Gene Expression Data Analysis in Partek® Genomics Suite™

HANDS-ON TRAINING

National Institutes of Health

September 2014

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

Contents

Copyright © 2013 Partek Incorporated. All rights reserved.

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

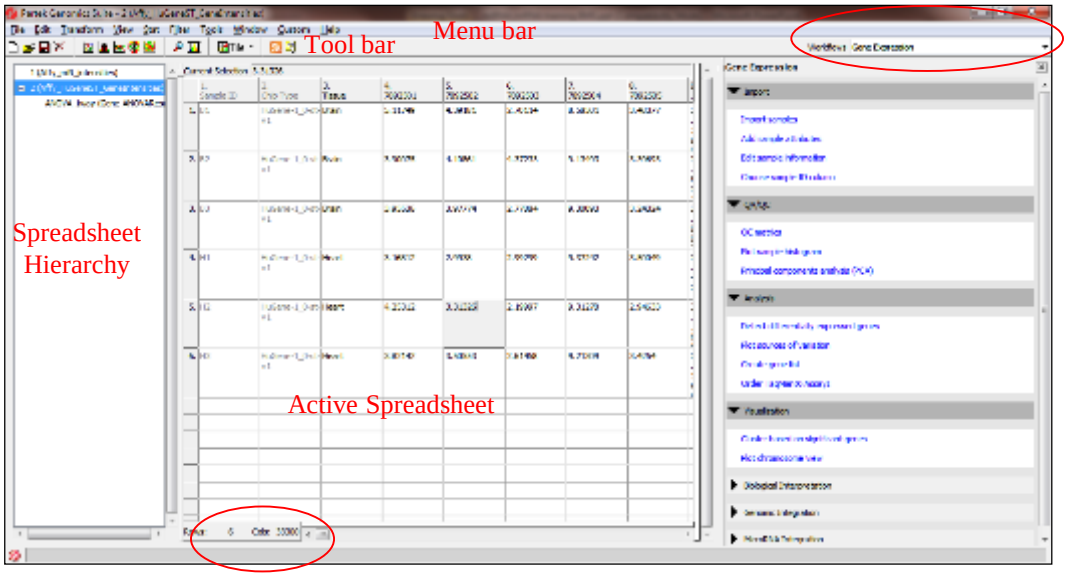
Tool bar: Accelerator buttons allow quick access to commonly used commands

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



Notes:_____

_____

_____

_____

_____

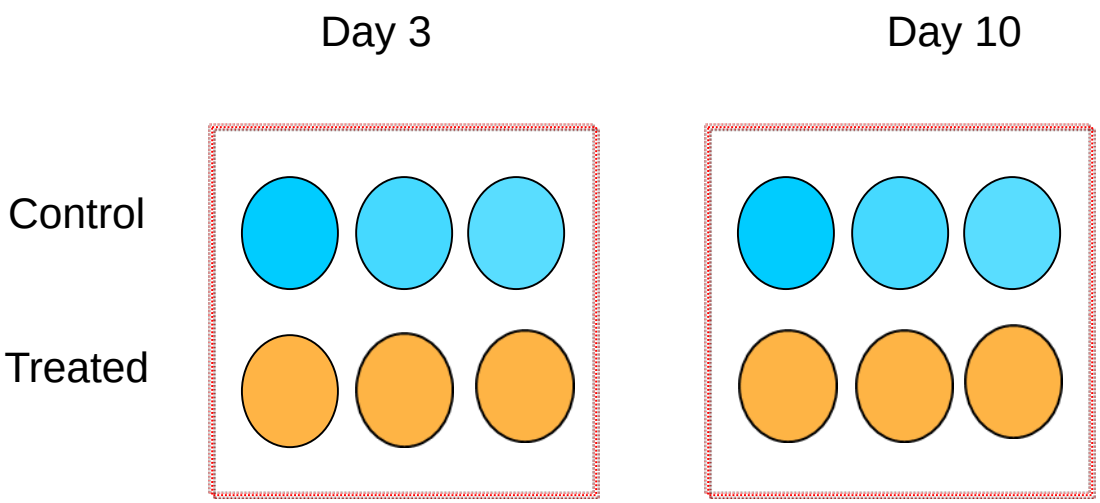
_____

Copyright © 2013 Partek Incorporated. All rights reserved.

Training Data

Data files in the project:

- 12 samples with 2 treatment on two time points
- Affymtrix HG-U133_Plus_2 array



