# Gene Expression Data Analysis in Partek® Genomics Suite®

HANDS-ON TRAINING

**NCI Workshop**

April 26th, 2017


**Eric Seiser, PhD**

*Field Application Scientist*

*Partek Incorporated*

*support@partek.com*

# Contents





Copyright © 2017 Partek Incorporated. All rights reserved.

# Partek® Genomics 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

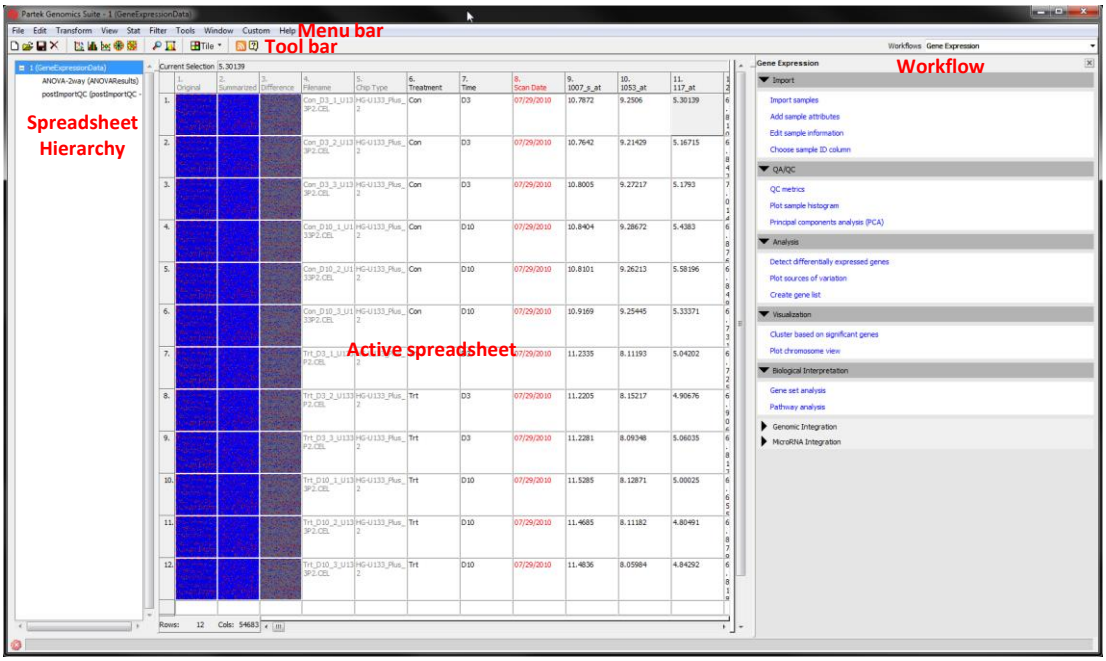
**Tool bar:** Accelerator buttons allow quick access to commonly used commands

**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:**

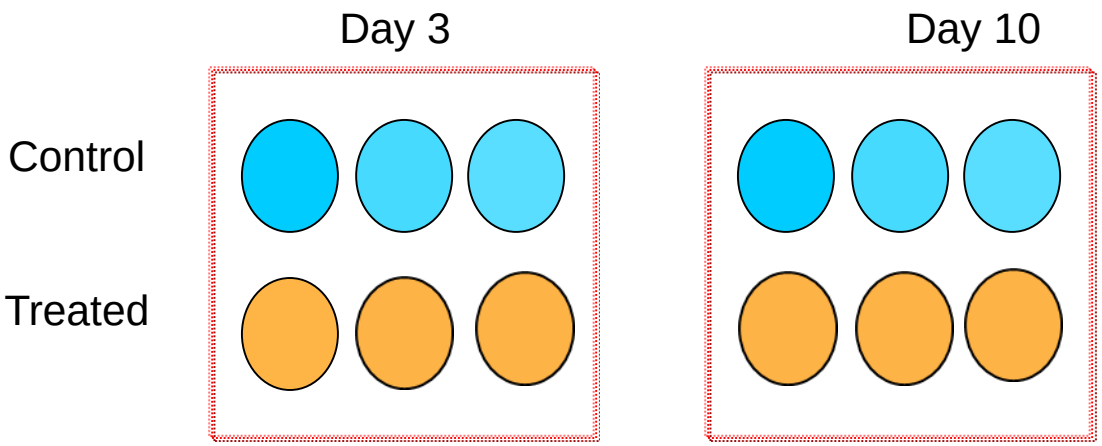


Copyright © 2017 Partek Incorporated. All rights reserved.

