Gene Expression Data Analysis in Partek® Genomics Suite®

HANDS-ON TRAINING

NCI Workshop

December 13th, 2017

Eric Seiser, PhD
Field Application Scientist
Partek Incorporated
support@partek.com

Contents

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

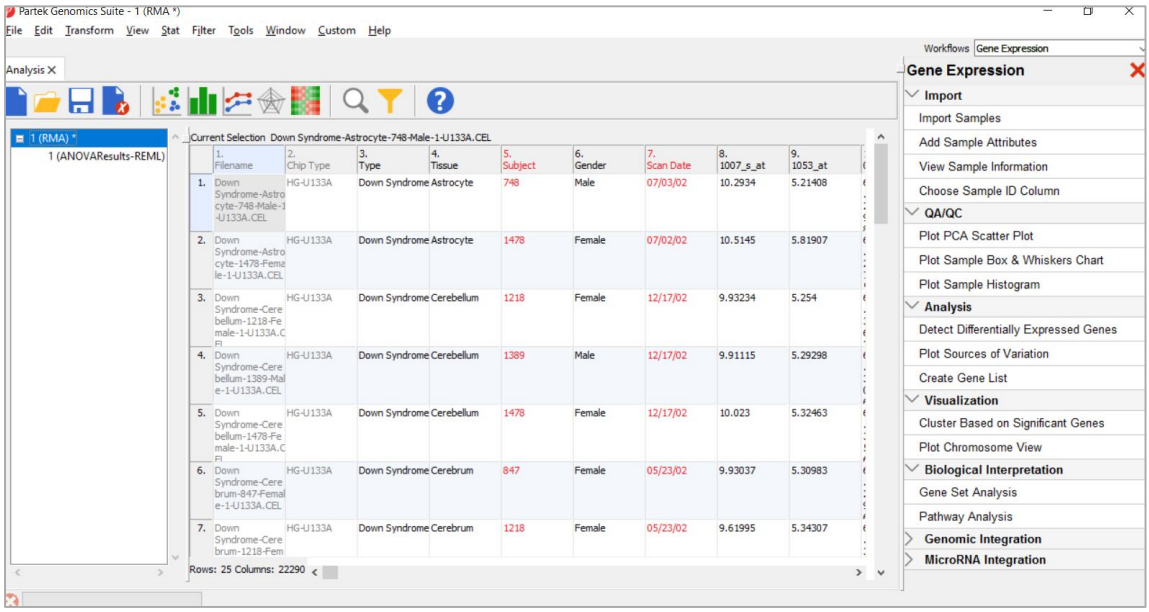
Tool bar: Accelerator buttons allow quick access to commonly used commands

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

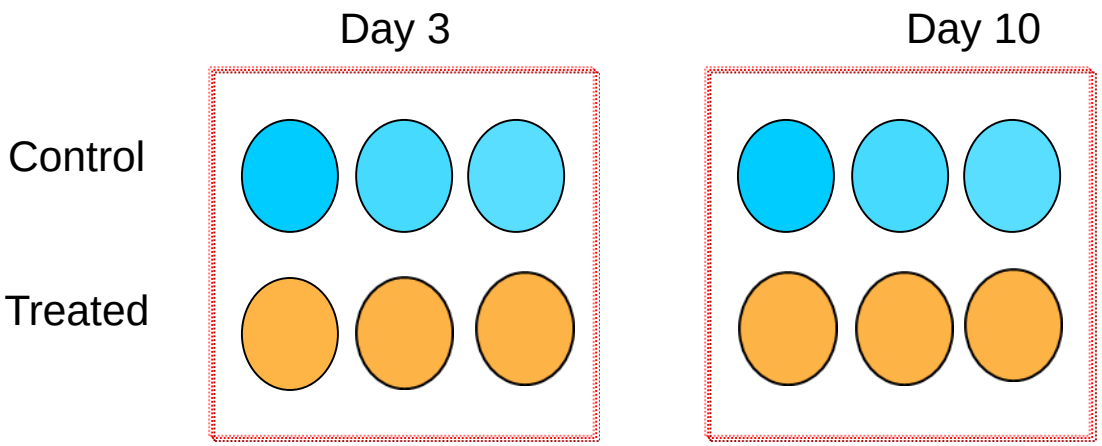


Notes:_____

Training Data

Data files in the project:

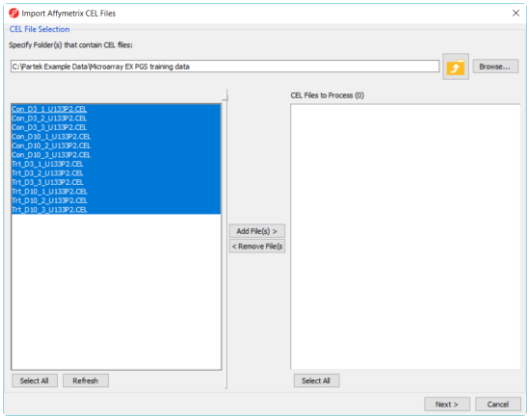
- Download the training dataset
  - https://s3.amazonaws.com/PartekLibraryFiles/training_data/GX_training_data.zip
- Glioma stem cells (GSC) cultured in control or differentiation media
- 12 samples with 2 treatment on two time points
  - Control and Treated
  - 3 and 10 days
- Affymetrix HG-U133_Plus_2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

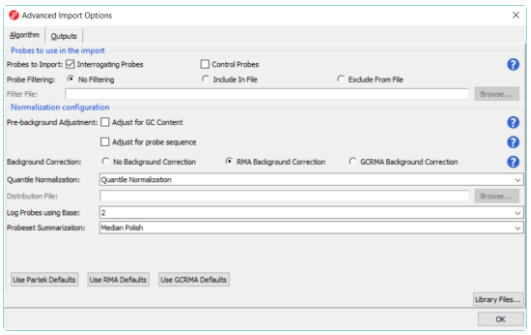


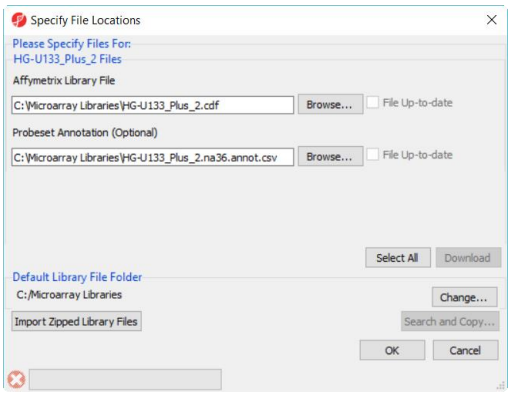
Notes:

Importing Data from Affymetrix® CEL Files

- Choose **Gene Expression** workflow
- Click on **Import Samples** and select **Import from Affymetrix CEL Files** option
- Browse to the folder that contains the CEL files
- Select all the default CEL files, and add 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.





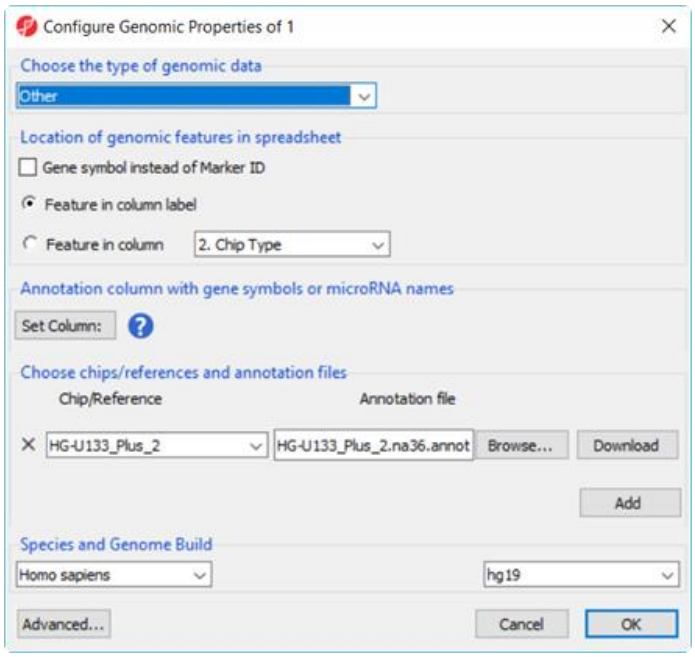


Notes:_____

Spreadsheet Properties

Analysis tab: contains tabular format of the data

- Each row is a sample
- Each column is a probe set ID with RMA normalized intensity value
- Annotation of the probe set is linked, to add/edit annotation, choose **File>Properties**
  - Gene symbol field is required for biological interpretation
  - Species information is required for biological interpretation



