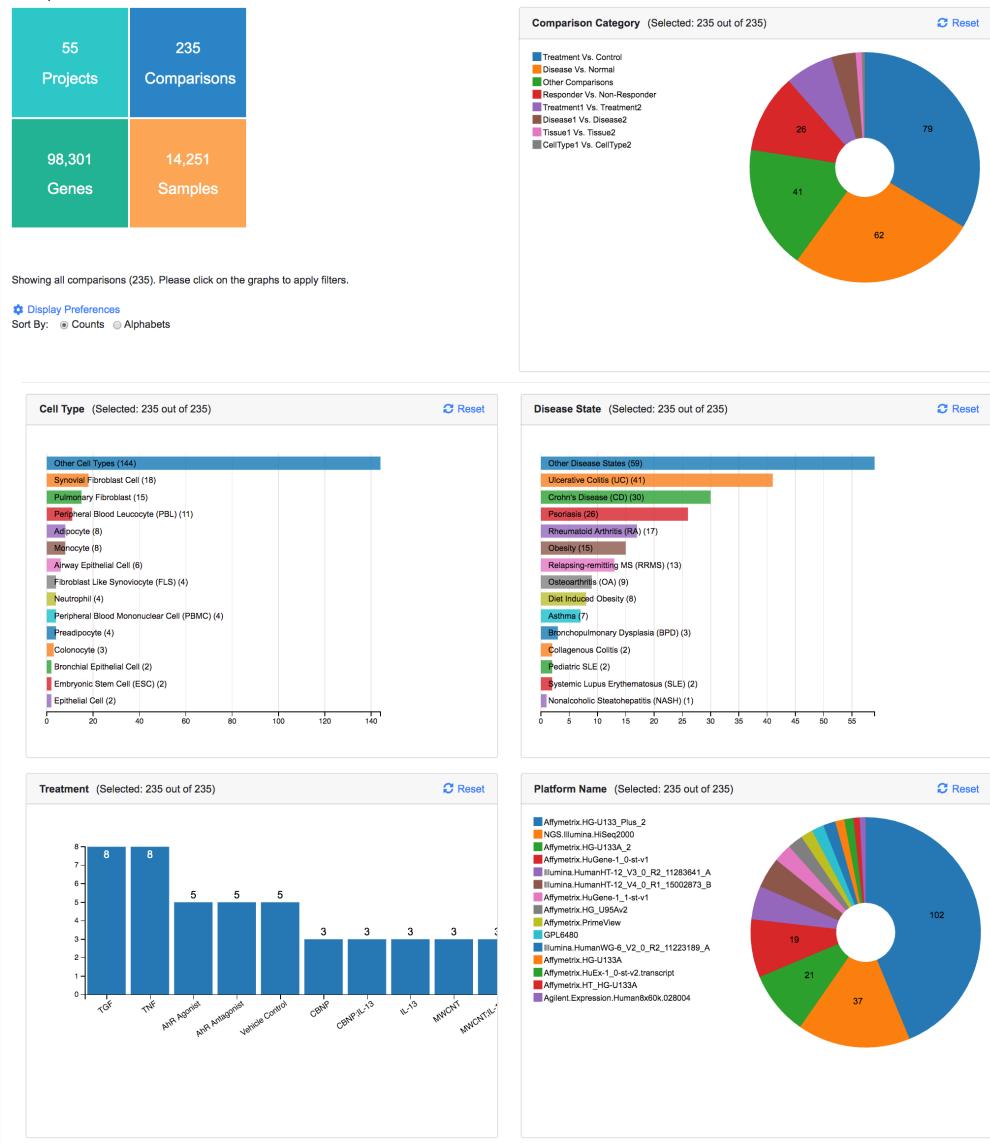


# OmicsView Supplementary



<http://omicsview.org/>

From the login page, users can use their email to register an account (recommended), as this enables users to save results and upload their own data. Otherwise, a guest account can be used to view public data.

# 1 Contents

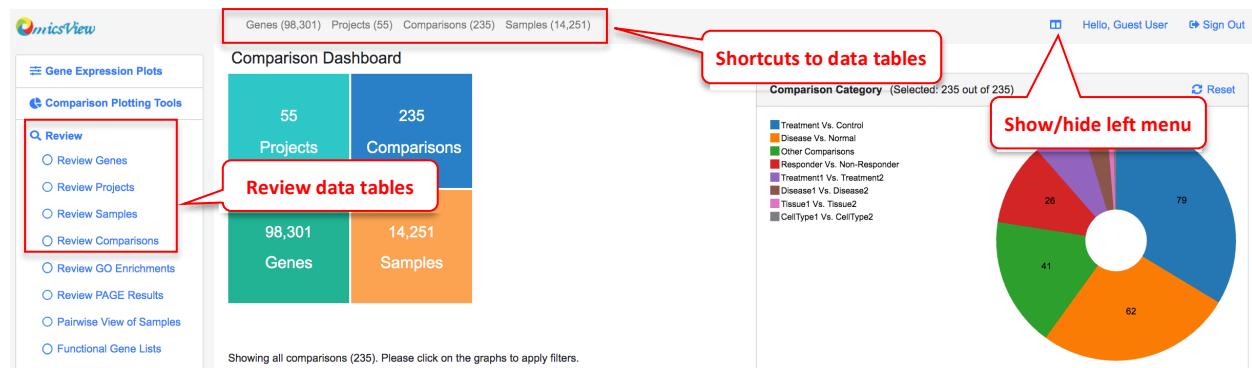
1.	Overview of Data in OmicsView .....	5
1.1	Projects.....	5
1.2	Samples .....	6
1.3	Comparisons.....	6
1.4	Genes.....	7
1.5	Saved Genes and Comparisons .....	8
2	Visualize Gene Expression .....	10
2.1	View Gene Expression from RNA-Seq .....	10
2.2	Change Sample Grouping in Gene Expression Plot .....	11
2.3	View Gene Expression from Microarray Data .....	12
2.4	View Gene Expression from Multiple Genes .....	15
2.5	View Gene Expression in Heatmap .....	17
2.6	Export Genes and Samples.....	19
2.7	Similar comparisons .....	19
2.7.1	Similar comparisons (GO).....	19
2.7.2	Similar comparisons (PAGE).....	20
2.8	Comparisons Venn Diagrams .....	21
2.8.1	Comparisons Venn Diagram (GO) .....	21
2.8.2	Comparisons Venn Diagram (PAGE).....	22
3	Visualize Comparison Data .....	23
3.1	Dashboard View of Comparison.....	23
3.2	Set Dashboard Preference.....	24
3.3	Dynamic Filtering of Comparisons on Dashboard .....	25
3.4	Bubble Plot of Comparisons Associated with a Single Gene .....	26
3.5	Data Filter and Advanced Settings in Bubble Plot .....	28
3.6	Bubble Plot of Sets of Genes and Comparisons .....	29
3.7	Get significant genes from comparisons .....	31
3.8	View Significantly Changed Genes in Bubble Plot .....	34
3.9	Volcano Plot of a Comparison .....	35
3.10	View Multiple Volcano Plots Together .....	36
3.11	Overlap Comparison Data to Pathway Graph .....	37

3.12	Pathway Plot from Several Comparisons .....	38
4	Functional Enrichment from Comparisons .....	40
4.1	Enrichment from Up and Down Regulated Genes .....	40
4.2	View Changed Genes from a Functional Term in Volcano Plot .....	42
4.3	View Enriched Pathways Directly from Comparison Details .....	44
4.4	Gene Set Enrichment from Ranked Genes .....	45
4.5	Pathway Heatmap From Comparisons .....	46
5	Data Upload .....	48
5.1	Data Type .....	48
5.1.1	Supported Data File Format.....	49
5.1.2	Expression Data Format.....	49
5.2	Data Upload options.....	50
5.3	Source File format .....	50
5.3.1	Project Info file.....	51
5.3.2	Sample Info file .....	52
5.3.3	Comparison Info file.....	52
5.4	Data Formats .....	53
5.4.1	Gene Expression Data .....	53
5.4.2	Comparison Data .....	53
6	Advanced Analyses .....	55
6.1	Correlation Tools .....	55
6.2	PCA Analysis .....	58
6.3	Meta-Analysis .....	60
6.3.1	Meta-Analysis (Comparisons) .....	60
6.3.2	Meta-Analysis (Gene Expression).....	64
7	Step-by-step guide of running the Crohn's disease meta-analysis .....	68
8	Methods .....	79
8.1	Data processing .....	79
8.2	Functional and Pathway Enrichment.....	79
8.2.1	HOMER workflow.....	79
8.2.2	Gene Set Enrichment Analysis (GSEA) workflow (PAGE) .....	81

8.3	Meta-Analysis .....	81
8.3.1	Meta-analysis statistics .....	82
9	Supplementary Tables .....	82
9.1	Table S1. Feature comparison of tools.....	82
9.2	Table S2. Mapping between use cases and OmicsView menu items.....	84

## 1. Overview of Data in OmicsView

The homepage of OmicsView gives an overview of the data curated. The Pie-charts and Bar-graphs are reactive and change depending on the selection. Users can access the different data tables from the left menu under search, or from the shortcuts at the top menu bar. Left menu can be hidden to provide more space for tables and graphs.



### 1.1 Projects

Individual research project is typically associated with a published study and has NCBI GEO accession number. Users can search projects from the home page. In the search result page, click each Project ID to view the full description of a project.

Projects (55)

The figure shows a table of 55 projects. The columns include: Project ID, Experiment Type, Platform, Title, PubMed, PubMed Authors, Contact Name, Release Date, Description, Disease, and Study Type. The table includes rows for GSE10500, GSE12251, and GSE12815. A red box highlights the 'Create a Project List' button at the top left of the table area.

Show [ 10 ] entries									
<input type="button" value="Copy"/> <input type="button" value="CSV"/> <input type="button" value="Excel"/> <input type="button" value="PDF"/>									
	Project ID	Experiment Type	Platform	Title	PubMed	PubMed Authors	Contact Name	Release Date	Description
<input type="checkbox"/>	GSE10500	Expression profiling by array	HG_U95Av2	Gene expression in rheumatoid arthritis synovial macrophages	18345002	Yarilina A,Park-Min KH,Antoniv T,Hu X,Ivashkiv LB	Taras, Antoniv	02/14/2008	Macrophages from RA synovial fluids were compared to primary human blood-derived macrophages.
<input type="checkbox"/>	GSE12251	Expression profiling by array	HG-U133_Plus_2	A Predictive Response Signature to Infliximab Treatment in Ulcerative Colitis	19700435	Arijs I,Li K,Toedter G,Quintens R,Van Lommel L,Van Steen K,Leemans P,De Hertogh G,Lemaire K,Ferrante M,Schnitzler F,Thorez L,Ma K,Song XY,M...	K. Li	08/25/2009	Infliximab, an anti-TNF $\alpha$ monoclonal antibody, is an effective treatment for ulcerative colitis (UC) inducing over 60% of patients to respond...
<input type="checkbox"/>	GSE12815	Expression profiling by array	HT_HG-U133A; HuEx-1_0-st-v2;transcript; HG-U133A_2; HG-U133A	PI3K pathway activity in the normal airway of smokers with lung cancer, and in smokers with airway dysplasia	20375364	Gustafson AM,Soldi R,Anderlind C,Scholand MB,Oian J,Zhang X,Cooper K,Walker D,McWilliams A,Liu G,Szabo E,Brody J,Mission PP,Lenburg ME,Lam S...	Adam, Gustafson	04/12/2010	Cytologically normal airway epithelial samples were collected during bronchoscopy of current and former smokers. Subjects enrolled in this s...

## 1.2 Samples

A project may include many samples. Each sample has its own description (full details available by clicking each Sample ID). Each sample has a gene expression profile.

### Review Samples

Advanced Search   Browse All Records   Display Preferences

Samples (14,251)

Create a Sample List   Sample Dashboard

Show 100 entries

<input type="checkbox"/>	Sample ID	Project Name	Platform (GPL)	Platform Name	Cell Type	Description	Disease State	Gender	Response	Tissue	Treatment
<input type="checkbox"/>	GSM265020	GSE10500	GPL8300	Affymetrix.HG_U95Av2	macrophage	Synovial fluid macrophages from individual RA donor isolated by positive selection of CD14+ cells	rheumatoid arthritis (RA)	No Info	No Info	synovial fluid	none
<input type="checkbox"/>	GSM265036	GSE10500	GPL8300	Affymetrix.HG_U95Av2	macrophage	Synovial fluid macrophages from individual RA donor isolated by positive selection of CD14+ cells	rheumatoid arthritis (RA)	No Info	No Info	synovial fluid	none
<input type="checkbox"/>	GSM265361	GSE10500	GPL8300	Affymetrix.HG_U95Av2	macrophage	Synovial fluid macrophages from individual RA donor isolated by positive selection of CD14+ cells	rheumatoid arthritis (RA)	No Info	No Info	synovial fluid	none
<input type="checkbox"/>	GSM265363	GSE10500	GPL8300	Affymetrix.HG_U95Av2	macrophage	Synovial fluid macrophages from individual RA donor isolated by positive selection of CD14+ cells	rheumatoid arthritis (RA)	No Info	No Info	synovial fluid	none
<input type="checkbox"/>	GSM265366	GSE10500	GPL8300	Affymetrix.HG_U95Av2	macrophage	Synovial fluid macrophages from individual RA donor isolated by positive selection of CD14+ cells	rheumatoid arthritis (RA)	No Info	No Info	synovial fluid	none

Users can search for specific samples using the search box.

To change columns displayed in the table, go to the Display Preferences.

Users can create lists by selecting specific samples using the “Create a Sample List” button. Users can then specifically download this list as a CSV.Excel/PDF. Samples from your collection can be loaded to other tools like heatmap.

## 1.3 Comparisons

Two groups of samples are compared, and the statistic values (fold change, p-value, adjusted p-values) are computed.

There are a lot of meta data available for each comparison. See the dashboard for an overview of key categories, and the detailed description of each comparison has the full information.