# Training Data

**Data files in the project:**

- Download the training dataset
  - https://s3.amazonaws.com/PartekLibraryFiles/training_data/GX_training_data.zip
- Glioma stem cells (GSC) cultured in control or differentiation media
- 12 samples with 2 treatment on two time points
  - Control and Treated
  - 3 and 10 days
- Affymetrix HG-U133_Plus_2 array



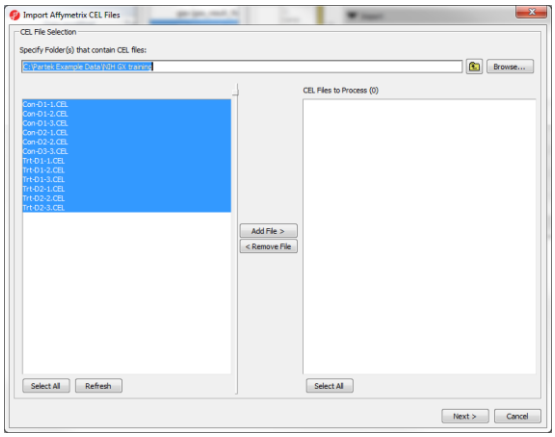

**Notes:**

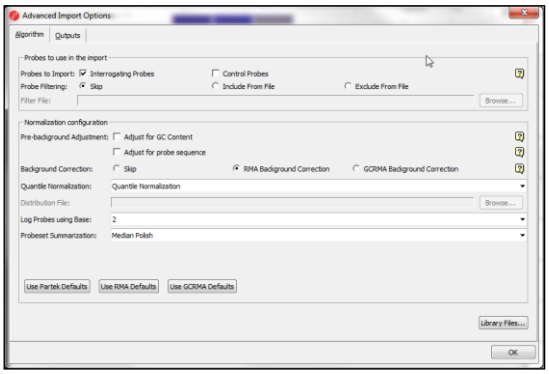


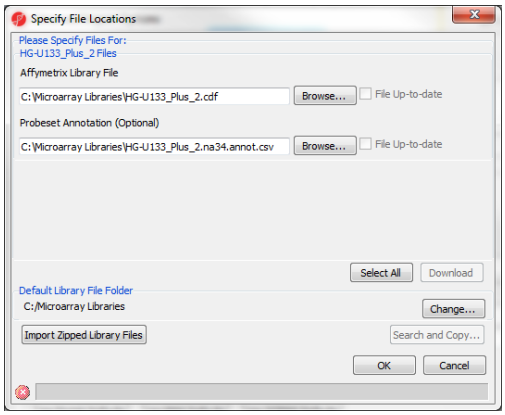
Copyright © 2017 Partek Incorporated. All rights reserved.

# 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**
- **Customize** allows you change the algorithm parameters, and verify library files
- Partek Genomics Suite will automatically download the Affymetrix library file




**Notes:**_____



Copyright © 2017 Partek Incorporated. All rights reserved.

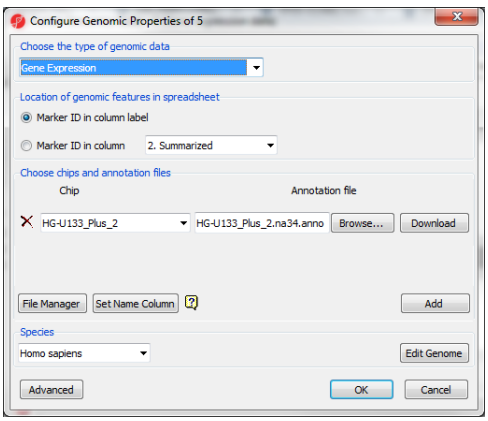
# Spreadsheet Properties

### Spreadsheet

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


### Annotation

- To link an annotation to a spreadsheet, select: File > Properties
- If importing directly from .CEL file, resulting spreadsheet is associated with annotation
- Microarray Libraries: Stores automatically downloaded annotations files




**Notes:**



Copyright © 2017 Partek Incorporated. All rights reserved.

