Creating gene lists from ANOVA results

- · Creating a gene list with the ANOVA Streamlined list manager
- Creating a gene list from a volcano plot

Creating a gene list with the ANOVA Streamlined list manager

Now that you have obtained statistical results from the microarray experiment, you can create new spreadsheets containing just those genes that pass certain criteria. This will streamline data management by focusing on just those genes with the most significant differential expression or substantial fold change. The *List Manager* can be used to specify numerous conditions for selecting genes of interest. In this tutorial, we are going to create a gene list of gene with a fold change between -1.3 to 1.3 that has an unadjusted p-value of < 0.0005.

- Invoke the List Manager dialog by selecting Create Gene List in the Analysis section of the Gene Expression workflow
- Ensure that the 1/ANOVA-3way (ANOVAResults) spreadsheet is selected as this is the spreadsheet we will be using to create our new gene list as shown (Figure 1)
- Select the ANOVA Streamlined tab.
- Set Contrast: find genes that change between two categories panel, to Down Syndrome vs. Normal and select Have Any Change from the Setti ng drop-down menu

This will find genes with different expression levels in the different types of samples.

- In the Configuration for "Down Syndrome vs Normal" panel, check that Include size of the change is selected and enter 1.3 into Change > and -1.3 in OR Change <
- Select Include significance of the change, choose unadjusted p-value from the dropdown menu, and < 0.001 for the cutoff

The number of genes that pass your cutoff criteria will be shown next to the # Pass field. In this example, 30 genes pass the criteria.

- Set Save the list as A
- Select Create to generate the new list A
- · Select Close to view the new gene list spreadsheet

| 🥑 Lie | st Manager | | | | | | × | | | | |
|-------|----------------------|--|--|--|--------------------|-----------------------|----------|--|--|--|--|
| List | Spreadsheet Name | ^ | List Info Venn Diagram | ANOVA Streamlin | ed Advanced | | | | | | |
| Α | 1 (Down_Syndrome-GE) | | Single factor: find genes t | ingle factor: find genes that vary across all samples upon single factor | | | | | | | |
| В | 1/ANOVA-3way (ANOVA | | Name | Setting | # Pass | | | | | | |
| | | | 🔘 Туре | | | | | | | | |
| | | | O Tissue | | | | | | | | |
| | | | | | | | | | | | |
| | | | Interaction: find genes the | at vary across all sa | mples upon interac | tion 🕜 | | | | | |
| | | | Name | | | Setting | # Pass | | | | |
| | | | O Type * Tissue | | | | | | | | |
| | | | Contrast: find genes that | | | | | | | | |
| | | | Name | | | Setting | # Pass | | | | |
| | | | Own Syndrome vs. | Normal | | Have Any Change 🗸 🗸 | 30 | | | | |
| | | | | | | | | | | | |
| | | | Configuration for "Down Syndrome vs. Normal" # Pass: | | | | | | | | |
| | | ✓ Include size of the change Fold change > 1.3 OR Fold change < -1.3 | | | | | | | | | |
| | | | Include significance of | of the change | unadjusted p-val | ue ~ < 0.001 ? | | | | | |
| | | | 30 genes passed the spec | fied criteria. | | | <u>_</u> | | | | |

Copyright © 2018 by Partek Incorporated. All Rights Reserved. Reproduction of this material without express written consent from Partek Incorporated is strictly prohibited.

| | You are about to create a list of genes that Have Any Change in Down Syndrome relative to Normal with pvalue < 0.001, Fold change > hange < -1.3 | 1.3 or Fold c |
|-----|--|---------------|
| < > | Save list as: A Browse Create | Configure |
| | • | Close |

Figure 7. Creating a gene list from ANOVA results

The spreadsheet Down_Syndrome_vs_Normal (A) will be created as a child spreadsheet under the Down_Syndrome-GE spreadsheet.

This gene list spreadsheet can now be used for further analysis such as hierarchical clustering, gene ontology, integration of copy number data, or be exported into other data analysis tools such as pathway analysis.

You can practice creating new gene list criteria of your own to become familiar with the List Manager tool. For more information, you can always click on the (12) buttons.

Creating a gene list from a volcano plot

Next, we will generate a list of genes that passed a p-value threshold of 0.05 and fold-changes greater than 1.3 using a volcano plot.

- Select the 1/ANOVA-3way (ANOVAResults) spreadsheet in the Analysis tab. This is the spreadsheet our gene list will be drawn from Select View > Volcano Plot from the Partek Genomics Suite main menu (Figure 2) •

| 🥬 Partek Genomics Sui | te - 1/ANOVA-3wa | y (ANO | VAResults) | | | | | | | | | - 0 | × | |
|------------------------------|---|--------|------------------|---|-------------------|-------------------|-----------------|-------------------------------|---------------------|-----------------------|----------|---------------------------------------|----------|---|
| File Edit Transform | View Stat Filte | r Too | ls Window | Custom Help | | | | | | | | Warkflaura Cone Everagion | | |
| Analysis M. Castles Dia | Box & Whiskers | | | | | | | | | | | Cana Expression | | ; |
| Analysis X Scatter Pic | Histograms | • | rces of variatio | es of variation X Interaction Plot X Dot Plot X | | | | | | | | | | • |
| Lì 🧀 🖵 📘 | Star Plot | | 🛞 🏭 🔍 🍸 🕐 | | | | | | | | | ✓ Import | | |
| | Scatter Plot Dot Plot(s) Probability Plot | | | | | | | | | Import Samples | ~ | ٢ | | |
| I (Down_Syndrome | | | Selection 212 | | | | | | | Add Sample Attributes | ~ | 1 | | |
| ANOVA-3way (Down_Syndron | Intensity Plot | rt | 1. Column # | 2. Probeset ID | 3. Entrez Gene | 4. Gene Symbol | Gene Title | 6. RefSeq Transcript ID | 7. p-value(Type) | 8. p-v alue | | View Sample Information | | |
| | Venn Diagram | | 212 | 200677 at | 754 | PTTG1IP | pituitary | NM 001286822 | 1.21194e-05 | 1.9 | | Choose Sample ID Column | | |
| | Violin Plot | | 275 | 200740_s_at | 6612 | SUMO3 | small | NM_001286416 | 2.35057e-05 | 4.2 | | ∕ QA/QC | | |
| | Cluster Genome | | 3169 | 203635_at | 10311 | DSCR3 | Down | NM_006052 /// | 3.16516e-05 | 0.0 | | PCA Scatter Plot | ~ | , |
| | Chromosome View | w | 2284 | 202749_at | 7485 | WRB | tryptophan rich | NM_001146218 | 4.02563e-05 | 7.6 | | Sample Box & Whiskers Chart | | |
| | Karyogram View | | 1860 | 202325_s_at | 522 | ATP5J | ATP synthase, | NM_001003696 | 4.60628e-05 | 0.0 | | Sample Histogram | | , |
| | MA Plot Volcano Plot | | 621 | 201086_x_at | 6651 | SON | SON DNA | NM_001291411 | 4.63436e-05 | 0.0 | | | ~ | ~ |
| | | | 19790 | 220419_s_at | 29761 | USP25 | ubiquitin | NM_001283041 | 6.29939e-05 | 0.0 | | ~ Analysis | | _ |
| | Forest Plot | 2 | 736 | 201201_at | 1476 | CSTB | cystatin B | NM_000100 | 6.76374e-05 | 0.0 | | Detect Differentially Expressed Genes | | _ |
| l | Sources of Varia | tion | 20293 | 220922_s_at | 30014 /// | LOC105369237 | sperm protein | NM_013453 /// | 7.13086e-05 | 7.0 | | View Sources of Variation | ~ | ٢ |
| | | 10. | 1378 | 201843_s_at | 2202 | EFEMP1 | EGF containing | NM_001039348 | 7.75104e-05 | 4.0 | | Create Gene List | ~ | 1 |
| | | 11. | 1752 | 202217_at | 8209 /// | C21orf33 /// | chromosome 21 | NM_004649 /// | 7.78656e-05 | 0.0 | | > Visualization | | |
| | | 12. | 353 | 200818_at | 539 | ATP50 | ATP synthase, | NM_001697 | 8.95525e-05 | 0.0 | | Biological Interpretation | | |
| | | 13. | 17758 | 218386_x_at | 10600 | USP16 | ubiquitin | NM_001001992 | 0.000115354 | 0.0 | | > Genomic Integration | | |
| | | 14. | 9059 | 209560_s_at | 8788 | DLK1 | delta-like 1 | NM_001032997 | 0.000118621 | 2.3 | | > miRNA Integration | | |
| | | 15. | 1177 | 201642_at | 3460 | IFNGR2 | interferon | NM_005534 /// | 0.000140321 | 0.0 | | | | |
| | | 16. | 17930 | 218558_s_at | 54148 | MRPL39 | mitochondrial | NM_017446 /// | 0.000140577 | 0.0 | | | | |
| | | 17. | 17490 | 218118_s_at | 100287932 | TIMM23 | translocase of | NM_006327 /// | 0.000151811 | 2.3 | | | | |
| | | 18. | 3956 | 204422_s_at | 2247 | FGF2 | fibroblast | NM_002006 | 0.000175323 | 5.3 | | | | |
| | | 19. | 21057 | 221689_s_at | 51227 | PIGP | phosphatidylino | NM_153681 /// | 0.000188366 | 4.0 | | | | |
| | | 20. | 20134 | 220763_at | 259217 | HSPA12A | heat shock | NM_025015 /// | 0.000209582 | 0.0 | | | | |
| | ~ | 21. | 1982 | 202447_at | 1666 | DECR1 | 2,4-dienoyl-Co/ | NM_001359 /// | 0.000211018 | 3.7 | | | | |
| < | > | Rows: | 22283 Columns: | 22 < | | | | | | > | ~ | | | 1 |
| A | | | | | | | | | | | |] | | |

Figure 8. Generating a Volcano Plot from ANOVA results

- Set X Axis (Fold-Change) to 12. Fold-Change(Down Syndrome vs. Normal), and the Y axis (p-value) to be 10. p-value(Down Syndrome vs. Normal)
- Select OK to generate a Volcano Plot tab for genes in the ANOVA spreadsheet (Figure 3)



Figure 9. Volcano plot generated from ANOVA spreadsheet

In the plot, each dot represents a gene. The X-axis represents the fold change of the contrast (*Down syndrome* vs. *Normal*), and the Y-axis represents the range of p-values. The genes with increased expression in Down syndrome samples are on the right side of the N/C (no change) line; genes with reduced expression in Down syndrome statistically significant with increasing Y-axis position. The genes that have larger and more significant changes between the Down syndrome and normal groups are on the upper right and upper left corner.

In order to select the genes by fold-change and p-value, we will draw a horizontal line to represent the p-value 0.05 and two vertical lines indicating the -1.3 and 1.3-fold changes (cutoff lines).

- Select Rendering Properties ()
- Choose the Axes tab
- Check Select all points in a section to allow Partek Genomics Suite to automatically select all the points in any given section
- Select the Set Cutoff Lines button and configure the Set Cutoff Lines dialog as shown (Figure 4)

| 🤣 Set Cutoff Lines | × | | | | | | |
|------------------------------------|----------|--|--|--|--|--|--|
| Vertical Line(s) X-Axis Value(s) | -1.3 1.3 | | | | | | |
| Horizontal Line(s) Y-Axis Value(s) | 0.05 | | | | | | |
| Select all points in a section | | | | | | | |
| Line Color Line Width | | | | | | | |
| ОК | Cancel | | | | | | |

Figure 10. Setting cutoff lines for -1.3 to 1.3 fold changes and a p-value of 0.05

• Select OK to draw the cutoff lines

· Select OK in the Plot Rendering Properties dialog to close the dialog and view the plot

The plot will be divided into six sections. By clicking on the upper-right section, all genes in that section will be selected.

• Right-click on the selected region in the plot and choose Create List to create a list including the genes from the section selected (Figure 5). Note that these p-values are uncorrected



Figure 11. Creating a gene list from a volcano plot

Note: If no column is selected in the parent (ANOVA) spreadsheet, all of the columns will be included in the gene list; if some columns are selected, only the selected columns will be included in the list.

• Specify a name for the gene list (example: volcano plot list) and write a brief description about the list.

The description is shown when you right-click on the spreadsheet > Info > Comments. Here, I have named the list "volcano plot list" and described it as "Genes with >1.3 fold change and p-value <0.05" (Figure 6). The list can be saved as a text file (*File > Save As Text File*) for use in reports or by downstream analysis software.

| P Create List from Selections | | | | | | |
|-------------------------------|---|---|--|--|--|--|
| Save as | k Training Data/new GE/volcano plot list Browse | | | | | |
| Description | Genes with >1.3 fold change and p-value <0.05 | • | | | | |
| | | | | | | |
| | | _ | | | | |
| | ▲ | - | | | | |
| | OK Cancel | | | | | |

Figure 12. Saving a list created from a volcano plot

« Identifying differentially expressed genes using ANOVA Performing hierarchical clustering »

Additional Assistance

If you need additional assistance, please visit our support page to submit a help ticket or find phone numbers for regional support.