Notes:_____

Add Sample Attributes

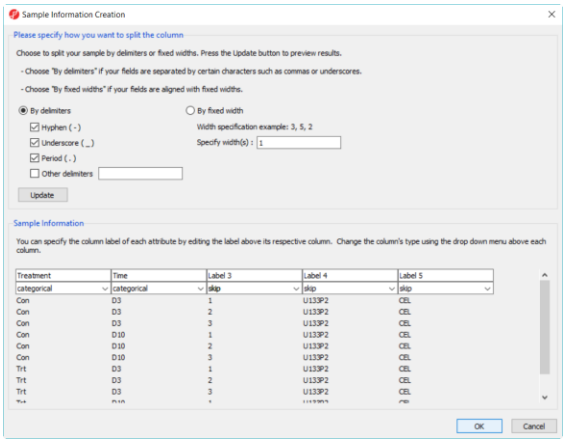
Select **Add Sample Attributes** on the workflow

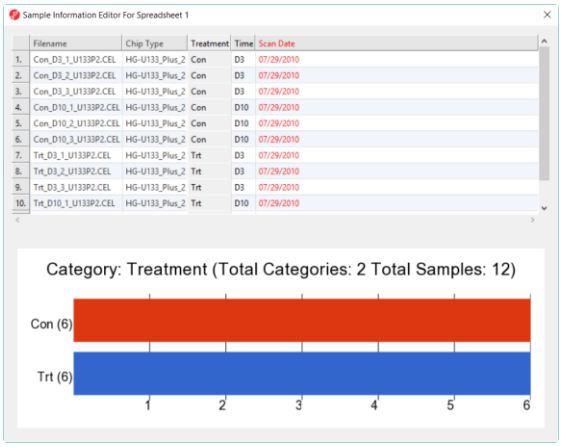
- Choose **Add attributes from an existing column**
- Specify **Treatment** and **Time** on the first 2 columns respectively, and skip the rest columns
- Click **OK** and **Save** the spreadsheet
- Another way to specify same attribute is to add one categorical attribute at a time

Select **View Sample Information**

- Click on each categorical column to view the histogram of the subgroups

Choose Sample ID Column: default is the file name, unique ID of each sample





Notes:_____

QA/QC - PCA Scatter Plot

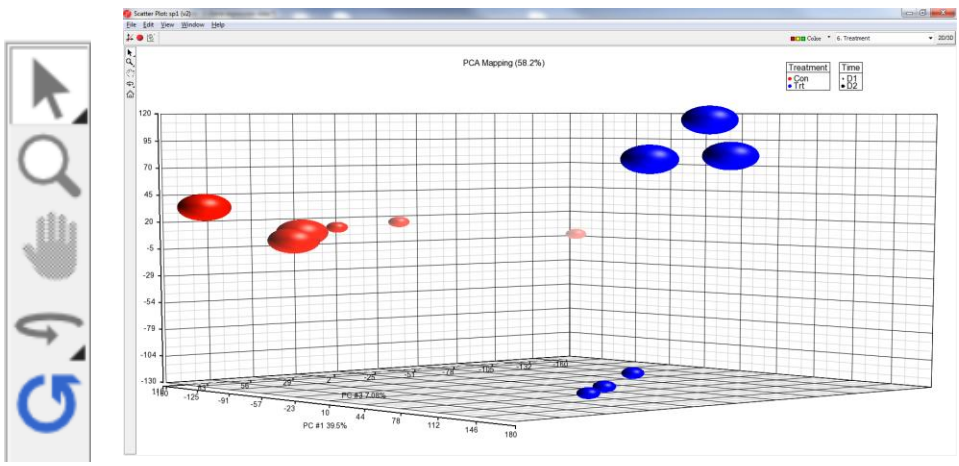
PCA scatter plot is one way to identify clustering patterns and outliers