# Add Sample Attributes

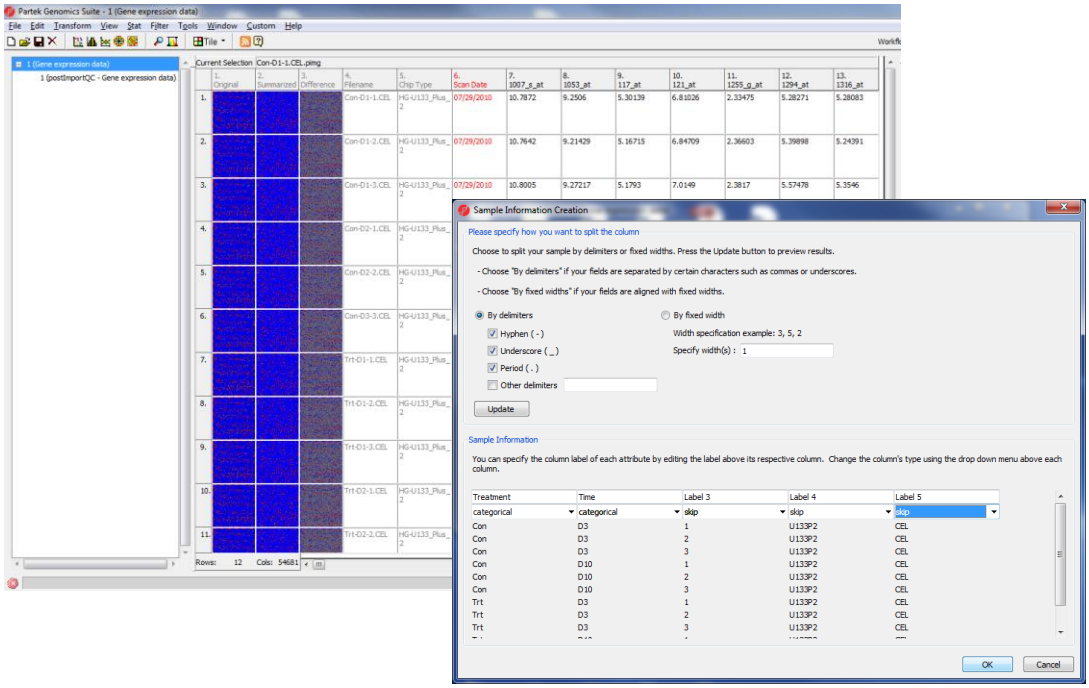
**Two spreadsheets are generated:**

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 © 2017 Partek Incorporated. All rights reserved.

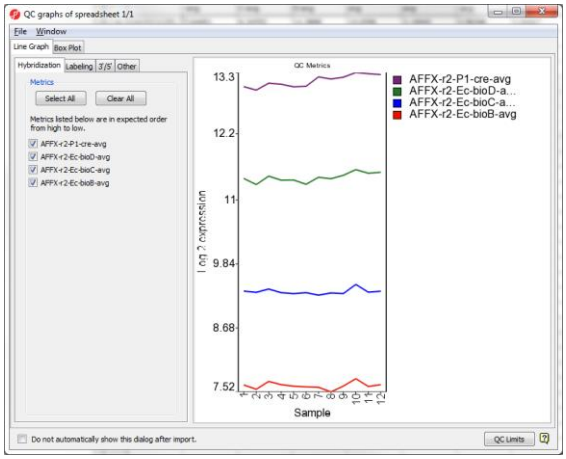
# QA and QC

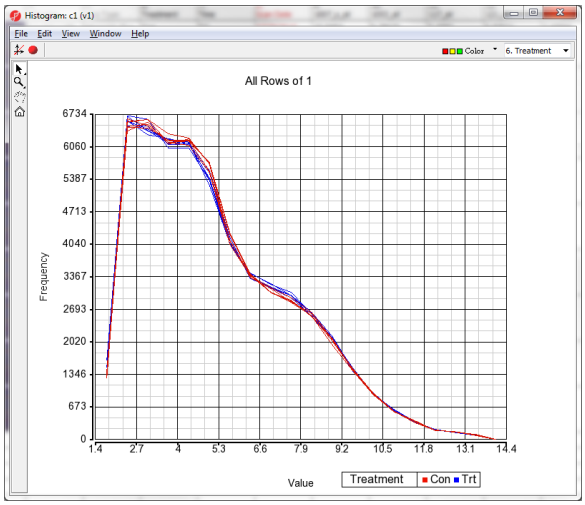
QA/QC is an exploratory analysis, assessing the preparation of the samples and can be used to identify outliers

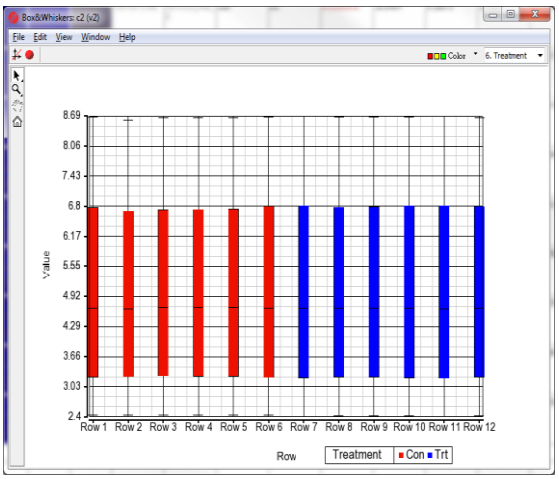
QC metrics are only available when you import Affymetrix .CEL files, it checks the quality of the chips based on control probesets

Histogram display the distribution of the samples

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




**Notes:**_____



Copyright © 2017 Partek Incorporated. All rights reserved.