Notes:_____

_____

_____

_____

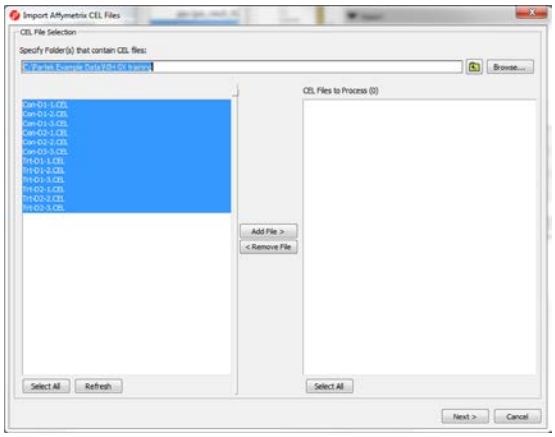
_____

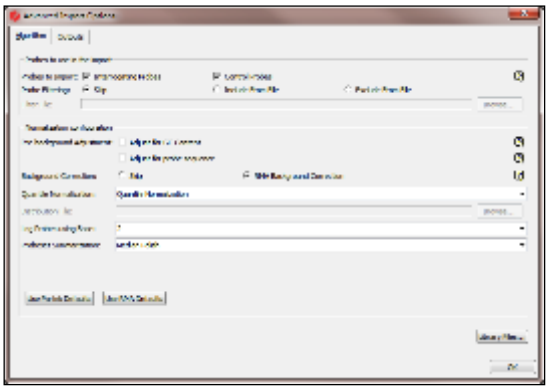
_____

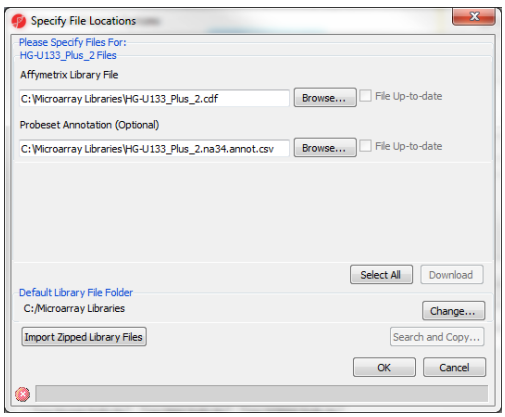
Copyright © 2013 Partek Incorporated. All rights reserved.

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:_____

Copyright © 2013 Partek Incorporated. All rights reserved.

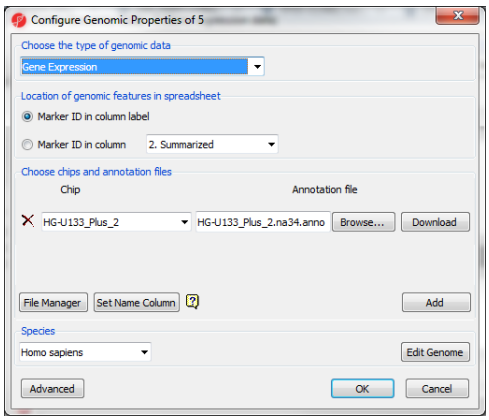
Spreadsheet Properties

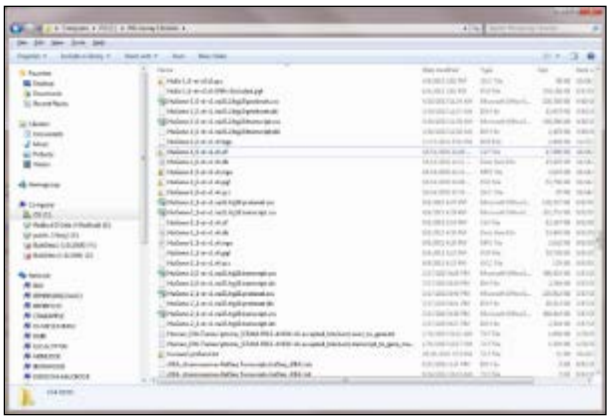
Spreadsheet

- 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

Annotation

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





Notes:_____

Copyright © 2013 Partek Incorporated. All rights reserved.

Add Sample Attributes

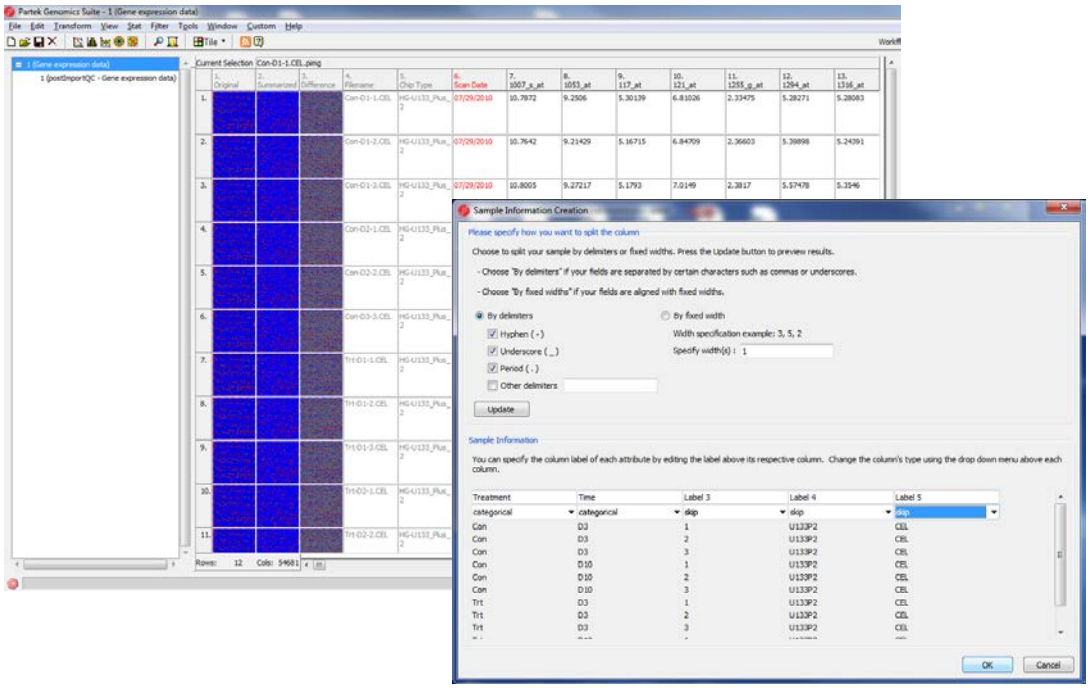
Two spreadsheets are generated:

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



Notes:_____

Copyright © 2013 Partek Incorporated. All rights reserved.

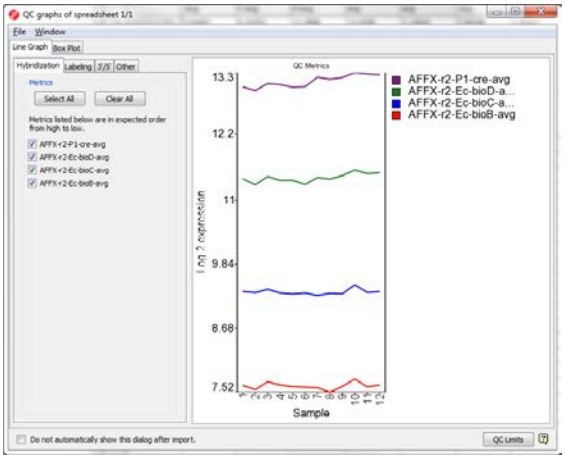
QA and QC

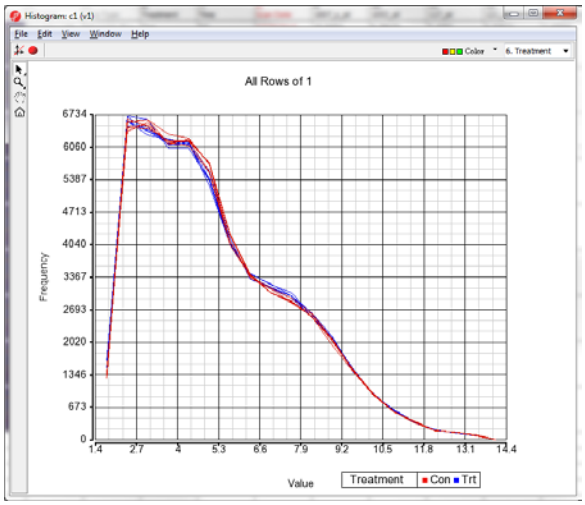
QA/QC is exploratory analysis, it is checking the preparation of the samples and indentify outliers

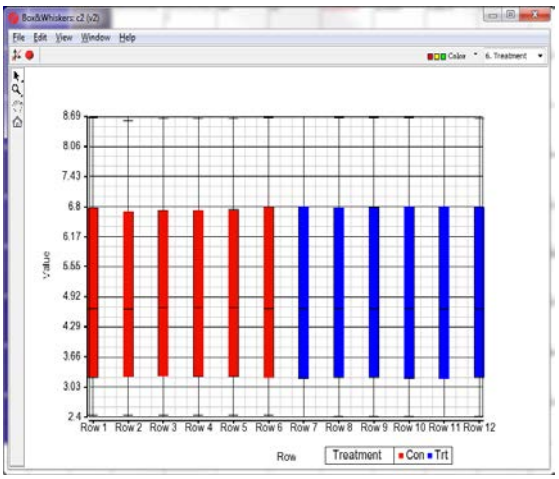
QC metrics is only available when you import Affymetrix .cel files, it checks the quality of the chips based on control probesets

Histogram display the distribution of the samples

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







Notes:_____

_____

_____

_____

_____

_____

Copyright © 2013 Partek Incorporated. All rights reserved.

PCA Scatter Plot

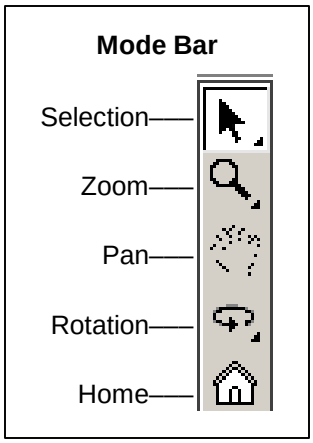
PCA scatter plot is another way to identify clustering patterns and outliers

