Gene Expression Data Analysis in Partek® Genomics Suite™

HANDS-ON TRAINING

National Institutes of Health

September 2015

Xiaowen Wang
Field Application Specialist
Partek Incorporated
support@partek.com
Contents

Partek Main Dialog.........................................................................................3
Importing Data from Affymetrix Cel Files.........................................................5
QA & QC..........................................................................................................8
Detect Differential Expressed gene.................................................................10
Create Gene List............................................................................................13
Hierarchical Clustering....................................................................................14
Biological Interpretation ................................................................................16
Filter Options................................................................................................21
**Partek® Genomic Suite™ Main Dialog**

**Analytical spreadsheet:** Central repository of data
- No limitation on number of rows or columns
- Rows represent observations of interest (experiments, samples, chips)
- Columns represent measures of the observations (variables, features, genes,)

**Menu bar:** Execute commands from a graphical user interface
- When spreadsheet is empty, most of the menu items are not displayed

**Tool bar:** Accelerator buttons allow quick access to commonly used commands

**Spreadsheet hierarchy:** Open multiple datasets and see the hierarchy
- Original spreadsheet: parent
- Result spreadsheet: child

**Active spreadsheet:** The active spreadsheet is shown highlighted in blue, and the spreadsheet name and associated file name are shown at the top of the dialog

**Workflow:** Used to guide you through a typical analysis of a specific assay
Training Data

Data files in the project:
• 12 samples with 2 treatment on two time points
• Affymtrix HG-U133_Plus_2 array

![Day 3 and Day 10 data comparison diagram with control and treated samples for Day 3 and Day 10.](image-url)
Importing Data from Affymetrix® CEL Files

• Choose **Gene Expression** workflow
• Browse to the folder that contains the CEL files
• Select all the default CEL files, and drag them to the right panel
• Click **Next**
• Specify the output file name—"**Gene expression data**" and use the default settings, then click **Import**
• **Customized** allows you change the algorithm parameters, and verify library files
• PGS will automatically download the library files.

---

**Notes:**

_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
Spreadsheet Properties

Spreadsheet
• Each spreadsheet consists of two files with the same name
• Spreadsheet linked to annotation
• * implies change is not saved, ptmp is unsaved temporary file
• Saving the project will preserve the hierarchy

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene expression data</td>
<td>File</td>
<td>2,567 KB</td>
</tr>
<tr>
<td>Gene expression data.fmt</td>
<td>Partek Data File</td>
<td>634 KB</td>
</tr>
</tbody>
</table>

Annotation
• To link the annotation to the spreadsheet select: File > Properties
• Microarray Libraries: Stores automatically downloaded annotations files

Notes:_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
Add Sample Attributes

Two spreadsheets are generated:
The data spreadsheet contains the RMA value for all probesets
The QC spreadsheet contains the control probe sets

- Select the **Gene expression data** spreadsheet
- Choose **Add attributes from an existing column**
- Specify **Treatment** and **Time** on the first 2 columns respectively, and skip the rest columns
- Click **OK** and **Save** the spreadsheet

- Another way to specify same attribute is to add one categorical attribute at a time
QA and QC

QA/QC is exploratory analysis, it is checking the preparation of the samples and identify outliers.

QC metrics is only available when you import Affymetrix .cel files, it checks the quality of the chips based on control probesets.

Histogram display the distribution of the samples:

- **QC metrics – PostImportQC**
- **Plot sample histogram**
- **View>Box & Whiskers>Rows(response)**

![Histogram graph](image)

**Notes:**

_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
PCA Scatter Plot

PCA scatter plot is another way to identify clustering patterns and outliers

- Select **Principal components analysis** on the workflow

**Notes**

- Each point in the scatter plot corresponds to a specific row in the spreadsheet
- Points that are close together in the plot are similar in the original high-dimensional space
- Points that are far apart in the plot are dissimilar

- Click on **Plot Properties** (red ball), to configure color by **Treatment**, size by **Time**
- Click on **Ellipsoid** to put the ellipsoid on each treatment type
- Select mode:
  - left click to select; scroll mouse wheel to zoom; drag mouse wheel to rotate
  - right click after select a point to filter/clear filter
Detect Differentially Expressed Gene

- Select **Treatment** and **Time**, Click **Add Factor**
- Click **Add Interaction**
- Click **Contrast**
- Add contrast of
  - Trt vs. Con
  - Trt* D3 vs Con * D3
  - Trt * D10 vs Con * D10
- Click **OK**
- Output file: **ANOVAResults**