# PCA Scatter Plot

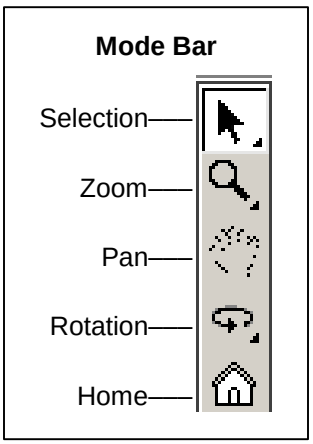
The principal components analysis (PCA) scatter plot is another way to assess relatedness between samples and identify outliers

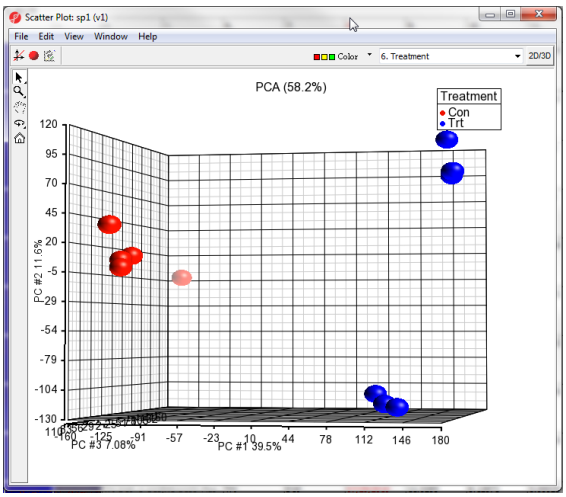
- Select **Principal components analysis** on the workflow

**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:**_____

_____

_____

_____

_____

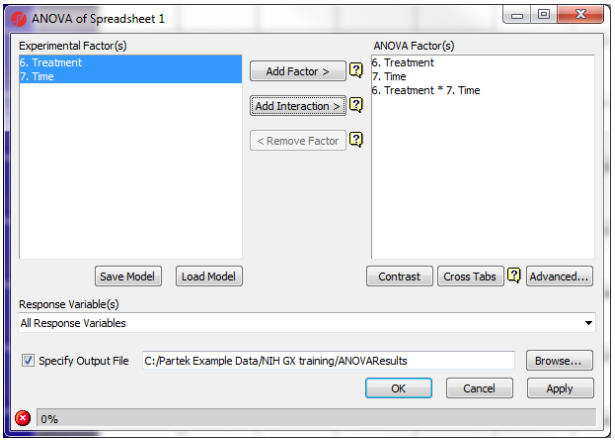
_____



Copyright © 2017 Partek Incorporated. All rights reserved.

# Detect Differentially Expressed Genes

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



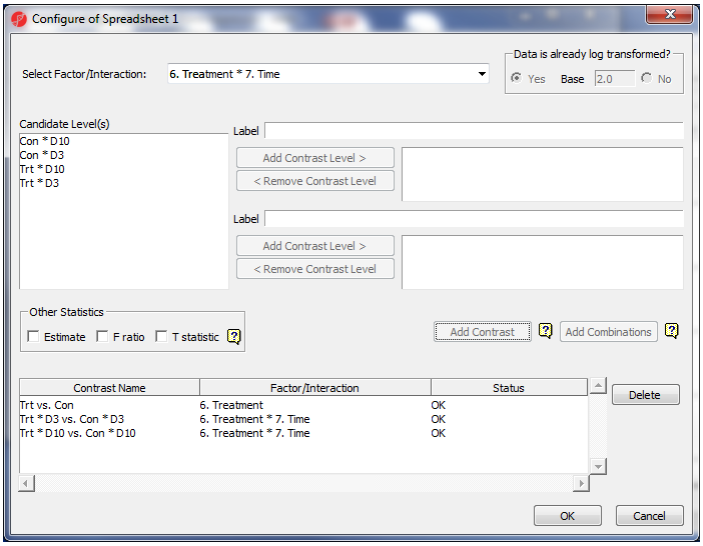

- Click **Contrast**
- Add contrast of

  **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 © 2017 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

Partek Genomics Suite automatically detects crossed/nested factors and automatically performs mixed model ANOVA when a random effect is 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 © 2017 Partek Incorporated. All rights reserved.