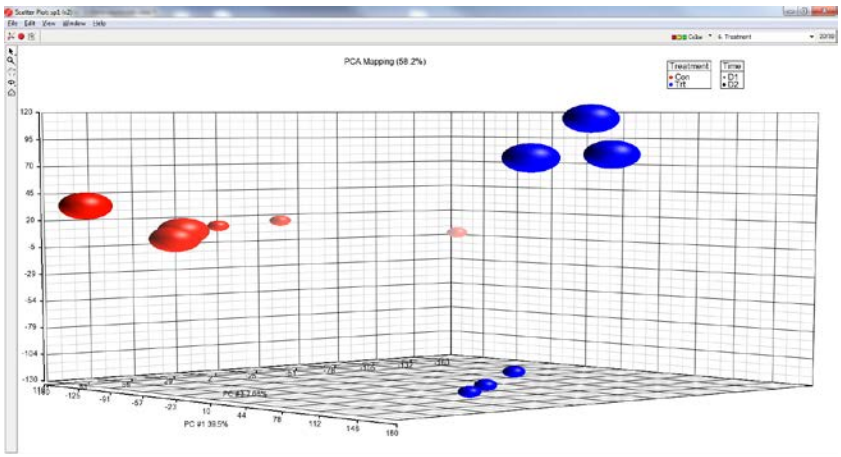
- Select **Principal components analysis** on the workflow

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





Notes:_____

_____

_____

_____

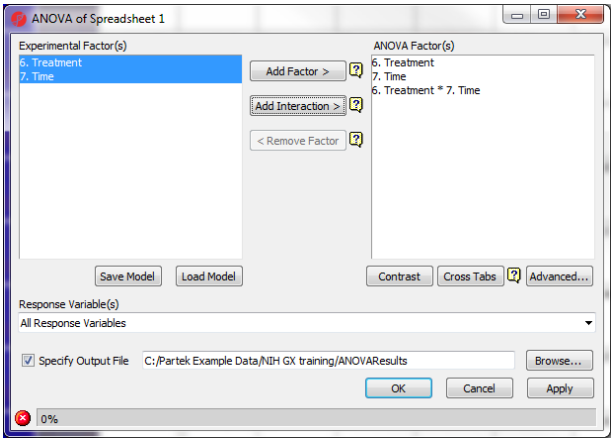
_____

_____

Copyright © 2013 Partek Incorporated. All rights reserved.

Detect Differentially Expressed Gene

- Select **Treatment** and **Time,** Click **Add Factor**
- Click **Add Interaction**



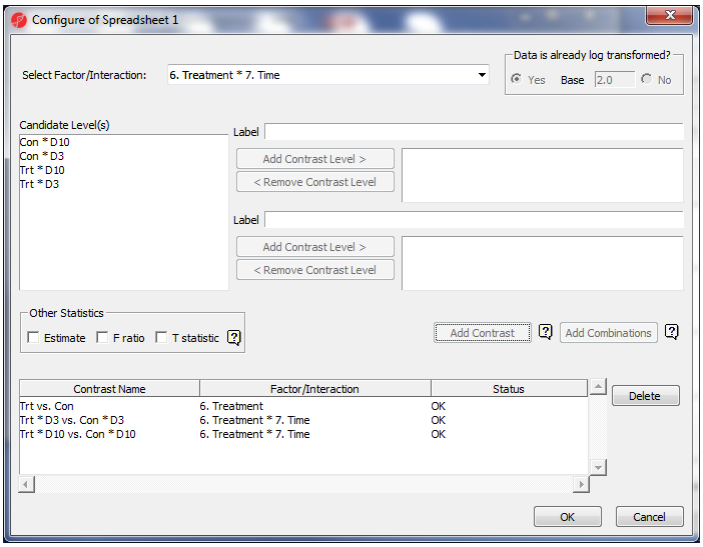
- Click **Contrast**
- Add contrast of

  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:_____

Copyright © 2013 Partek Incorporated. All rights reserved.

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

Treatment\Time	D10	D3	Total
Con	3	3	6
Trt	3	3	6
Total	6	6	12

Notes:_____

_____

_____

_____

_____

_____

Copyright © 2013 Partek Incorporated. All rights reserved.

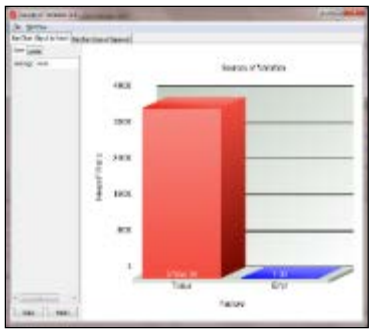
Result of ANOVA

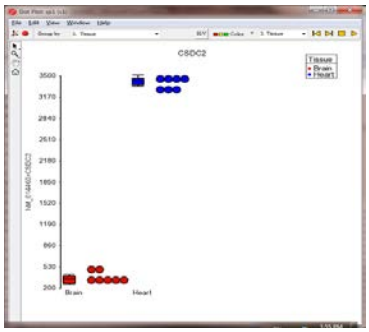
Each row is a gene with its p-value and fold change and any other statistical information. The spreadsheet is sorted by the first p-value column.

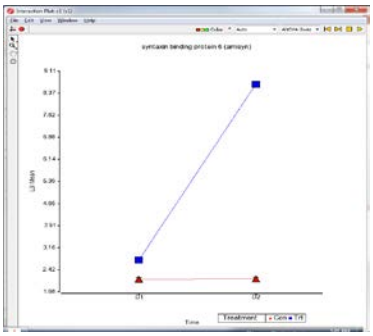
Right click on a row header to get details

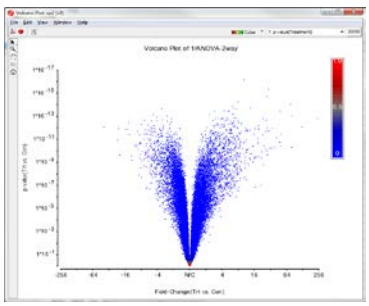
- Select **HML Report**
- Select **Dot Plot**
- Select **Source of Variation**
- Select **ANOVA Interaction Plot**
- Select **View>Volcano Plot**

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









Notes:_____

_____

_____

_____

_____

_____

Copyright © 2013 Partek Incorporated. All rights reserved.

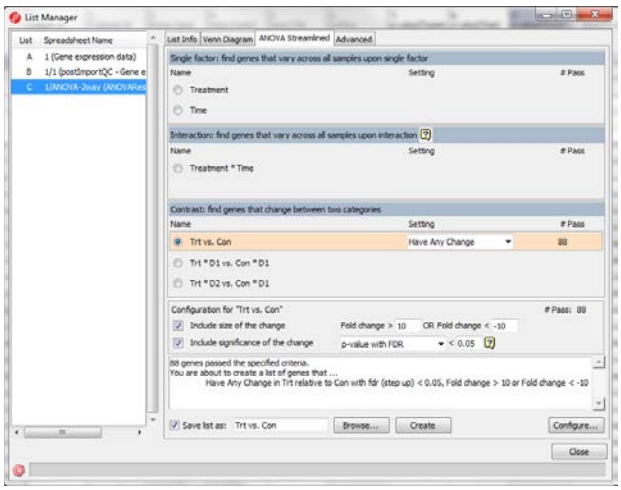
Create List

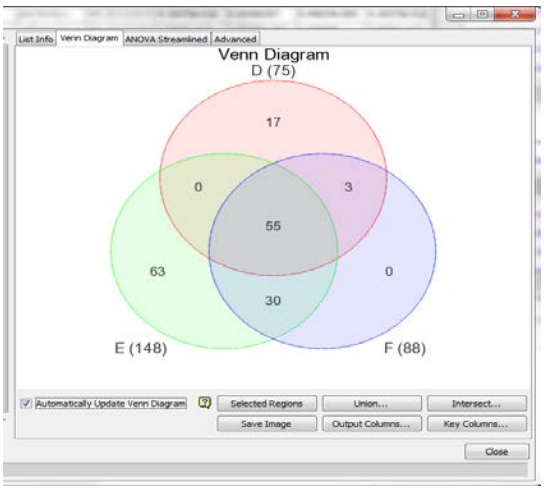
Generate a list of miRNA that is showing differential expression between brain and heart.

- 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:_____

_____

_____

_____

_____

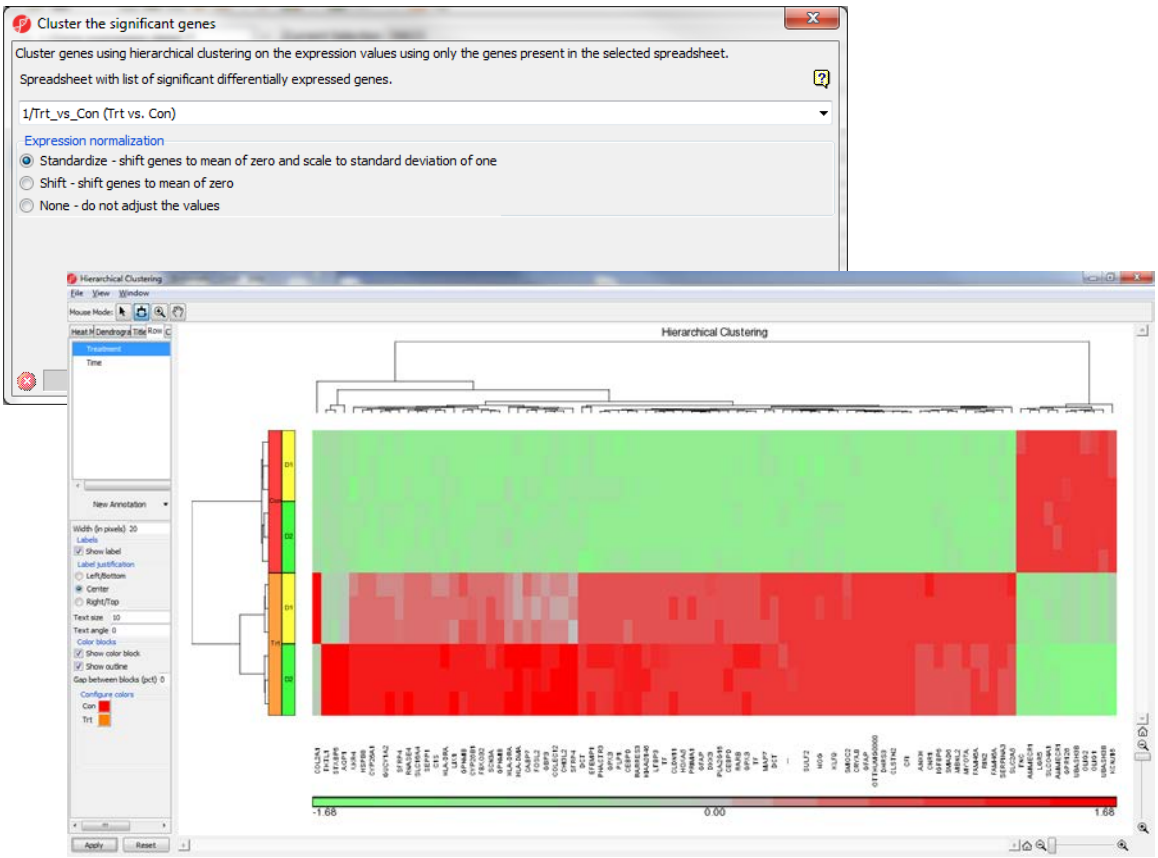
_____

Copyright © 2013 Partek Incorporated. All rights reserved.

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:_____

Copyright © 2013 Partek Incorporated. All rights reserved.

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

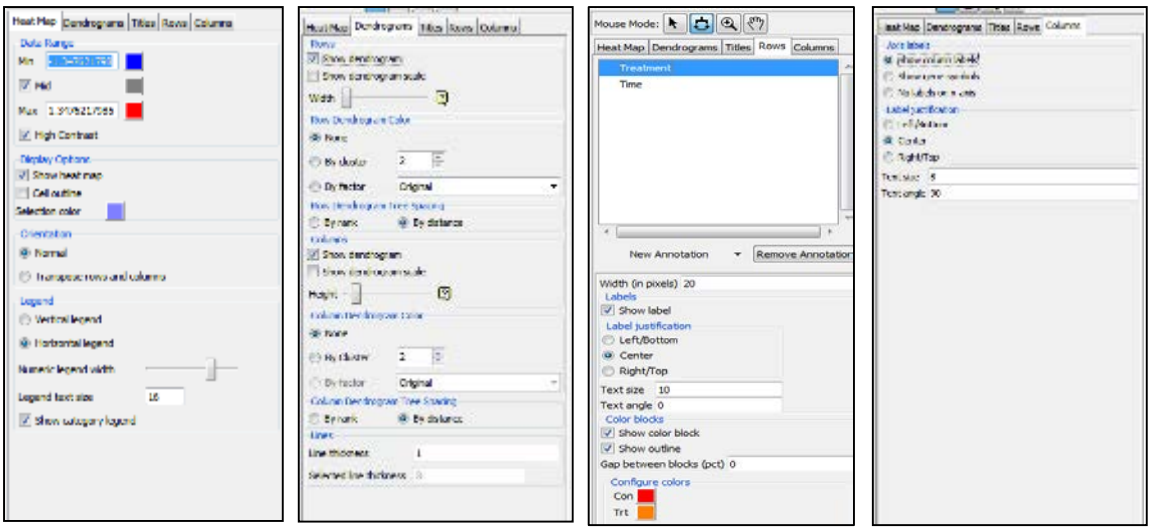
Rows

- 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:_____

Copyright © 2013 Partek Incorporated. All rights reserved.

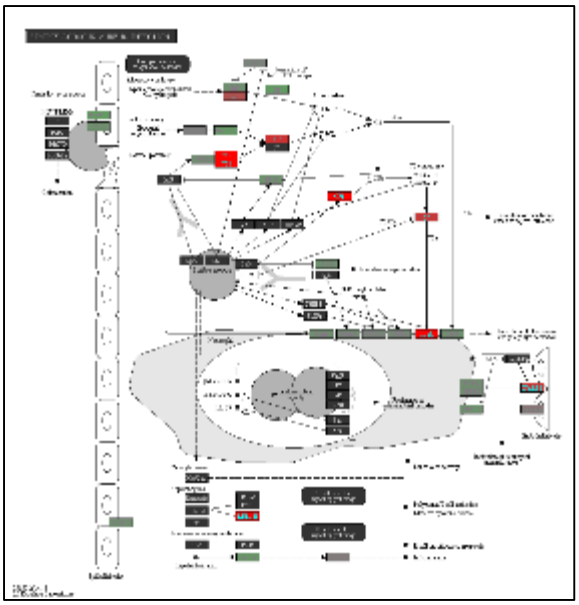
Biological Interpretation—Pathway Enrichment

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



Notes:

Copyright © 2013 Partek Incorporated. All rights reserved.

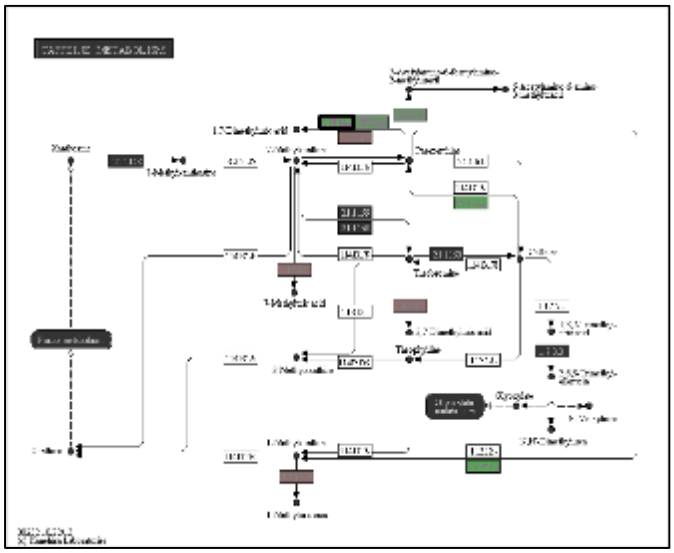
Biological Interpretation—Pathway ANOVA

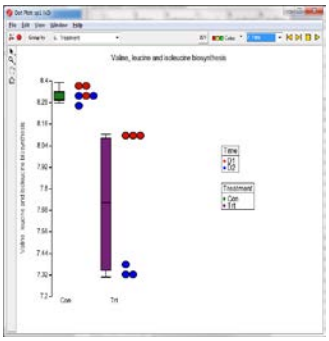
Pathway ANOVA: Detect differentially expressed pathway

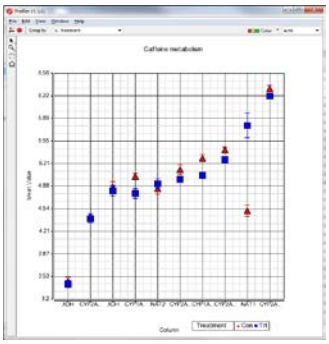
- Select *Gene expression data* spreadsheet
- Select **Pathway analysis> Partek Pathway> Pathway ANOVA**
- Change *Restrict analysis to pathways with fewer than* **100** genes to save time

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:_____

_____

_____

_____

_____

_____

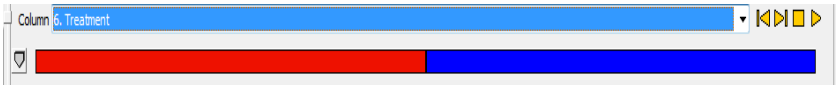
Copyright © 2013 Partek Incorporated. All rights reserved.

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

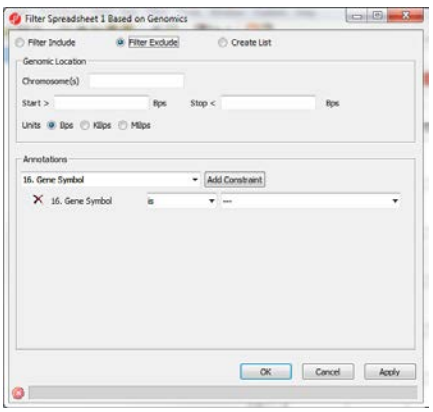
- Click **Filter>Filter Column> Column Filter Manager>Filter based on Max < cutoff**

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,**

1. Column #	2. Probeset ID	3. Entrez Gene	4. Gene Symbol	5. Gene Title	6. RefSeq Transcript ID	7. p-value(Treatment)	8. p-value(Time)	9. p-value(Treatment * Time)	10. p-value(Trt * D10 vs. Con * D10)	11. Ratio(Trt * D10 vs. Con * D10)	12. Fold-Change(Trt * D10 vs. Con * D10)	13. Fold-Change(Trt * D10 vs. Con * D10) (Description)	14. T(Trt * D10 vs Con * D10)
38992	229729_at	51754	TMEM8B	transmembrane	NM_001042589	0.950793	0.912729	0.950759	0.999976	0.999998	-1	No change	-3.11819e-005
54452	48659_at	60672	MIIP	migration and	NM_001025374	0.00278673	0.13946	0.00278609	0.999907	1	1	No change	0.000120175
40678	231415_at	114049	WBSCR22	Williams Beuren	NM_001202560	0.433187	0.529872	0.433314	0.99987	0.999987	-1.00001	Trt * D10 down	-0.000167871
41763	232500_at	57186	RALGAPA2	Ral GTPase	NM_020343	0.147737	0.0513168	0.147676	0.999849	0.999981	-1.00002	Trt * D10 down	-0.000194731
6555	1562352_at	---	OTTHUMG0000	NULL /// NULL	---	0.995768	0.402719	0.99601	0.999829	0.999984	-1.00002	Trt * D10 down	-0.000220599
53537	244278_at	---	OTTHUMG0000	NULL /// NULL	---	0.377193	0.0284717	0.377372	0.999798	0.999981	-1.00002	Trt * D10 down	-0.000261593
50327	241069_at	---	---	---	---	0.529504	0.314524	0.529275	0.999795	1.00003	1.00003	Trt * D10 up vs	0.000265193
51883	242625_at	91543	RSAD2	radical	NM_080657	0.793957	0.422403	0.793649	0.999773	0.999981	-1.00002	Trt * D10 down	-0.000293268
13990	204534_at	7448 /// 645832	SEBOX /// VTN	SEBOX	NM_000638 ///	0.399002	0.784391	0.398742	0.999719	1.00002	1.00002	Trt * D10 up vs	0.000362979
41514	232251_at	152195	NUDT16P1	nudix	NM_194289 ///	0.645337	0.156583	0.644896	0.999646	1.00004	1.00004	Trt * D10 up vs	0.000457128
3571	1557311_at	---	---	---	---	0.389896	0.176269	0.389556	0.999626	1.00003	1.00003	Trt * D10 up vs	0.000484005
45545	236287_at	---	---	---	---	0.232624	0.14951	0.232364	0.999569	1.00004	1.00004	Trt * D10 up vs	0.00055723
5767	1561096_at	285419	LOC285419	uncharacterized	NR_027105 ///	0.193401	0.223396	0.193145	0.999502	0.99997	-1.00003	Trt * D10 down	-0.000643604
8942	1568906_at	728196	LOC728196	uncharacterized	XR_040780	0.0591978	0.997538	0.059109	0.999474	1.00002	1.00002	Trt * D10 up vs	0.000680539
16880	207429_at	6582	SLC22A2	solute carrier	NM_003058 ///	0.560648	0.35898	0.559946	0.999392	1.00005	1.00005	Trt * D10 up vs	0.000786024
7569	1564294_at	---	---	---	---	0.119095	0.206348	0.118881	0.99935	0.999952	-1.00005	Trt * D10 down	-0.000840022
12486	203028_s_at	1535	CYBA	cytochrome	NM_000101	0.376107	0.0988519	0.375511	0.999327	1.00007	1.00007	Trt * D10 up vs	0.000870332
43921	234663_at	---	---	---	---	0.640129	0.0618251	0.639282	0.999317	1.00006	1.00006	Trt * D10 up vs	0.000883042
9688	1570173_at	25896	INTS7	integrator	NM_001199809	0.434894	0.782388	0.435655	0.999226	0.999888	-1.00011	Trt * D10 down	-0.00100068
37640	228377_at	57565	KLHL14	kelch-like family	NM_020805	0.877886	0.0459619	0.876764	0.999196	1.00009	1.00009	Trt * D10 up vs	0.00104009

Notes:_____

_____

_____

_____

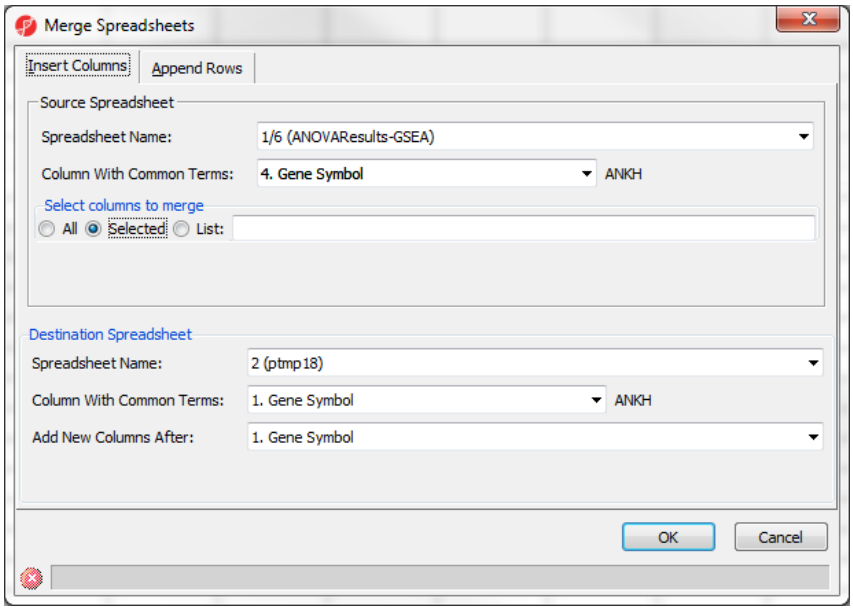
_____

_____

Copyright © 2013 Partek Incorporated. All rights reserved.

Output gene list for GSEA analysis - Continue

- Right click on **Gene symbol** column > **Create list with occurrence counts**, a new spreadsheet is generated
- Choose **File>Merge Spreadsheet**
  - Select **T(Trt * D10 vs. Con * D10)** column on **ANOVAResults-GSEA** spreadsheet
  - Source: ANOVAResults-GSEA; Destination: new spreadsheet
  - Column with common terms: gene symbol
  - Merge selected columns
- 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**



Notes:_____

_____

_____

_____

_____

_____

Copyright © 2013 Partek Incorporated. All rights reserved.

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:_____

_____

_____

_____

_____

_____

Copyright © 2013 Partek Incorporated. All rights reserved.