### Review Comparisons

Advanced Search   Browse All Records   Display Preferences

Comparisons (235)

Create a Comparison List   Create a Sample List   Significantly Changed Genes   Comparison Dashboard

Show 10 entries

<input type="checkbox"/>	Actions	Comparison ID	Case Disease State	Comparison Type	Platform Name	Project Name
<input type="checkbox"/>		GSE10500.GPL8300.test1	rheumatoid arthritis (RA)	glm	Affymetrix.HG_U95Av2	GSE10500
<input type="checkbox"/>		GSE12251.GPL570.test1	ulcerative colitis (UC)	glm	Affymetrix.HG-U133_Plus_2	GSE12251
<input type="checkbox"/>		GSE12815.GPL3921.test1	normal control	glm	Affymetrix.HG-U133A	GSE12815
<input type="checkbox"/>		GSE12815.GPL3921.test2	normal control	glm	Affymetrix.HG-U133A	GSE12815
<input type="checkbox"/>		GSE12815.GPL5175.test1	bronchopulmonary dysplasia (BPD)	glm	Affymetrix.HuEx-1_0-st-v2.transcript	GSE12815

View full details of a comparison by clicking the  (Review Comparison) button.

View volcano plot by clicking the  (View Volcano Chart) button.

Users can also map the comparison data onto WikiPathways by clicking the  (View Pathway) button.

The selected comparisons can be saved to your collection by creating a list using the “Create a Comparison List” button for easy loading into the plotting tools.

## 1.4 Genes

The genome-wide gene expression values were detected in each sample using RNA-Seq or microarrays. All the human genes that have expression values are listed in gene table. The gene annotation from different platforms were all mapped to NCBI gene ID (EntrezID) for consistency across platforms.

### Review Genes

Q Advanced Search    B Browse All Records    D Display Preferences

Genes (98,301)



Show / 250 entries	Copy	CSV	Excel	PDF	Search:
Actions	Gene Name	Entrez ID	Transcript Number	Description	Alias
	ZZZ3	26009	11	zinc finger ZZ-type containing 3	ATAC1 ZZZ3
	ZZEF1	23140	11	zinc finger ZZ-type and EF-hand domain containing 1	ZZZ4 ZZEF1
	ZYXP1	106480342	1	zyxin pseudogene 1	ZYXP1
	ZYX	7791	10	zyxin	ESP-2 HED-2 ZYX
	ZYG11B	79699	2	zyg-11 family member B, cell cycle regulator	ZYG11 ZYG11B
	ZYG11AP1	100131879	1	zyg-11 family member A, cell cycle regulator pseudogene 1	ZYG11AP1
	ZYG11A	440590	3	zyg-11 family member A, cell cycle regulator	ZYG11 ZYG11A

To find a gene, users can specifically search in each field like gene symbol, gene description, gene alias, NCBI gene ID, Ensembl gene ID or Uniprot ID in the Advanced Search option.

For some common genes, the symbols used in publications are often not the official symbol, and users can try search alias field in Advanced Search. For example, TP53 is often referred to as P53 in publication. Users need to search P53 in alias or tumor protein p53 in description to find it if you don't know its official symbol.

The NCBI Gene search <https://www.ncbi.nlm.nih.gov/gene> is a good source to get official gene symbols and IDs.

Users can view full details of a gene by clicking the  (Review Gene) button.

## Gene Details: TP53

Search Genes Create Gene List Gene Expression Levels Bubble Plot

Gene Summary

Gene ID:	TP53	Entrez ID:	7157
Gene Name:	TP53	Transcript Number:	28
Strand:	-	Chromosome:	17
Start:	7668401	End:	7675493
Exon Length:	3922	Source:	Entrez_Gene_20181011
Description:	tumor protein p53	Alias:	BCC7, LFS1, P53, TRP53, TP53
Ensembl:	ENSG00000141510	UniGene:	Hs.437460, Hs.740601
UniProt:	K7PPA8, P04637, Q53G5, H2EHT1, A0A087X1Q1, A0A087WXZ1, A0A087WT22	AccNum:	NM_000546, NM_001126112, NM_001126113, NM_001126114, NM_001126115, NM_001126116, NM_001126117, NM_001126118, NM_001276695, NM_001276696, NM_001276697, NM_001276698, NM_001276699, NM_001276760, NM_001276761, NP_000537, NP_001119584, NP_001119585, NP_001119586, NP_001119587, NP
Biotype:	protein_coding		

From gene details, Users can access RNA-Seq data in a box plot or view all comparisons including this gene in a bubble plot.

## 1.5 Saved Genes and Comparisons

Users can save selected genes or comparison for future use (e.g., multiple gene and multiple comparisons bubble plot). From gene search, check the genes you want to save, and click the yellow button “save selected genes”.

### Review Genes

Advanced Search Browse All Records Display Preferences

Genes (98,301)

Create a Gene List 3. Create list

Show 100 entries 1. Search Term

Search: kinase

Actions	Gene Name	Entrez ID	Transcript Number	Description	Alias
<input checked="" type="checkbox"/>	XYLB	9942	5	xylulokinase	XYLB
<input checked="" type="checkbox"/>	FAM20B	9917	2	FAM20B, glycosaminoglycan xylosylkinase	gxk1 FAM20B
<input checked="" type="checkbox"/>	NUAK1	9891	4	NUAK family kinase 1	ARK5 NUAK1
<input type="checkbox"/>		12		tousled like kinase 1	PKU-beta TLK1
<input type="checkbox"/>		42		membrane associated guanylate kinase, WW and PDZ domain containing 2	ACVRIP1 AIP-1 AIP1 ARIP1 MAGI-2 NPHS15 SSCAM MAGI2
<input type="checkbox"/>	CDK11B	984	9	cyclin dependent kinase 11B	CDC2L1 CDK11 CDK11-p110 CDK11-p46 CDK11-p58 CLK-1 PITSREAI PK58 p58 p58CDC2L1 p58CLK-1 CDK11B
<input type="checkbox"/>	MELK	9833	14	maternal embryonic leucine zipper kinase	HPK38 MELK

2. Check genes of interests

Similarly, you can save comparisons.

## Comparisons (235)

1. Search Term

3. Create list

2. Check comparisons of interests

Comparisons (235)						
		Comparison ID	Case Disease State	Comparison Type	Platform Name	
Actions					Project Name	
<input checked="" type="checkbox"/>		GSE72819.GPL11154.DESeq2.test2	ulcerative colitis (UC)	DESeq2.v1.10.1.os.v101316	NGS.Illumina.HiSeq2000	GSE72819
<input checked="" type="checkbox"/>		GSE72819.GPL11154.DESeq2.test1	ulcerative colitis (UC)	DESeq2.v1.10.1.os.v101316	NGS.Illumina.HiSeq2000	GSE72819
<input checked="" type="checkbox"/>		GSE6731.GPL8300.test4	ulcerative colitis (UC)	glm	Affymetrix.HG_U95Av2	GSE6731
<input checked="" type="checkbox"/>				glm	Affymetrix.HG_U95Av2	GSE6731
<input checked="" type="checkbox"/>				glm	Affymetrix.HuGene-1_0-st-v1	GSE59071
<input checked="" type="checkbox"/>		GSE59071.GPL6244.test1	ulcerative colitis (UC)	glm	Affymetrix.HuGene-1_0-st-v1	GSE59071
<input checked="" type="checkbox"/>		GSE57945.GPL11154.DESeq2.test7	ulcerative colitis (UC)	DESeq2.v1.10.1.os.v101316	NGS.Illumina.HiSeq2000	GSE57945
<input checked="" type="checkbox"/>		GSE57945.GPL11154.DESeq2.test2	ulcerative colitis (UC)	DESeq2.v1.10.1.os.v101316	NGS.Illumina.HiSeq2000	GSE57945

To view saved genes, click “My Results” link on the left menu and then “Gene Lists”.

Click to switch to different saved lists

Similarly, to view the saved comparisons, select “Comparison Lists” under “My Results”:

## 2 Visualize Gene Expression

For each gene, Users can view its expression levels across multiple samples. Most data in OmicsView are from microarrays, consisting of more than 50K samples, and nearly 3000 samples are from RNA-Seq.

### 2.1 View Gene Expression from RNA-Seq

Choose the Gene Expression from RNA-Seq -> Single Gene from left menu and enter the official symbol of gene. Alternatively, in the gene details page, click View Gene Expression link.

The screenshot shows the 'Gene Expressions from RNA-Seq' interface. A red box labeled '1. Select from menu' points to the left sidebar under 'Gene Expression Plots'. A red box labeled '2. Enter gene symbol and search' points to the 'Gene Name:' input field containing 'INS'. A red box labeled '3. (Optional) Adjust data filter, to be applied to attributes' points to the 'Data Filter' button. A red box labeled '4. (Optional) Choose attributes' points to the 'Sample Attributes' section. A red box labeled '5. Click to view gene expression in Boxplot' points to the 'Lat. Plot' button.

Gene Expressions from RNA-Seq

Data Options

Gene Name: INS

Search Options:

- Data Filter
- Refresh Filter
- Search with Project IDs
- Search with Sample IDs
- Search with Comparison IDs

Cell Type: Set Filter Disease Stage: Set Filter Disease State: Set Filter Gender: Set Filter

Sample Source: Set Filter Tissue: Set Filter Platform Name: Set Filter

Plots Options

Sample Attributes:  Select All (Selected: 9)  
 Disease State  Tissue  Gender  Cell Type  Project Name  Response  Sampling Time  Subject ID  Treatment (More Attributes)  
 Update Disease State with Comparison Category (Help):

Lat. Plot Advanced Options Reset

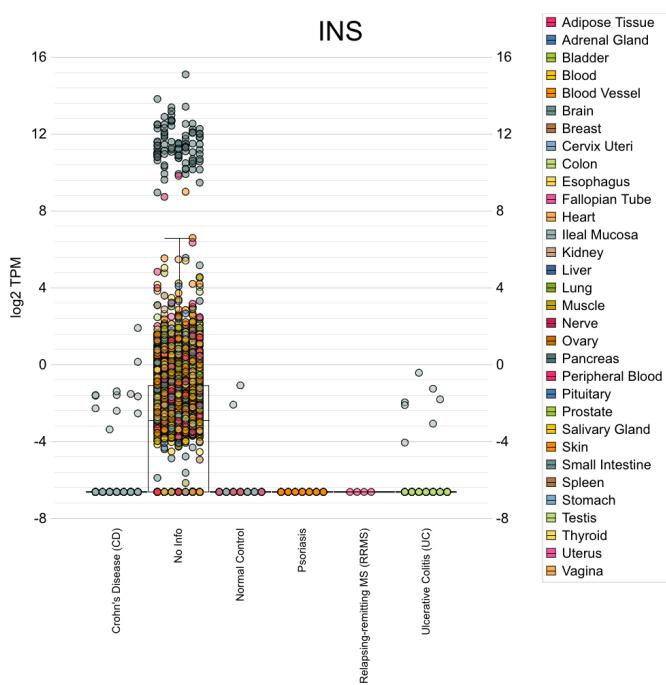
As an optional step, Users can choose what sample attributes to pass to the plot and use data filter to choose only a subset of data points.

The Data Filter can be very useful if there are too many data points, and you can focus on a few diseases or tissue types.

The screenshot below shows default boxplot displaying data points from all diseases, and users can focus on diseases of interest by applying the above "Disease State" filter to narrow down the search. By default, 5000 randomly selected data points are shown.

### Gene Expression Levels for INS

**⚠** The following plot contains 5000 randomly selected data points (out of 9,538) from the search result.



Plot after data filtering was applied to select a few diseases. Now 410 out of 9538 data points are shown. The data filter pop-up window was shown to the left of the box plot in the screen shot below.

Disease State

Category	# of Sample
Adenoma	15
Alzheimer's Disease (AD)	2
<input checked="" type="checkbox"/> Asthma	503
Bronchopulmonary Dysplasia (BPD)	26
Chronic Obstructive Pulmonary Disease (COPD)	2
Collagenous Colitis	2
Colorectal Adenocarcinoma	15
<input checked="" type="checkbox"/> Crohn's Disease (CD)	348
Dermatitis	23
Diet Induced Obesity	46
Discoid Lupus Erythematosus (DLE)	7
Inflammatory Bowel Disease (IBD)	15
Irritable Bowel Syndrome (IBS)	20
Lung Cancer	447

Search Summary

Q. The search result contains 410 out of 9,538 data points, which matches all of the conditions below:

1. Disease State is: asthma or crohn's disease (CD) or psoriasis or ulcerative colitis (UC)

**Download: Raw Data - Plot Data**

Gene Expression Levels for INS

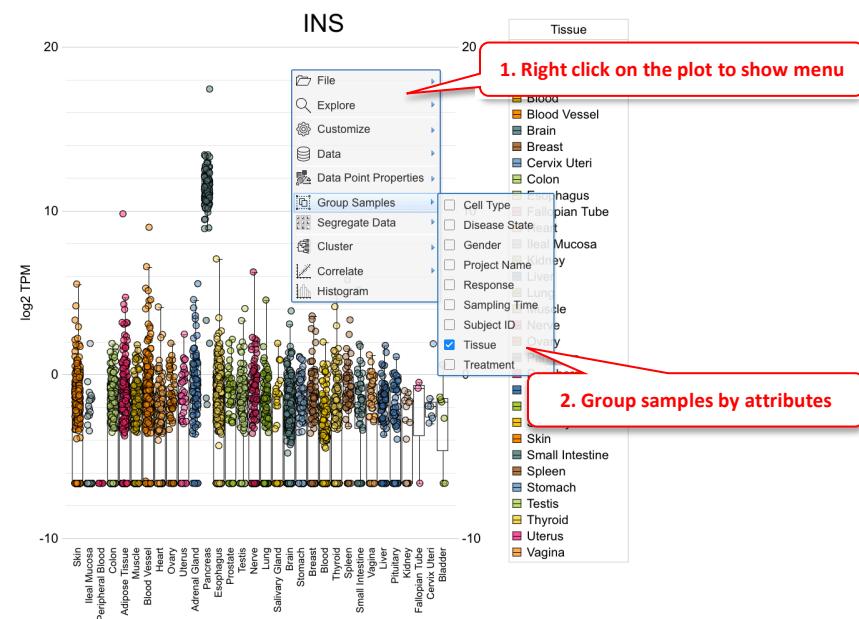
This is a zoomed-in view of the box plot for the "INS" gene. The y-axis is "log2 TPM" ranging from -7.5 to 2.5. The x-axis categories are "Crohn's Disease (CD)", "Psoriasis", and "Ulcerative Colitis (UC)". A legend indicates three tissues: Colon (pink), Ileal Mucosa (blue), and Skin (green). The plot shows a clear separation between the three groups. The "Colon" group has the highest expression levels, followed by "Skin" and then "Ileal Mucosa".

## 2.2 Change Sample Grouping in Gene Expression Plot

The boxplot is created using CanvasXpress (<https://canvasxpress.org>) plug-in, and sample grouping and coloring can be customized by the user. In the example below, we show how sample grouping can be changed.

