DEBrowser:

Interactive Differential Expression Analysis Tool

Introduction

Differential gene expression analysis has become an increasingly popular tool in determining and viewing up and/or down experssed genes between two sets of samples. The goal of Differential gene expression analysis is to find genes or transcripts whose difference in expression, when accounting for the variance within condition, is higher than expected by chance. DESeq2 is an R package available via Bioconductor and is designed to normalize count data from high-throughput sequencing assays such as RNA-Seq and test for differential expression (Love et al. 2014). With multiple parameters such as padjust values, log fold changes, plot styles, and so on, altering plots created with your DE data can be a hassle as well as time consuming. The Differential Expression Browser uses DESeq2 (Love et al., 2014), EdgeR (Robinson et al., 2010), and Limma (Ritchie et al., 2015) coupled with shiny (Chang, W. et al., 2016) to produce real-time changes within your plot queries and allows for interactive browsing of your DESeq results. In addition to DESeq analysis, DEBrowser also offers a variety of other plots and analysis tools to help visualize your data even further.

DEBrowser

DEBrowser utilizes Shiny, a R based application development tool that creates a wonderful interactive user interface (UI) combined with all of the computing prowess of R. After the user has selected the data to analyze and has used the shiny UI to run DESeq2, the results are then input to DEBrowser. DEBrowser manipulates your results in a way that allows for interactive plotting by which changing padj or fold change limits also changes the displayed graph(s). For more details about these plots and tables, please visit our quick start guide for some helpful tutorials.

For comparisons against other popular data visualization tools, see the comparison table below (Figure 41).

Quick start

Before you start;

First, you will have to install R and/or RStudio. (On Fedora/Red Hat/CentOS, these packages have to be installed; openssl-devel, libxml2-devel, libcurl-devel, libpng-devel) Running these simple commands will launch the DEBrowser within your local machine:

```
# Installation instructions:
# 1. Install DEBrowser and its dependencies by running the lines below
# in R or RStudio.

source("http://www.bioconductor.org/biocLite.R")

biocLite("debrowser")

# 2. Load the library

library(DEBrowser)

# 3. Start DEBrowser

startDEBrowser()
```

Browsing your Data

Data via TSV file

Once you have the DEBrowser running, a page will load asking to choose a TSV file or to load the demo data. In order to run DESeq2, we are going to need gene quantifications for genes contained in a tab-separated values (TSV) format. Gene quantifications table can be obtained running standard software like HTSeq (Anders,S. et al, 2014) or RSEM (Li and Dewey, 2011). The file values must contain the gene, transcript(s), and the sample raw count values you wish to enter into DEBrowser.

It's important to note that if your rows contain duplicate gene names, DEBrowser will reject your TSV file. Please try to keep unique gene names. A sample file looks like:

gene	transcript	exper_rep1	exper_rep2	control_rep1	control_rep2
DQ714826	uc007tfl.1	0.00	0.00	0.00	0.00
DQ551521	uc008bml.1	0.00	0.00	0.00	0.00
AK028549	uc011wpi.1	2.00	1.29	0.00	0.00

You can also view/use the demo data by clicking the 'Load Demo!' text as an example. For the case study demo data, feel free to download our case study demo files at http://bioinfo.umassmed.edu/pub/debrowser/advanced_demo.tsv or a simplified version http://bioinfo.umassmed.edu/pub/debrowser/simple_demo.tsv.

Data via JSON objects

DEBrowser also accepts JSON objects via hyperlink by following a few conversion steps.
First, using the API provided by Dolphin, we will convert a TSV file into a JSON object using this web api:

```
http://dolphin.umassmed.edu/public/api/
```

The Two parameters it accepts (and examples) are:

```
1. source=http://bioinfo.umassmed.edu/pub/debrowser/advanced_demo.tsv
2. format=JSON
```

Leaving you with a hyperlink for:

```
http://dolphin.umassmed.edu/public/api/?source=http://bioinfo.umassmed.edu/pub/debrowser/
advanced_demo.tsv&format=JSON
```

Next you will need to encode the URL so you can pass it to the DEBrowser website. You can find multiple URL encoders online, such as the one located at this web address: http://www.url-encode-decode.com/.

Encoding our URL will turn it into this:

```
http%3A%2F%2Fdolphin.umassmed.edu%2Fpublic%2Fapi%2F%3Fsource%3Dhttp%3A%2F%2Fbioinfo
.umassmed.edu%2Fpub%2Fdebrowser%2Fadvanced_demo.tsv%26format%3DJSON
```

Now this link can be be used in debrowser as:

```
http://debrowser.umassmed.edu:443/debrowser/R/
```

It accepts two parameters:

1. `jsonobject=http%3A%2F%2Fdolphin.umassmed.edu%2Fpublic%2Fapi%2F%3Fsource%3Dhttp%3A%2F%2Fbioinfo.umassmed.edu%2Fpub%2Fdebrowser%2Fadvanced_demo.tsv%26format%3DJSON`
2. `title=no`

The finished product of the link will look like this:

`http://debrowser.umassmed.edu:443/debrowser/R/?jsonobject=http://dolphin.umassmed.edu/public/api/?source=http://bioinfo.umassmed.edu/pub/debrowser/advanced_demo.tsv&format=JSON&title=no`

Entering this URL into your web browser will automatically load in your data as a JSON object, allowing you to start browsing your data right away.

Batch Effect Conditioning

In addition to the sample TSV file you will provide; you can also correct for batch effects or any other normalizing conditions you might want to address that might be within your results. To handle for these conditions, simply create a TSV file such as the one located below:

sample	batch	condition
s1_b1_cA	1	A
s2_b1_cA	1	A
s3_b2_cB	2	B
s4_b2_cB	2	B
s5_b1_cB	1	B

This meta data file is custom made TSV created by the user and is used in order to establish different batch effects for multiple conditions. You can have as many conditions as you may require, as long as all of the samples are present. Once the TSV file has been loaded in along with your data TSV file, DEBrowser uses ComBat (part of the SVA bioconductor package) to adjust for possible batch effect or conditional biases. For more information about ComBat within the SVA package you can visit here: https://bioconductor.org/packages/release/bioc/vignettes/sva/inst/doc/sva.pdf.

To load in the specific file that contains the batch meta data, at the start of the DEBrowser there will be a "Choose Meta Data File (Optional)" which you can then select the batch meta data file to use for this analysis (Figure 1). Upon meta-data loading, you will then be able to select from a drop down box that will specify which column you want to use for analysis.

After obtaining and loading in the gene quantifications file, and if specified the meta data file containing your batch correction fields, you then have the option to view QC information of your quantifications or you can continue on to running DESeq2 (Figure 2).

Quality Control Information:

Upon selection of QC information, you will be shown an all-to-all plot of your samples (Figure 3). This sample-by-sample comparison will help you visualize possible discrepancies between replicate samples, in case you may want to omit them for further analysis. This graph includes sample-to-sample dotplot correlations as well as sample histograms. To the left of this plot are various plot-shaping options you can alter to more easily view the all-to-all plot.

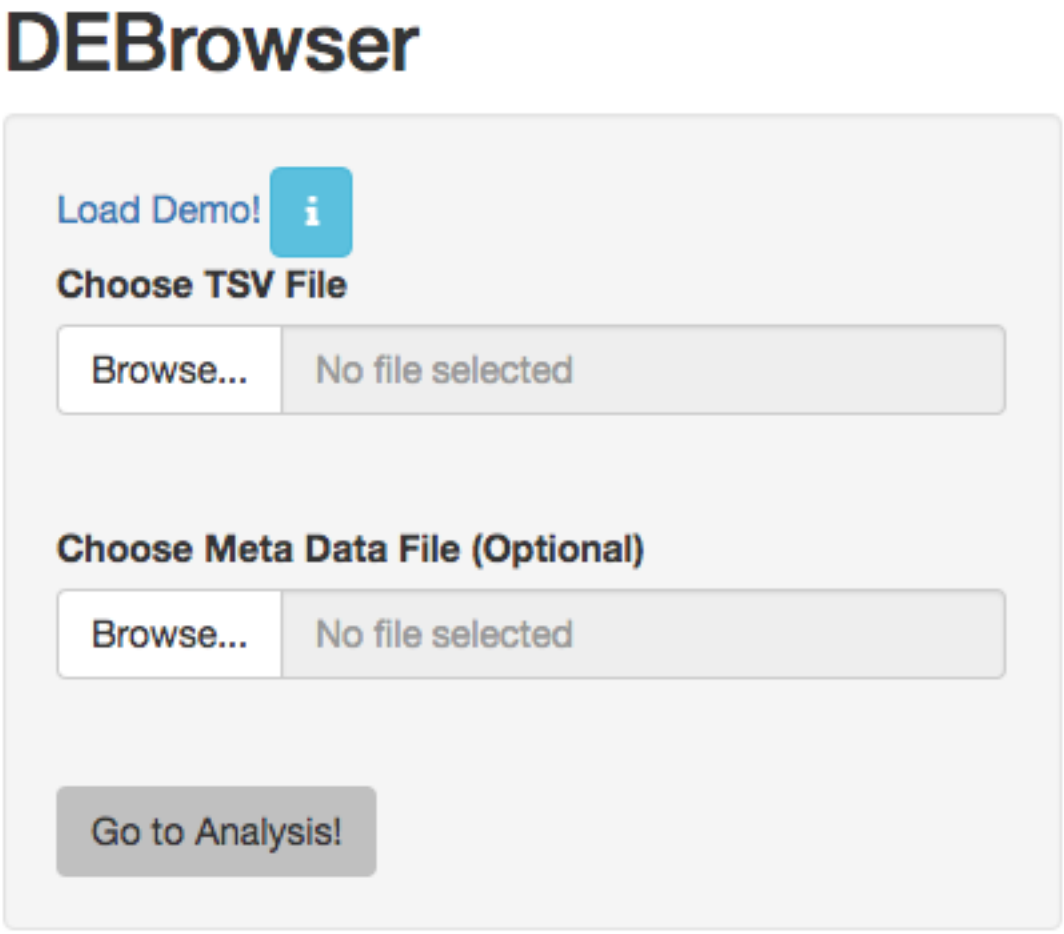

Figure 1: *Example of a data file/meta data file being loaded into DEBrowser*

Figure 2: *The initial options selection.*

Additionally, two more QC plots are available for you to use: heatmap and PCA plots. The heatmap (Figure 4) will display genes for each sample within your dataset in the form of a heatmap based on your dataset selection and PCA (Figure 5) will display Principal component analysis of your dataset. Additionally, you can view the IRQ (Interquartile Range) for both your raw data and your data after normalization (Figure 6). You can also view a density plot for your sample data for your raw data and the data after normalization (Figure 7). IQR and Density plots are another great visualization too to help you spot outliers within your sample data incase you want to remove or look into any possible discrepancies.

All of these plots will aid in viewing your preliminary data to see if there are any potential errors between replicates or batch effects (Reese et. al, 2013; Risso et al., 2014). You have the option of viewing an interactive heatmap by selecting the 'Interactive' checkbox in the left side panel when you have selected the Heatmap option. This Interactive heatmap will display genes as you hover over them for a more in-depth understanding. You can select these various plot options by selecting the type of plot you wish to view on the left panel.

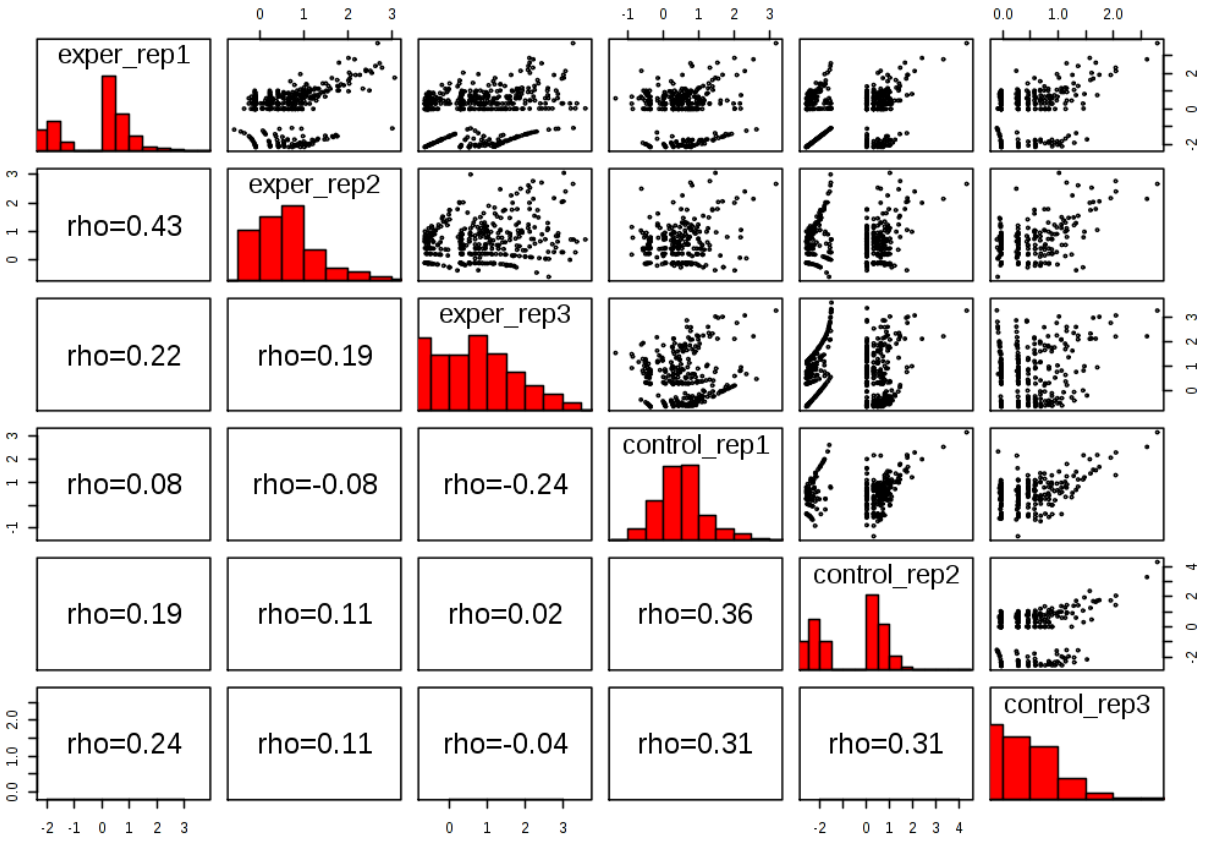

Figure 3: *Display of the all-to-all plot in the initial QC plots page.*

You can also view the genes within your quantification file in various ways. The 'Tables' tab will bring you to a table setup based on the dataset you have selected on the left options panel. The 'All detected' option lists all of the genes present within your file. The 'Selected' option lets you browse your gene selection based on your interactive heatmap selection. The Last option, 'Most Varied' (Figure 8), will display your top N varied genes. You can alter the value of N by selecting 'most-varied' from the dropdown menu on the left.

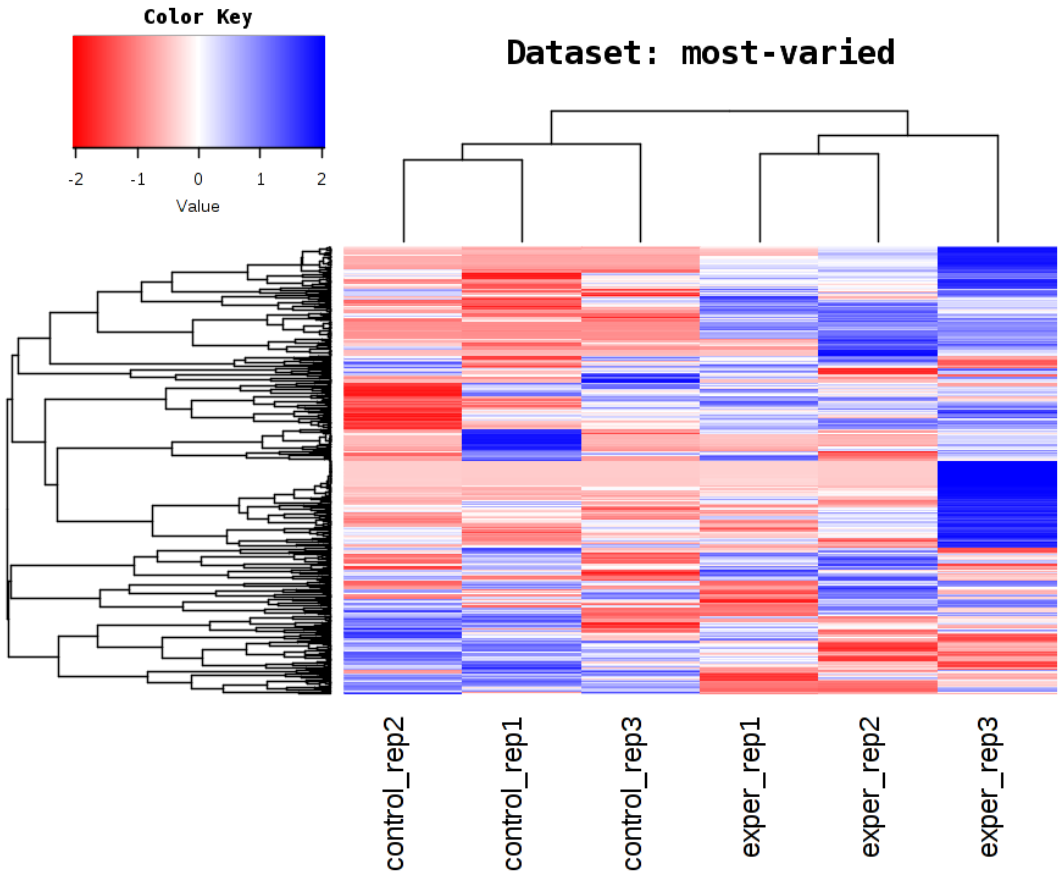

Figure 4: *Display of the most varied genes heatmap in the initial QC plots page.*

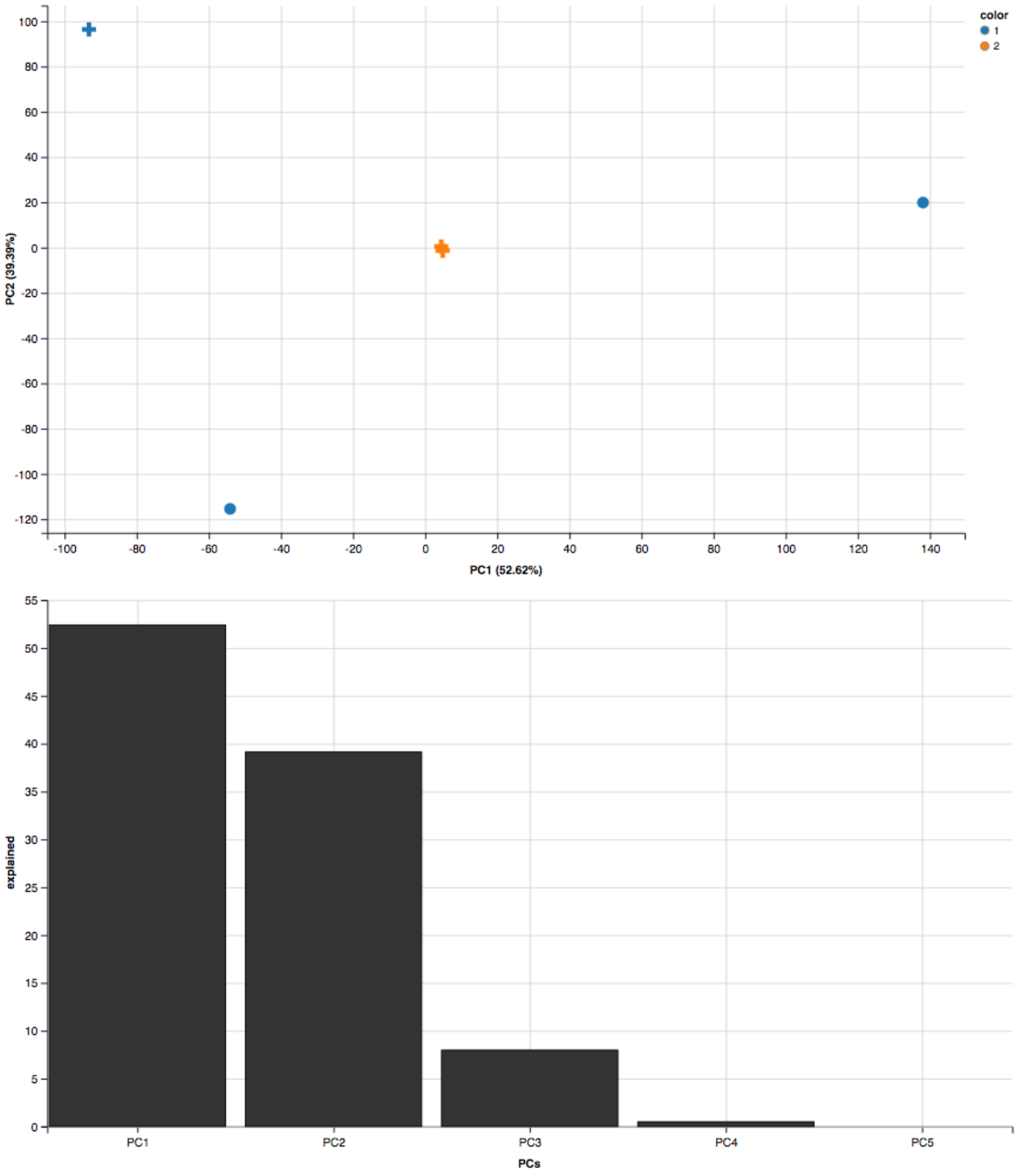

Figure 5: *Display of the PCA plot in the initial QC plots page.*

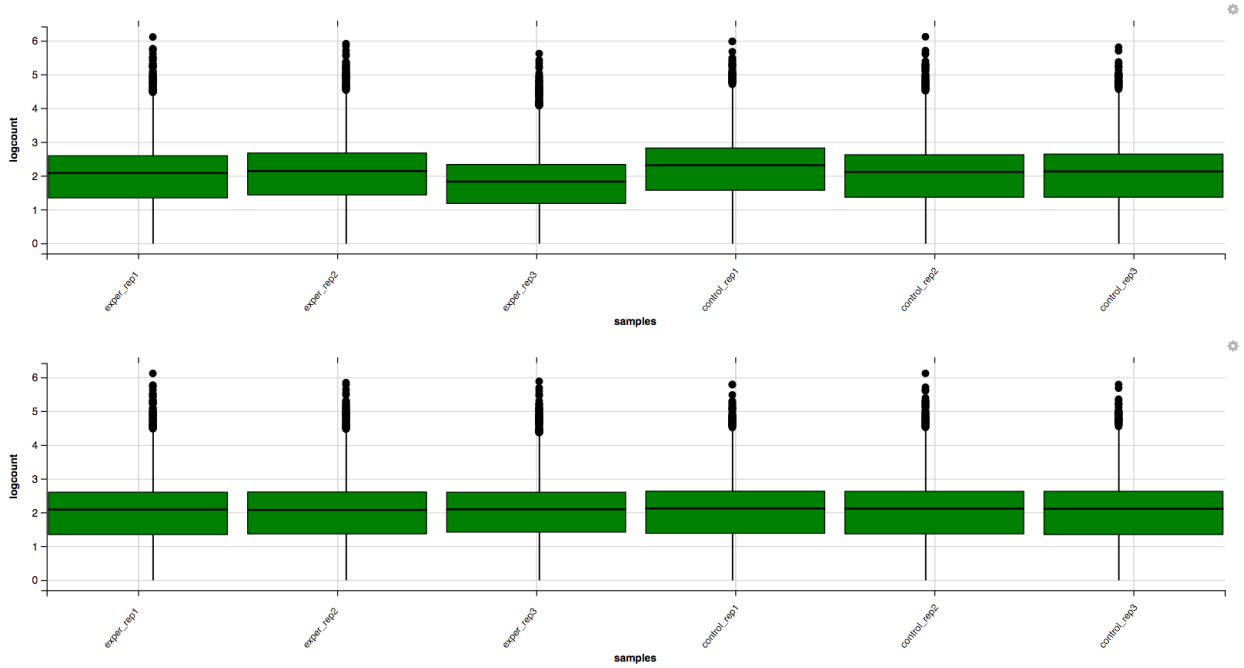

Figure 6: *Display of the IQR in the initial QC plots page.*

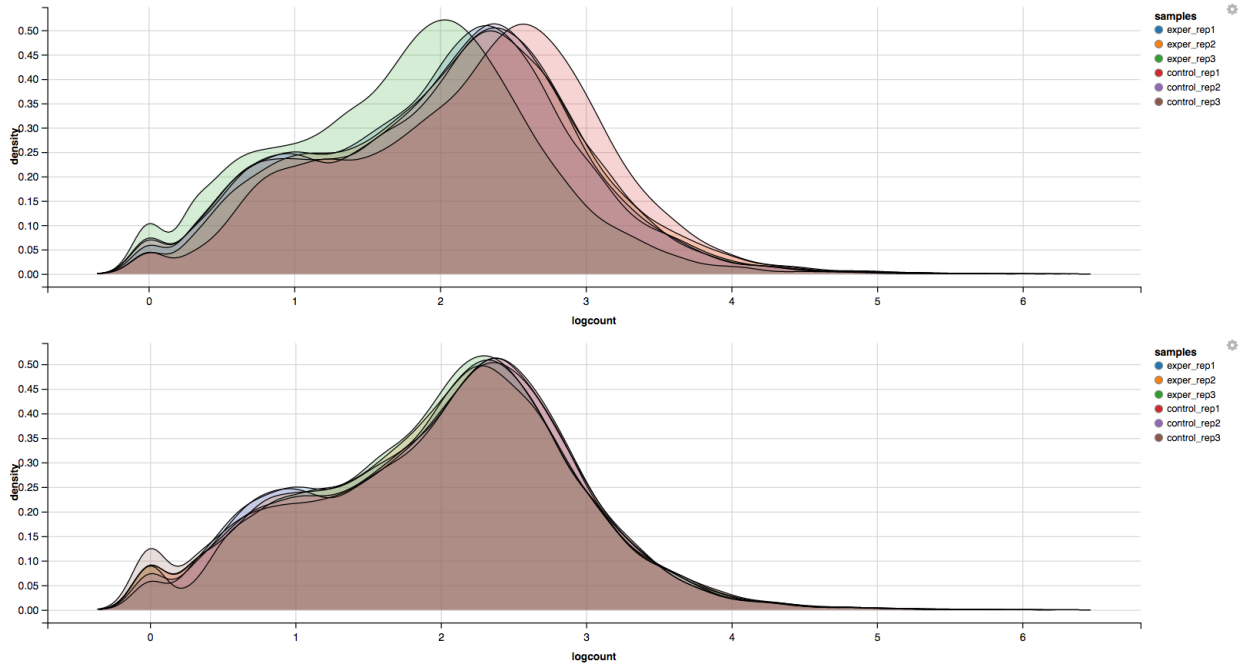

Figure 7: *Display of the density plot in the initial QC plots page.*

	exper_rep1	exper_rep2	exper_rep3	control_rep1	control_rep2	control_rep3
Tcap	0.07083901	0.2562075	417.23030	0	0.023871750	0.8085033
Actn2	0.05498247	0.3610054	198.81985	0	0.018557236	0.0000000
Xirp2	0.05143369	0.0000000	172.91841	0	0.017365545	0.0000000
Myot	0.04180303	0.0000000	117.65273	0	0.014127335	0.0000000
Myl3	0.04110437	0.0000000	114.38192	0	0.013892180	0.0000000
Myl2	0.04002795	0.0000000	109.49594	0	0.013529812	0.0000000
Trdn	0.03323063	0.0000000	82.31962	0	0.011239779	0.0000000
Smtnl1	0.02910461	0.0000000	68.33547	0	0.009848215	0.0000000
Tnni1	0.02510072	0.0000000	56.17734	0	0.008496756	0.0000000
3425401B19Rik	0.02396547	0.6121776	51.83526	0	0.008113375	0.0000000

Showing 1 to 10 of 500 entries Previous 1 2 3 4 5 ... 50 Next

Figure 8: *Displayed table of most varied genes.*

Starting DESeq:

Upon selecting to run DESeq, you are then able to select which samples will be used within your differential expression analysis By clicking the 'Add New Comparison' button, you can add as many different sample comparisons as you want. To alter the samples within each comparison, you can click on a sample and press delete to remove it or you can click the empty whitespace within the tab to bring a dropdown of samples to select from. Sample names are created based on the column headers within your data file. Once you've selected your comparisons, you are then ready to run DESeq2 to calculate differential expression by clicking on the 'Submit!' button (Figure 9).

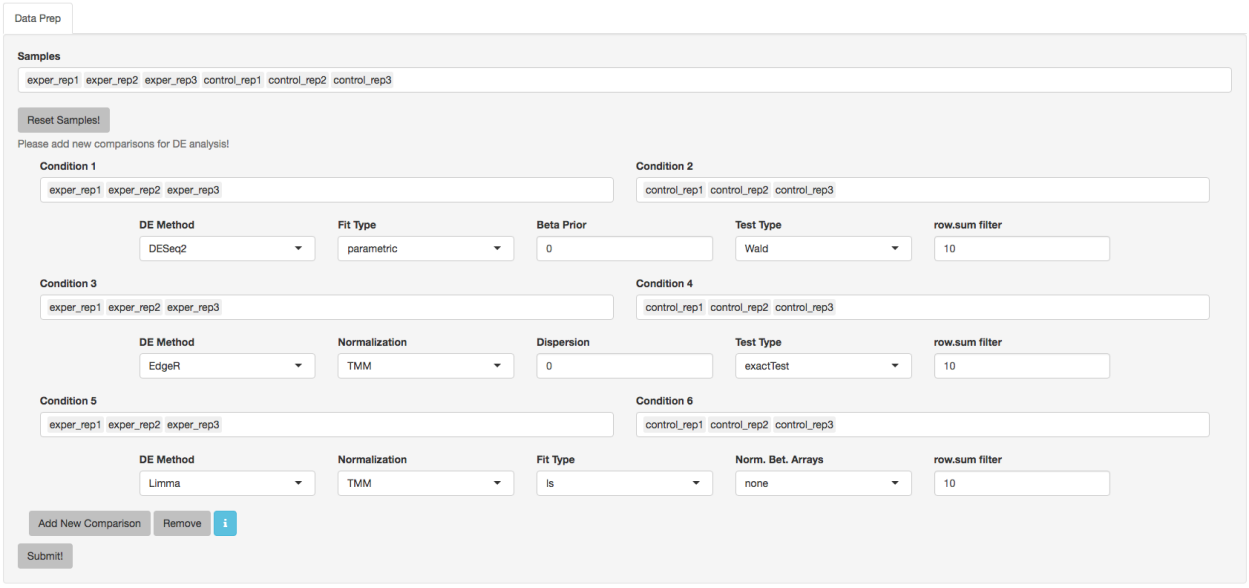

Figure 9: *Menus after loading in a sample.*

Differential Expression Calculations

DESeq2

For the details please check the user guide at this location: https://www.bioconductor.org/packages/release/bioc/vignettes/DESeq2/inst/doc/DESeq2.pdf

DESeq2 performs multiple steps in order to analyze the data you've provided for it. The first step is to indicate the condition that each column (experiment) in the table represent. You can group multiple samples into one condition column. DESeq2 will compute the probability that a gene is differentially expressed (DE) for ALL genes in the table. It outputs both a nominal and a multiple hypothesis corrected p-value (padj) using a negative binomial distribution.

Un-normalized counts

DESeq2 requires count data as input obtained from RNA-Seq or another high-throughput sequencing experiment in the form of matrix values. Here we convert un-integer values to integer to be able to run DESeq2. The matrix values should be un-normalized, since DESeq2 model internally corrects for library size. So, transformed or normalized values such as counts scaled by library size should not be used as input. Please use edgeR or limma for normalized counts.

Used parameters for DESeq2

- **fitType:** Either "parametric", "local", or "mean" for the type of fitting of dispersions to the mean intensity. See estimate Dispersions for description.
- **betaPrior:** Whether or not to put a zero-mean normal prior on the non-intercept coefficients See nbinomWaldTest for description of the calculation of the beta prior. By default, the beta prior is used only for the Wald test, but can also be specified for the likelihood ratio test.
- **testType:** Either "Wald" or "LRT", which will then use either Wald significance tests (defined by nbinomWaldTest), or the likelihood ratio test on the difference in deviance between a full and reduced model formula (defined by nbinomLRT)
- **rowsum.filter:** Regions/Genes/Isoforms with total count (across all samples) below this value will be filtered out

EdgeR

For the details please check the user guide at this location: https://www.bioconductor.org/packages/release/bioc/vignettes/edgeR/inst/doc/edgeRUsersGuide.pdf

Used parameters for EdgeR

- **Normalization:** Calculate normalization factors to scale the raw library sizes. Values can be "TMM","RLE","upperquartile", or "none".
- **Dispersion:** Either a numeric vector of dispersions or a character string indicating that dispersions should be taken from the data object.

- **testType:** exactTest or glmLRT. exactTest: Computes p-values for differential abundance for each gene between two samples, conditioning on the total count for each gene. The counts in each group are assumed to follow a binomial distribution. glmLRT: Fits a negative binomial generalized log-linear model to the read counts for each gene and conducts genewise statistical tests.
- **rowsum.filter:** Regions/Genes/Isoforms with total count (across all samples) below this value will be filtered out

Limma

For the details please check the user guide at this location: https://bioconductor.org/packages/release/bioc/vignettes/limma/inst/doc/usersguide.pdf

Limma is a package to analyze of microarray or RNA-Seq data. If data is normalized with spike-in or any other scaling, transformation or normalization method, Limma can be ideal. In that case, prefer limma rather than DESeq2 or EdgeR.

Used parameters for Limma

- **Normalization:** Calculate normalization factors to scale the raw library sizes. Values can be "TMM","RLE","upperquartile", or "none".
- **Fit Type:** Fitting method: "ls" for least squares or "robust" for robust regression
- **Norm. Bet. Arrays:** Normalization Between Arrays; Normalizes expression intensities so that the intensities or log-ratios have similar distributions across a set of arrays.
- **rowsum.filter:** Regions/Genes/Isoforms with total count (across all samples) below this value will be filtered out

Analyzing the Results

After clicking on the 'Submit!' button, DESeq2 will analyze your comparisons and store the results into separate data tables. Shiny will then allow you to access this data, with multiple interactive features, and at the click of a button. It is important to note that the resulting data produced from DESeq is normalized. Upon finishing the DESeq analysis, a tab-based menu will appear with multiple options (Figure 10) at the top-center of the page.

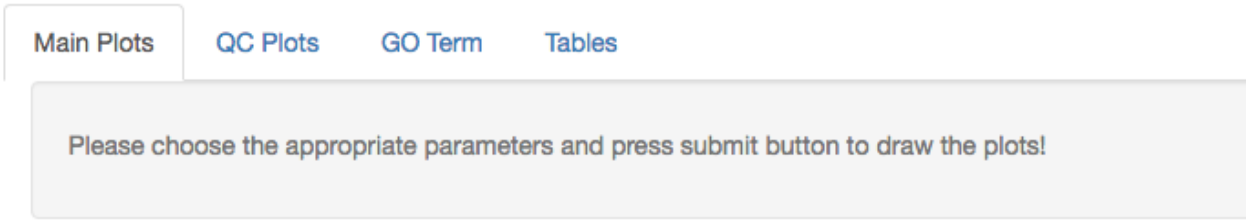

Figure 10: *List of the tabbed menus in DEBrowser.*

The first tab, the 'Main Plots' section, is where you will be able to view the interactive results plots. On the left hand side of the screen will be the options panel used to alter the padj and fold change cutoff values, what specific data set to use such as up or down regulated genes, what comparison dataset you would like to use to plot, and what type of plot you would like to view your results in (Figure 11).

Comparisons are ordered based on how they were entered by the user and for information about samples within each comparison are displayed right under the option tabs. Plot choices include:

- Scatter plot (Figure 12)
- Volcano plot (Figure 13)
- MA plot (Figure 14)

Once you have selected your values, you can hit the 'Submit!' button to create your interactive plots! Depending on whether you are using the local install or the web-based application, rendering times my vary.

The top left plot is the main plot of whichever plot you have selected to use to analyze your results. Up-regulated genes are displayed in green while down-regulated genes are displayed in red (Figure 15). Hovering over a gene on this plot will display the bottom two bar graphs: the genes normalized variation and colored by condition in the left graph, and the normalized variation between conditions within the right graph. Hovering over a gene will also display information about that gene in regards to both conditions you have selected. By clicking and dragging your mouse to create a selection over the main graph, you will create the top right plot, or the zoomed in version of your selection. If you are going to change any of the parameters on the left, please make sure to re-click the 'Submit!' button to update the graphs. You can also change which type of dataset to use within the main plots by selecting from the drop down dataset box such as looking at the N most varied genes which are displayed in yellow(Figure 16). Additionally, you can further filter these datasets by typing in the genes of interest, or regex for specific genes, to search for those specific genes within the dataset (Figrue 14). Specific gene searches are displayed in blue. All of these filtration options can be located on the left side panel which will also change based on the plot or dataset you wish to view/manipulate. It's also worth noting that the plots are resizable as well as downloable.

Selecting the 'QC Plots' tab will take you to the quality control plots section. These QC plots are very similar to the QC plots shown before running DESeq, however the dataset being used here depends on the one you select on the left menu. In addition to the all-to-all plot shown within the previous QC analysis, users can also view a heatmap and PCA plot of their analyzed data by selecting the proper plot on the left menu.

The heatmap is a great way to analyze replicate results of genes all in one simple plot (Figure 18). Users have the option to change the clustering method used as well as the distance method used to display their heatmap. In addition, you can also change the size of the heatmap produced and adjust the p-adjust and fold change cut off for this plot as well. Once all of the parameters have been set, click the 'Submit!' button at the bottom of the left menu to generate your heatmap.

Used clustering and linkage methods in heatmap

- **complete:** Complete-linkage clustering is one of the linkage method used in hierarchical clustering. In each step of clustering, closest cluster pairs are always merged up to a specified distance threshold. Distance between clusters for complete link clustering is the maximum of the distances between the members of the clusters.
- **ward D2:** Ward method aims to find compact and spherical clusters. The distance between two clusters is calculated by the sum of squared deviations from points to centroids. "ward.D2" method uses criterion (Murtagh and Legendre 2014) to minimize ward clustering method. The only difference ward.D2 and ward is the dissimilarities from ward method squared before cluster updating. This method tends to be sensitive to the outliers.
- **single:** Distance between clusters for single linkage is the minimum of the distances between the members of the clusters.
- **average:** Distance between clusters for average linkage is the average of the distances between the members of the clusters.

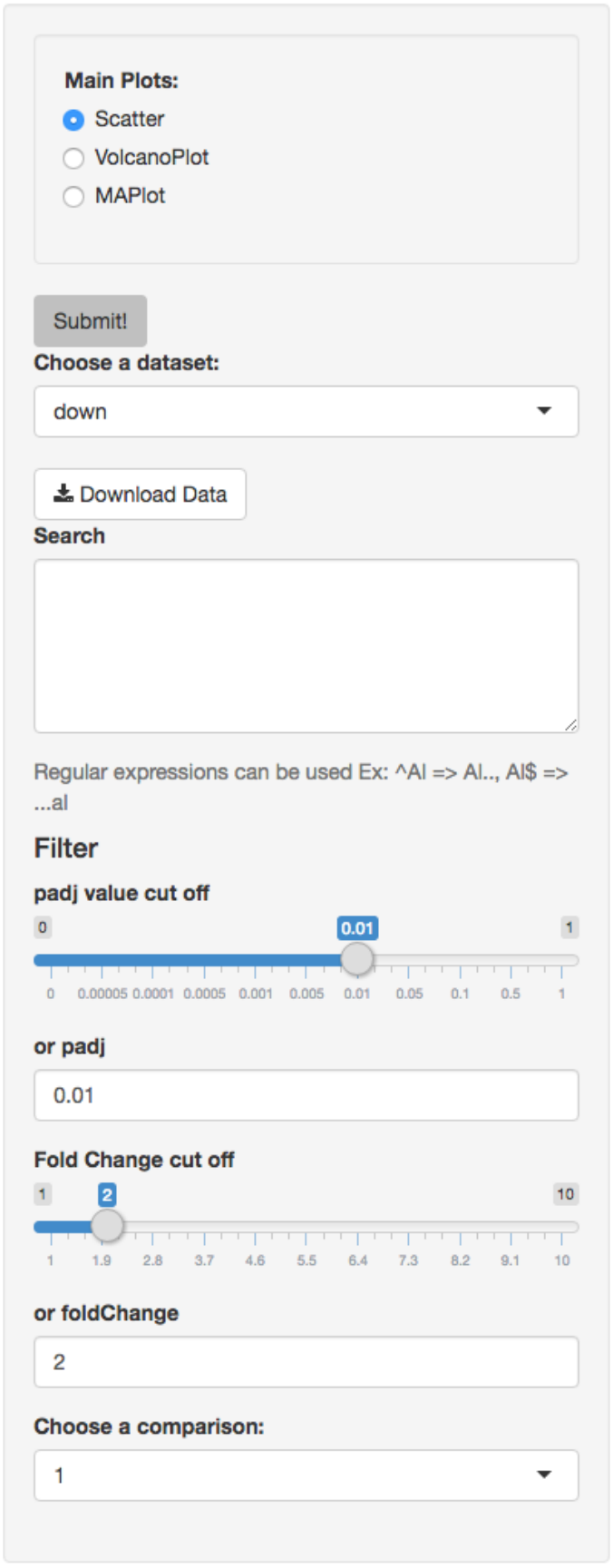

Figure 11: *The Left parameter options panel*

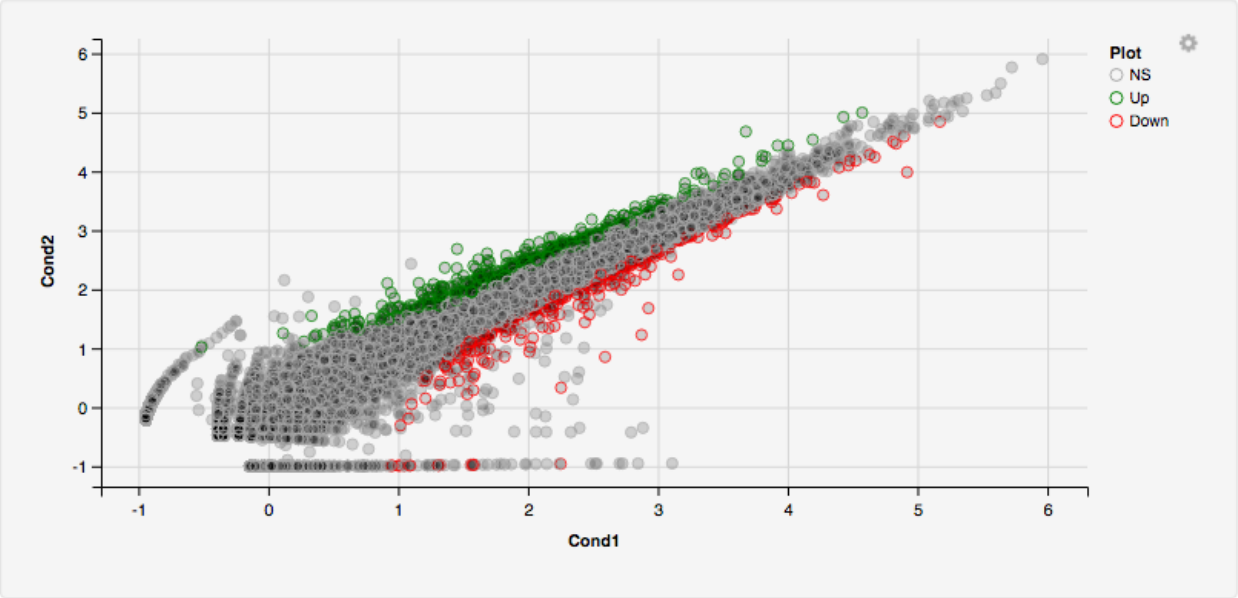

Figure 12: *Main scatter plot.*

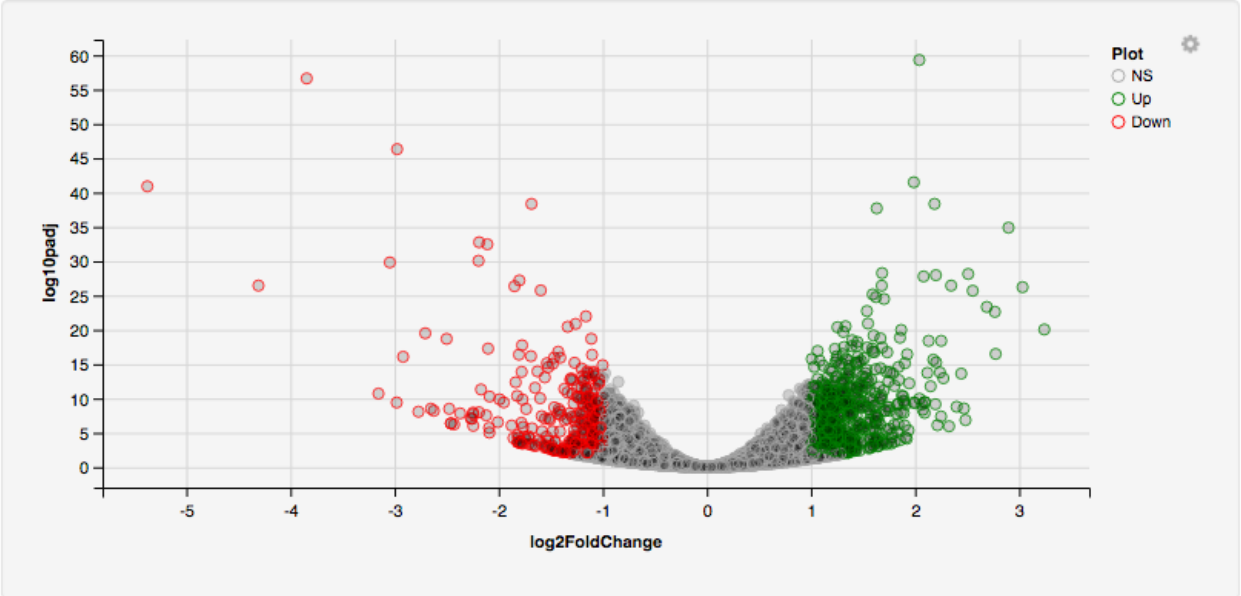

Figure 13: *Main volcano plot.*

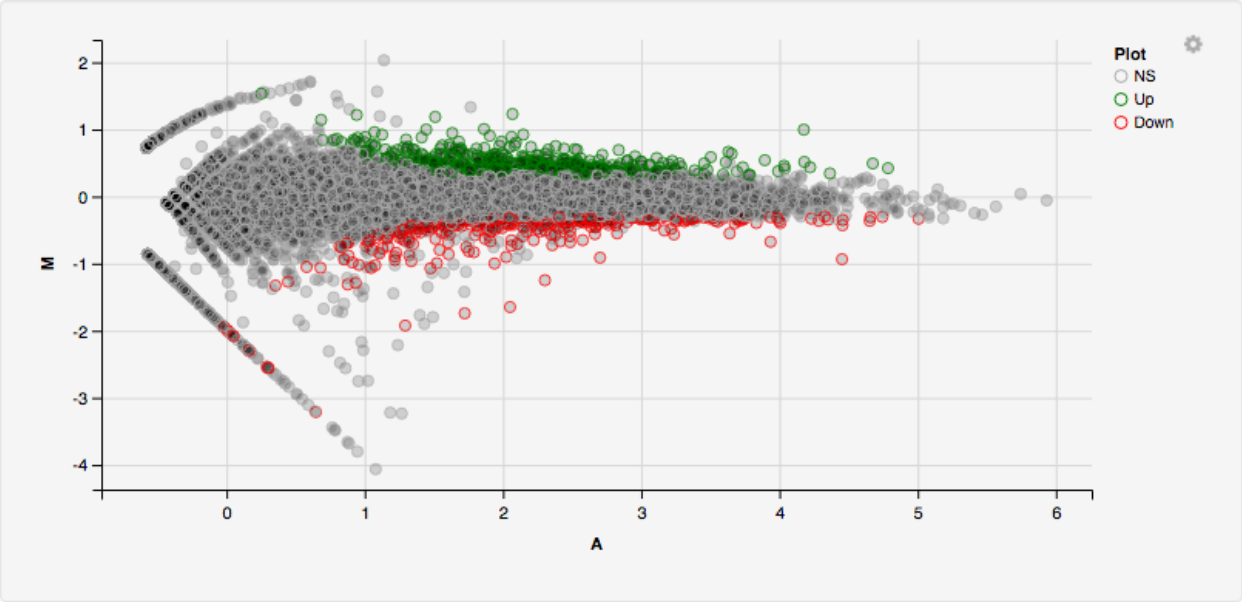

Figure 14: *Main MA plot.*

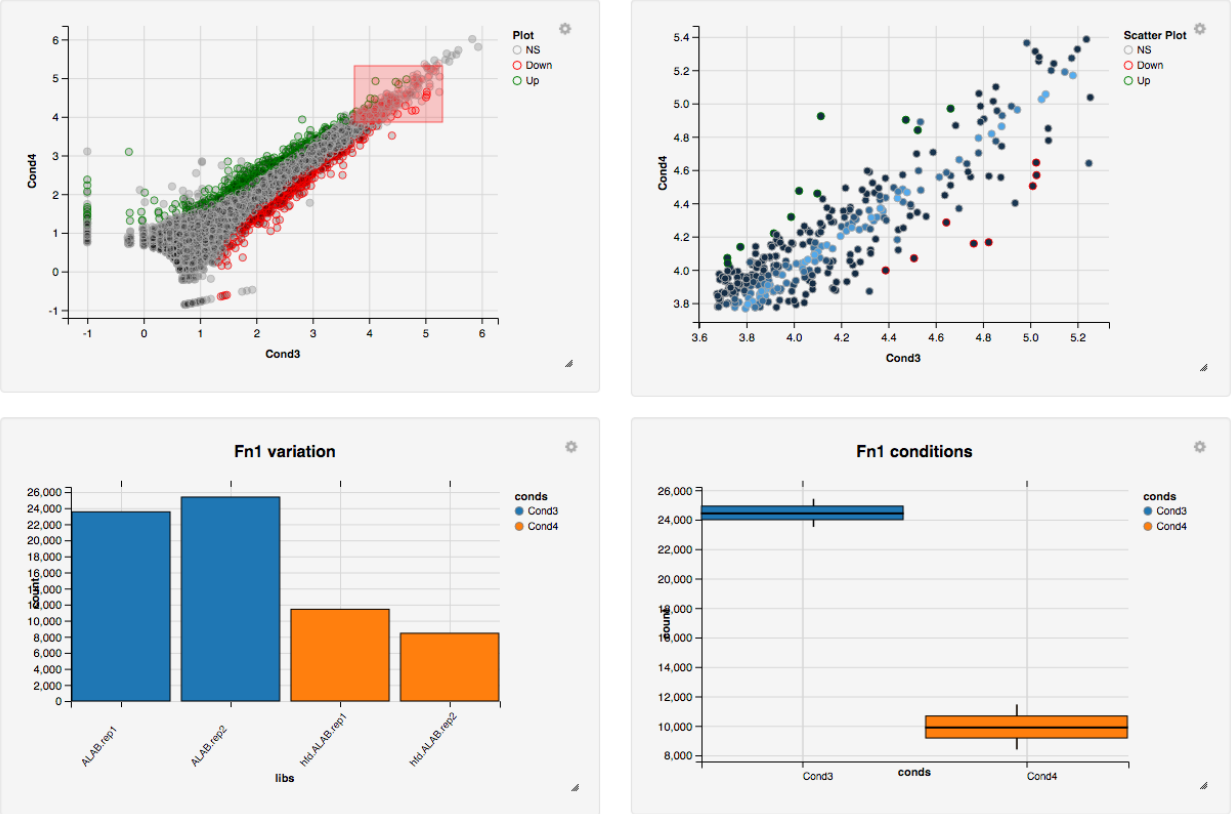

Figure 15: *The main plots page within DEBrowser.*

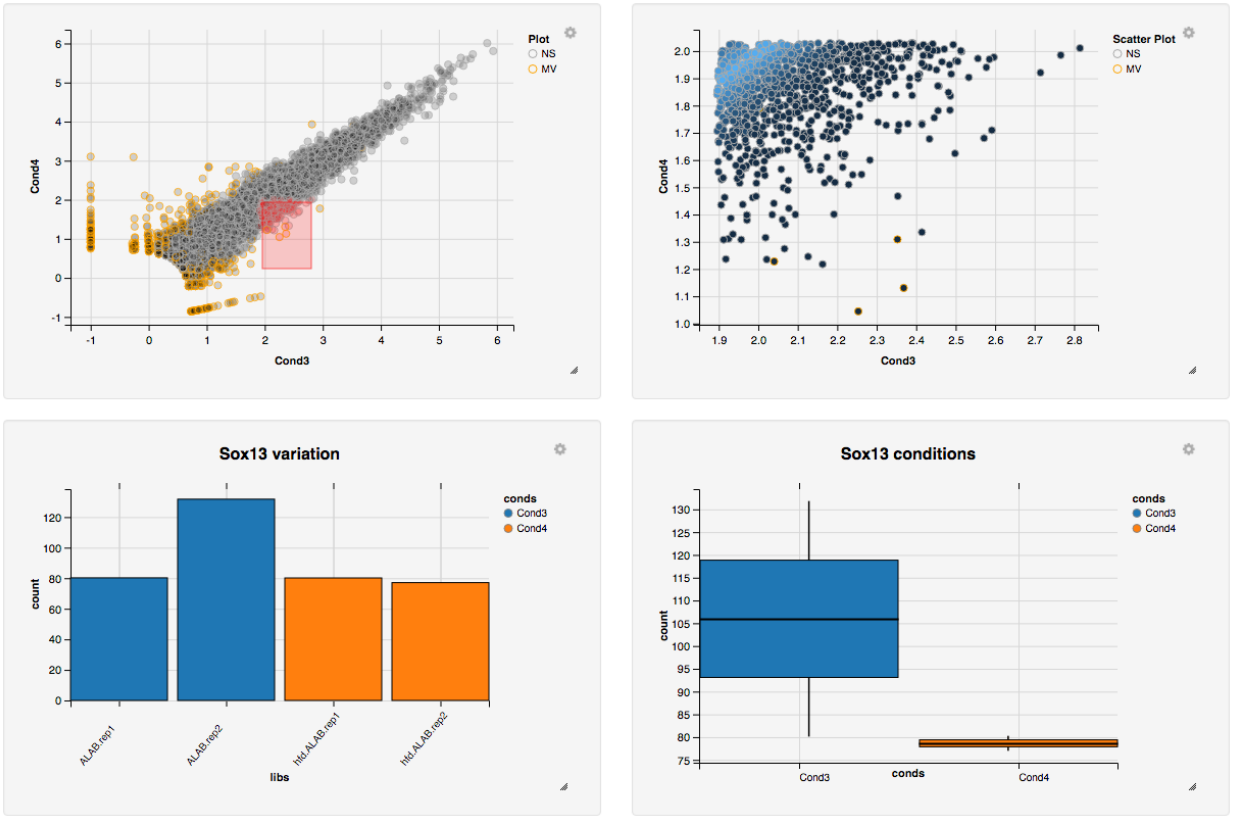

Figure 16: *Display of the most varied genes as a scatter plot.*

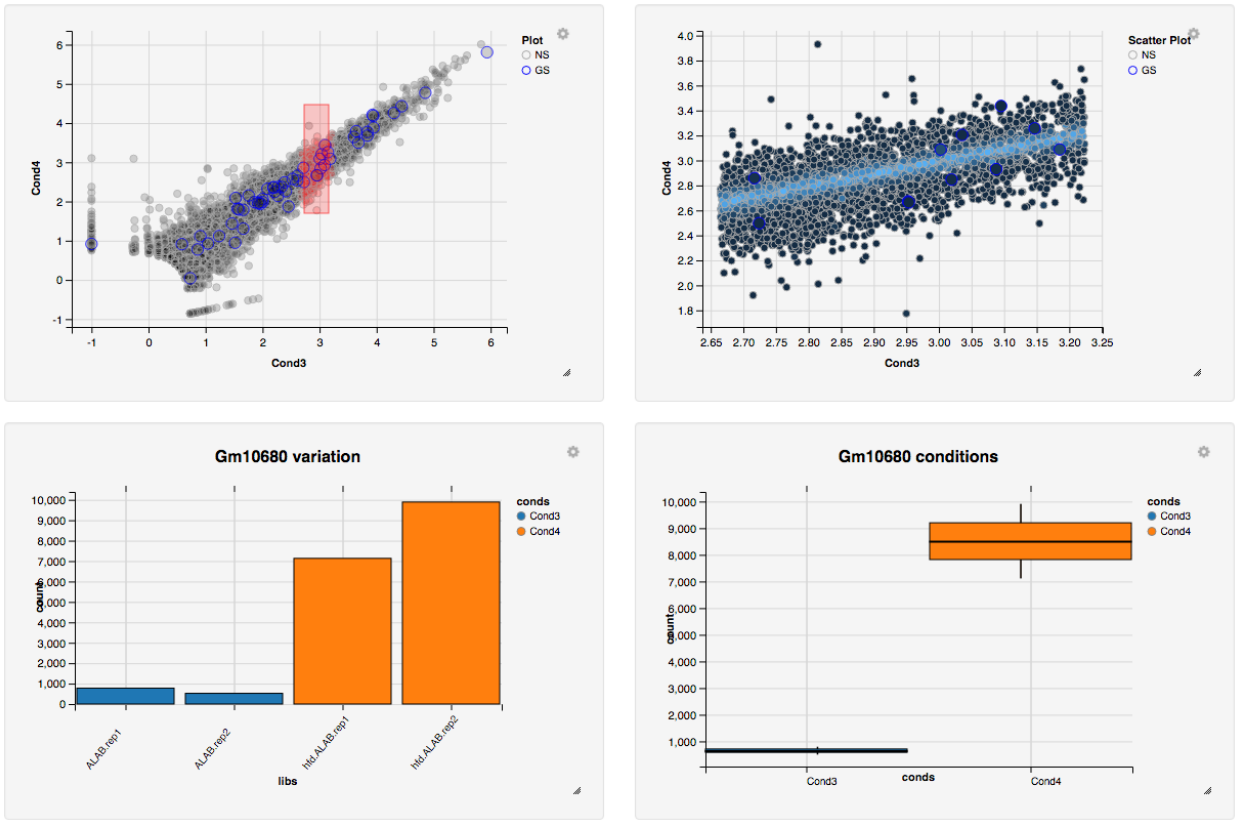

Figure 17: *Display of the geneset list as a scatter plot.*

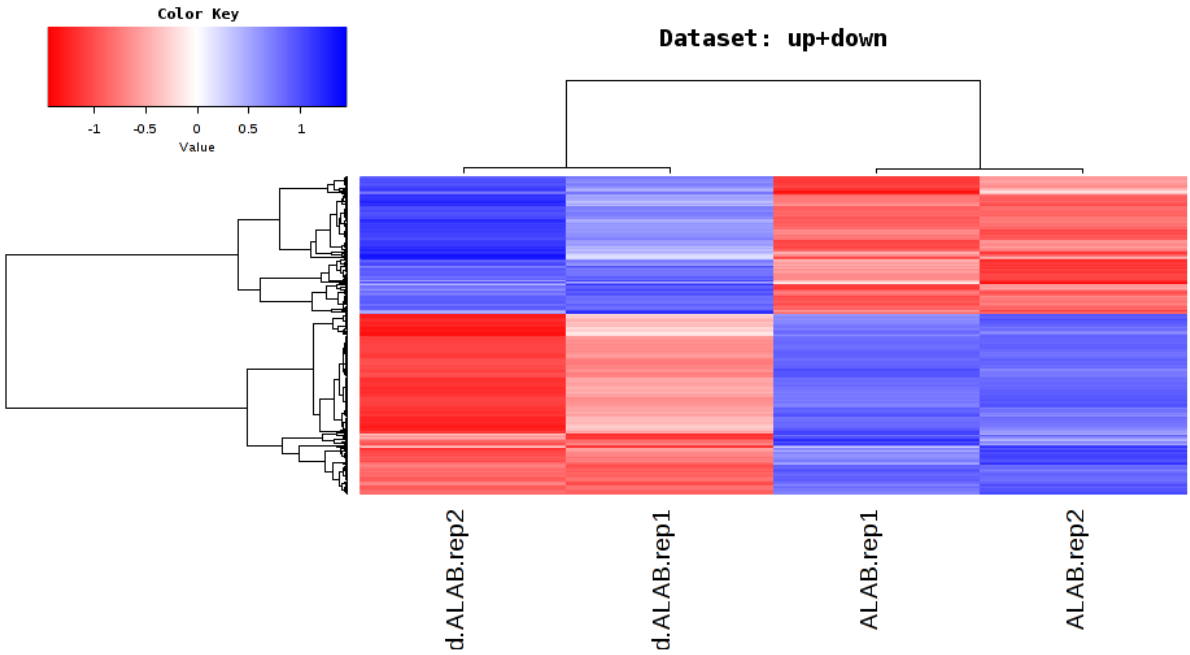

Figure 18: *Display of the heatmap within DEBrowser.*

- **mcquitty:** mcquitty linkage is when two clusters are joined, the distance of the new cluster to any other cluster is calculated by the average of the distances of the soon to be joined clusters to that other cluster.
- **median:** This is a different averaging method that uses the median instead of the mean. It is used to reduce the effect of outliers.
- **centroid:** The distance between cluster pairs is defined as the Euclidean distance between their centroids or means.

Used distance methods in heatmap

- **cor:** 1 - cor(x) are used to define the dissimilarity between samples. It is less sensitive to the outliers and scaling.
- **euclidean:** It is the most common use of distance. It is sensitive to the outliers and scaling. It is defined as the square root of the sum of the square differences between gene counts.
- **maximum:** The maximum distance between two samples is the sum of the maximum expression value of the corresponding genes.
- **manhattan:** The Manhattan distance between two samples is the sum of the differences of their corresponding genes.
- **canberra:** Canberra distance is similar to the Manhattan distance and it is a special form of the Minkowski distance. The difference is that the absolute difference between the gene counts of the two genes is divided by the sum of the absolute counts prior to summing.
- **minkowsky:** It is generalized form of euclidean distance.

You can also select to view an interactive version of the heatmap by clicking on the 'Interactive' checkbox on the left panel under the height and width options. Selecting this feature changes the heatmap into an interactive version with two colors, allowing you to select specific genes to be compared within the GO term plots (Figure 19). In order to use the interactive heatmap selection within your GO term query, you must use either the up+down dataset or the most varied dataset for the heatmap display. This will allow you to compare interesting clusters found within the heatmap within our GO Term analysis section which will be discussed later within the materials.

Principal Component Analysis (PCA) is another excellent method of checking replicates (Figure 20). PCA calculates the variance between all of the samples genes within your current comparison set and creates a two-dimensional graph to represent the proportion of variance explained in different components. Within the PCA plot section you can select the p-adjust value, fold change cut off value, which comparison set to use, which dataset to use, the height and width of the corresponding plots, as well as which principal components to analyze by changing the appropriate values on the left menu. If loaded with a batch meta data file, you can select from a specific dropdown within the PCA QC plots in order to group specific points on color or shape based on specified groups within your meta data file.

The next tab, 'GO Term', takes you to the ontology comparison portion of DEBrowser. From here you can select the standard dataset options such as p-adjust value, fold change cut off value, which comparison set to use, and which dataset to use on the left menu. In addition to these parameters, you also can choose from the 4 different ontology plot options: 'enrichGO' (Figure 21-22), 'Disease' (Figure 23-24), 'enrichKEGG' (Figure 25), and 'compareCluster'.
Selecting one of these plot options queries their specific databases with your current DESeq results. By selecting the 'selection' dataset on the left panel after selecting specific genes from the interactive heatmap, you will be able to compare your specific gene selection within the various GO Term databases.

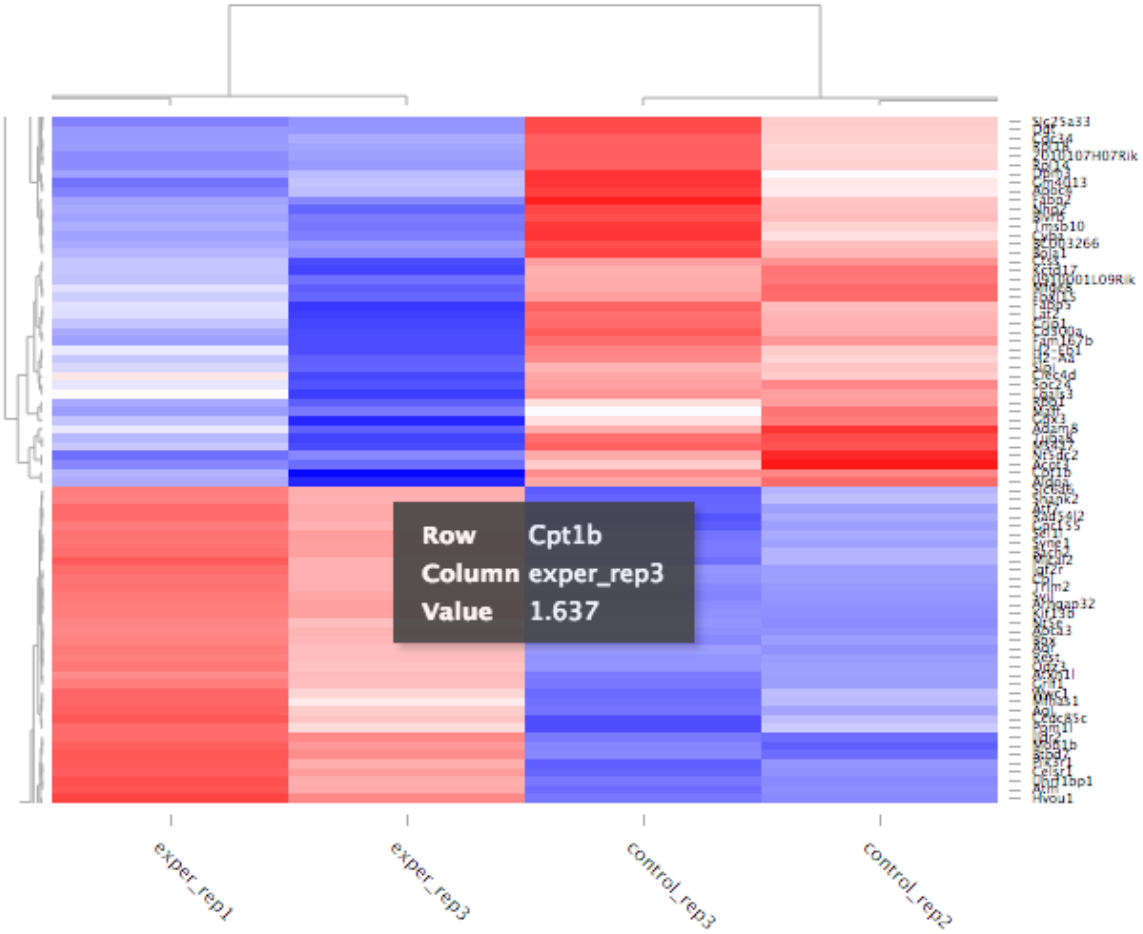

Figure 19: *View of the interactive Heatmap.*

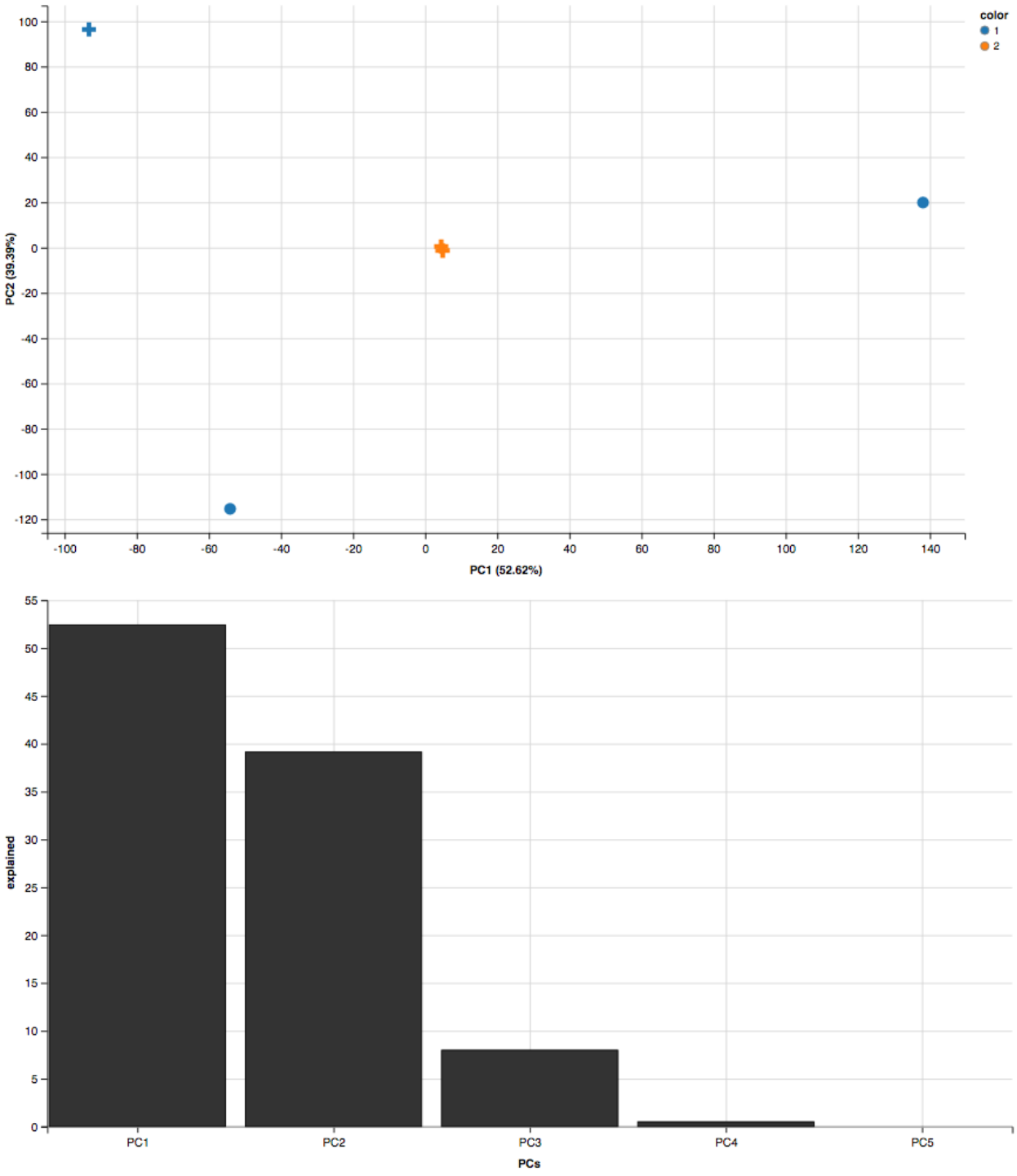

Figure 20: *Display of the PCA plot within DEBrowser.*

In order to use your selected genes from the interactive heatmap, you must first make your selection within the interactive heatmap. Next you will want to switch to the GO Terms tab and use the 'selected' dataset (Figure 26). Once all your other parameters have been selected, hit submit and you will use your selected genes from the interactive heatmap in your GO Term analysis.

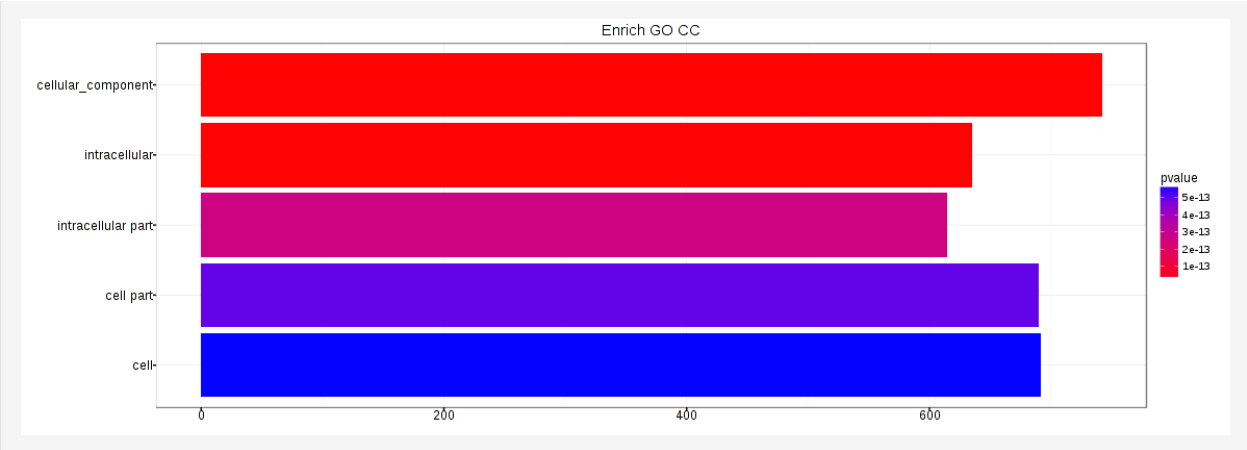

Figure 21: *Display of the GO Plot section within DEBrowser.*

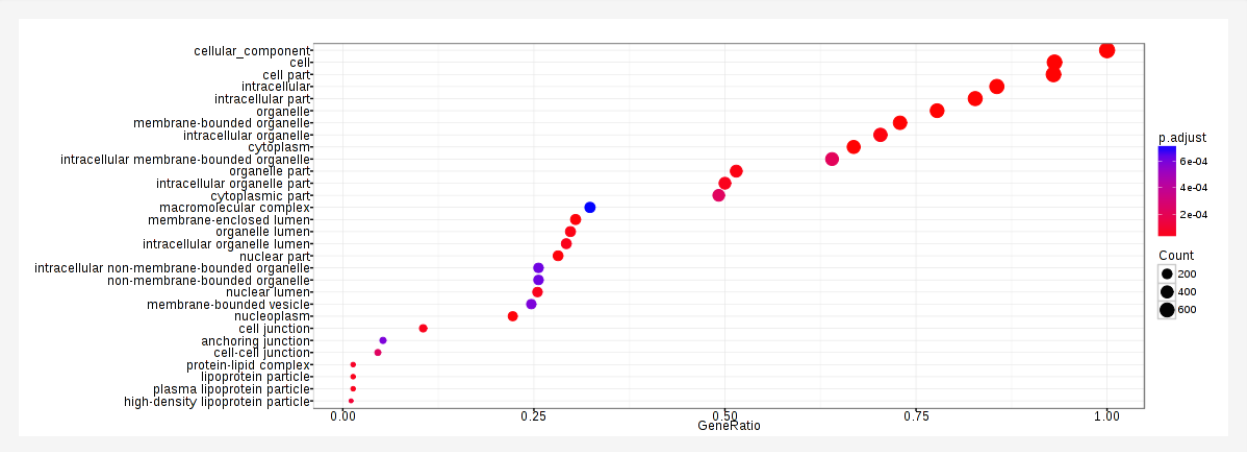

Figure 22: *Display of the GO dotplot section within DEBrowser.*

The last tab, 'Tables', contains various result information in table formats. The 'All Detected' option contains the list of all the genes within the TSV/CSV provided with the corresponding DESeq analyses. Up-regulated values are shown in green while down-regulated values are displayed in red. To view any particular dataset's custom options, the dataset type must be selected.

The 'Up+Down' option contains a list of all the up and down-regulated genes based on the options selected on the left panel (Figure 27). The 'Up' option contains a list of all the up-regulated genes based on the options selected on the left panel. The 'Down' option contains a list of all the down-regulated genes based on the options selected (Figure 28). The 'Selected' option contains the list of genes selected from the main plots section. By clicking and dragging your mouse on the main plot within the 'Main Plots' tab, you will then be able to see that selection in list form within the 'Selected' option. The 'Gene Set' option allows you to filter out gene data based on genes selected via a text box. To create a gene set, simply type the names of the genes you wish to view in the text box on the left panel in a comma- seperated format. You can also use regular expressions in order to search for specific gene sets (Figure 29-30). The 'Most Varied' option, much like the original QC 'Most Varied' tab, allows you to view the list of most varied genes based on user

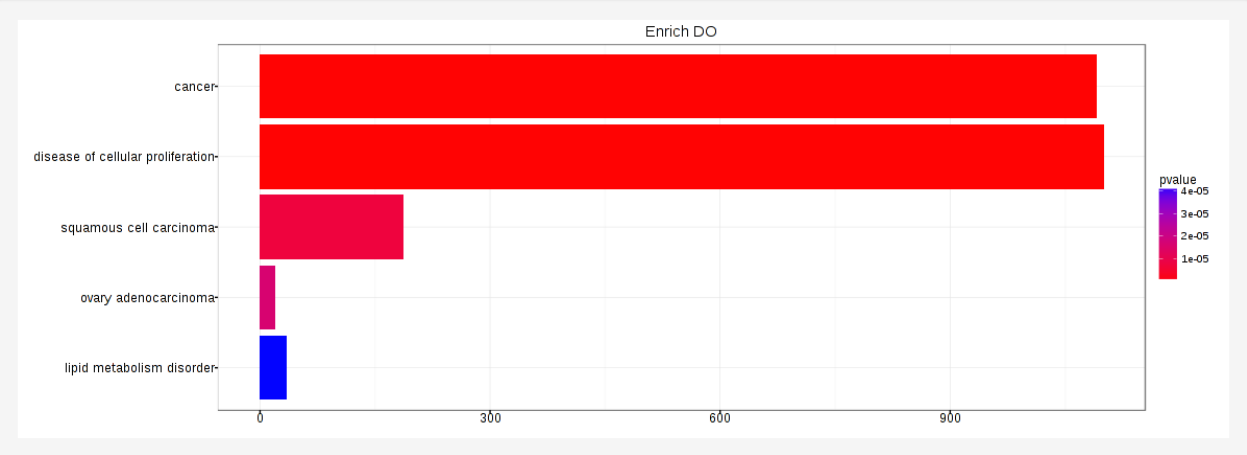

Figure 23: *Display of the DO plot section within DEBrowser.*

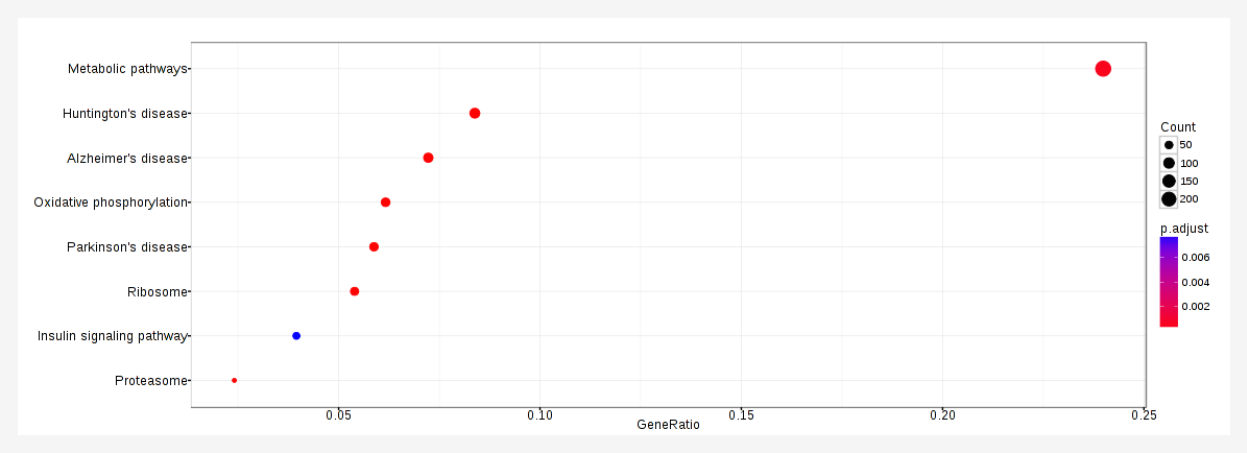

Figure 24: *Display of the DO dotplot section within DEBrowser.*

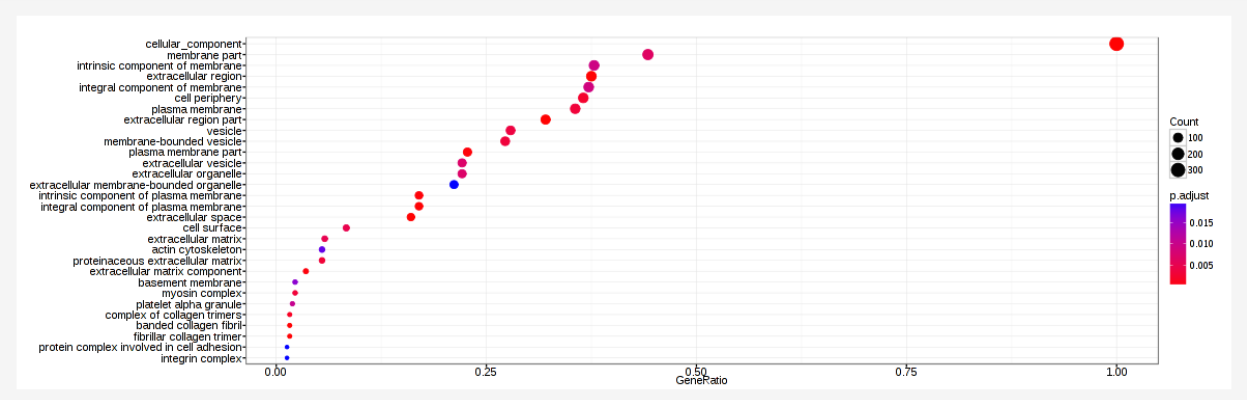

Figure 25: *Display of the KEGG dotplot section within DEBrowser.*

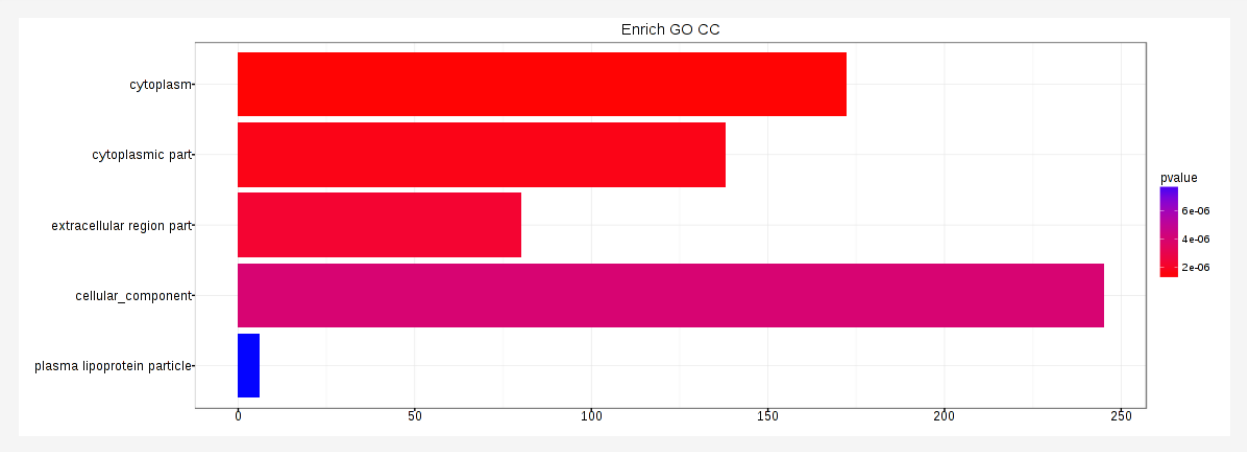

Figure 26: *Display of Heatmap selected enriched GO Term search.*

input parameters on the left panel. The 'Comparisons' option allows you to view the differences between your different condition comparisons (Figure 31). Comparisons between datasets are shown if at least one of the conditional comparisons has passed the padj value or fold change cut off.

It is also important to note that comparisons with only one sample cannot create statistically significant p-adjust values. The more replicates you have within a condition, the greater the statistical significance of your comparisons.

	exper_rep1	exper_rep2	control_rep1	control_rep2	padj	log2FoldChange	pvalue	foldChange	log10padj
Elf1	180.8138	143.9636	398.53221	366.8527	2.782311e-05	1.147924	1.284437e-06	2.2159486	4.555594
Arid5b	403.7134	306.8225	797.19256	764.4420	9.775937e-06	1.070373	3.857649e-07	2.0999760	5.009842
Mylk	779.6438	559.0541	1575.20152	1426.0087	2.519808e-06	1.100451	8.081353e-08	2.1442174	5.598633
Zmiz1	319.5362	250.2194	739.33626	533.4005	6.155624e-05	1.079484	3.320013e-06	2.1132799	4.210728
Acly	661.0852	536.2167	1341.36859	1255.0917	1.524147e-06	1.061008	4.435316e-08	2.0863886	5.816973
Rgs3	216.1112	179.7580	81.60404	86.5441	5.658512e-04	-1.118956	4.935626e-05	0.4604268	3.247298
0910001L09Rik	721.9910	1023.8724	383.93756	416.7290	2.360061e-05	-1.066351	1.062554e-06	0.4775253	4.627077
Ndufs7	1435.9452	2054.6520	757.84355	907.2896	3.076469e-05	-1.017440	1.458435e-06	0.4939921	4.511947
Ggnbp1	212.9142	224.4536	88.84741	115.4885	1.533815e-03	-1.008454	1.665031e-04	0.4970785	2.814227
Eif4ebp2	517.6240	316.7467	1172.23876	1068.9623	2.883636e-07	1.329759	6.260655e-09	2.5136065	6.540060

Showing 1 to 10 of 942 entries

Previous 1 2 3 4 5 … 95 Next

Figure 27: *Display of the up+down-regulated genes table.*

Lastly, the tables have a bunch of features that allow you to view your DESeq results more conveniently. By clicking on a column header, you can sort the data within the table based either alphabetical or numeric sorting. You can also enter a term, or regex query, within the search box on the left panel to filter for a specific gene within the table.

With that, you've now successfully navigated the DEBrowser and are ready to start inserting your own data files and browsing your own experiments. Enjoy the DEBrowser!

	exper_rep1	exper_rep2	control_rep1	control_rep2	padj	log2FoldChange	pvalue	foldChange	log10padj
Rgs3	216.11118	179.75800	81.604037	86.544103	5.658512e-04	-1.118956	4.935626e-05	0.4604268	3.247298
0910001L09Rik	721.99102	1023.87236	383.937559	416.728959	2.360061e-05	-1.066351	1.062554e-06	0.4775253	4.627077
Ndufs7	1435.94520	2054.65204	757.843548	907.289571	3.076469e-05	-1.017440	1.458435e-06	0.4939921	4.511947
Ggnbp1	212.91419	224.45360	88.847407	115.488503	1.533815e-03	-1.008454	1.665031e-04	0.4970785	2.814227
D330041H03Rik	92.58962	98.35942	8.247317	12.314260	5.569122e-09	-2.478989	6.745566e-11	0.1793700	8.254213
Ly6d	395.22784	388.31105	128.107546	156.367738	2.651719e-07	-1.370498	5.733473e-09	0.3867578	6.576472
Uap1l1	233.44308	369.46491	132.803304	101.824883	1.496916e-04	-1.227011	9.590205e-06	0.4272017	3.824803
Spc25	202.39166	255.24959	77.752635	116.657064	6.860687e-04	-1.123414	6.213270e-05	0.4590064	3.163632
Cyp4f18	36.06257	35.77581	6.468167	7.420492	4.936523e-03	-1.494130	7.111270e-04	0.3549949	2.306579
Fbxo31	1841.10380	1925.54094	859.437493	901.547298	1.106378e-07	-1.062080	2.115495e-09	0.4789410	6.956096

Showing 1 to 10 of 299 entries

Previous 1 2 3 4 5 ... 30 Next

Figure 28: *Display of the down-regulated genes table.*

	exper_rep1	exper_rep2	control_rep1	control_rep2	padj	log2FoldChange	pvalue	foldChange	log10padj
Aldh1l1	8115.545541	7267.081087	12052.495187	12176.759321	0.0014043816	0.62746129	1.498849e-04	1.5448441	2.8525149
Aldh6a1	4825.900008	4904.631608	7850.761026	7405.502504	0.0014870637	0.62119018	1.596155e-04	1.5381436	2.8276704
Aldh2	25663.337351	28503.354981	30879.056233	26706.294510	0.7834425633	0.07864782	6.452793e-01	1.0560278	0.1059928
Alg6	125.842970	110.216239	132.847750	176.298482	0.3969523503	0.34124412	2.262510e-01	1.2668486	0.4012616
Alkbh6	182.158696	218.504831	83.080049	81.933358	0.0003466188	-1.15863086	2.643143e-05	0.4479374	3.4601479
Alkbh8	14.852163	19.841149	58.743998	56.431901	0.0073881026	1.29690349	1.176288e-03	2.4570096	2.1314671
Alg5	688.766722	764.662513	433.130892	521.065485	0.0195940518	-0.58532319	4.025684e-03	0.6665000	1.7078757
Aldh1a7	1182.931572	1364.427715	2333.701065	2159.759108	0.0002193157	0.77987503	1.517016e-05	1.7169821	3.6589303
Als2cl	532.409227	432.957083	628.729357	442.004044	0.7097486112	0.13707623	5.484711e-01	1.0996743	0.1488954
Alox15	4.680895	1.981245	5.298022	2.412806	1.0000000000	0.04664030	9.084498e-01	1.0328568	

Showing 1 to 10 of 58 entries

Previous 1 2 3 4 5 6 Next

Figure 29: *Display of the geneset input box.*

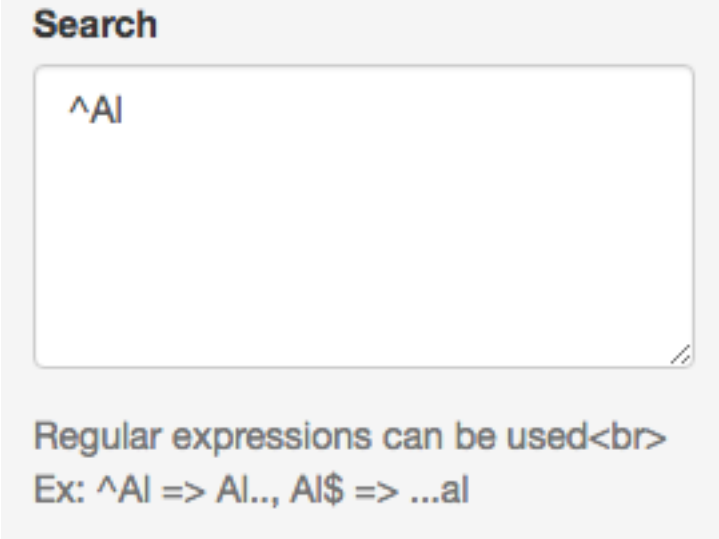

Figure 30: *Display of the gene set search of the term '^al'.*

	exper_rep1	exper_rep2	exper_rep3	control_rep1	control_rep2	control_rep3	foldChange.C1.vs.C2	padjC1.vs.C2
0610005C13Rik	4221.00	5483.01	2072.16	3242.50	2283.14	2548.00	0.4908259	1.167991e-07
0910001L09Rik	609.00	1031.00	422.00	527.00	358.00	429.00	0.4775253	2.360061e-05
1110007C09Rik	145.00	211.01	77.00	75.00	52.00	67.00	0.3422499	2.914530e-06
1110008P14Rik	340.00	523.00	185.00	238.00	221.00	198.00	0.4902069	4.067816e-04
1110038B12Rik	247.00	339.83	126.57	152.00	140.97	88.00	0.4614585	2.277593e-04
1300002K09Rik	796.00	1009.00	401.00	2932.57	1765.00	1637.15	2.0837019	1.055394e-07
1700020O03Rik	83.05	57.90	58.02	228.06	152.47	126.64	2.0081597	6.102020e-03
1810044D09Rik	73.00	90.00	36.00	32.00	27.00	22.00	0.3810179	9.854926e-04
2010003K11Rik	885.00	826.00	403.00	375.00	271.39	334.00	0.3339000	1.219213e-11
2010107H07Rik	372.00	547.00	196.00	240.00	228.00	164.00	0.4705770	1.437859e-04

Showing 1 to 10 of 942 entries Previous 1 2 3 4 5 ... 95 Next

Figure 31: *Condition comparisons table within DEBrowser.*

Case Study

Taking a look at the case study (Vernia S. et al 2014), Multiple heatmaps were created to display findings within the research. The heatmaps generated for the study were customized to a high level of specificity. However, using a sample dataset generated from this study, it is possible to recreate similar heatmaps (Figure 32-33) displayed within the studies findings.

The main difference between the heatmaps created within DEBrowser and the heatmaps created within the research paper is that the clustering method used within the paper was a k-means method with k equaling 6.

The JNK2 Knock-out versus JNK1 Knock-out:

High fat diet JNK1 knock-out and JNK2 knock-out samples compared against high fat diet wild type samples showed a stronger effect from JNK2 KO.
From the figures below, JNK2 KO has a stronger effect than JNK1 KO samples. There are 177 genes (Figure 34) that have padj < 0.01 and $|\log2 \text{ foldchange}| > 1$ in the JNK2 KO comparison while there are only 17 genes (Figure 35) detected in the JNK1 KO comparison with the same cutoffs.

JNK1 and JNK2 serve partially redundant functions:

High fat diet JNK1 and High fat diet JNK2 double KO has 1018 significantly different genes. When we compare HFD JNK1 KO only (177 Genes) and HFD JNK2 KO only (17 genes) with HFD wild type side-by-side, most of the up and down regulated genes are not overlapping. Up regulated genes (Figure 36.) and down regulated (Figure 37.) in JNK1 KO was easy to analyze for these gene comparisons. There is only 1 gene overlapping out of the 17 that were significantly different in JNK1 KO comparisons with padj < 0.01 and $|\log2\text{foldchange}| > 1$ cutoffs. It shows that both individual KO might have individual functions in addition to their redundant functions. When we looked at the genes in JNK1 KO in the KEGG database, they are enriched in "Fatty acid elongation". JNK2 KO are enriched in "PPAR signaling pathway" and "Biosynthesis of unsaturated fatty acids". DEBrowser's powerful comparison function makes different condition comparisons and running GO Term analysis on selected genes much easier.

Comparing the HFD wild type and the normal chow wild type also shows significant differences between regulated genes (Figure 38). Expanding on the analysis further, the upregulated genes analyzed are then compared to KEGG and Disease ontologies to show a variety of metabolism related correlations (Figures 39-40).

Using the 'advanced demo' dataset we mentioned earlier, you too can recreate these tables using the same data. Browsing, changing parameters, and creating unique plots to view and analyze data can be a creative way to recreate the same analytical results produced. DEBrowser can be used in multiple ways to check the reproducibility of research results in a highly interactive format!

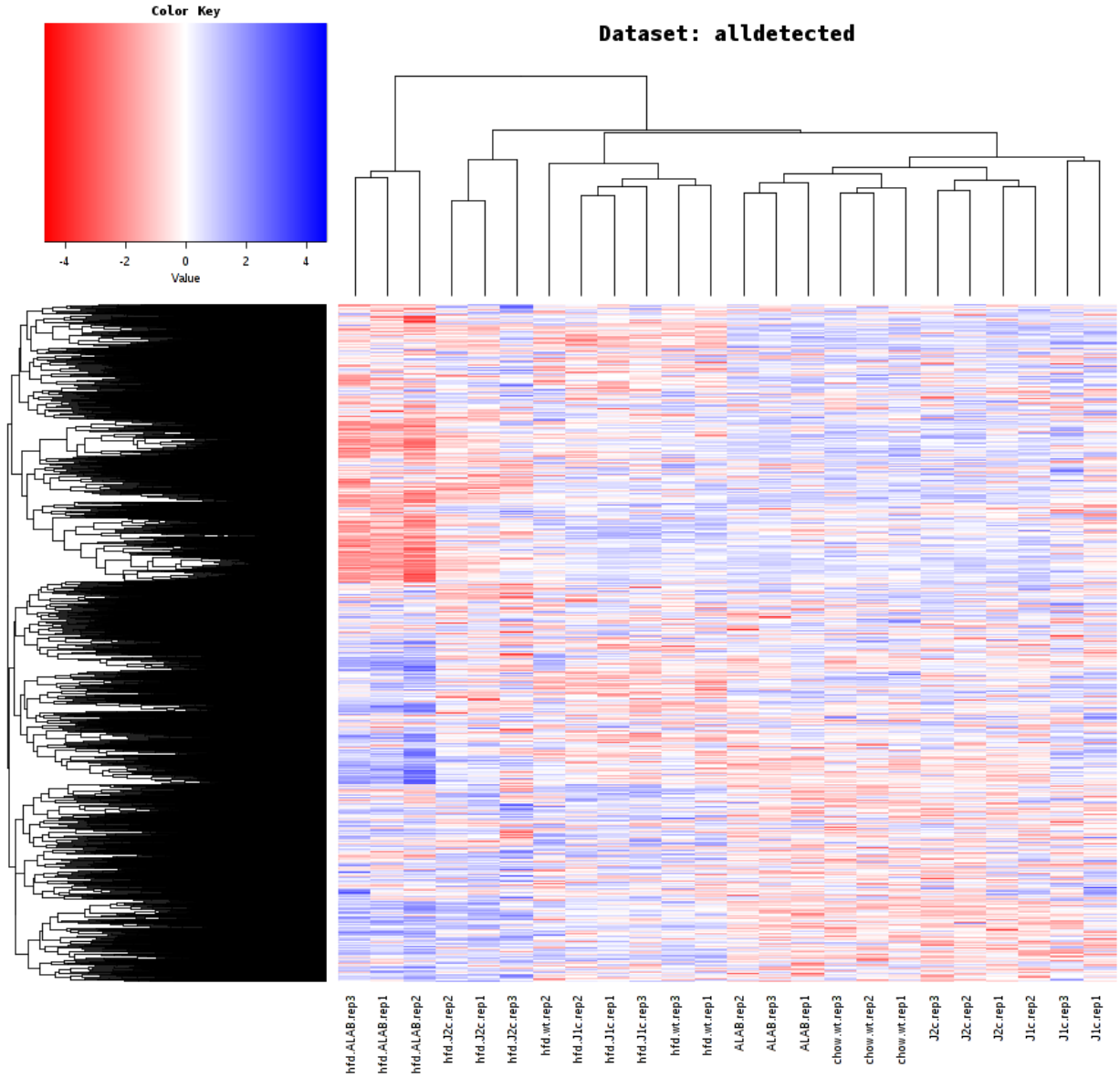

Figure 32: *All detected genes heatmap using case study data.*

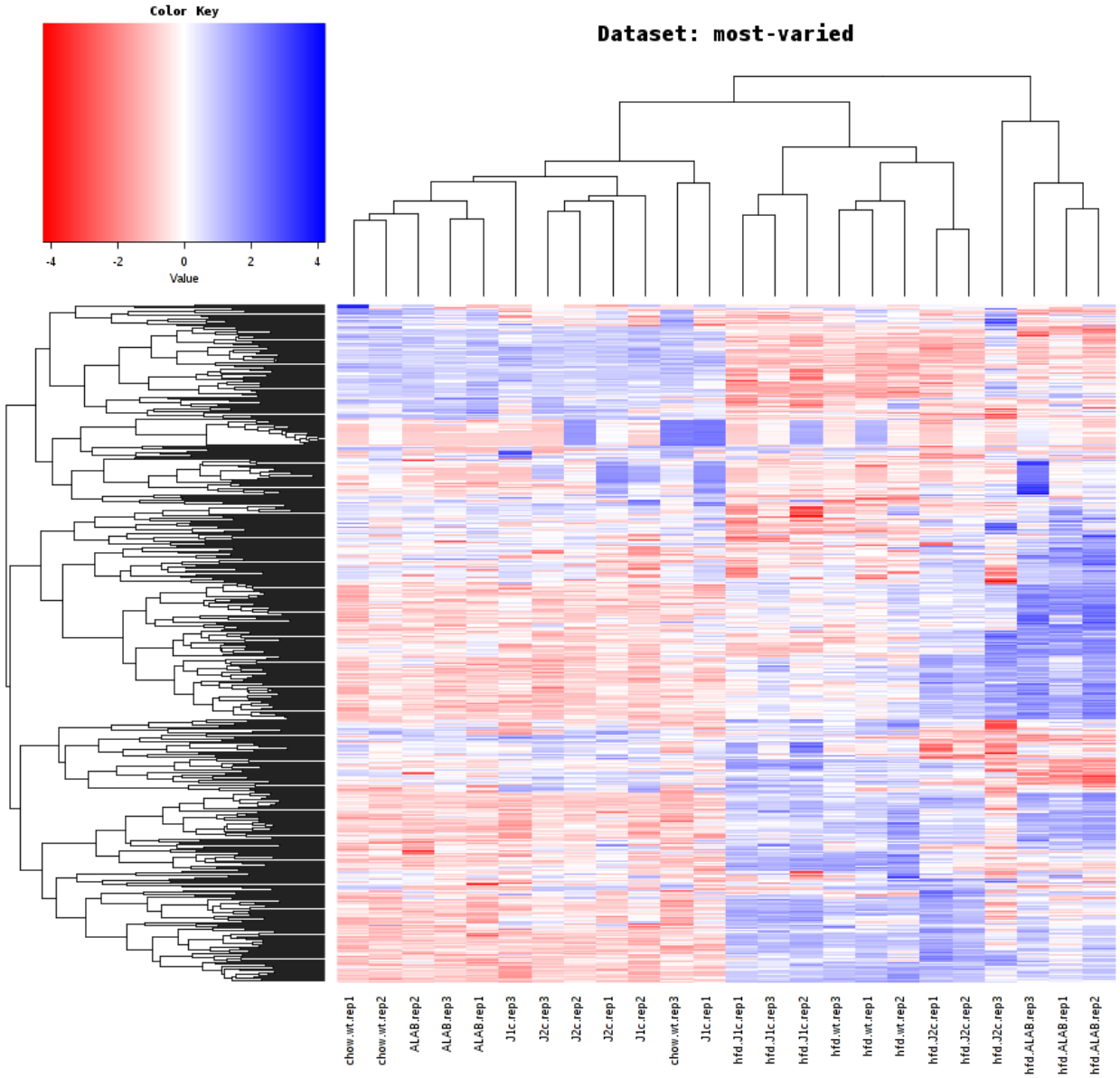

Figure 33: *Most varied genes heatmap using case study data.*

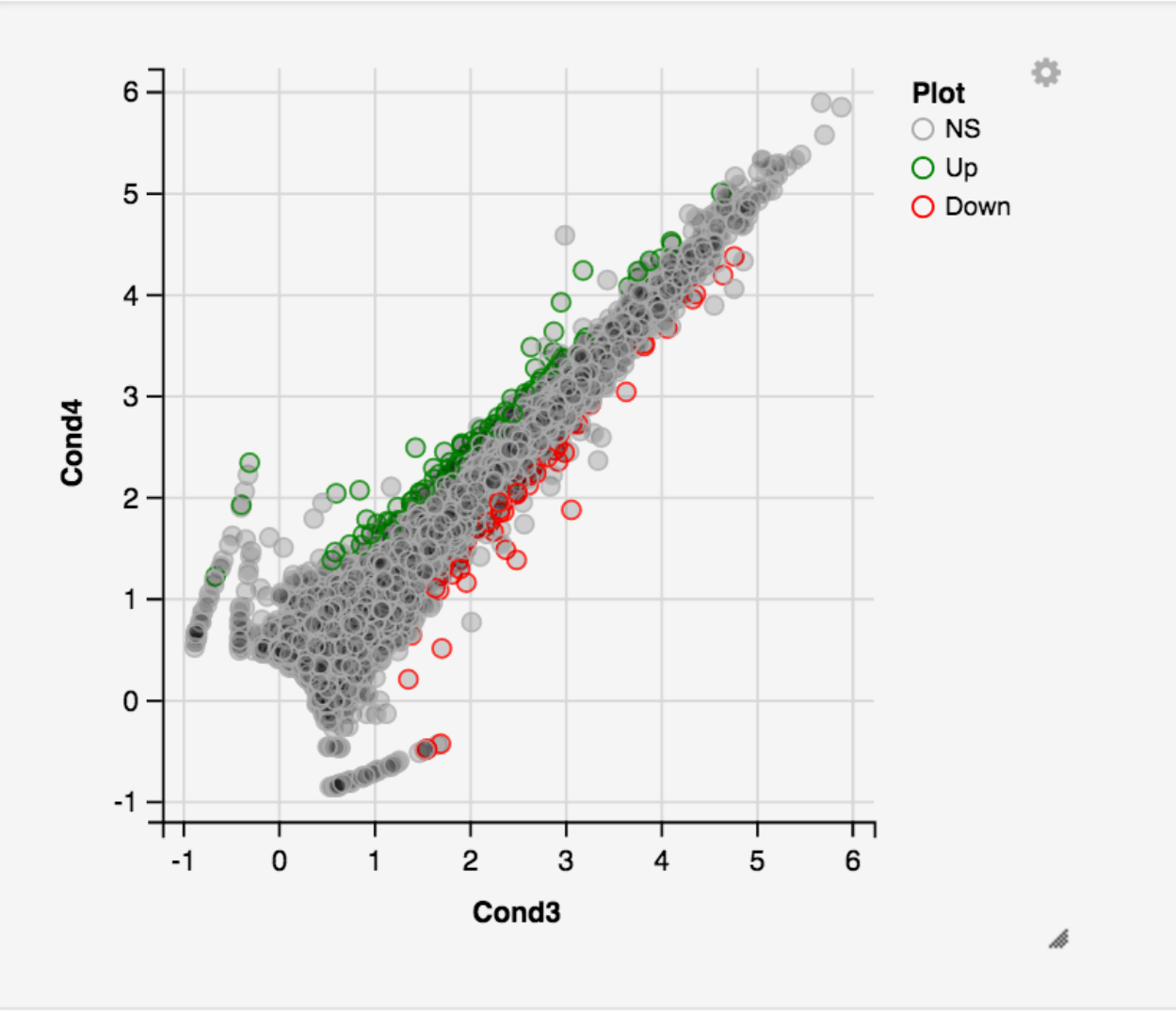

Figure 34: *High fat diet JNK2 vs. High fat diet wild type.*

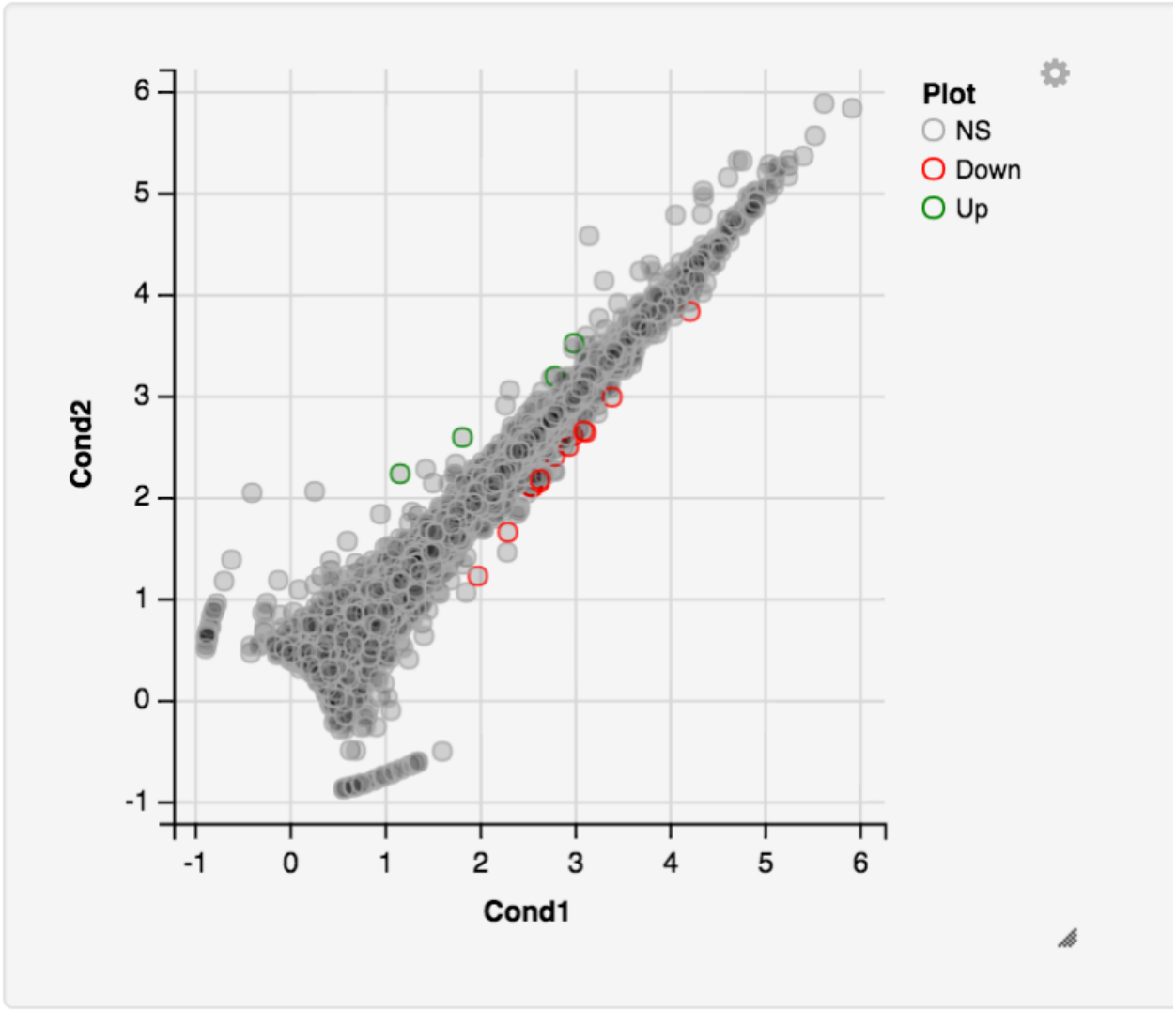

Figure 35: *High fat diet JNK1 vs. High fat diet wild type.*

Show 100 entries Search:

	hfd.J1c.rep1	hfd.J1c.rep2	hfd.J1c.rep3	hfd.J2c.rep1	hfd.J2c.rep2	hfd.J2c.rep3	hfd.wt.rep1	hfd.wt.rep2	hfd.wt.rep3	foldChange.C1.vs.C2	padjC1.vs.C2	foldChange.C3.vs.C4	padjC3.vs.C4
Fam25c	15.00	15.00	12.00	124.00	89.00	92.00	216.00	168.00	141.00	4.6493331	1.400870e-31	1.6028993	1.512849e-02
Gm3219	68.25	69.28	56.68	362.35	311.87	173.43	509.29	297.76	401.43	4.0244553	2.817700e-35	1.2950809	4.094627e-01
D14Ertd449e	1240.75	976.74	735.96	3918.61	3128.40	3210.89	4265.32	3299.84	2634.10	2.8667412	2.886716e-31	0.9789110	9.500865e-01
Elovl3	831.00	580.00	461.00	1477.65	1081.22	2136.59	2132.00	1628.80	1095.92	2.1355465	8.229894e-12	1.0287723	9.461544e-01
Avpr1a	235.00	159.00	112.00	110.00	143.00	183.00	463.00	295.99	289.00	1.7441710	1.195739e-05	2.0517755	9.158149e-05
Slc30a10	307.47	266.48	256.44	266.84	217.63	267.56	923.58	383.01	472.37	1.7256018	1.195739e-05	2.0709712	1.665522e-06
Orm2	754.78	665.65	1314.24	724.06	306.71	340.54	4410.06	2834.87	2062.15	1.5809712	2.091627e-02	5.1922785	3.825391e-25
Sdf2l1	226.20	206.25	135.01	158.90	98.20	64.25	348.52	392.51	300.08	1.5625632	5.993858e-03	2.5996620	3.499584e-06
BC023105	127.09	106.17	84.00	86.18	62.03	78.66	200.62	270.87	137.07	1.5218190	3.221938e-02	2.2481111	1.673786e-04
AK050336	30.97	29.00	27.00	23.00	12.00	23.20	74.00	48.04	53.00	1.4573142	9.563531e-02	2.2950538	3.692708e-04
Saa1	6335.89	9875.58	3470.68	2526.71	1144.84	1115.50	23878.23	16527.70	12201.84	1.4531444	9.965658e-02	7.8738119	6.125568e-40

Figure 36: *Upregulated genes in hfd JNK1 KO (C1) vs. hfd wt (C2) DE comparison shows 4 upregulated genes (padj <0.01 and |log2foldchange| > 1).*

	hfd.J1c.rep1	hfd.J1c.rep2	hfd.J1c.rep3	hfd.J2c.rep1	hfd.J2c.rep2	hfd.J2c.rep3	hfd.wt.rep1	hfd.wt.rep2	hfd.wt.rep3	foldChange.C1.vs.C2	padjC1.vs.C2	foldChange.C3.vs.C4	padjC3.vs.C4
Osbpl5	192.00	232.00	157.00	213.00	108.00	50.34	65.00	34.00	43.00	0.3506937	1.526990e-15	0.5139084	1.729129e-02
Col27a1	995.04	905.55	673.71	786.17	470.92	102.52	471.94	238.36	295.84	0.4146506	6.105474e-20	0.8719148	7.926571e-01
Acot3	1662.52	1251.57	1028.69	2253.47	1443.69	607.08	674.99	240.57	482.20	0.4146852	1.759343e-13	0.5444698	5.661891e-02
Sik1	464.02	414.89	394.99	212.07	269.00	500.03	233.00	105.00	140.00	0.4153451	1.526990e-15	0.5417091	8.707424e-03
Cyp2a4	130.16	94.46	65.30	117.80	46.75	43.74	41.18	11.36	6.16	0.4332577	6.448177e-08	0.4570055	5.922631e-03
Fmo2	403.51	364.96	262.00	286.99	116.00	152.00	132.10	122.76	129.87	0.4489962	5.222289e-11	0.7665225	3.781453e-01
Lrtm1	645.05	442.07	227.01	116.91	76.01	3.35	213.95	91.23	143.56	0.4506913	8.928251e-09	1.1833620	6.842123e-01
Wee1	398.88	327.00	342.98	507.00	312.00	106.00	119.00	101.00	156.98	0.4507056	1.574416e-09	0.6675139	2.853198e-01
Ces1d	19488.06	15719.94	13625.63	18003.98	11312.56	4566.26	9995.39	4547.92	6915.28	0.4558083	1.648869e-16	0.7623613	5.439395e-01
Usp2	1440.01	1229.24	1023.00	1123.60	641.98	329.20	364.52	366.97	592.00	0.4716093	3.463335e-08	0.8055423	6.551508e-01
4931408D14Rik	791.30	493.67	560.07	606.41	368.68	648.13	275.74	221.04	266.18	0.4719165	4.929441e-11	0.5166766	1.037990e-05
Thrsp	2653.49	3055.07	1652.57	2164.19	942.13	81.00	999.07	942.00	975.00	0.4730771	1.369220e-09	0.9863414	9.828469e-01
Per2	1211.25	803.89	864.96	651.00	430.00	272.00	460.37	428.00	339.99	0.4755592	9.196677e-12	0.9243280	8.548310e-01
Pcsk9	390.00	412.00	142.00	20.00	43.00	20.00	120.01	115.00	95.00	0.5042294	1.195739e-05	2.4750695	2.670572e-04
Osbpl3	390.00	313.00	376.00	576.00	305.00	479.00	341.00	92.00	107.00	0.5660530	4.624714e-04	0.4446031	8.402982e-05
Unc5b	177.31	128.57	131.14	320.22	169.86	149.89	68.28	52.00	85.07	0.5827592	9.233318e-04	0.4185191	4.791351e-05
AB343393	19.54	37.22	10.37	47.26	26.59	37.84	0.00	0.00	0.00	0.5915025	1.000000e+00	0.2005701	3.163872e-11
AB343394	19.54	37.22	10.37	47.26	26.59	37.84	0.00	0.00	0.00	0.5915025	1.000000e+00	0.2005701	3.163872e-11
Timp3	368.00	392.46	361.18	584.72	324.05	510.44	268.08	153.40	229.66	0.6116067	3.864526e-04	0.4959968	1.447491e-06
Hist1h2ba	9.58	23.08	26.80	39.71	45.37	66.22	0.00	0.00	0.00	0.6131840	1.000000e+00	0.1648057	2.818024e-14
Pla2g6	941.00	837.00	795.99	1689.80	1086.00	783.00	851.46	375.05	427.00	0.6298882	1.300244e-04	0.4988761	3.323478e-04
Maff	90.00	99.00	105.00	211.99	111.02	117.00	42.00	37.00	67.00	0.6533288	4.014431e-02	0.4475340	5.029121e-04
Lamb3	438.64	399.06	249.71	843.91	564.02	1203.94	354.71	158.72	190.53	0.6624739	6.427730e-03	0.3070825	1.003725e-14
Fads2	4757.16	4011.96	3782.38	6731.47	4543.52	9614.60	4060.33	2620.51	2709.97	0.7201717	2.103450e-04	0.4825753	1.925965e-07
Acot2	417.04	436.58	297.37	1022.44	668.78	1364.51	423.82	235.13	198.02	0.7346508	7.816503e-02	0.3147538	1.177744e-15
Cxcl1	448.00	719.47	443.00	216.00	110.19	148.00	418.00	318.00	392.02	0.7350129	1.536542e-01	2.2167020	4.240696e-07
Ppp1r3b	6718.30	5063.11	3485.45	2420.50	1322.96	1383.59	4181.64	3248.83	3658.24	0.7463974	4.692243e-02	2.0449169	1.975298e-06
Nt5dc2	100.47	76.57	78.48	181.78	109.09	218.67	94.56	33.50	51.78	0.7464653	2.734181e-01	0.4104510	3.850738e-06
Sparcl1	104.00	79.00	69.00	180.00	86.00	216.00	78.00	37.00	59.00	0.7478406	2.626941e-01	0.4393275	5.256841e-05

Figure 37: *Downregulated genes in hfd JNK1 KO (C1) vs. hfd wt (C2) DE comparison shows 13 downregulated genes (padj <0.01 and |log2foldchange| > 1). Only one of them is in JNK2 KO (C3) vs. hfd wt (C4) DE comparison.*

DEBrowser vs other Differential Expression analysis software:

The comparison table (Figure 41) displays multiple comparisons between debrowser and other various methods of viewing Differential Expression Analysis results.

Some of the comparisons can be viewed either within the tool itself or within some of the figures provided. A multiple tool comparisons can be observed within figure 35, an interactive visualization of gene highlighting can be observed for figures 13-15, and an interactive visualization of biological variation or condition comparisons can also be observed for figure 35.

For more information on MeV (Howe et al., 2011) please visit this link: MeV

For more information about Chipster (Kallio et al., 2011), please visit this link: Chipster

For more information about Galaxy (Giardine et al., 2005), please visit this link: Galaxy

For more information about CummeRBund (Trapnell et al., 2012), please visit this link: CummeRbund

Batch Effect Correction Example

Batch effects can have negative effects on your overall data as a whole. If samples were not handled at the same time, or just do not have the sample level of technical variance as the rest of your samples, this can lead to specific batch effects that might skew test results. In Figures 42-45, we show how forgetting to deposit one enzyme can have skewing effects as well as the results of batch effect correction using ComBat.

Future Plan:

Future plans will include the following:

```
* Venn Diagrams to compare overlapping differentially expressed genes in different condition comparison
* Increase in the number of used clustering methods.
* GO term analysis gene lists will be added for found GO categories.
```

References

1. Anders,S. et al. (2014) HTSeq - A Python framework to work with high-throughput sequencing data.
2. Chang,W. et al. (2016) shiny: Web Application Framework for R.
3. Chang,W. and Wickham,H. (2015) ggvis: Interactive Grammar of Graphics.
4. Giardine,B. et al. (2005) Galaxy: a platform for interactive large-scale genome analysis. Genome Res., 15, 1451–1455.
5. Howe,E.A. et al. (2011) RNA-Seq analysis in MeV. Bioinformatics, 27, 3209–3210.
6. Kallio,M.A. et al. (2011) Chipster: user-friendly analysis software for microarray and other high-throughput data. BMC Genomics, 12, 507.
7. Li,B. and Dewey,C.N. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics, 12, 323.
8. Love,M.I. et al. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol., 15, 550.
9. Reese,S.E. et al. (2013) A new statistic for identifying batch effects in high-throughput genomic data that uses guided principal component analysis. Bioinformatics, 29, 2877–2883.
10. Reich,M. et al. (2006) GenePattern 2.0. Nat. Genet., 38, 500–501.
11. Risso,D. et al. (2014) Normalization of RNA-seq data using factor analysis of control genes or samples. Nat. Biotechnol., 32, 896–902.
12. Ritchie,M.E. et al. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res., 43, e47–e47.
13. Trapnell,C. et al. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat. Protoc., 7, 562–578.
14. Vernia,S. et al. (2014) The PPARα-FGF21 hormone axis contributes to metabolic regulation by the hepatic JNK signaling pathway. Cell Metab., 20, 512–525.
15. Murtagh, Fionn and Legendre, Pierre (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? Journal of Classification 31 (forthcoming).
16. Johnson et al. (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics, 8, 118-127.

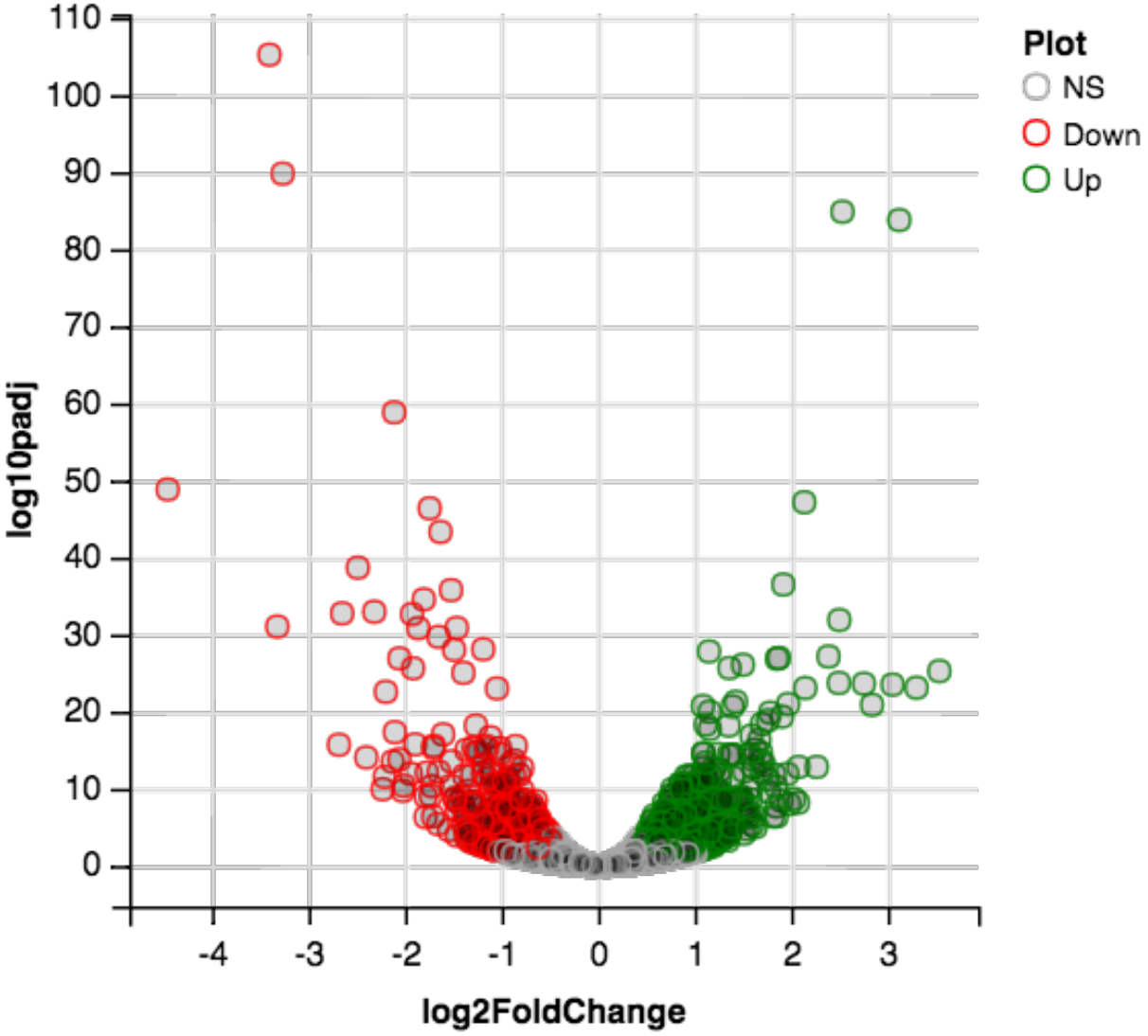

Figure 38: *Up and Down regulated genes volcano plot of HFD WT vs Chow WT.*

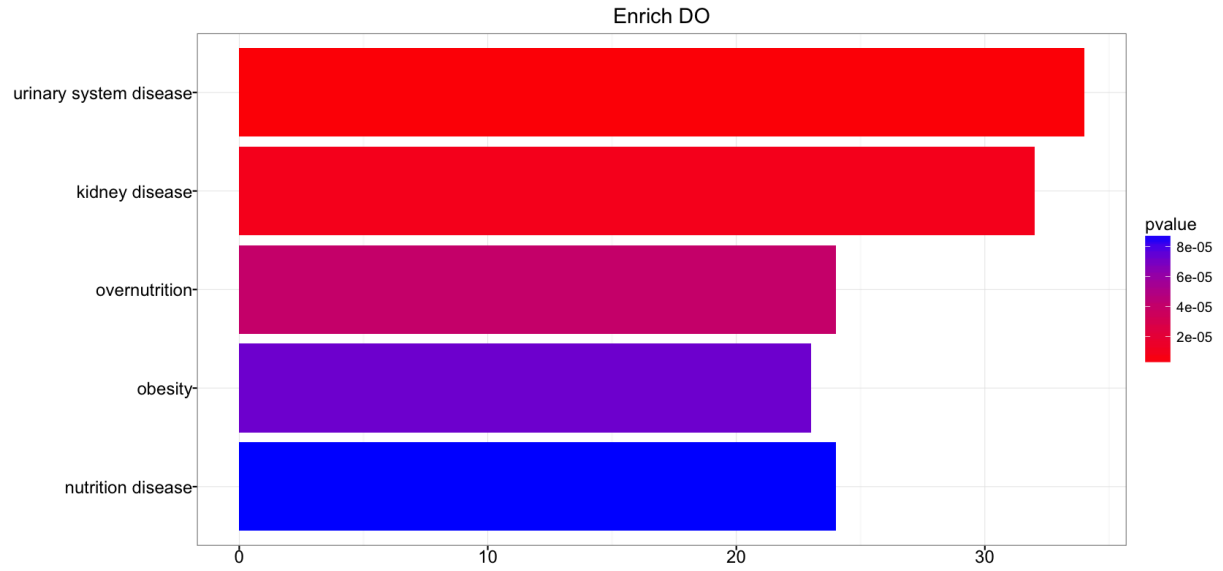

Figure 39: *HFD upregulated gene list used for DO enrichment*

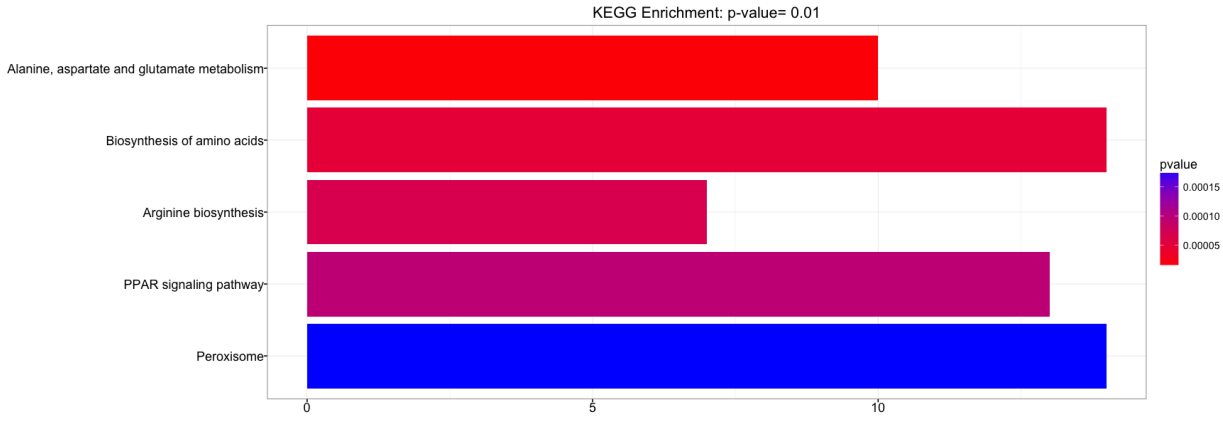

Figure 40: *HFD upregulated gene list used for KEGG enrichment*

	DEBrowser	MeV	Chipster	Galaxy	CummeRBund
Easy Install	✓	✓	✗	✓	✓
Full R Support	✓	✗	✗	✗	✓
Web Interface	✓	✓	✓	✓	✗
Local Install	✓	✓	✓	✗	✓
DESeq2	✓	✓	✓	✓	✗
EDGeR	✓	✓	✓	✓	✗
Limma	✓	✓	✗	✓	✗
CuffDiff	✗	✗	✓	✓	✓
Gene Searching with Regular Expression	✓	✓	✗	✗	✗
Gene Subset Selection Using Interactive Heat Map	✓	✗	✗	✗	✗
Ability to Zoom on Interactive Plots	✓	✗	✗	✗	✗
JSON Object Reading Capabilities	✓	✗	✗	✗	✗
Interactive All-to-All Scatter Plots	✓	✗	✗	✗	✗
Multiple DE Tool Comparison Tables	✓	✗	✗	✗	✗
Cluster Selection on Heatmap for GO Term Analysis	✓	✗	✗	✗	✗
On-The-Fly Gene details for Plots	✓	✗	✗	✗	✗
Interactive Visualization of Biological Variation/Condition Comparison	✓	✗	✗	✗	✗

Figure 41: *Comparison table of DEBrowser, MeV, Chipster, Galaxy, and CummeRBund*

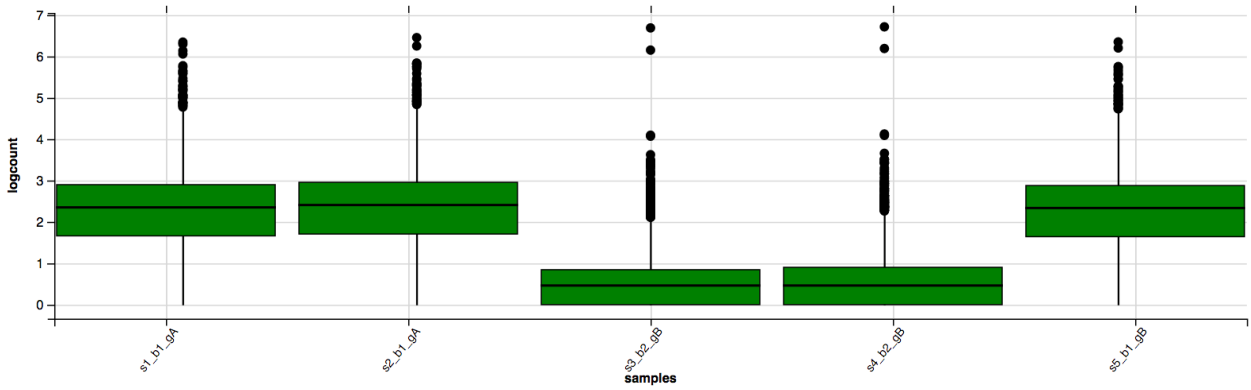

Figure 42: *IQR of samples before batch correction*

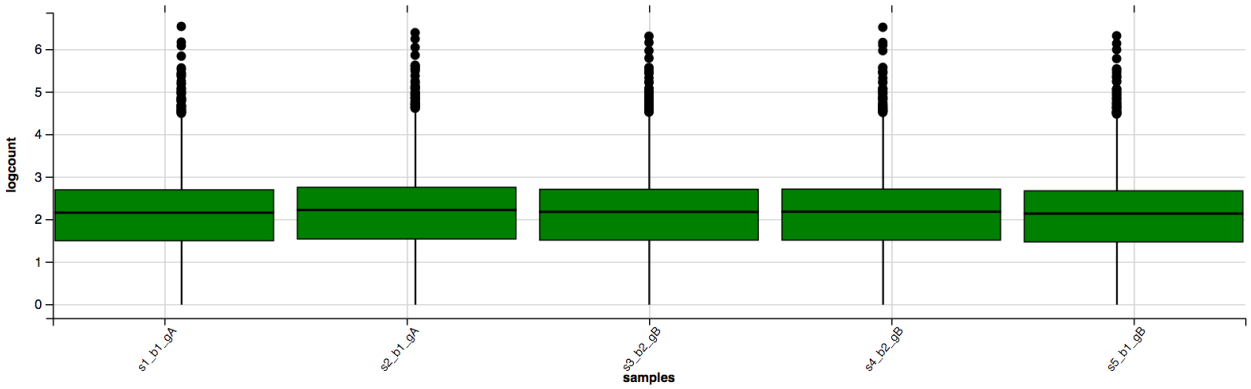

Figure 43: *IQR of samples after batch correction*

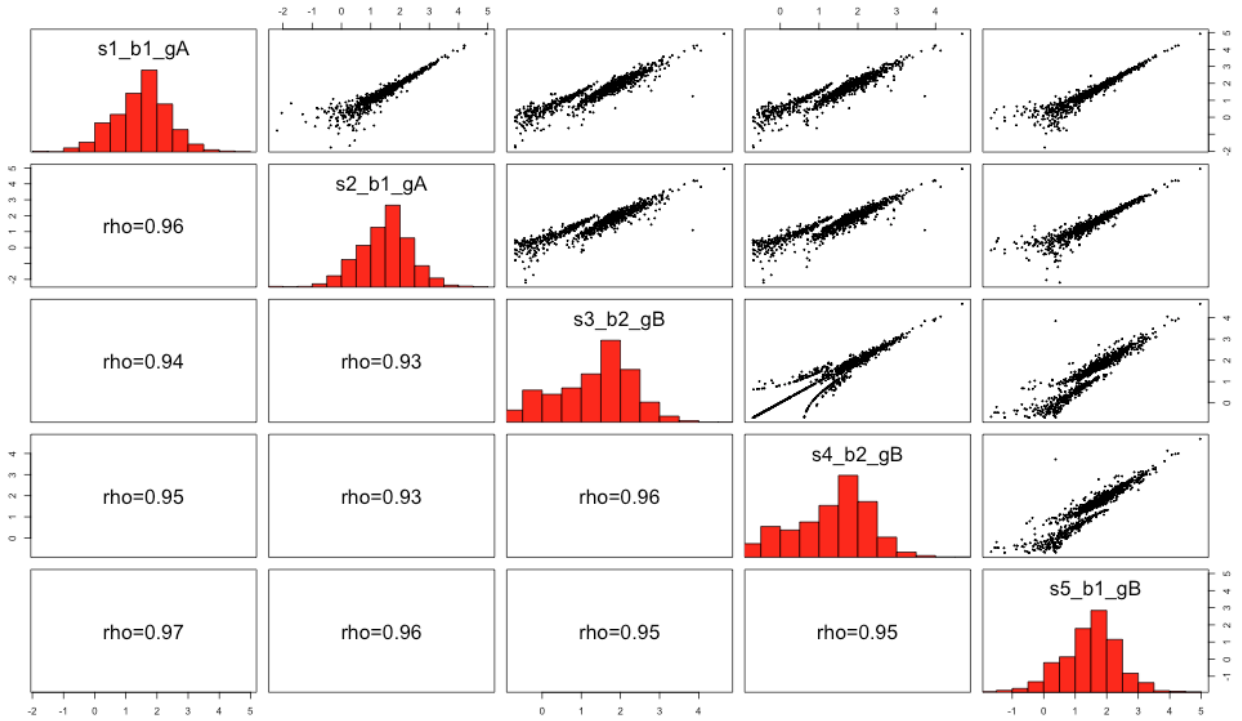

Figure 44: *All-2-All scatter plots showing QC without batch correction.*

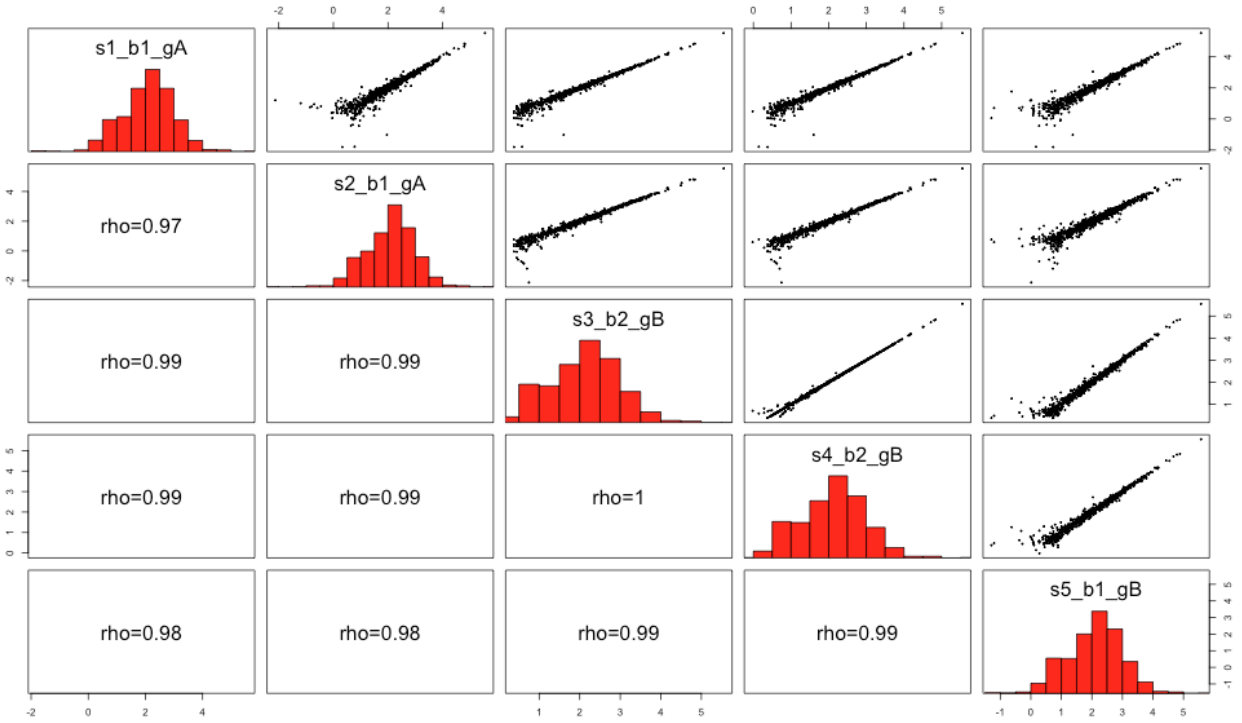

Figure 45: *All-2-All scatter plots showing QC with batch correction.*