- Go to the *QA/QC* setion of the workflow > **Plot PCA Scatter Plot**

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

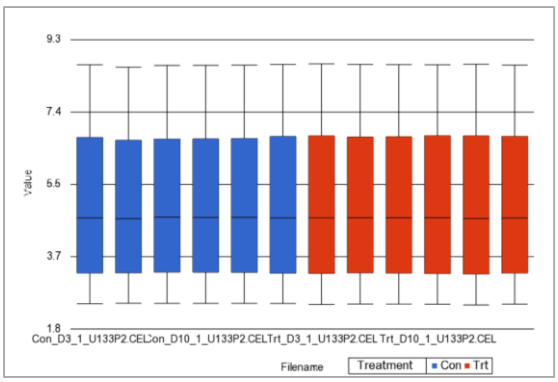
- Click on **Plot Properties** to configure color by **Treatment**, size by **Time**
- Click on **Ellipsoid** to put the ellipsoid on each treatment type
- Select mode:
    - left click to select; scroll mouse wheel to zoom; drag mouse wheel to rotate
    - right click after select a point to filter/clear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

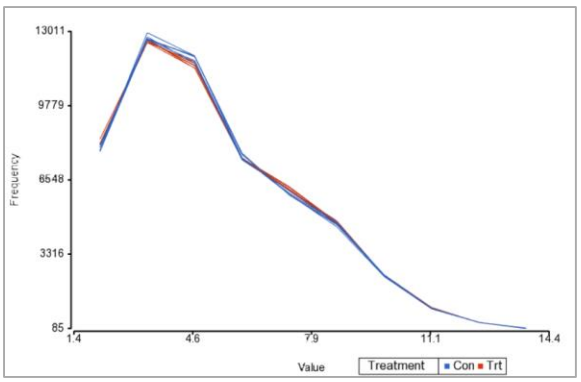


Notes:_____

QA/QC – Histogram and Box plot

- Select **Plot Sample Box & Whiskers Chart**
  - Each box is a sample
  - Line inside the box is the median (2nd quartile)
  - Box represent the first and third quartiles
  - Whiskers represent 10th percentile and 90the percentile by default, can be configured
- Select **Plot sample histogram**
  - Each line is a sample
  - X-axis is the range of the values
  - Default 20 bins on X-axis, can be configured from **Plot Properties**