### Gene Expression Levels for INS

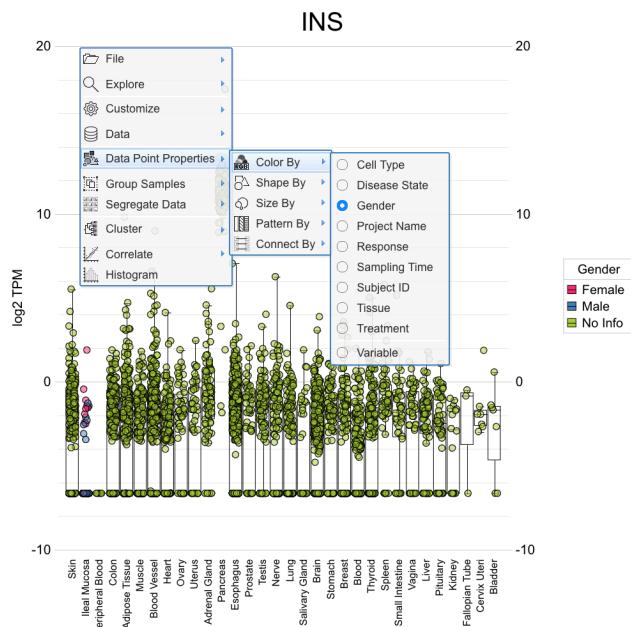
**⚠** The following plot contains 5000 randomly selected data points (out of 9,538) from the search result.



Once the sample grouping was changed to tissue, the box plot shows that insulin is only expressed in pancreatic islets. Users can also change how the data points are colored (default is by disease).

### Gene Expression Levels for INS

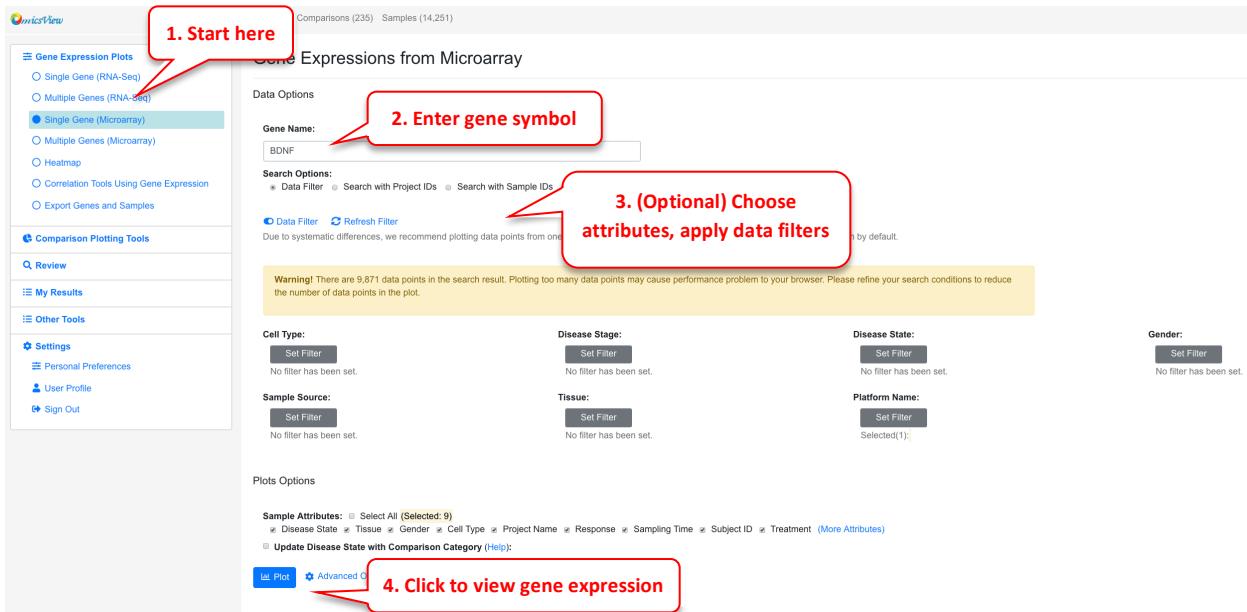
**⚠** The following plot contains 5000 randomly selected data points (out of 9,538) from the search result.



## 2.3 View Gene Expression from Microarray Data

The way to view microarray data is very similar to RNA-Seq data. However, since the expression values from different array platforms are typically not directly comparable, the system by default will choose the platform with the largest number of data points. The user can override this filter if needed.

In addition, we recommend the user to add data filter because typically there are too many data points from microarray data. Data filters will help the user to focus on the most important tissues or diseases. If there are still more than 5000 data points after data filer, only the first 5000 are shown in the boxplot.

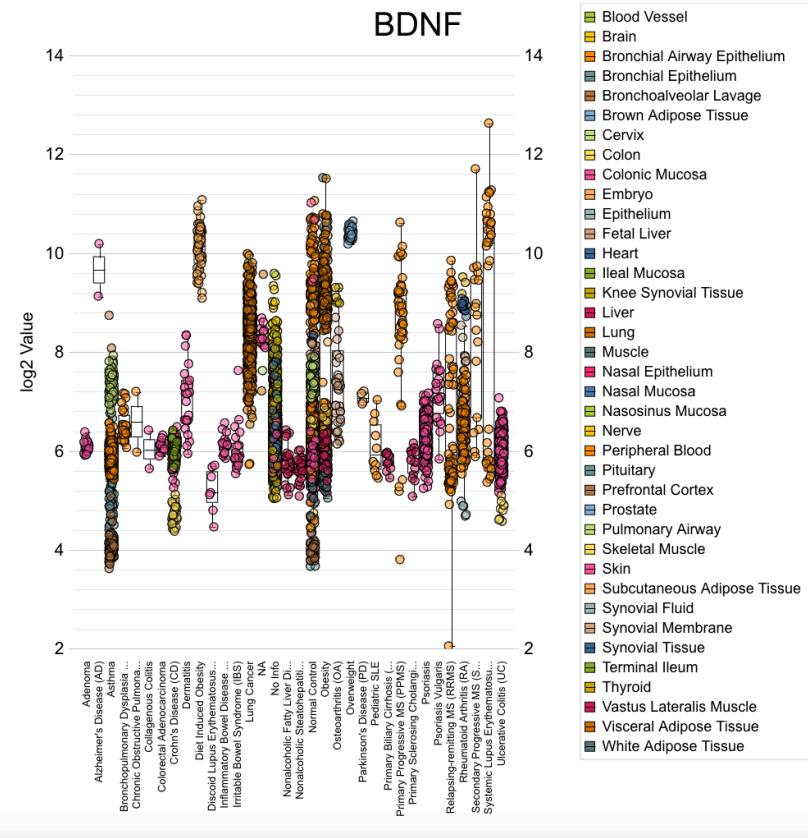


## Search Summary

Q The search result contains 4,504 out of 4,504 data points.

 Download: [Raw Data - Plot Data](#)

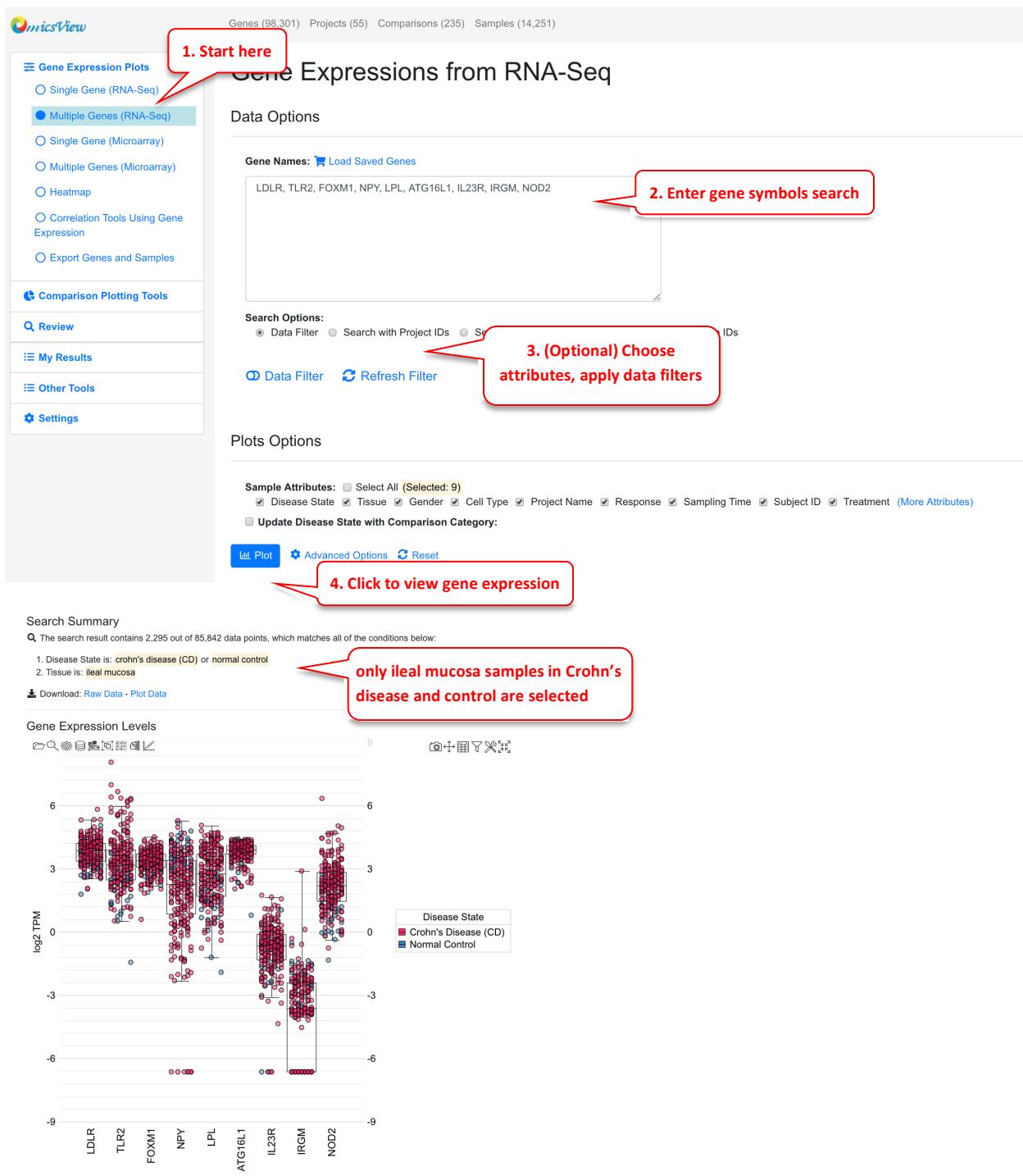
## Gene Expression Levels for BDNF



The sample grouping and coloring can be changed by the user. See the section 2.2 for details.

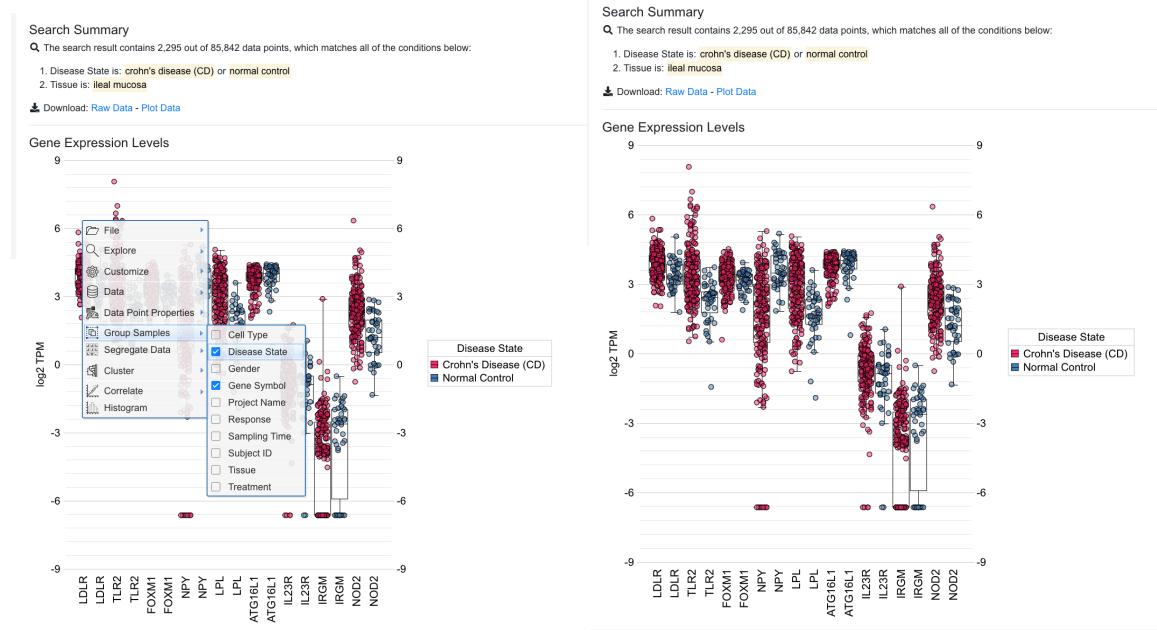
## 2.4 View Gene Expression from Multiple Genes

Users can view multiple genes on the same boxplot. The interface resembles single gene, here you need to enter a list of gene symbols. Users can also load the saved gene lists from your collection. In multiple gene plot, you may want to filter the data to come up with a reasonable number of data points.



From multiple gene plot, you can use the built-in function of CanvasXpress to change grouping and coloring.

In the example plot above, we can change the setting to use group by to distinguish Crohn's disease samples and control samples.



In the updated plot above, it's easy to see that many of these genes are expressed slightly higher in disease (red) vs. control (blue).