**Note:** Fold change calculation is different on linear vs. log data.

**Notes:**

__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________
__________________________________________________________________________________
ANOVA in Partek Genomics Suite

Different Types of ANOVA

• Equal variance t-Test
• Paired t-Test
• Repeated Measurement ANOVA
• ANCOVA
• Mixed Model ANOVA
• Correlation

Automatically detects crossed/nested factors
Automatically performs mixed model when random effect are included

6. Treatment vs. 7. Time

<table>
<thead>
<tr>
<th>Treatment\Time</th>
<th>D10</th>
<th>D3</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Con</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Trt</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>6</td>
<td>12</td>
</tr>
</tbody>
</table>
Result of ANOVA

Each row is a gene with its p-value and fold change and any other statistical information. The spreadsheet is sorted by the first p-value column.

Right click on a row header to get details

• Select HML Report
• Select Dot Plot
• Select Source of Variation
• Select ANOVA Interaction Plot
• Select View>Volcano Plot

Right click on the ANOVA spreadsheet > Info > Comments to access the ANOVA model details
Create List

Generate a list of miRNA that is showing differential expression between brain and heart.

- Click **Create List** on the workflow
- Click **Configure** to change the default fold change cutoff as 10
- Create the following 3 gene list with default settings:
  - **Treatment vs Control**
  - **Treatment * D3 vs Control * D3**
  - **Treatment * D10 vs Control * D10**

A new child spreadsheet will be generated for each gene list

- Click on **Venn Diagram** tab to and select the three gene list—PGS allow 5 way Venn diagram
- Select any section in venn diagram to generate a new gene list

---

**Notes:**

________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________
______________________________________________________________________________

---
Hierarchical Clustering

To visualize the heatmap and cluster of the significant list of gene:

- Select **Treatment vs Control** gene list
- Choose **Cluster Based on Significant miRNAs** on the workflow
- Select the **Hierarchical Clustering** option
- Choose the **Treatment vs Control** spreadsheet with default settings
- Click **OK**

Notes:

_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
Hierarchical Clustering Configuration

Heatmap
- Click on the color square to change the heatmap color

Dendrograms
- Uncheck Show dendrogram scale
- Change the width/height of the dendrogram
- Color dendrogram
- Change dendrogram spacing

Rows
- Change the width of annotation
- Check show label
- Change color
- Add new annotation

Columns
- Column header or miRNA name/gene symbol

Mode: mouse over, select, zoom, and flip

Notes:
Biological Interpretation—Pathway Enrichment

**Pathway enrichment:** Test if lead genes are over represented in any pathway

- Select **Treated vs Control** gene list spreadsheet
- Select **Pathway analysis> Partek Pathway> Pathway Enrichment**
- Leave all parameters as default options
- Select **ANOVAResult** as additional list to send to pathway

Pathway enrichment result spreadsheet:

- Right click on a row header to create gene list
  - Export genes in pathway will output all the genes in that pathway from KEGG database
  - Export genes in list and in pathway
- Color genes based on ANOVAResult fold change value
**Biological Interpretation—Pathway ANOVA**

**Pathway ANOVA:** Detect differentially expressed pathway
- Select *Gene expression data* spreadsheet
- Select *Pathway analysis* > *Partek Pathway* > *Pathway ANOVA*
- Change *Restrict analysis to pathways with fewer than 100* genes to save time

Pathway ANOVA result spreadsheets:
- Two spreadsheet—pathway level result and gene level result
- On Pathway ANOVA result spreadsheet
  - Right click on a row header to draw profile and dot plot

---

**Notes:**

[Images of pathway analysis and spreadsheet results are included here.]

---

Copyright © 2013 Partek Incorporated. All rights reserved.
Filter Options

Filter samples
Interactive row filter – create subset of samples based on group information

• Click Filter>Filter Rows>Interactive Filter
  • Right click on a group bar to filter include only selected group
  • Left click on a group bar to toggle the filter status

Filter genes
Filter out low intensity genes

• Click Filter>Filter Column>Column Filter Manager>Filter based on Max < cutoff

Filter out probesets without annotation

• Click Filter>Filter based on genomic location
  • Select Annotation field > Add Constraint

Filter include genes of interest

• Click Filter>Filter Column>Filter column based on a list

Notes:________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________
________________________________________________________________________

Copyright © 2013 Partek Incorporated. All rights reserved.
Output gene list for GSEA analysis