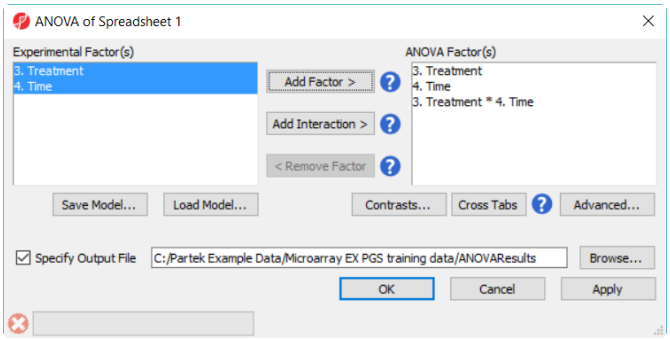




Notes:_____

Detect Differentially Expressed Genes

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



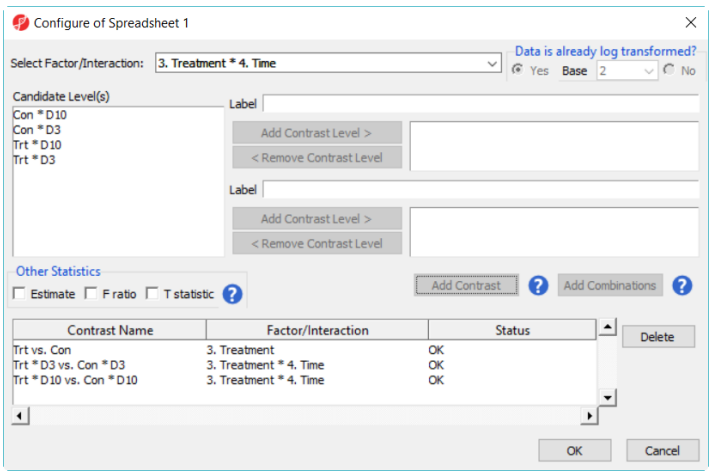
- Click **Contrast**
- Add contrast of

  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:_____

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

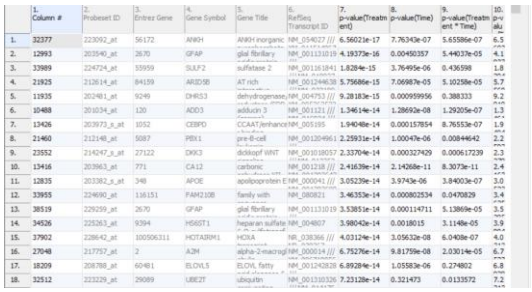
Notes:

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.

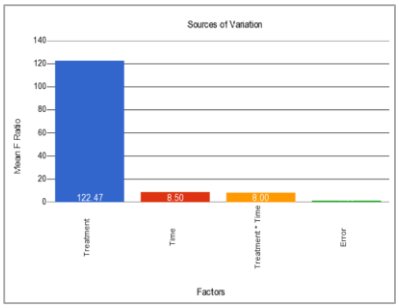
Right click on a row header to get details

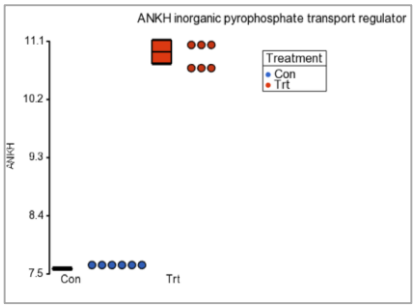
- Select **HML Report**
- Select **Dot Plot**
- Select **Source of Variation**
- Select **ANOVA Interaction Plot**
- Select **View>Volcano Plot**

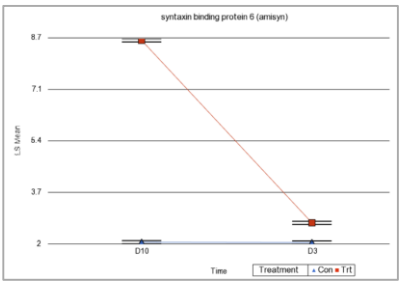


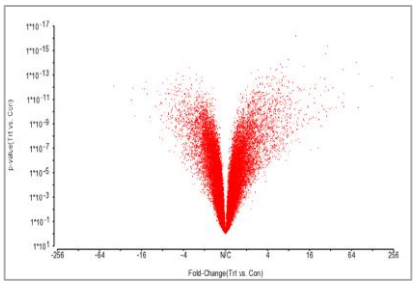
	1. Column #	2. Probeset ID	3. Entrez Gene	4. Gene Symbol	5. Gene Title	6. RefSeq Transcript ID	7. p-value(Treatment)	8. p-value(Time)	9. p-value(Treatment * Time)	10. p-value
1.	32377	223092_at	56172	ANKH	ANKH inorganic	NM_054027 ///	6.56021e-17	7.76343e-07	5.65586e-07	6.5
2.	12993	203540_at	2670	GFAP	glial fibrillary	NM_001131019	4.19373e-16	0.00450357	5.44037e-05	4.1
3.	33989	224724_at	55959	SULF2	sulfatase 2	NM_001161841	1.8284e-15	3.76495e-06	0.436598	1.8
4.	21925	212614_at	84159	ARID5B	AT rich	NM_001244638	5.75686e-15	7.06987e-05	5.10258e-05	5.7
5.	11935	202481_at	9249	DHRS3	dehydrogenase,	NM_004753 ///	9.28183e-15	0.000959956	0.388333	9.2
6.	10488	201034_at	120	ADD3	adducin 3	NM_001121 ///	1.34614e-14	1.28692e-08	1.29205e-07	1.3
7.	13426	203973_s_at	1052	CEBPD	CCAAT/enhancer	NM_005195	1.9404e-14	0.000157854	8.76553e-07	1.9
8.	21460	212148_at	5087	PBX1	pre-B-cell	NM_001204961	2.25931e-14	1.00047e-06	0.00844642	2.2
9.	23552	214247_s_at	27122	DKK3	dickkopf WNT	NM_001018057	2.33704e-14	0.000327429	0.000617239	2.3
10.	13416	203963_at	771	CA12	carbonic	NM_001218 ///	2.41639e-14	2.14268e-11	8.3073e-11	2.4
11.	12835	203382_s_at	348	APOE	apolipoprotein E	NM_000041 ///	3.05239e-14	3.9743e-06	3.84003e-07	3.0
12.	33955	224690_at	116151	FAM210B	family with	NM_080821	3.46353e-14	0.000802534	0.0470829	3.4
13.	38519	229259_at	2670	GFAP	glial fibrillary	NM_001131019	3.53851e-14	0.000114711	5.13869e-05	3.5
14.	34526	225263_at	9394	HS6ST1	heparan sulfate	NM_004807	3.98042e-14	0.0018015	3.1148e-05	3.9
15.	37902	228642_at	100506311	HOTAIRM1	HOXA	NR_038366 ///	4.03124e-14	3.05632e-08	6.0408e-07	4.0
16.	27048	217757_at	2	A2M	alpha-2-macrogl	NM_000014 ///	6.75276e-14	9.81759e-08	2.03014e-05	6.7
17.	18209	208788_at	60481	ELOVL5	ELOVL fatty	NM_001242828	6.89284e-14	1.05583e-06	0.274802	6.8
18.	32512	223229_at	29089	UBE2T	ubiquitin	NM_001303326	7.23128e-14	0.321473	0.0133572	7.2