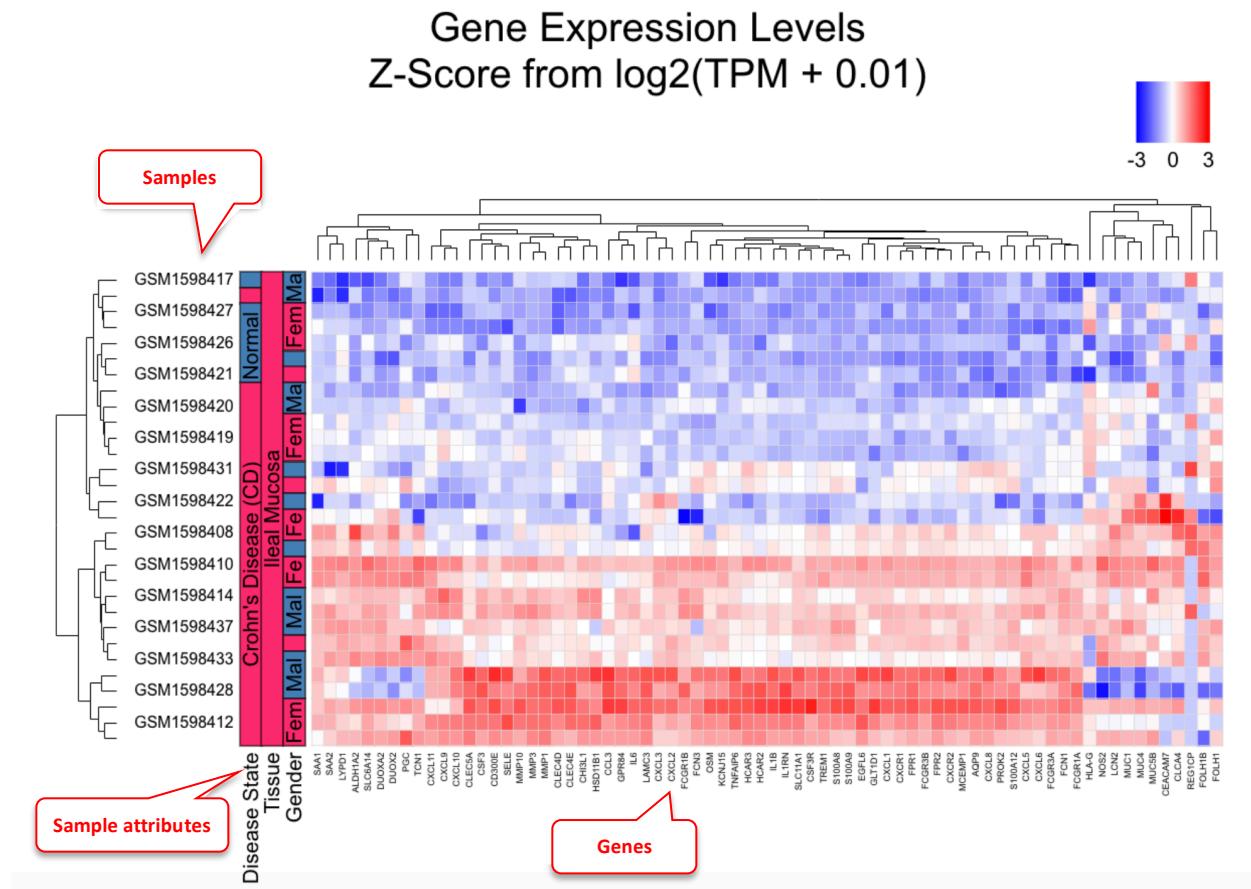
## 2.5 View Gene Expression in Heatmap

Heatmap can be useful to visualize gene profiles from multiple samples. It can also provide information about how genes and samples cluster. This data is from PMID: 25003194

The screenshot shows the OmicsView software interface. On the left, there's a sidebar with various tools: Gene Expression Plots (Single Gene (RNA-Seq), Multiple Genes (RNA-Seq), Single Gene (Microarray), Multiple Genes (Microarray), Heatmap, Correlation Tools Using Gene Expression, Export Genes and Samples), Comparison Plotting Tools, Review, My Results, Other Tools, and Settings. A red box labeled "1. Start here" points to the "Heatmap" option. The main area is titled "Heatmap". It has a search bar "Names:" with a dropdown menu showing "OXA2", "AQP9", "IL8", "MMP3", "DUOX2", "OSM", "HCAR3", and "CXCR1". To the right is a list of "Sample ID" with entries: "GSM1598429", "GSM1598430", "GSM1598431", "GSM1598432", "GSM1598433", "GSM1598434", "GSM1598435", "GSM1598436", and "GSM1598437". A red box labeled "2. Enter or Load Saved Genes" points to the search bar. Another red box labeled "3. Enter or Load Saved Samples" points to the sample ID list. Below the search bar is a section for "Sample Attributes" with checkboxes for Disease State (unchecked), Tissue (checked), Gender (checked), Cell Type (unchecked), Project Name (unchecked), Response (unchecked), Sampling Time (unchecked), Subject ID (unchecked), and Treatment (unchecked). A red box labeled "4. (Optional) Choose attributes overlaying on heatmap" points to the "Advanced Options" button. At the bottom left is a "Plot" button with a red box labeled "6. Plot" pointing to it. In the center, there's a note: "⚠ Some of the genes you entered do not exist in the database. Please click [here](#) for details." At the bottom right is a "Raw data" link with a red box labeled "5. (Optional) Change data transformation, clustering options" pointing to it. There's also a "Download: Raw Data - Heatmap Data" link.

Users can enter genes and samples in the box or load pre-saved genes and samples quickly from your collection. Be default, we will log2 transform the gene expression data, perform scaling of the data across samples for each gene, and limit the scaled value to -3 to 3 before displaying the data in heatmap. This works well in most situations. However, advanced users can change the options. For example, if you want to keep the order of samples as you entered, just uncheck “Enable Clustering Samples”.

The heatmap is rendered by CanvasXpress. Users can change the plot size if needed.



In the example heatmap, we entered a few significantly changed genes in Crohn's disease, and choose a few disease and control samples. As shown in the heatmap, most of these genes show distinct patterns between disease and control.

One of the samples GSM1598422 is labeled as disease, but its gene expression signature match those with normal control very well. Therefore, from heatmap clustering, we can decide that this sample is likely an outlier, the sample may be mislabeled, or this patient had very different gene expression patterns form all other patients.

## 2.6 Export Genes and Samples

This is a useful feature to download raw (counts) and processed (TPM values) expression matrices from the datasets. This gives the user the opportunity to do any downstream analysis on the raw reads independently.

1. Start here

2. Enter or Load Saved Genes

3. Enter or Load Saved Samples

4. Hit submit

5. Download the Gene TPM, Counts or All the data associated with it

For this functionality, the samples need to belong to the same platform type and the requested limit is to 100 genes and 20 samples. The files are exported as csv and can be loaded and used by the user to perform their own downstream analysis.

## 2.7 Similar comparisons

For a given comparison, this functionality helps identify other similar comparisons.

### 2.7.1 Similar comparisons (GO)

To identify similar comparisons based on GO terms in up or downregulated genes.

1. Start here

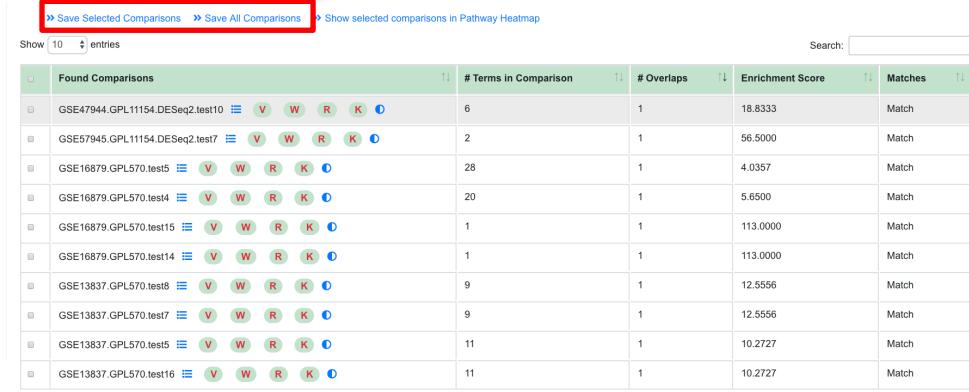
2. Enter comparison ID

3. Pick GO tree to be probed

4. Option to probe up or down regulated

4. Hit search

Save most similar comparisons



	Found Comparisons	# Terms in Comparison	# Overlaps	Enrichment Score	Matches
1	GSE47944.GPL11154.DESeq2.test10	6	1	18.8333	Match
2	GSE57945.GPL11154.DESeq2.test7	2	1	56.5000	Match
3	GSE16879.GPL570.test5	28	1	4.0357	Match
4	GSE16879.GPL570.test4	20	1	5.6500	Match
5	GSE16879.GPL570.test15	1	1	113.0000	Match
6	GSE16879.GPL570.test14	1	1	113.0000	Match
7	GSE13837.GPL570.test8	9	1	12.5556	Match
8	GSE13837.GPL570.test7	9	1	12.5556	Match
9	GSE13837.GPL570.test5	11	1	10.2727	Match
10	GSE13837.GPL570.test16	11	1	10.2727	Match

## 2.7.2 Similar comparisons (PAGE)

Similarly, the similar comparisons can be identified using PAGE results

omicsView

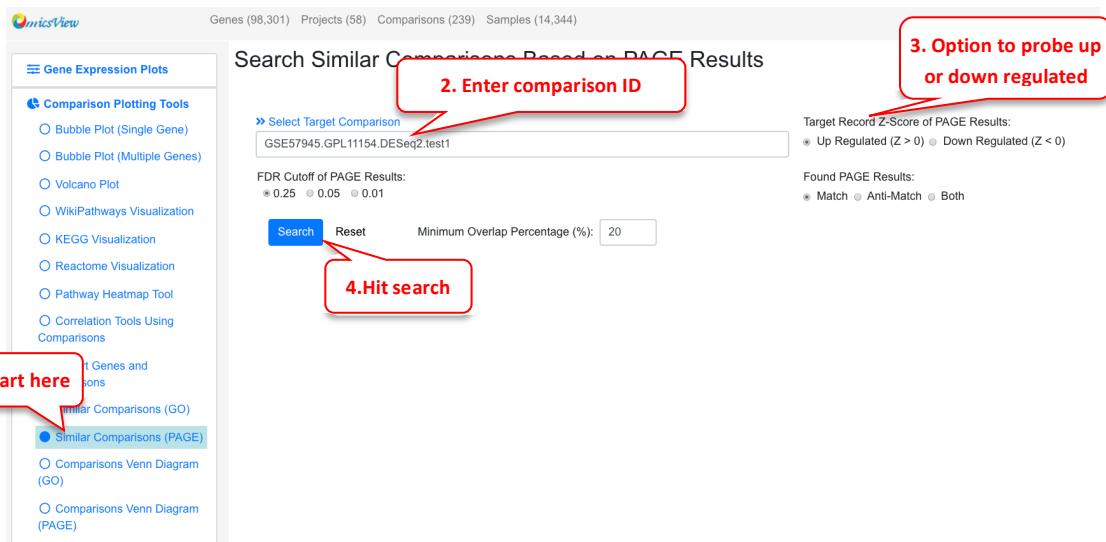
Genes (98,301) Projects (58) Comparisons (239) Samples (14,344)

**1. Start here**

**2. Enter comparison ID**

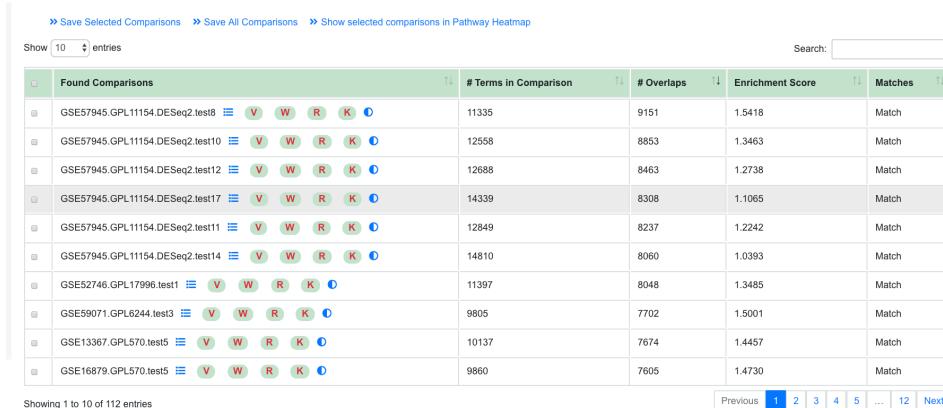
**3. Option to probe up or down regulated**

**4. Hit search**



A list of comparisons that were found similar to the target comparison are displayed. They can be sorted on different attributed like # Terms in Comparison, # Overlaps and. Enrichment score

**4. Hit search**



	Found Comparisons	# Terms in Comparison	# Overlaps	Enrichment Score	Matches
1	GSE57945.GPL11154.DESeq2.test8	11335	9151	1.5418	Match
2	GSE57945.GPL11154.DESeq2.test10	12558	8853	1.3463	Match
3	GSE57945.GPL11154.DESeq2.test12	12688	8463	1.2738	Match
4	GSE57945.GPL11154.DESeq2.test17	14339	8308	1.1065	Match
5	GSE57945.GPL11154.DESeq2.test11	12849	8237	1.2242	Match
6	GSE57945.GPL11154.DESeq2.test14	14810	8060	1.0393	Match
7	GSE52746.GPL17996.test1	11397	8048	1.3485	Match
8	GSE59071.GPL6244.test3	9805	7702	1.5001	Match
9	GSE13367.GPL570.test5	10137	7674	1.4457	Match
10	GSE16879.GPL570.test5	9860	7605	1.4730	Match

## 2.8 Comparisons Venn Diagrams

A good way to identify similarities and difference between different comparisons is by using the Comparisons Venn Diagram functionality. There are two ways to use this tool: By GO terms and PAGE results.

### 2.8.1 Comparisons Venn Diagram (GO)

For identifying similar GO terms between comparisons either in the up and down regulated categories, this functionality can be used. In this example, 3 comparisons from Psoriasis are probed, the user also has the option of probing 2 comparisons by leaving the third field empty.

1. Start here

2. Enter comparison IDs

3. Choose Up or Down regulation

4. Select GO tree and P-value cut off

5. Submit

Compare Comparisons Based on GO Analysis Results (Venn Diagram)

A GSE47944.GPL11154.DESeq2.test10  
B GSE47944.GPL11154.DESeq2.test11  
C GSE47944.GPL11154.DESeq2.test12

Select Comparison 1: \* Up Regulated    Down Regulated

Select Comparison 2: \* Up Regulated    Down Regulated

Select Comparison 3: \* Up Regulated    Down Regulated

GO Tree: KEGG  
P-Value Cutoff:  $10^{-10}$   $10^{-6}$   0.01

Submit   Reset

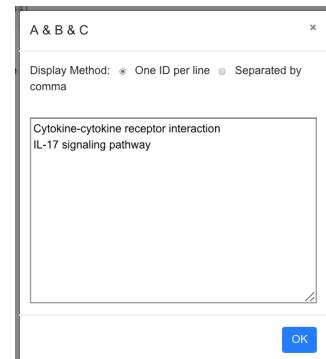
Upon finishing, an overlap summary is generated. This contains a Venn Diagram of the intersection of all the GO terms between the 3 comparisons. The count number can be clicked, as shown below, to view which terms are included.

#### Overlap Summary

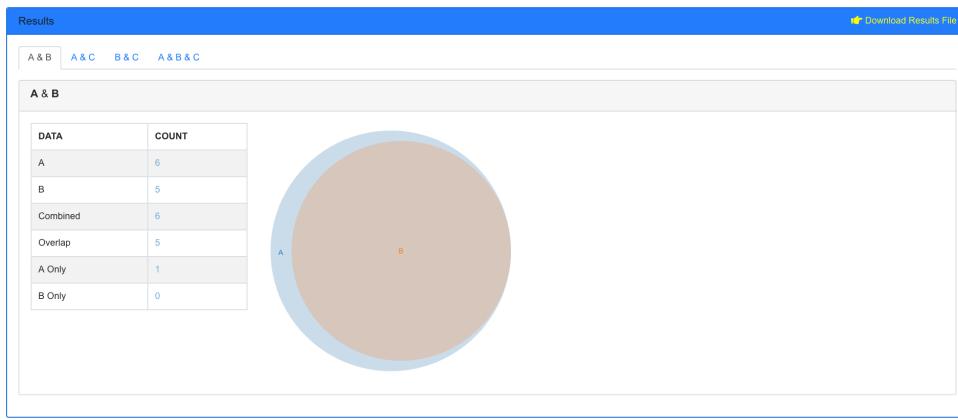
A: GSE47944.GPL11154.DESeq2.test10 (6)  
B: GSE47944.GPL11154.DESeq2.test11 (5)  
C: GSE47944.GPL11154.DESeq2.test12 (3)

#### Download Overlap Results

C	3
A & B	5
A & C	2
B & C	2
A, B & C	2
A only	1
B only	0
C only	1
Combined	7



Furthermore, there is another section below this where pairs of the comparisons can be probed. Here, just A & B are displayed.



## 2.8.2 Comparisons Venn Diagram (PAGE)

Similarly, the user can probe the similarities between PAGE results or 2 or 3 comparisons. In this example we probe comparisons from Psoriasis again.

Upon finishing an overlap summary is generated which shows the intersection of the PAGE results between the comparisons.

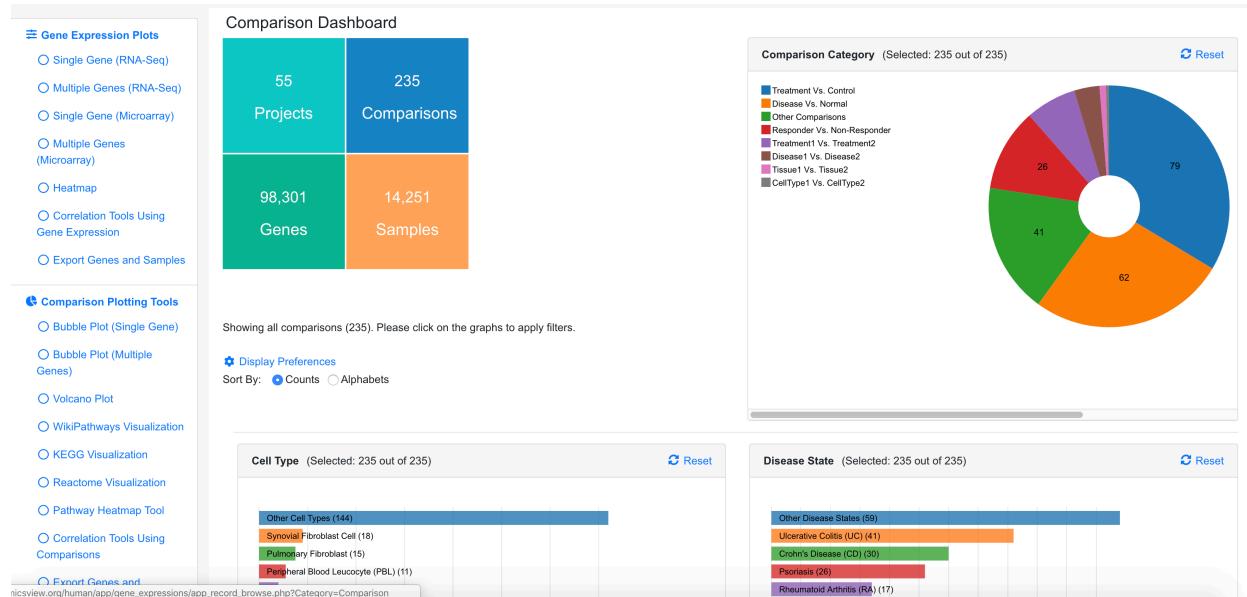


## 3 Visualize Comparison Data

There are 235 comparisons in the demo system, covering different categories including Disease vs. Normal, Treatment vs. Control, Tissue 1 vs. Tissue 2 etc. The system offers many ways to visualize the data and help users find the most interesting data related to their study.

### 3.1 Dashboard View of Comparison

The dashboard shows a summary of all the comparisons.



The dashboard shows Comparison Categories, Cell Type, Disease State, Treatment, Platform summaries. A table is also available below the dashboard that lists all the comparisons.

#### Comparisons (235)

[Create a Comparison List](#) [Create a Sample List](#) [Significantly Changed Genes](#)

Show 100 entries

Search:

Actions	Comparison ID	Case Disease State	Comparison Type	Platform Name	Project Name
<a href="#"></a> <a href="#"></a> <a href="#"></a>	GSE10500.GPL8300.test1	rheumatoid arthritis (RA)	glm	Affymetrix.HG_U95Av2	GSE10500
<a href="#"></a> <a href="#"></a> <a href="#"></a>	GSE12251.GPL570.test1	ulcerative colitis (UC)	glm	Affymetrix.HG-U133_Plus_2	GSE12251
<a href="#"></a> <a href="#"></a> <a href="#"></a>	GSE12815.GPL3921.test1	normal control	glm	Affymetrix.HT_HG-U133A	GSE12815
<a href="#"></a> <a href="#"></a> <a href="#"></a>	GSE12815.GPL3921.test2	normal control	glm	Affymetrix.HT_HG-U133A	GSE12815
<a href="#"></a> <a href="#"></a> <a href="#"></a>	GSE12815.GPL5175.test1	bronchopulmonary dysplasia (BPD)	glm	Affymetrix.HuEx-1_0-st-v2.transcript	GSE12815
<a href="#"></a> <a href="#"></a> <a href="#"></a>	GSE12815.GPL5175.test2	bronchopulmonary dysplasia (BPD)	glm	Affymetrix.HuEx-1_0-st-v2.transcript	GSE12815
<a href="#"></a> <a href="#"></a> <a href="#"></a>	GSE12815.GPL5175.test3	bronchopulmonary dysplasia (BPD)	glm	Affymetrix.HuEx-1_0-st-v2.transcript	GSE12815

## 3.2 Set Dashboard Preference

Users can change how the comparison summary is displayed by following the red numbers.

The screenshot shows the 'Personal Preferences' page in OmicsView. The left sidebar has a red circle around the 'Settings' option. The main content area has three red circles labeled 1, 2, and 3. Circle 1 is over the 'Settings' link in the sidebar. Circle 2 is over the 'Gene' tab in the top navigation bar. Circle 3 is over the 'Show Top 15' checkbox in the 'Comparison Dashboard' section.

Personal Preferences

General Gene Sample Project Comparison

General

Expand Left Menu:  Opened  Closed

Gene Data Type:  Value  TPM

Comparison Dashboard

Cell Type:  Hide Unknown  Hide Others  Show Top 15 (Uncheck to show all)

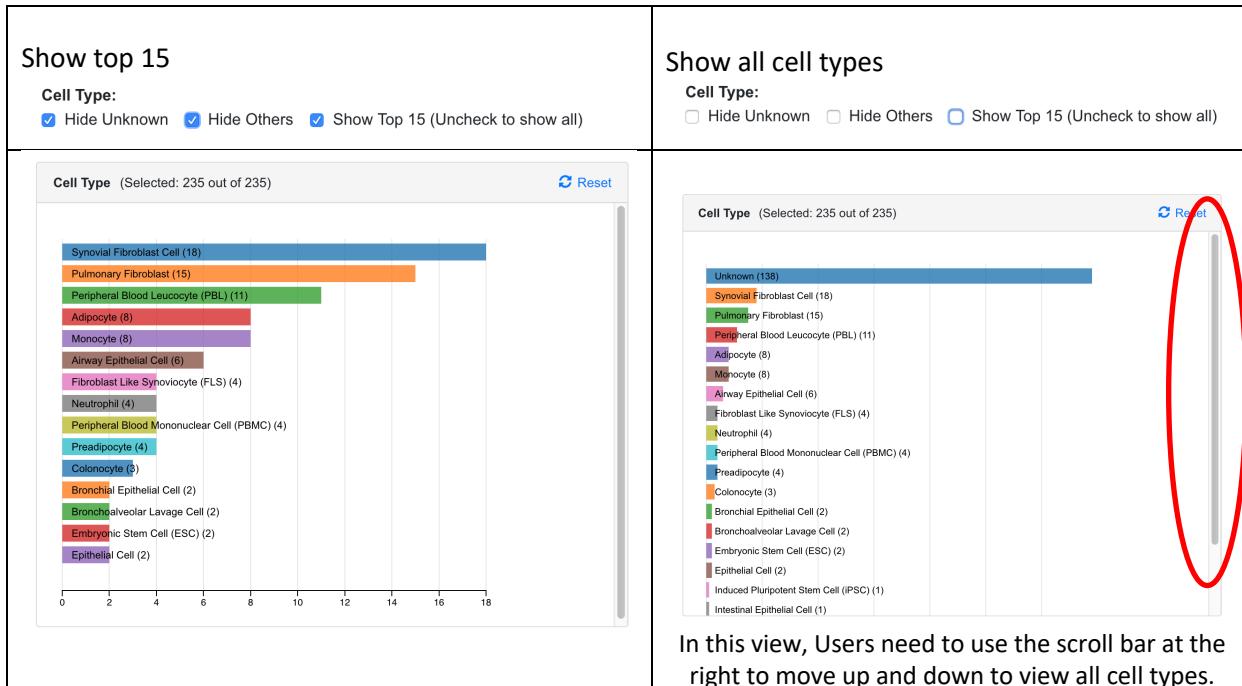
Disease State:  Hide Unknown  Hide Normal Control  Hide Others  Show Top 15 (Uncheck to show all)

Treatment:  Hide Unknown  Hide Others

Platform Name:  Hide Others

Save

Since there are many Cell Types and Disease States, the user can choose to display only the top 15 categories, or all categories. See example screenshots below.

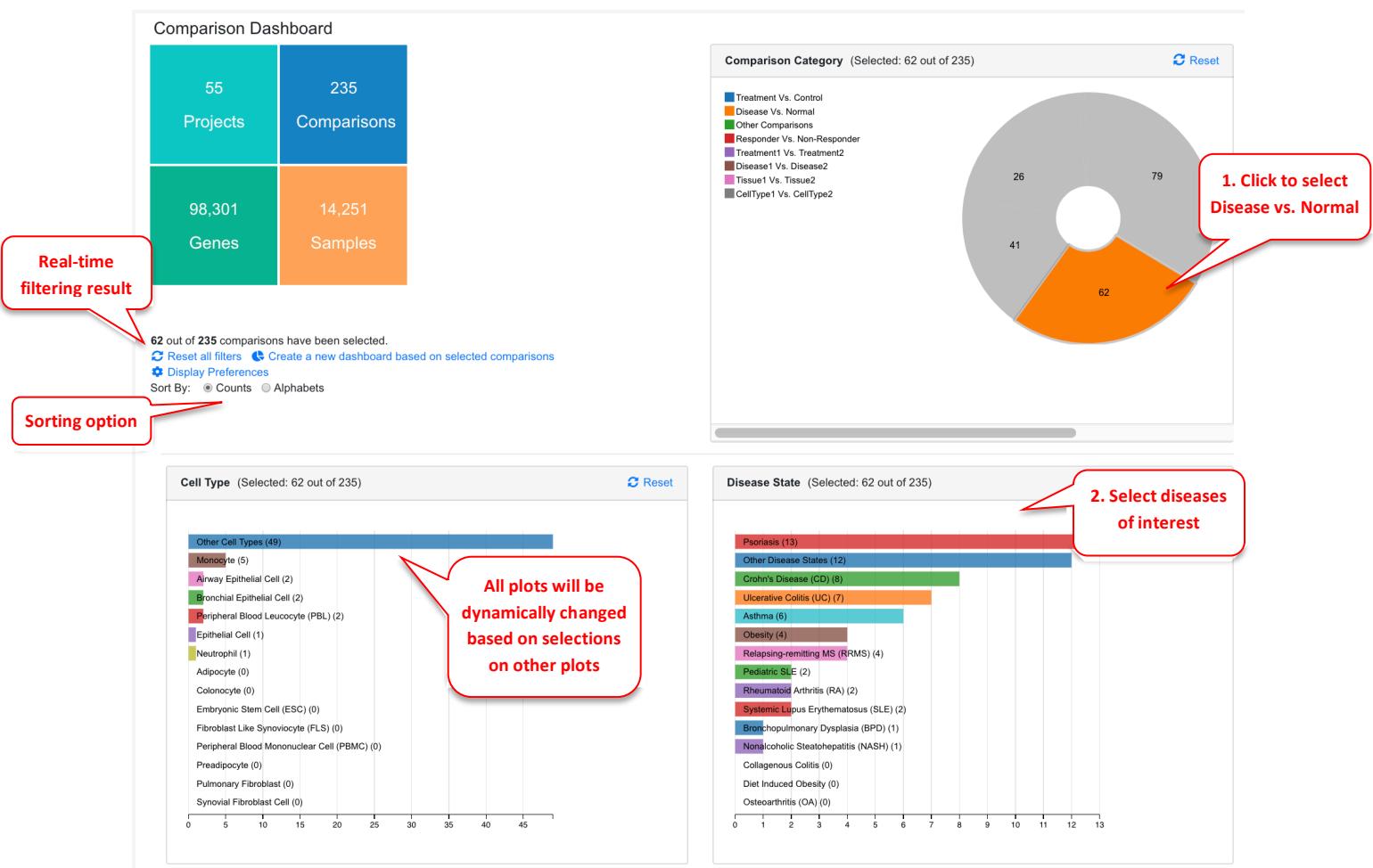


When the user chooses only the top 15 categories, all other categories are shown as Other Cell Type. There are other data points where the Cell Type is unknown from the study. Two categories denoted as "Other" and "Unknown" can be hidden by the user in the preference.

