed by Affymetrix [1]. GCN scales the intensities with respect to the difference in probe affinity associated with Guanine and Cytosine (GC) content. On the other hand, the *Scale intensity transform* was developed improve inter-platform comparisons between microarray and RNA-seq data. Specifically, the algorithm stretches the intensity distribution to a common range with a power law mapping that decompresses fold change ratios . This ultimately simplifies fold change comparisons between different technologies.

The *RMA background transform* check box applies robust multi-array average (RMA) normalization to your microarray samples [2]. This is a well-accepted method of background correction for microarray intensities.

The *Quantile normalization* check box performs quantile normalization on all microarray intensities. This is important in making sure that signals from any two arrays are comparable.

Any combination of these advanced options can be saved as its own option set. Once the option set has been created, it will be available in the *Option set* drop-down menu for use in future projects.

After you have finished configuring all microarray conversion options. Click the **Next** button to configure the parameters for the aligner you have chosen to use. Please refer to the documentation on *Alignment* for more information about configuring aligners.

Visualizing Microarray Probes

To visualize microarray probes in Partek Flow's *Chromosome viewer*, click the **Select tracks** button at the top left corner of the viewer. Select the **Probe intensities** check box to display microarray probes (Figure 8).

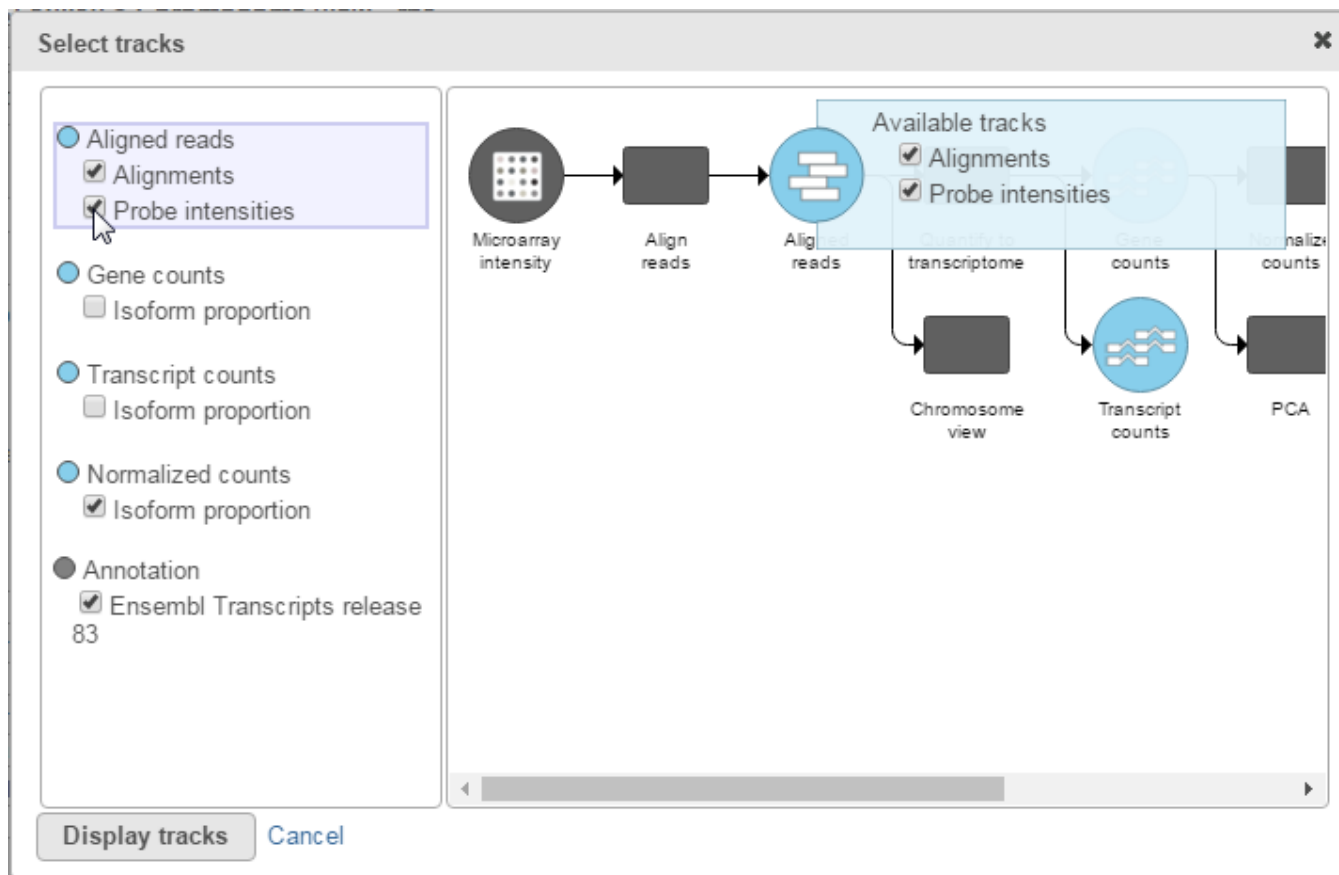


Figure 17. Displaying microarray probes in Chromosome viewer

This will display microarray probes aligned to the reference genome (Figure 9). By default, the probes are colored by their intensities with darker shades signifying higher probe intensities.

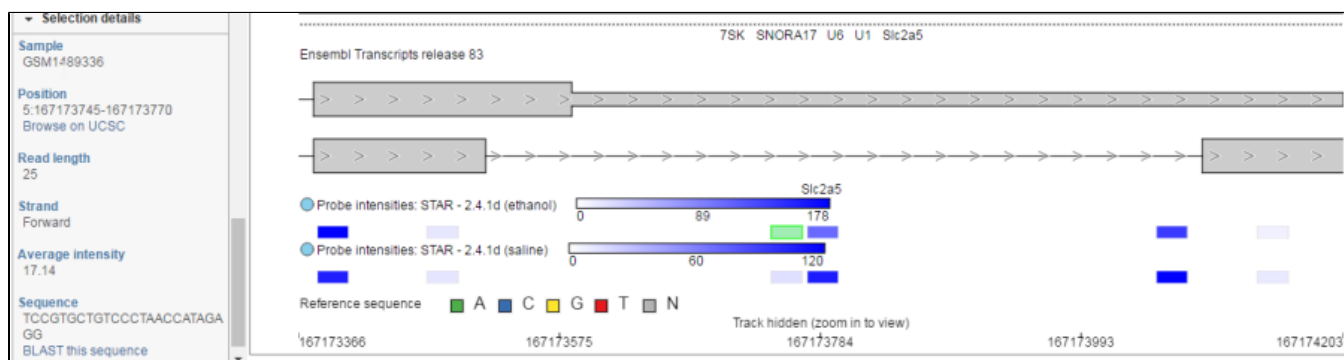


Figure 18. Viewing probe intensities

Clicking a specific probe will display additional information about that probe including mapping position and sequence length. If the probe intensities are consolidated for specific categorical attributes, the average probe intensity is also displayed. For more details on how to use the *Chromosome viewer*, please consult its documentation.

Downstream Tasks

The *Convert to aligned reads* task generates an *Aligned reads* data node. This then allows you to perform most tasks that can be performed on that data node type in Partek Flow. These include *Quantify to transcriptome*, *Normalize counts* and *Differential gene expression (GSA) analysis*. Please consult the documentation on a specific task for specific instructions on their use.

References

1. Affymetrix Whitepaper on [Microarray normalization using Signal Space Transformation with probe Guanine Cytosine Count Correction](#). Accessed February 17, 2016
2. Irizarry RA, Hobbs B, Collin F et al. [Exploration, normalization, and summaries of high density oligonucleotide array probe level data](#). *Biostat.* 2003; 4:249-264.

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:       45 rates