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









Notes:_____

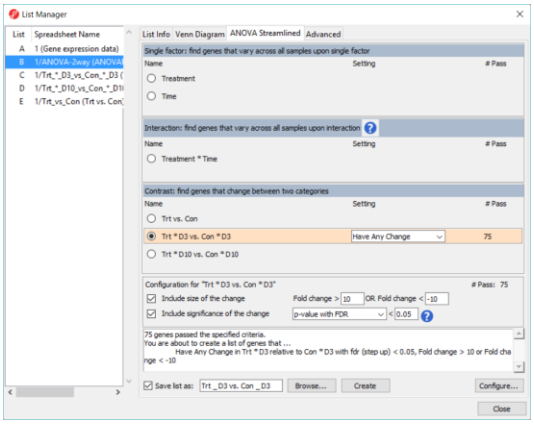
Create List

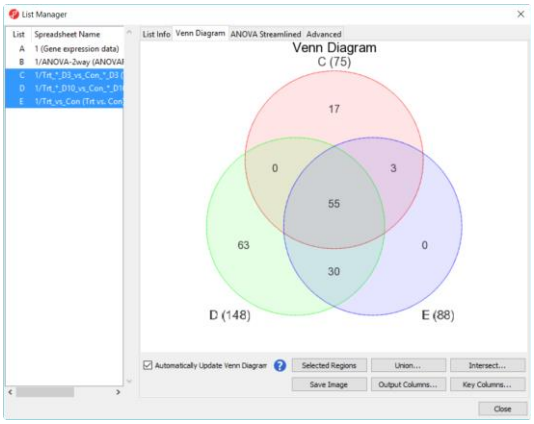
Generate a list of miRNA that is showing differential expression between brain and heart.

