

Importing Exon Data into Partek® Genomics Suite™

This tutorial describes how to import data from Affymetrix GeneChip® Human Exon 1.0 ST Arrays. The example data used in this tutorial includes 20 paired normal and colon cancer samples obtained from 10 subjects. Importing exon arrays is basically the same process as importing Affymetrix gene expression arrays, although you may choose to import only a subset of “well annotated” exons.

System requirements: The minimum configuration is 512MB RAM and 3 GB free disk space.

NOTE: Partek is capable of importing all ~1.5 million exons on the array, and capable of running RMA on any number of exon arrays without pre-filtering using a fixed amount of memory. Partek’s import from .CEL files is highly optimized and requires constant memory no matter how many chips are imported (even using multi-chip normalization algorithms such as RMA). Although the import and normalization can handle any number of chips, the software performance will be reduced when the memory of the computer is filled completely. For exon arrays, Partek can efficiently process up to 340 Exon chips in 3 hours on a laptop PC with a 2 GHz Pentium Centrino CPU and 2GB RAM. This study includes only 20 exon arrays and can be imported and normalized with RMA in approximately 10 minutes and use less than 512 MB of RAM.

For this tutorial, 1.5 million exons will be filtered to a “well annotated” core subset for a “quick start”. However, importing and analyzing all 1.5 million exons can be done similarly without pre-filtering.

This tutorial will demonstrate how to:

- Import Affymetrix Exon CEL files and normalize using the RMA algorithm (Irizarry, et al 2003)
- Import and merge sample information stored in a text file with the RMA imported Exon expression data

Importing CEL Files from Exon Arrays

The data and library files for this experiment can be downloaded by going to **Help > On-line Tutorials** from the Partek main menu and searching for the *Colon Cancer Study* listed under the **Exon** heading.

- Download the zip file *Colon_Cancer_CELFilesAndSampleInfo-Exon.zip* to your local disk. For this example, assume the data will be stored at *C:\Partek Example Data\Colon Cancer (Exon)*
- Download the *Library and Annotation Files*. The zip file includes

HuEx-1_0-st-v2.clf, *HuEx-1_0-st-v2.pgf*,
HuEx-1_0-st-probeset-annot.csv, *HuEx-1_0-st-transcript-annot.csv*, and
HuEx-1_0-st-v2.r2.dt1.hg18.core.mps. Extract the files to the
C:\Microarray Libraries\Affymetrix\HuEx-1_0-st-v2 folder on your local
computer Note: To get the up-to-date library and annotation files, please
go to Affymetrix' web site:

<http://www.affymetrix.com/support/technical/libraryfilesmain.affx>

- Select **Exon** from the workflow combo box in the Partek main window
- Select **Import Samples**
- Select the **Browse...** button to select the *C:\Partek Example Data\Colon Cancer (Exon)* folder. By default, all the files with a .cel extension are selected (Figure 1)

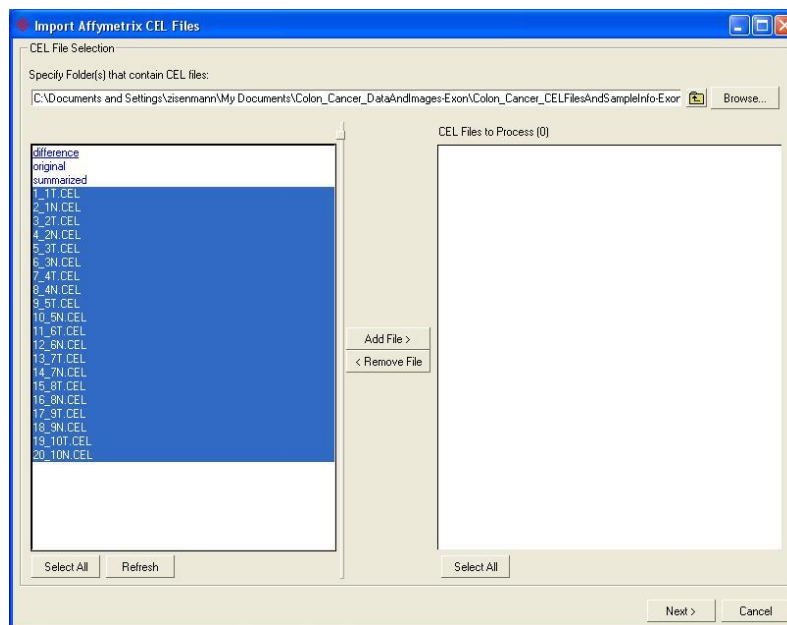


Figure 1: Selecting the folder that contains .cel files of the experiment

- Select the **Add File >** button to move all the .cel files to the right panel. 20 files will be processed
- Select **Next >**. The dialog shown in Figure 2 will appear. Leave the dialog's default files

By default, Partek uses RMA to normalize the data; to change the algorithm parameters and configure to get more outputs, select the **Modify...** button (Figure 2).