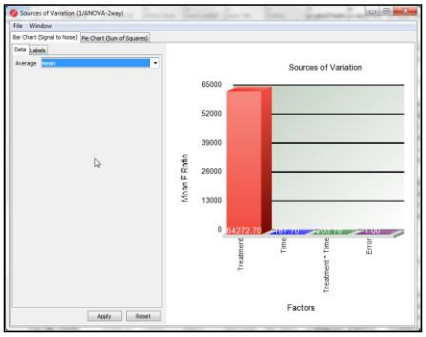
# ANOVA Results

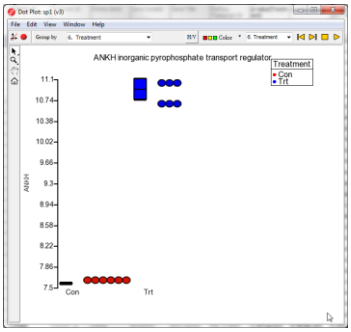
Each row is a gene with its p-value and fold change and any other statistical information. The spreadsheet is sorted by the first p-value column.

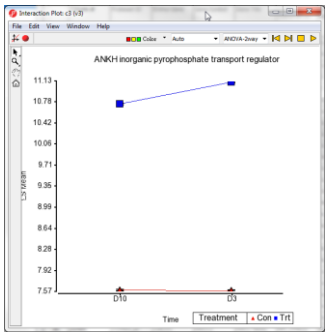
**Right click on a row header to get details on each gene**

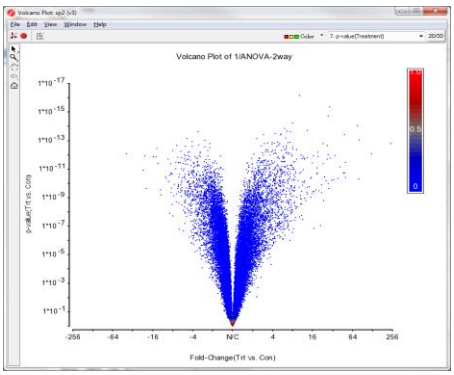
- Select **HTML Report**
- Select **Dot Plot**
- Select **Source of Variation**
- Select **ANOVA Interaction Plot**
- Select **View>Volcano Plot**

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




**Notes:**_____



Copyright © 2017 Partek Incorporated. All rights reserved.

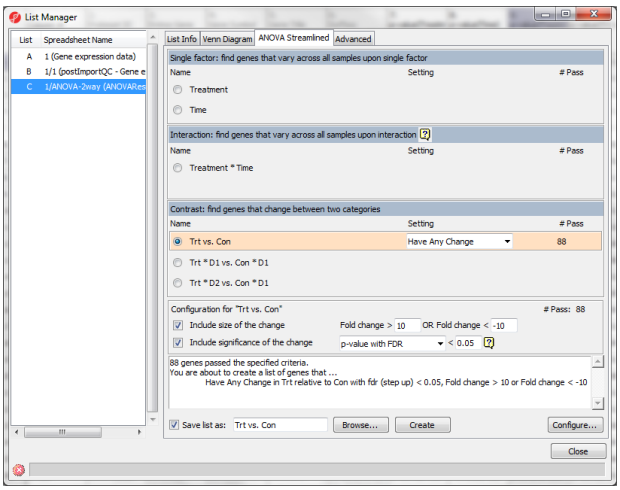
# Create Gene List

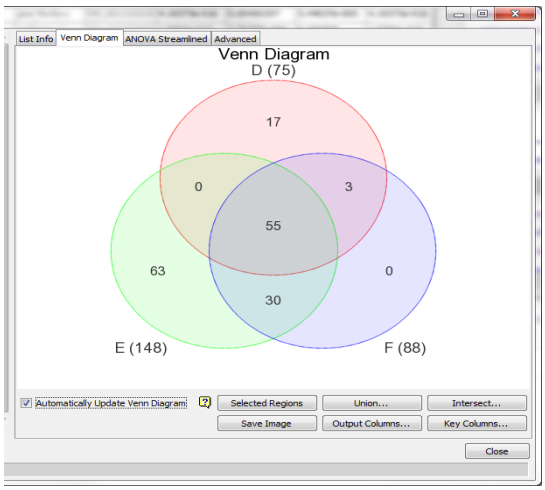
Generate a list of genes that is showing differential expression between treated and untreated cells.

- 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 and select the three gene lists—PGS allow 5 way Venn diagram
- Select any section in Venn diagram to generate a new gene list




**Notes:**_____

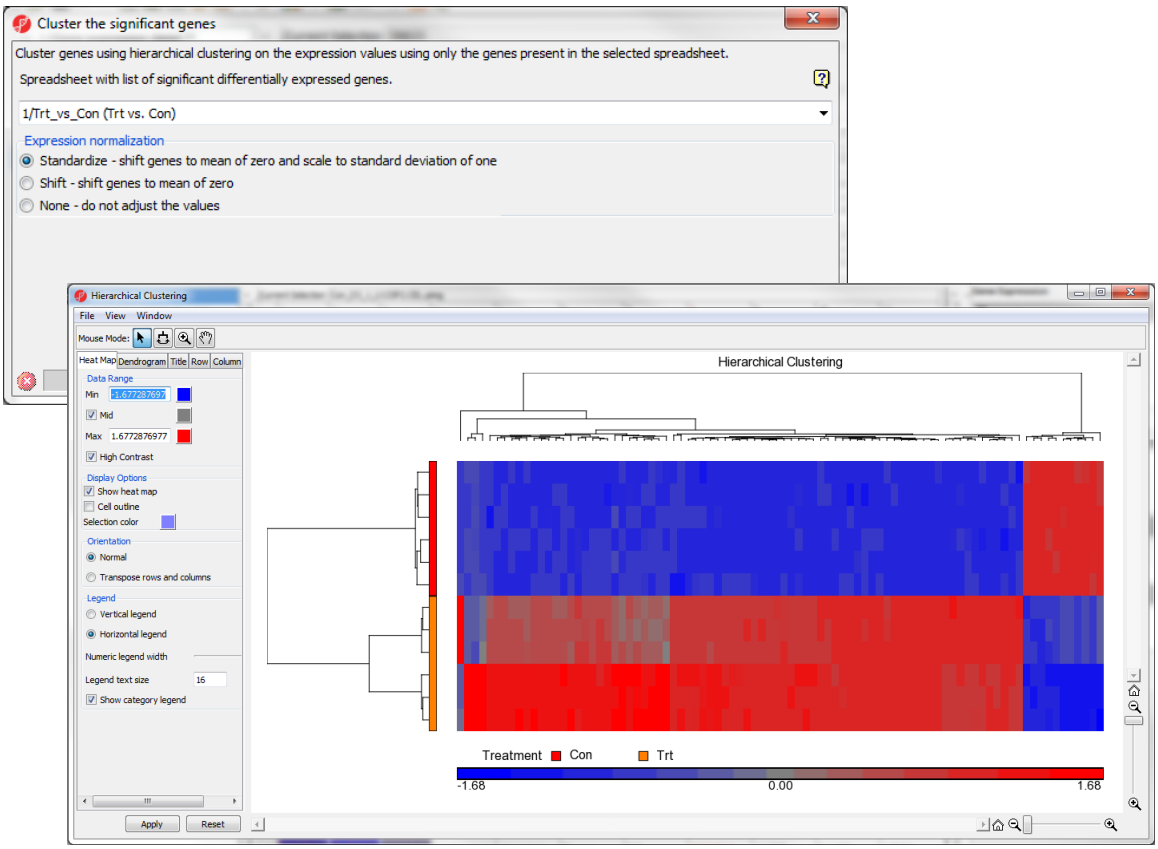


Copyright © 2017 Partek Incorporated. All rights reserved.

# Hierarchical Clustering

To visualize the heatmap and cluster a significant list of genes:

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




**Notes:**_____



Copyright © 2017 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

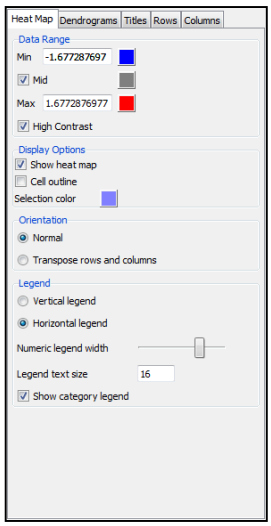
### Rows

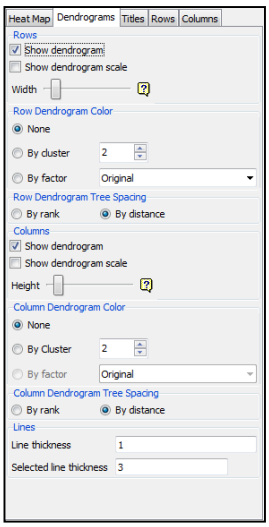
- Change the width of annotation
- Check show label
- Change color
- Add new annotation

### Columns

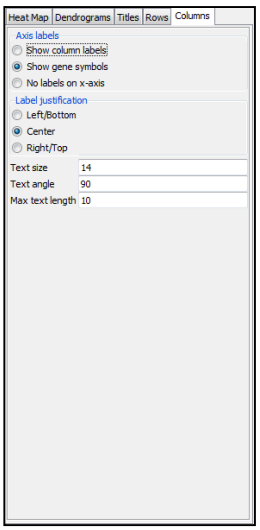
- Column header or mRNA name/gene symbol

**Modes:** mouse over, select, zoom, and flip








**Notes:**



Copyright © 2017 Partek Incorporated. All rights reserved.