### 3.3 Dynamic Filtering of Comparisons on Dashboard

The user can click any chart in the dashboard to focus on one or more categories that are of interest to the study.

In the example below, we are only interested in Disease vs. Normal comparisons, and we further narrow down the data to three disease areas: ulcerative colitis, Crohn's disease, and rheumatoid arthritis. Now we are looking at 36 out of 199 comparisons. The table at the bottom of the page shows details of these 36 comparisons. The "Reset All Charts" link above the statistics panel is used to show all comparisons again.



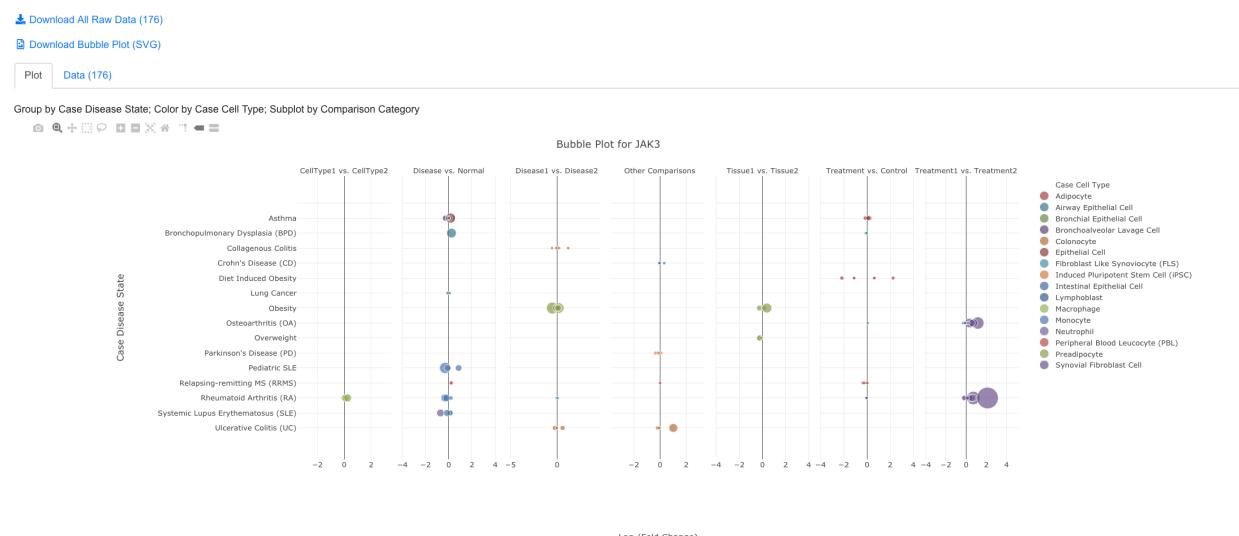
## 3.4 Bubble Plot of Comparisons Associated with a Single Gene

For each gene, users can view all the available comparisons in a bubble chart.

The screenshot shows the OmicsView software interface with the following details:

- Header:** Genes (98,301) Projects (55) Comparisons (235) Samples (14,251)
- Left Sidebar (Comparison Plotting Tools):**
  - Bubble Plot (Single Gene)** (selected, circled 1)
  - Bubble Plot (Multiple Genes)
  - Volcano Plot
  - WikiPathways Visualization
  - KEGG Visualization
  - Reactome Visualization
  - Pathway Heatmap Tool
  - Correlation Tools Using Comparisons
  - Export Genes and Comparisons
  - Similar Comparisons (GO)
  - Similar Comparisons (PAGE)
  - Comparisons Venn Diagram (GO)
  - Comparisons Venn Diagram (PAGE)
- Data Options:**
  - Gene Name:** JAK3 (circled 2)
- Plots Options:**
  - Y-Axis:** Case Disease State
  - Y-axis Settings:** Top 10 (radio button selected)
  - Color By:** Case Cell Type
  - Color By Settings:** Top 10 (radio button selected)
  - Subplot By:** Comparison Category
  - Subplot By Settings:** Show All Values (radio button selected)
  - Marker Area:** Adjusted p-value
  - Marker Shape By:** None
- Buttons:** Plot (highlighted), Advanced Options, Reset

The default settings work for most users. After clicking the Plot button, you will see a plot like below:



In the bubble plot, the X-axis shows log2 Fold Change of the comparison, the Y-axis shows disease state. Each dot represents the comparison result of this gene from one comparison. The color of the dot represents cell type, and the size of the dot represents significance (-log10(FDR), larger is more significant).

The user can click and unclick the color legend at right to select or deselect cell types. When mouse over a dot, more details are shown. And the user can also click the dot to link to other graphs.

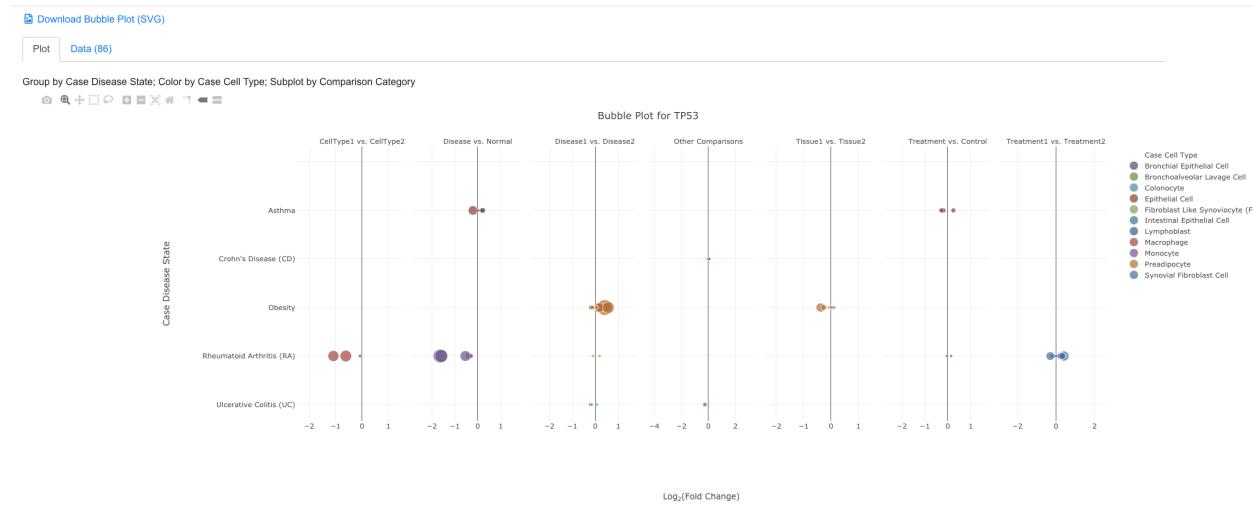
The tool bar at top right corner allows the user to zoom and pan the graph.

### 3.5 Data Filter and Advanced Settings in Bubble Plot

In addition, advanced users can change settings by click "Modify Settings Button". For example, the user may want to show a selected list of diseases. After clicking Customize in Case\_DiseaseState, user can select which diseases to display in the pop-up window.

Category	# of Comparison Data
Ulcerative Colitis (UC)	109
Normal Control	96
Crohn's Disease (CD)	63
Obesity	56
Psoriasis	46
Rheumatoid Arthritis (RA)	40
Osteoarthritis (OA)	25
Relapsing-remitting MS (RRMS)	24
Asthma	13
Overweight	9
Diet Induced Obesity	8
Bronchopulmonary Dysplasia (BPD)	6
Collagenous Colitis	6

After modifying the setting, the user can click plot button to view the new chart. The system will display how many data points are chosen based on the filter.



### 3.6 Bubble Plot of Sets of Genes and Comparisons

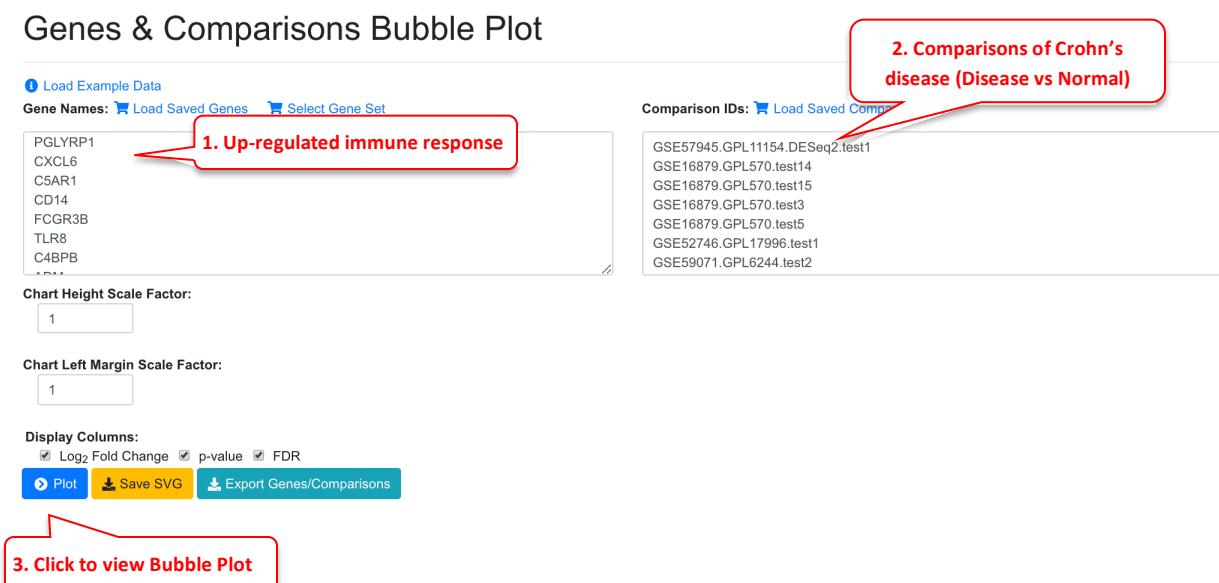
It can be useful to look at a set of genes (e.g., all differentially expressed genes, or genes from a certain pathways) in a set of related comparisons (e.g., all from the same disease).

To view this type of bubble plot, select Bubble Plot (Multiple Genes).

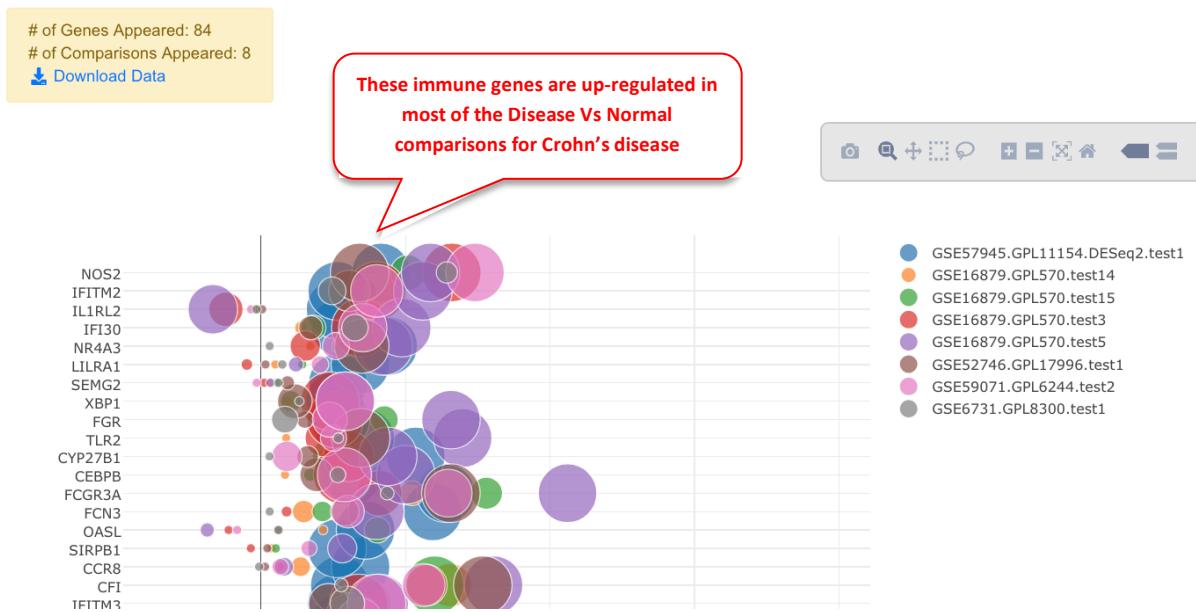


In the Genes and Comparisons Bubble plot window, users can now enter the symbols of the genes, and the comparison names. However, it is much easier to use the saved genes and saved comparisons features, or other tools from the system to quickly get a gene set.

In the example below, we use dashboard to select 8 comparisons that are for Disease vs. Normal in Crohn's disease (CD). We save the comparisons and load in the bubble plot tool. For gene list, we get the up-regulated immune response genes from comparison GSE57945.GPL11154.DESeq2.test1 and paste into the gene names fields.



In the bubble plot, the gene symbols are listed in Y-axis. The X-axis represents logFC, color of the bubble represents comparison, and the size of the bubble represents the significance.

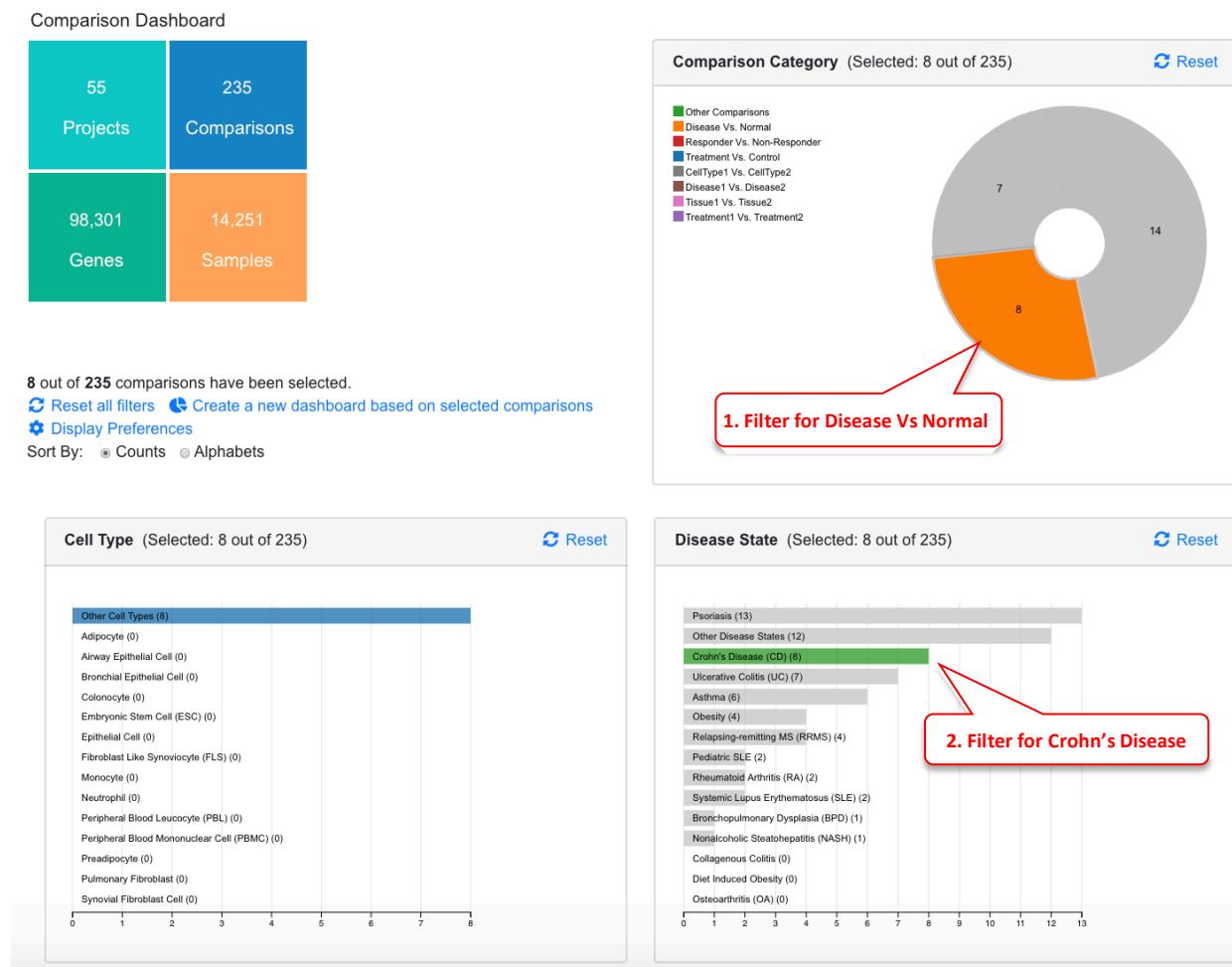


In the legend, the color keys for comparisons are shown. Users can click the color key in the legend to hide/show comparisons. The size of the color dot in the legend correlates to the largest bubble for that comparison, which is the most significant gene with the smallest FDR.

### 3.7 Get significant genes from comparisons

Another way to get a gene set to visualize in the genes/comparisons bubble plot is to filter for significantly changed genes. To do this, first select a few comparisons from the dashboard, and click the "View Significantly Changed Genes" button.

Dashboard filter:



In table, select comparisons and view significantly changed genes.

The screenshot shows a table of comparison entries. At the top, there are three buttons: 'Create a Comparison List' (green), 'Create a Sample List' (blue), and 'Significantly Changed Genes' (red). Red callout boxes labeled '1. Check all', '2. Create comparison list', and '3. View significantly changed genes' point to the first column header 'Actions', the 'Create a Comparison List' button, and the 'Significantly Changed Genes' button respectively. The table has columns: Actions, Comparison ID, Case Disease State, Comparison Type, Platform Name, and Project Name. Each row contains a checkbox in the Actions column and a small circular icon with letters C, V, and P. The table shows 8 entries, with the last one highlighted. At the bottom, it says 'Showing 1 to 8 of 8 entries' and has navigation buttons for Previous (1) and Next.

Actions	Comparison ID	Case Disease State	Comparison Type	Platform Name	Project Name
<input checked="" type="checkbox"/> C V P	GSE16879.GPL570.test14	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
<input checked="" type="checkbox"/> C V P	GSE16879.GPL570.test15	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
<input checked="" type="checkbox"/> C V P	GSE16879.GPL570.test3	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
<input checked="" type="checkbox"/> C V P	GSE16879.GPL570.test5	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
<input checked="" type="checkbox"/> C V P	GSE52746.GPL17996.test1	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE52746
<input checked="" type="checkbox"/> C V P	GSE57945.GPL11154.DESeq2.test1	crohn's disease (CD)	DESeq2.v1.10.1.os.v101316	NGS.Illumina.HiSeq2000	GSE57945
<input checked="" type="checkbox"/> C V P	GSE59071.GPL6244.test2	crohn's disease (CD)	glm	Affymetrix.HuGene-1_0-st-v1	GSE59071
<input checked="" type="checkbox"/> C V P	GSE6731.GPL8300.test1	crohn's disease (CD)	glm	Affymetrix.HG_U95Av2	GSE6731

In the significantly Changed Genes window, the comparisons from the previous page are already loaded. You can add or remove comparisons if needed.

Now select direction (up-, down-, or both), and use the logFC cutoff and FDR value to get a list of genes. Depending on the comparisons, sometimes you may need to adjust the logFC and FDR values to get a good list of genes. In general, for bubble plot, using <100 genes will make the graph easier to read.

Once you are happy with the gene list, you can save it. You can also export the list for later use.

# Significantly Changed Genes

## Comparison IDs:

 Load Saved  
Comparison IDs

GSE6731.GPL8300.test1  
GSE16879.GPL570.test3  
GSE16879.GPL570.test5  
GSE16879.GPL570.test14  
GSE16879.GPL570.test15  
GSE6731.GPL17006.test1

View significantly changed genes in  
these comparisons

## Direction:

Upregulated  Downregulated  Both

## Log2FC Cutoff:

1

Adjust the values if needed to  
increase or decrease number of genes

## Cutoff Category:

P-Value  FDR

## Cutoff Value:

0.05

Click submit to view genes

 Submit





List of Significantly Changed Genes (3,703)

Export to save file

Gene ID	Description	GSE6731.GPL8300.test1 Log2 Fold Change	GSE6731.GPL8300.test1 p-value	GSE6731.GPL8300.test1 Adjusted p-value	GSE16879.GPL570.test3 Log2 Fold Change	GSE16879.GPL570.test3 p-value	GSE16879.GPL570.test3 Adjusted p-value	GSE16879.GPL570.test3 Log2
7SK_35	NA							
A1CF	APOBEC1 complementation factor				-1.0692	2.318e-006	2.249e-004	-1.716
A2M	alpha-2-macroglobulin							1.504
AADAC	arylacetamide deacetylase							
AAED1	peroxiredoxin like 2C							1.712
AASS	amino adipate-semialdehyde synthase				1.1714	2.822e-005	1.195e-003	1.066
AB074160	NA							
AB074166	NA				1.3909	1.407e-004	3.447e-003	1.159
AB075489	NA				-1.1693	4.597e-007	7.521e-005	-1.098
ABAT	4-aminobutyrate aminotransferase				-1.2640	5.500e-005	1.835e-003	-1.460

Showing 1 to 10 of 3,703 entries

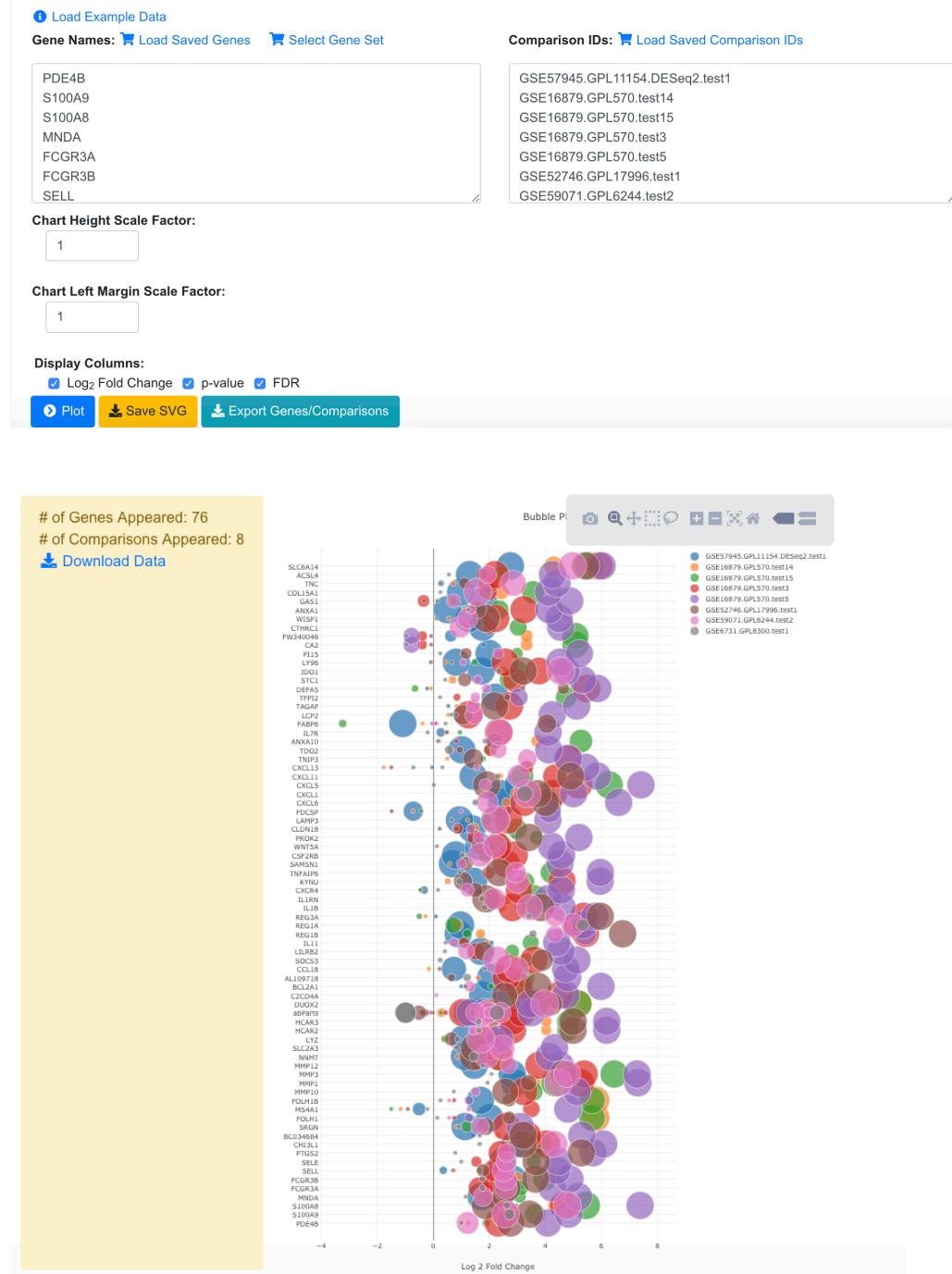
Previous 1 2 3 4 5 ... 371 Next

## 3.8 View Significantly Changed Genes in Bubble Plot

Back to the bubble plot, users can load the saved comparisons and saved genes and view the plot.

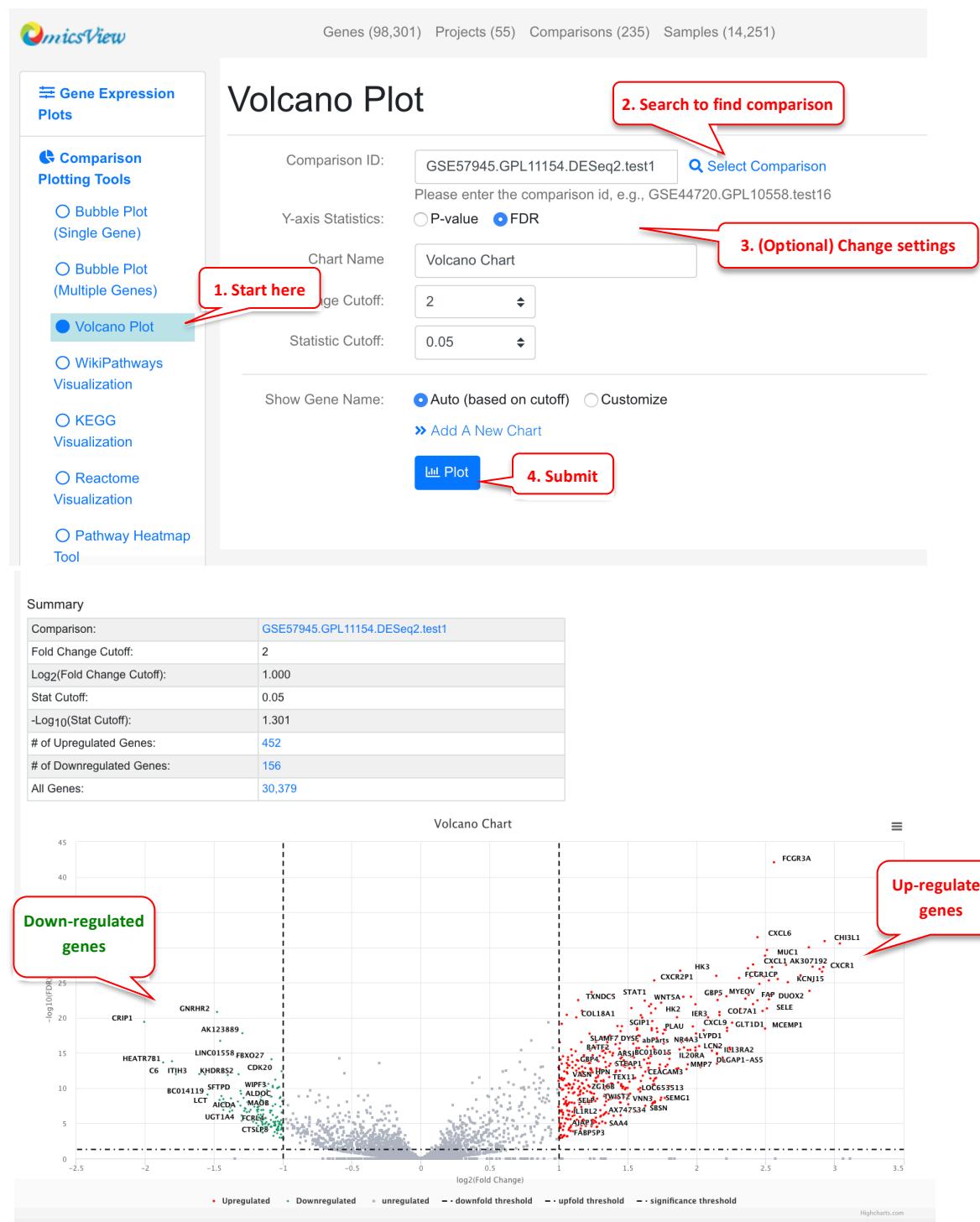
In the example below, most significant genes come from up-regulated direction.