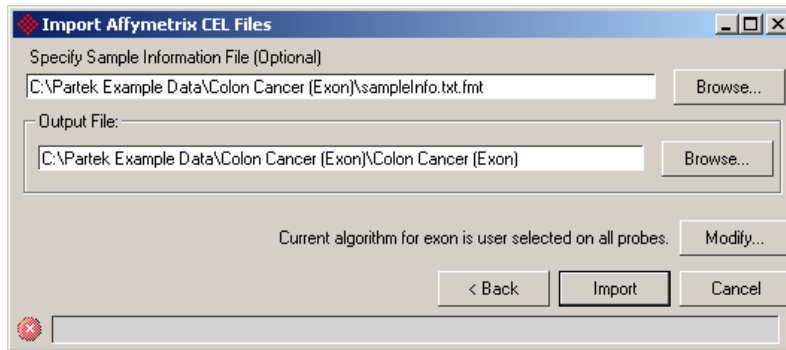


Figure 2: Specifying Input Files

- Leave the settings in the **Algorithm** tab at their default settings (Figure 3)

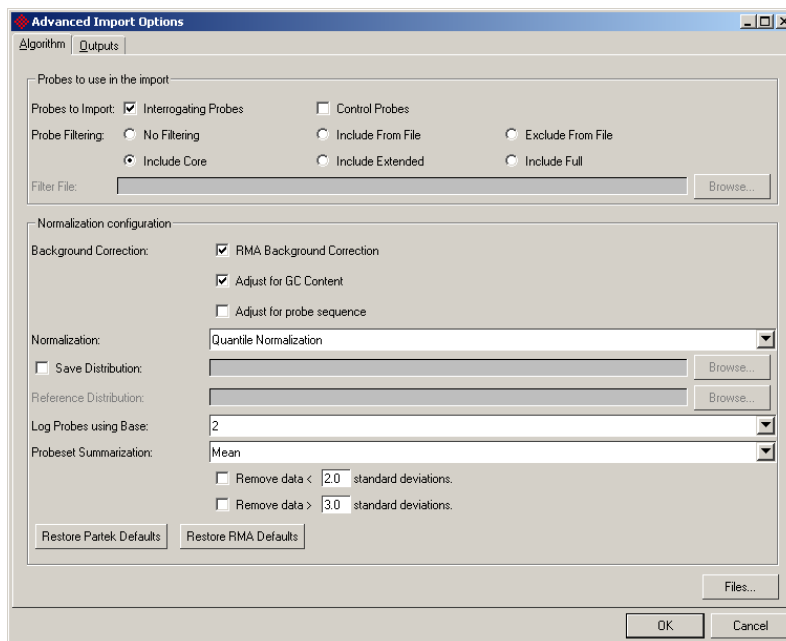


Figure 3: Specifying the normalization algorithm

- Select the **Outputs** tab
- Select to output chip images based on **Original**, **Summarized** and **Difference** values
- In the *Extract Time Stamp and Date From CEL File* panel, click **Date** button to extract the chip scan date. This information can help you to detect if there is batch effects caused by the process time
- *Add Statistics on Raw data* and choose **Median** and **Mean**. We will compute the medians and means based on the raw values for each chip
- Select **OK** (Figure 4)

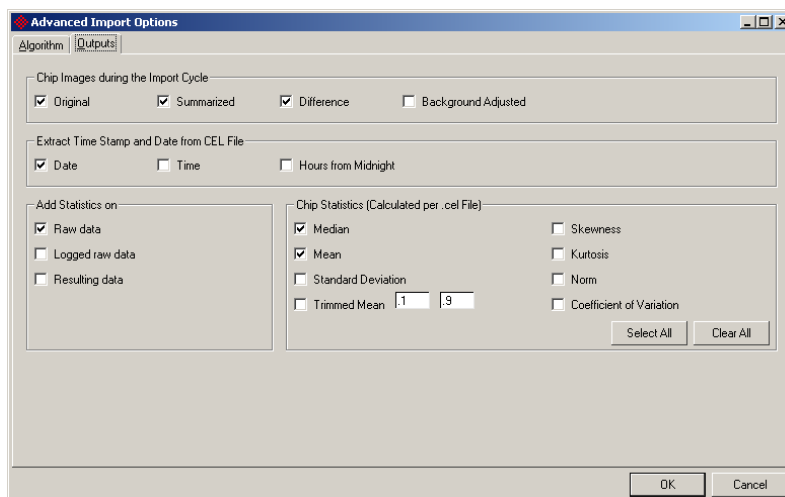


Figure 4: Specifying more outputs values

- Select **Import** on the *Import CEL File* dialog

If you have not already done so, you will be requested to specify the microarray library root folder (Figure 5).

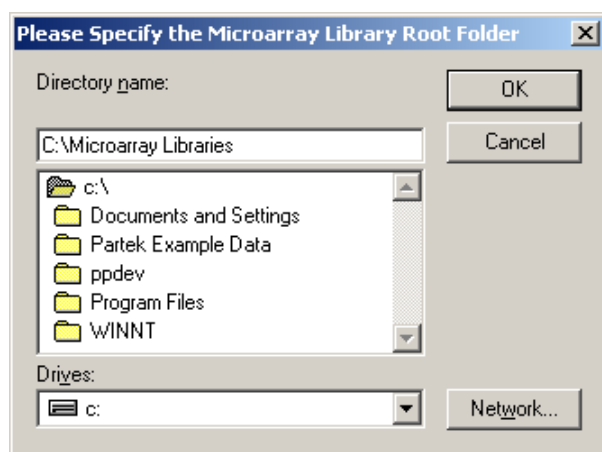


Figure 5: Specifying the Microarray Library Root Folder

- Select **OK** (Figure 6)

Partek will search your driver for the files and begin downloading any files that it cannot find. If you are unable to obtain the files through the automatic downloader then you will have an opportunity to manually specify the files (Figure 6**Error! Reference source not found.**).

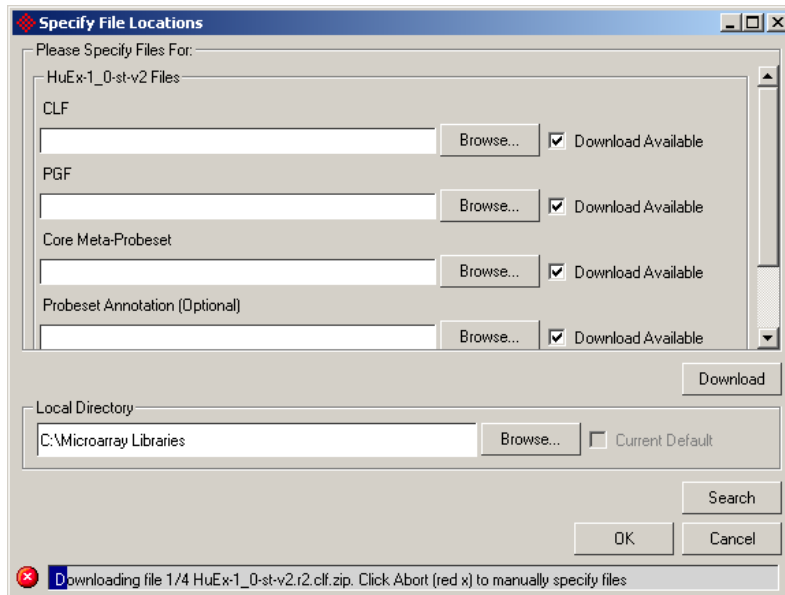


Figure 6: Configuring the Partek file downloader

- Select **OK** (Figure 6)

After the CEL file importing is finished, the result file will be automatically opened in the Partek Analytical Spreadsheet[®], where you will see 20 rows representing 20 chips and ~228K columns representing the exons (Figure 7).

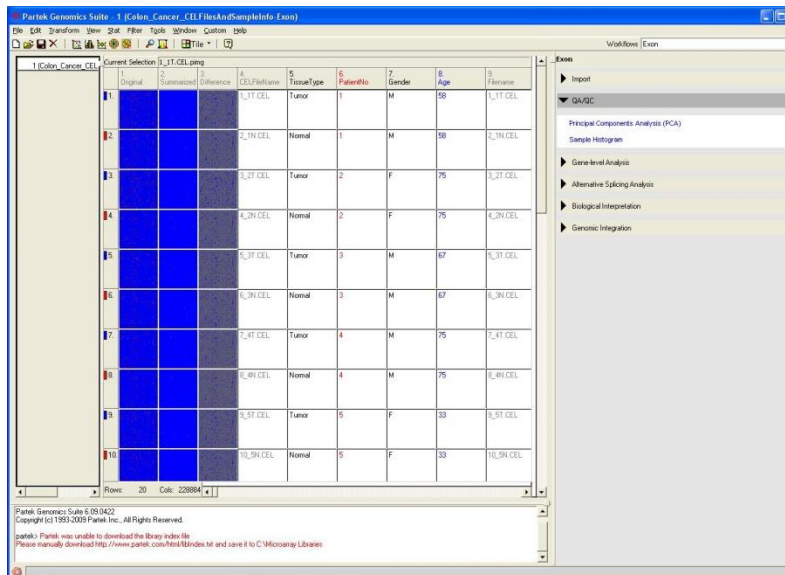


Figure 7: Viewing the Analytical Spreadsheet with CEL file images

End of Tutorial

This is the end of the exon data import tutorial. You can proceed to the Exon Colon Cancer Analysis tutorial. If you need additional assistance with this data set, you

can call our technical support staff at +1-314-878-2329 or email support@partek.com.

References

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., Speed, TP. (2003) Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data. *Biostatistics*, Vol. 4, Number 2: 249-264.