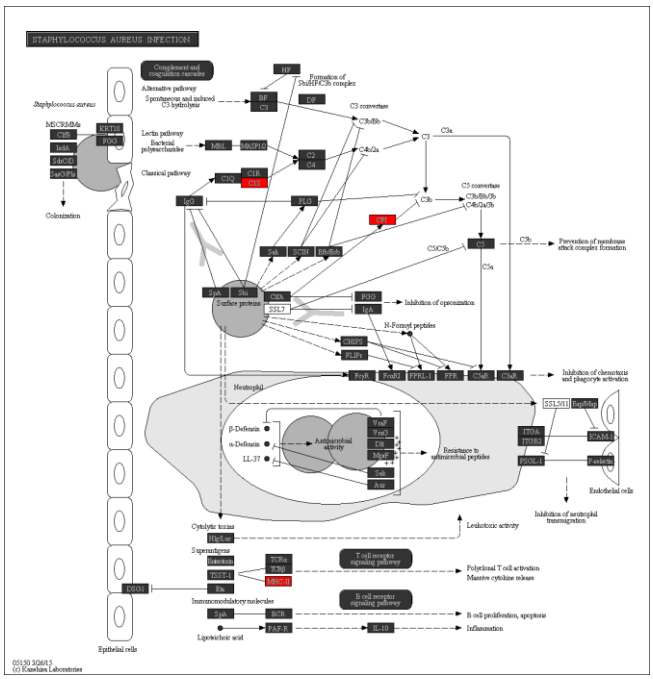
# Biological Interpretation—Pathway Enrichment

**Pathway enrichment:** Tests if filtered genes are overrepresented in any pathway

- Select **Treated vs Control** gene list spreadsheet
- Select **Pathway analysis> 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 © 2017 Partek Incorporated. All rights reserved.

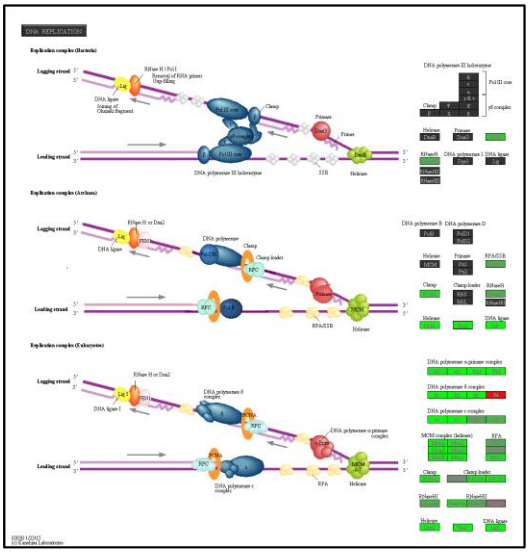
# Biological Interpretation—Pathway ANOVA

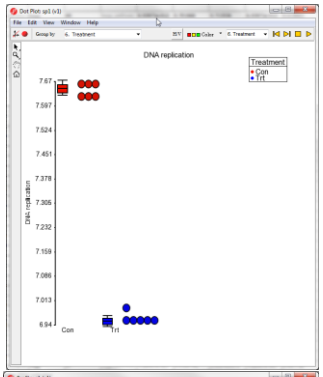
**Pathway ANOVA:** Detect differentially expressed pathway

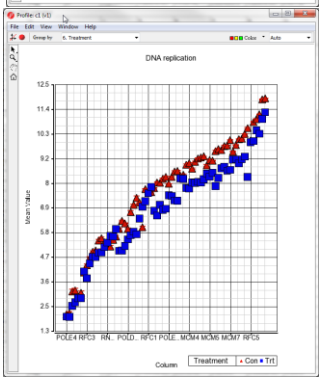
- Select *Gene expression data* spreadsheet
- Select **Pathway analysis> Pathway ANOVA**
- Change *Restrict analysis to pathways with fewer than* **100** genes to save time

Pathway ANOVA result spreadsheets:

- Analysis generates two spreadsheets: 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 © 2017 Partek Incorporated. All rights reserved.

# Advanced Analyses

**Detailed tutorials and data for the following advanced analysis demos can be found at the following location: http://www.partek.com/pgs-resources-microarray (TUTORIALS tab):**

**Batch correction:**

Tutorial:
http://www.partek.com/Tutorials/microarray/Gene_Expression/Breast_Cancer/Breast_Cancer_tutorial-GE.pdf

Data:
http://www.partek.com/Tutorials/microarray/Gene_Expression/Breast_Cancer/Breast_Cancer-GE.zip

**Survival analysis:**

Tutorial: http://www.partek.com/Tutorials/microarray/Survival_Analysis/Survival_Analysis.pdf

Data: http://www.partek.com/Tutorials/microarray/Survival_Analysis/Survival.zip

**Integration of genomic data:**

Tutorial: http://www.partek.com/Tutorials/microarray/microRNA/miRNA_tutorial.pdf

Data: http://www.partek.com/Tutorials/microarray/microRNA/miRNA_tutorial_data.zip

**Notes:**



Copyright © 2017 Partek Incorporated. All rights reserved.