Create unique gene symbol list with sorted t-statistics of Trt*D10 vs Con*D10

- Select **Gene expression data** spreadsheet
- Select **Detect differentially expressed genes** step on the workflow
- Select **Treatment** and **Time**, Click **Add Factor**, then click **Add Interaction**
- Click **Contrast** to add: **Trt** * **D10** vs **Con** * **D10**
- Make sure to check the **T statistics** button
- Click **Add contrast** then Click **OK**
- Output file: **ANOVAResults-GSEA**
- Right click on **p-value**(Trt* D10 vs. Con * D10) to sort Descending

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>30992</td>
<td>223792_at</td>
<td>S175H</td>
<td>THIOE</td>
<td>Transcriptome</td>
<td>0.051753</td>
<td>0.012729</td>
<td>0.013873</td>
<td>0.013873</td>
<td>0.999995</td>
<td>-1</td>
<td>No change</td>
<td>-3.11879e-05</td>
</tr>
<tr>
<td>54562</td>
<td>468535_at</td>
<td>50872</td>
<td>ASPP2</td>
<td>Expression</td>
<td>0.00278673</td>
<td>0.13446</td>
<td>0.06278609</td>
<td>0.000007</td>
<td>1</td>
<td>1</td>
<td>No change</td>
<td>0.00021075</td>
</tr>
<tr>
<td>33576</td>
<td>131415_at</td>
<td>121940</td>
<td>RB600D2</td>
<td>Williams</td>
<td>0.45187</td>
<td>0.529472</td>
<td>1.02541</td>
<td>0.999873</td>
<td>0.000007</td>
<td>-1.00061</td>
<td>Trt*D10 down</td>
<td>0.00046787</td>
</tr>
<tr>
<td>12526</td>
<td>225400_at</td>
<td>27196</td>
<td>RAGPA1</td>
<td>Other</td>
<td>0.147727</td>
<td>0.0122683</td>
<td>0.147767</td>
<td>0.009849</td>
<td>0.009991</td>
<td>-1.00002</td>
<td>Trt*D10 down</td>
<td>0.00029721</td>
</tr>
<tr>
<td>6555</td>
<td>1562552_at</td>
<td>165852</td>
<td>OTHM0202020</td>
<td>NULL</td>
<td>0.008788</td>
<td>0.042729</td>
<td>0.09081</td>
<td>0.008820</td>
<td>0.000005</td>
<td>-1.00002</td>
<td>Trt*D10 down</td>
<td>0.00026850</td>
</tr>
<tr>
<td>51537</td>
<td>2444370_at</td>
<td>2444370</td>
<td>OTHM0202020</td>
<td>NULL</td>
<td>0.377101</td>
<td>0.0284727</td>
<td>0.377732</td>
<td>0.000768</td>
<td>0.000001</td>
<td>-1.00002</td>
<td>Trt*D10 down</td>
<td>0.00026512</td>
</tr>
<tr>
<td>50327</td>
<td>2141992_at</td>
<td>2141992</td>
<td>--</td>
<td>--</td>
<td>0.529094</td>
<td>0.134524</td>
<td>0.529275</td>
<td>0.999795</td>
<td>1.00000</td>
<td>1.00000</td>
<td>Trt*D10 up</td>
<td>0.00021685</td>
</tr>
<tr>
<td>51083</td>
<td>2422525_at</td>
<td>21743</td>
<td>R3802D</td>
<td>Refseq</td>
<td>0.759957</td>
<td>0.422403</td>
<td>0.79049</td>
<td>0.999773</td>
<td>0.000005</td>
<td>-1.00002</td>
<td>Trt*D10 down</td>
<td>0.00029036</td>
</tr>
<tr>
<td>15596</td>
<td>504554_at</td>
<td>504554</td>
<td>--</td>
<td>--</td>
<td>0.399952</td>
<td>0.784091</td>
<td>0.585842</td>
<td>0.999979</td>
<td>1.00000</td>
<td>1.00000</td>
<td>Trt*D10 up</td>
<td>0.00055995</td>
</tr>
<tr>
<td>41114</td>
<td>2222241_at</td>
<td>2222241</td>
<td>PA1748P</td>
<td>Other</td>
<td>0.465827</td>
<td>0.161503</td>
<td>0.618909</td>
<td>0.999966</td>
<td>1.00004</td>
<td>1.00004</td>
<td>Trt*D10 up</td>
<td>0.00057138</td>
</tr>
<tr>
<td>3871</td>
<td>1257315_at</td>
<td>1257315</td>