- 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



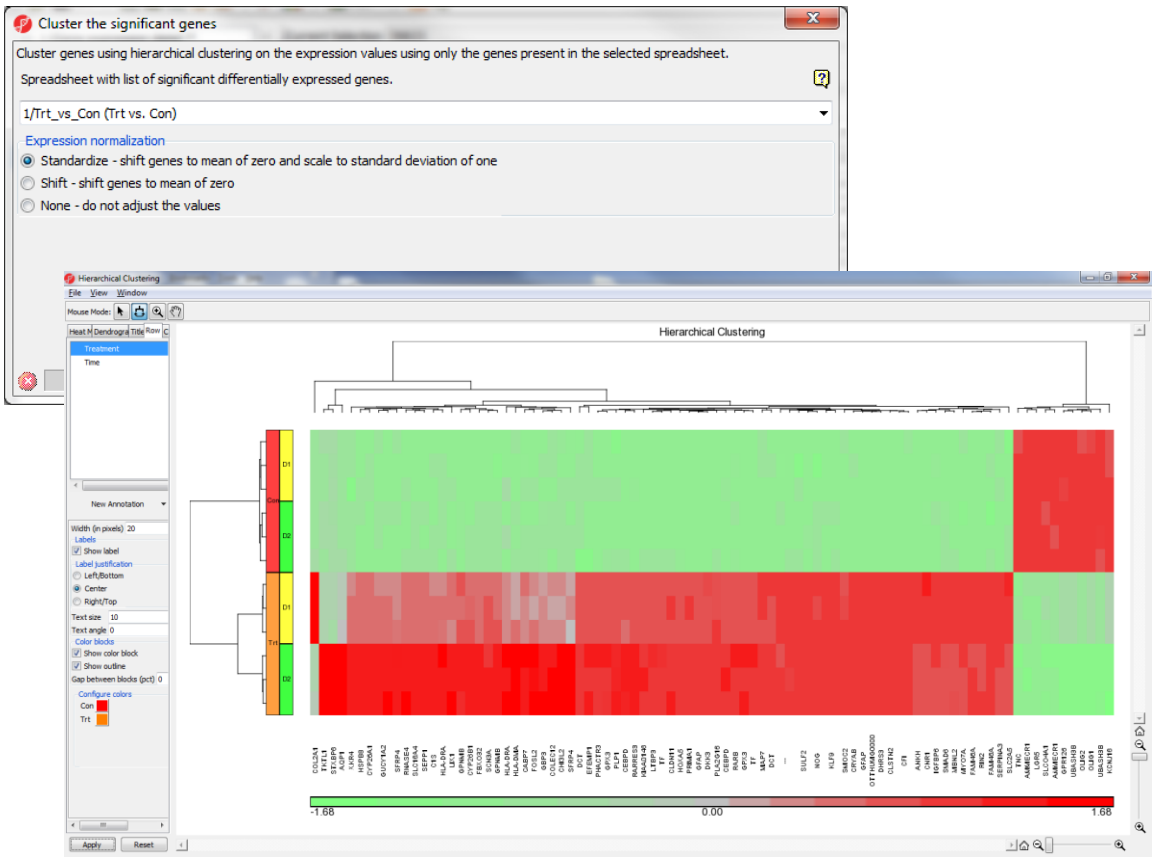


Notes:_____

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 Genes** 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
- Change the orientation

Dendrograms

- Change the width/height of the dendrogram
- Color dendrogram

Rows

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

Columns

- Label with column header or gene symbol

Save/Load: save or load configuration settings

Mode: mouse over, select, zoom, and flip

Notes:

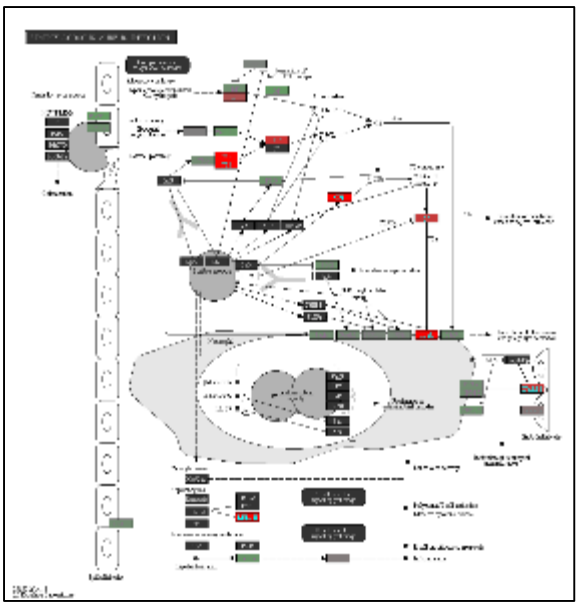
Biological Interpretation—Pathway Enrichment