# Independent Analysis

**The goal of this session is to obtain published microarray data from the Gene Expression Omnibus (GEO) and run independent analysis using the Gene Expression workflow. A list of goals will be provided as a point of reference for the analysis.**

Sci Rep. 2015 Sep 24;5:14273. doi: 10.1038/srep14273.

## Opposite Effects of M1 and M2 Macrophage Subtypes on Lung Cancer Progression.

Yuan A[1], Hsiao YJ[2], Chen HY[3], Chen HW[4], Ho CC[5], Chen YY[4], Liu YC[1], Hong TH[6,7], Yu SL[2,8,9,10], Chen JJ[11,12], Yang PC[5].

Author information

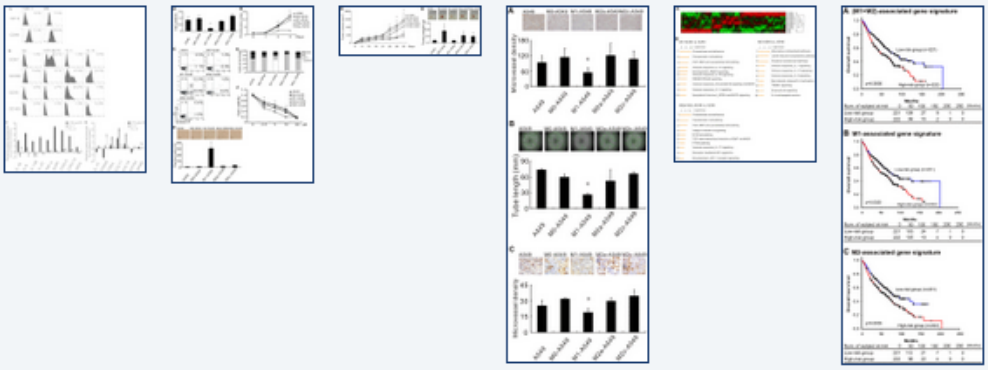
**Abstract**

Macrophages in a tumor microenvironment have been characterized as M1- and M2-polarized subtypes. Here, we discovered the different macrophages' impacts on lung cancer cell A549. The M2a/M2c subtypes promoted A549 invasion and xenograft tumor growth. The M1 subtype suppressed angiogenesis. M1 enhanced the sensitivity of A549 to cisplatin and decreased the tube formation activity and cell viability of A549 cells by inducing apoptosis and senescence. Different macrophage subtypes regulated genes involved in the immune response, cytoskeletal remodeling, coagulation, cell adhesion, and apoptosis pathways in A549 cells, which was a pattern that correlated with the altered behaviors of the A549 cells. Furthermore, we found that the identified M1/M2 gene signatures were significantly correlated with the extended overall survival of lung cancer patients. These results suggest that M1/M2 gene expression signature may be used as a prognostic indicator for lung cancer patients, and M1/M2 polarization may be a target of investigation of immune-modulating therapies for lung cancer in the future.

PMID: 26399191 PMCID: PMC4585843 DOI: 10.1038/srep14273

[Indexed for MEDLINE] Free PMC Article

Images from this publication. See all images (6) Free text



**Notes:**



Copyright © 2017 Partek Incorporated. All rights reserved.

# Independent Analysis Goals

**1.Download raw CEL files from GEO**

**2.Extract data and import into genomics suite**

**3.Add sample attributes**

**4.Explore the data using PCA**

**5.Identify differentially expressed genes between control and macrophage co-culture**

**6.Identify differentially expressed genes between control and each macrophage subtype**

**7.Plot expression for a significant gene**

**8.Create lists of significant genes**

**9.Use a venn diagram to look at overlap between macrophage subtypes and create a list**

**10.Perform hierarchical clustering on a significant gene list, overlaying control and subtype information on the plot**

**11.Perform GO and Pathway enrichment on a significant gene list**

**Optional: Attempt to replicate the results of the study following the methodology as closely as possible (differences in analysis may lead to differences in results)**

**Notes:**



Copyright © 2017 Partek Incorporated. All rights reserved.

# Further Training

**Self-learning**

- Help > Check for Updates
- Help > On-line tutorials
- Recorded webinars available on Partek Incorporated's YouTube page

**Regional Technical Support**

- Email: support@partek.com
- Phone: +1-314-878-2329

**Notes:**



Copyright © 2017 Partek Incorporated. All rights reserved.