<td>--</td>
<td>--</td>
<td>0.386568</td>
<td>0.178569</td>
<td>0.386568</td>
<td>0.999026</td>
<td>1.00000</td>
<td>1.00000</td>
<td>Trt*D10 up</td>
<td>0.00048803</td>
</tr>
<tr>
<td>41545</td>
<td>2201925_at</td>
<td>2201925</td>
<td>--</td>
<td>--</td>
<td>0.224624</td>
<td>0.14651</td>
<td>0.223641</td>
<td>0.999969</td>
<td>1.00004</td>
<td>1.00004</td>
<td>Trt*D10 up</td>
<td>0.00057293</td>
</tr>
<tr>
<td>5767</td>
<td>1501096_at</td>
<td>1501096</td>
<td>LOC382949</td>
<td>Undercharacter</td>
<td>0.193401</td>
<td>0.23396</td>
<td>0.193145</td>
<td>0.999802</td>
<td>0.000007</td>
<td>-1.00003</td>
<td>Trt*D10 down</td>
<td>0.00094804</td>
</tr>
<tr>
<td>8942</td>
<td>1508190_at</td>
<td>1508190</td>
<td>LOC382949</td>
<td>Undercharacter</td>
<td>0.051975</td>
<td>0.957538</td>
<td>0.051909</td>
<td>0.999974</td>
<td>1.00002</td>
<td>1.00002</td>
<td>Trt*D10 down</td>
<td>0.00050798</td>
</tr>
<tr>
<td>16068</td>
<td>207492_at</td>
<td>6562</td>
<td>SL2C22A2</td>
<td>Other</td>
<td>0.560548</td>
<td>0.38398</td>
<td>0.559946</td>
<td>0.999932</td>
<td>1.00005</td>
<td>1.00005</td>
<td>Trt*D10 up</td>
<td>0.000786924</td>
</tr>
<tr>
<td>7969</td>
<td>1562032_at</td>
<td>1562032</td>
<td>--</td>
<td>--</td>
<td>0.129235</td>
<td>0.204129</td>
<td>0.128061</td>
<td>0.99995</td>
<td>0.000005</td>
<td>-1.00003</td>
<td>Trt*D10 down</td>
<td>0.00055022</td>
</tr>
<tr>
<td>13846</td>
<td>203026_at</td>
<td>1562</td>
<td>CTF4A</td>
<td>cytochrome</td>
<td>0.376207</td>
<td>0.098859</td>
<td>0.375511</td>
<td>0.999927</td>
<td>1.00007</td>
<td>1.00007</td>
<td>Trt*D10 up</td>
<td>0.00078332</td>
</tr>
<tr>
<td>43021</td>
<td>214613_at</td>
<td>214613</td>
<td>--</td>
<td>--</td>
<td>0.640212</td>
<td>0.162651</td>
<td>0.630282</td>
<td>0.999917</td>
<td>1.00036</td>
<td>1.00036</td>
<td>Trt*D10 up</td>
<td>0.00088342</td>
</tr>
<tr>
<td>9008</td>
<td>325732_at</td>
<td>25096</td>
<td>RPS7AT2</td>
<td>Integrator</td>
<td>0.434984</td>
<td>0.782389</td>
<td>0.438353</td>
<td>0.999828</td>
<td>0.000888</td>
<td>-1.00011</td>
<td>Trt*D10 down</td>
<td>0.00120685</td>
</tr>
</tbody>
</table>

Notes: 
Output gene list for GSEA analysis - Continue

• Right click on Gene symbol column > Createlistwithoccurrence counts, a new spreadsheet is generated
  – Choose File > MergeSpreadsheet Select T(Trt* D10 vs. Con * D10) column on ANOVAResults-GSEA spreadsheet
  – Source: ANOVAResults-GSEA; Destination: new spreadsheet
  – Columnwithcommonterms: genesymbol
  – Mergeselectedcolumns

• Click OK

• Right click on T(Trt* D10 vs. Con * D10) column > Sort Ascending

• Delete all other columns except for gene symbol and t statistics columns

• Choose File > Save As Text > GSEA gene list.txt
Further Training

Self-learning
• Help > Check for Updates
• Help > On-line tutorials
• Recorded webinars

Regional Technical Support
• Email: support@partek.com
• Phone: +1-314-878-2329

Notes:
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________
_________________________________________________________________________