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



Notes:_____

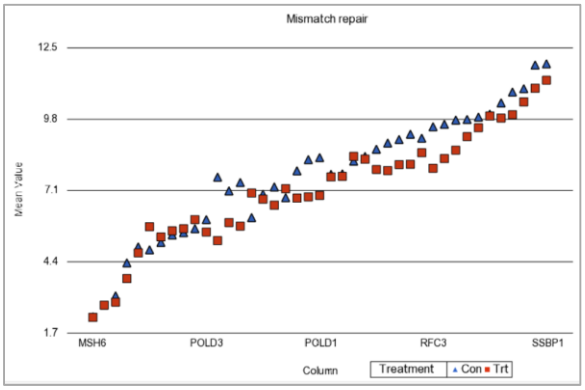
Biological Interpretation—Pathway ANOVA

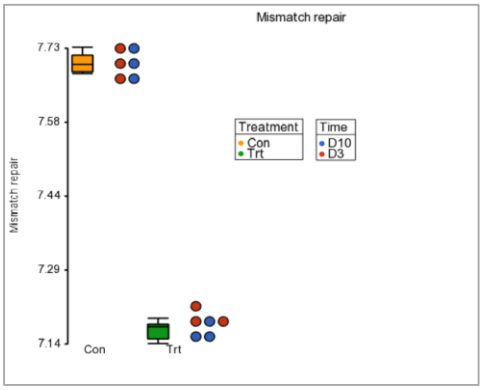
Pathway ANOVA: Detect differentially expressed pathway

- Select *Gene expression data* spreadsheet
- Select **Pathway analysis> Partek Pathway> Pathway ANOVA**
- Change *Restrict analysis to pathways with fewer than* **50** 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





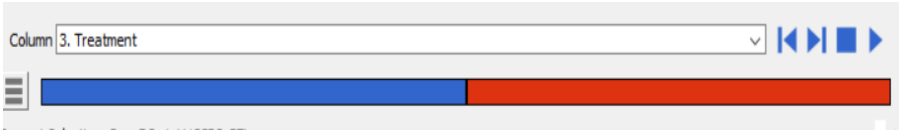
Notes:_____

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



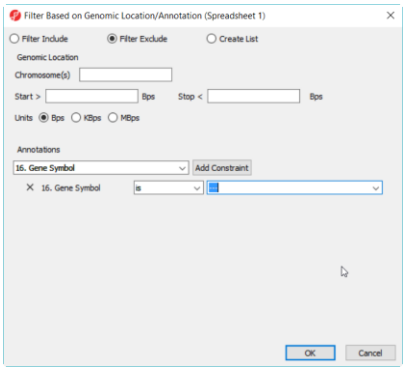
Filter genes

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 annotation**
- Select **Annotation field > Add Constraint**
- **Filter Exclude** if **Gene Symbol is ---**



Filter include genes of interest

- Click **Filter>Filter Column>Filter column based on a list**

Notes:_____

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:

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

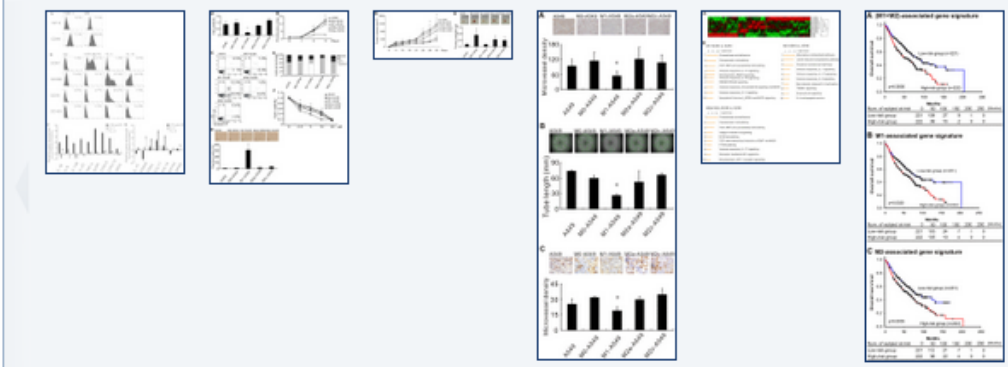
Abstract

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:

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:

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: