## Importing Exon Data into Partek<sup>®</sup> Genomics Suite<sup>TM</sup>

This tutorial describes how to import data from Affymetrix GeneChip<sup>®</sup> 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)

🚸 Import Affymetrix CEL Files			
CEL File Selection			
Specify Folder(s) that contain CEL files:			
CID comments and Cotting to Second Managements (Colors Co	Data Andreas	Even Verlag, Courses, CEL Electra (Courselate (c.	
C: \Documents and Settings\Zisenmann\My Documents\Colon_Ca	ncer_DataAndimage:	-Exon/Colon_Cancer_CELFilesAndSampleInro-	Stor Browse
	}	CEL Files to Process (0)	
difference	1		
original			
summarized			
2_1N.CEL			
3_2T.CEL			
4_2N.CEL 5_3T.CEL			
6_3N.CEL			
7_4T.CEL 8_4N.CEL			
9_5T.CEL			
10_5N.CEL	-		
12 6N.CEL	Add File >		
13_7T.CEL	< Remove File		
14_7N.CEL 15.ST.CEL			
16_8N.CEL			
17_9T.CEL			
19_10T.CEL			
20_10N.CEL			
Select All Refresh		Select All	
			Next > Cancel

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).

🚸 Import Affymetrix CEL Files	
Specify Sample Information File (Optional)	
C:\Partek Example Data\Colon Cancer (Exon)\sampleInfo.txt.fmt	Browse
Output File:	
C:\Partek Example Data\Colon Cancer (Exon)\Colon Cancer (Exon)	Browse
Current algorithm for exon is user selected on all probe	s. Modify
< Back Import	Cancel

Figure 2: Specifying Input Files

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

gorithm Output	3				
Probes to use in	the import				
Probes to Import:	🔽 Interrogating	g Probes	Control Probes		
Probe Filtering:	O No Filtering		O Include From File	C Exclude From File	
	Include Core	в	O Include Extended	C Include Full	
Filter File:					Browse
Normalization c	onfiguration				
Background Corr	ection:	🔽 RMA Bao	skground Correction		
		🔽 Adjust for	GC Content		
		Adjust for	probe sequence		
Normalization:		Quantile Norm	alization		
🔲 Save Distrib	ution:				Browse
Reference Distrit	oution:				Browse
Log Probes using	) Base:	2			
Probeset Summa	rization:	Mean			
		Remove	data < 2.0 standard deviations.		
		🗌 Remove	data > 3.0 standard deviations.		
Restore Partek	Defaults Re	store RMA Defa	ults		

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)

Advanced Import Options           Algorithm         Quitputs		_ <u> </u>
Chip Images during the Import Cycle           Images during the Import Cycle	Difference Background Adjusted	
Extract Time Stamp and Date from CEL File	Hours from Midnight	
Add Statistics on	Chip Statistics (Calculated per , cel File) Median Mean Standard Deviation Trimmed Mean	Skewness Kurtosis Norm Coefficient of Variation Select All Clear All
		OK Cancel

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).

Please Specify the Microarray Library Roo	t Folder 🛛 🗙
Directory <u>n</u> ame:	ОК
C:\Microarray Libraries	Cancel
<ul> <li>c:\</li> <li>Documents and Settings</li> <li>Partek Example Data</li> <li>ppdev</li> <li>Program Files</li> <li>WINNT</li> </ul>	
Drives:	Net <u>w</u> ork

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.**).

🊸 Specify File Locations				
Please Specify Files For:				
HuEx-1_0-st-v2 Files			— <b>–</b>	
CLF				
	Browse	🔽 Download Available		
PGF				
	Browse	🛛 🗹 Download Available		
Core Meta-Probeset		-		
	Browse	🛛 🗹 Download Available		
Probeset Annotation (Optional)				
	Browse	🛛 🗹 Download Available	<b>_</b>	
		C	ownload	
Local Directory				
C:\Microarray Libraries	Brows	se 🗖 Current Defau	lt	
			Search	
		ОК	Cancel	
🔕 Downloading file 1/4 HuEx-1_0-st-v2.r2.clf.zip. Click Abort (red x) to manually specify files				

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<sup>®</sup>, where you will see 20 rows representing 20 chips and ~228Kcolumns 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 <u>support@partek.com.</u>

## 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.

Copyright © 2010 by Partek Incorporated. All Rights Reserved. Reproduction of this material without expressed written consent from Partek Incorporated is strictly prohibited.