### Genes & Comparisons Bubble Plot



### 3.9 Volcano Plot of a Comparison

Volcano plot is useful to view a top-level summary of how many genes are significantly up- or down-regulated in a comparison.



Users can use mouse to drag over an area to zoom in. Mouse over a point will show the gene details. Click the data point will show you links to other graphs.

### 3.10 View Multiple Volcano Plots Together

Users can also show multiple comparisons side-by-side. If needed, the user can also highlight the same group of genes across the volcano plots.

#### Volcano Plot

Comparison ID: GSE57945.GPL11154.DESeq2.test1  The first comparison

Y-axis Statistics:  P-value  FDR

Chart Name: Volcano Chart

Fold Change Cutoff: 2

Statistic Cutoff: 0.05

Comparison ID: GSE16879.GPL570.test3  Inhouse  Online The second comparison

Please enter the comparison id, e.g., GSE44720.GPL10558.test16

Y-axis Statistics:  P-value  FDR

Chart Name: Volcano Chart

Fold Change Cutoff: 2

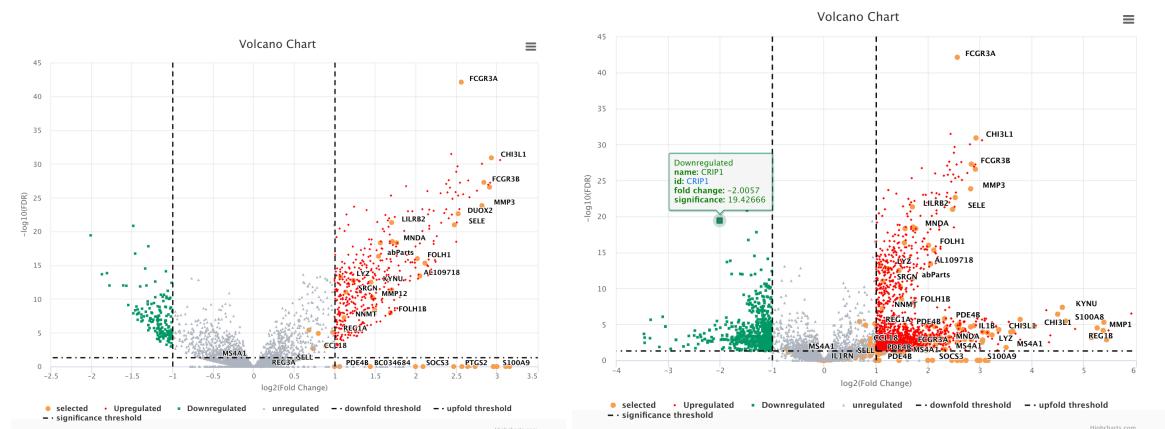
Statistic Cutoff: 0.05

Show Gene Name:  Auto (based on cutoff)  Customize Use "Customize" to highlight genes entered in the text box

Enter Genes Names:

Add more comparisons if needed

The resulting volcano plots are shown as below. Selected genes are shown as orange dots.

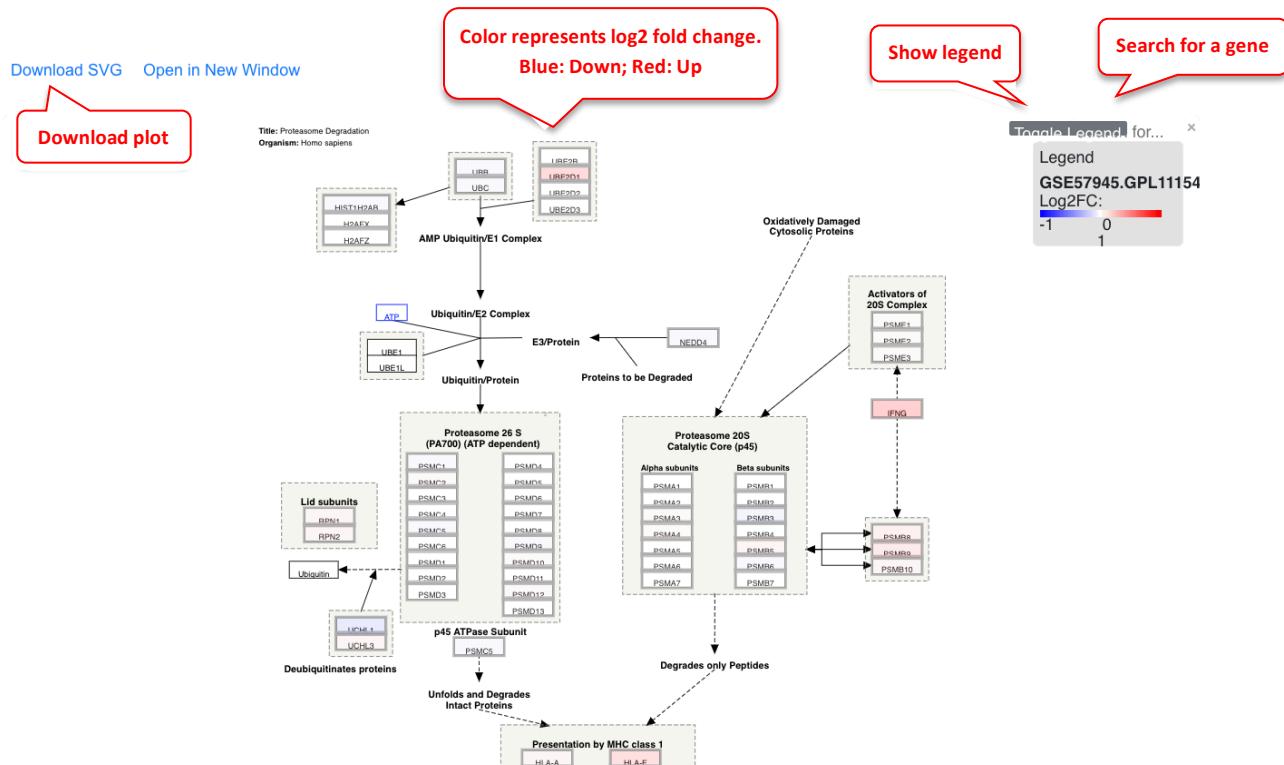


### 3.11 Overlap Comparison Data to Pathway Graph

If users are interested in a particular pathway, sometimes it is useful to map the RNA-Seq or microarray data to the pathway for visualization.

The screenshot shows the WikiPathway Visualization interface. On the left, there's a sidebar with various options: Gene Expression Plots, Comparison Plotting Tools (Bubble Plot (Single Gene), Bubble Plot (Multiple Genes), Volcano Plot), WikiPathways Visualization (selected), KEGG Visualization, Reactome Visualization, Pathway Heatmap Tool, Correlation Tools Using Comparisons, Export Genes and Comparisons, Similar Comparisons (GO), Similar Comparisons (PAGE), and Comparisons Venn. A red box labeled "1. Start here" points to the "WikiPathways Visualization" option. The main area has a title "Proteasome Degradation" with a "Download PDF" button and a "Start Over" link. Below it is a search bar for "Enter comparison names" with a placeholder "E11154.DESeq2.test1". A red box labeled "2. Choose comparison" points to the search bar. Further down is a "Process Above Comparison List" button. A red box labeled "3. Process the comparison" points to a modal window titled "Comparison Name: GSE57945.GPL11154.DESeq2.test1". Inside the modal, there's a dropdown for "Coloring of logFC:" set to "Gradient Blue-White-Red (-1,0,1)". A red box labeled "4. Submit to view the pathway" points to a "Submit" button at the bottom of the modal. At the very bottom of the main page, there are "Download in SVG" and "Open in New Window" links.

In the pathway plot, typically we use red-blue color scale to show the log2 Fold Change. Blue is down-regulated, red is up-regulated.



### 3.12 Pathway Plot from Several Comparisons

The user can add multiple comparisons from the pathway plot tool by clicking Add Comparison link. This is a good functionality to overlay differential genes from multiple comparisons along a pathway.

In this example we use data from Connor-Robson et al., Neurobiology of Disease, 2019. In this paper, the authors performed proteomics (Mass spec) and RNAseq analysis on diseased and normal iPSCs. They performed both analysis at 2 timepoints of differentiation: D35 and D56. This visualization will identify if the same genes are differentially expressed in the proteomic and the RNAseq datasets, as well as in the two timepoints on a WikiPathway.

The data was uploaded into OmicsView. For the proteomic dataset, the protein ID was converted to the gene ID. Here the 4 comparisons (2 Proteomics and 2 RNAseq) are loaded, and the P13k Akt signaling pathway is probed.

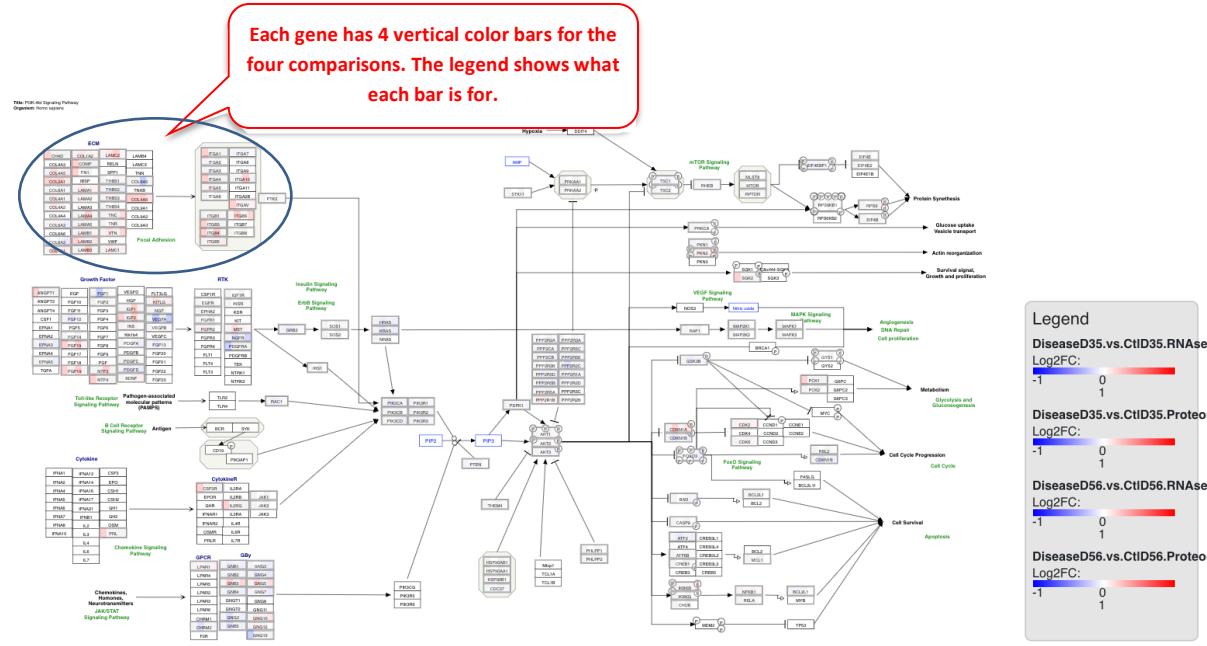
The screenshot shows the OmicsView interface with the following steps highlighted:

- 1. Select WikiPathways**: A red box highlights the "WikiPathways Visualization" option in the left sidebar.
- 2. Select pathway to probe**: A red box highlights the "Select Pathway" button in the main panel.
- 3. Add the comparisons**: A red box highlights the "Enter comparison names" input field where four comparisons are listed: "DiseaseD56.vs.CtlD56.RNAseq\_P4", "DiseaseD56.vs.CtlD56.Proteomics\_P3", "DiseaseD35.vs.CtlD35.RNAseq\_P4", and "DiseaseD35.vs.CtlD35.Proteomics\_P3".
- 4. Process the comparisons**: A red box highlights the "Process Above Comparison List" button below the comparison input field.

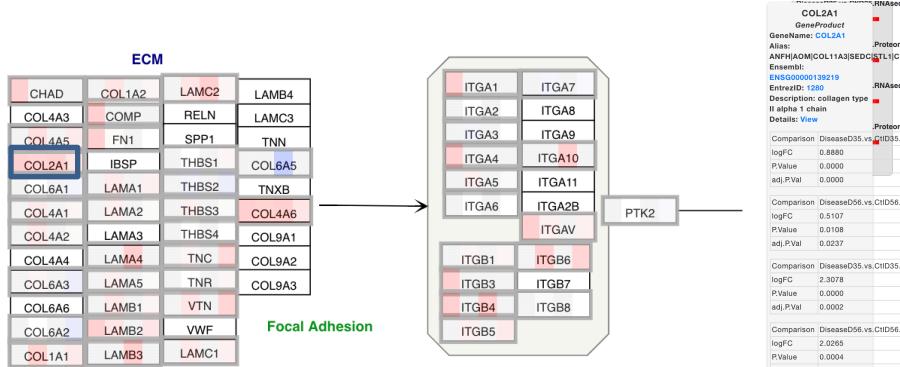
Below the comparison input field, there is a section for "Upload your comparison files:" with a "Choose File" button and a note: "You can keep uploading multiple files." Three separate configuration boxes are shown for each comparison, each with fields for "Comparison Name", "Coloring of logFC:", and "Coloring of P.Value" or "adj.P.Val".

Comparison Name	Coloring of logFC:	Coloring of P.Value / adj.P.Val
DiseaseD56.vs.CtlD56.RNAseq_P4	Gradient Blue-White-Red (-1,0,1)	P.Value
DiseaseD35.vs.CtlD35.Proteomics_P3	Gradient Blue-White-Red (-1,0,1)	P.Value
DiseaseD56.vs.CtlD56.Proteomics_P3	Gradient Blue-White-Red (-1,0,1)	P.Value

The pathway plot now has 4 color bars corresponding to the different comparisons.



Users can also zoom into the plot for better visibility of the genes. The figure below is zoomed into the circular section from the plot above.



In this example, many of the ECM genes are getting differentially regulated in the same direction across all 4 comparisons. If a particular gene is clicked on, like COL2A1 here, the details of changes in the gene in each comparison can be seen.

## 4 Functional Enrichment from Comparisons

### 4.1 Enrichment from Up and Down Regulated Genes

Let's first select a comparison to view.

The screenshot shows the 'Review Comparisons' page. On the left is a sidebar with various options like 'Review Genes', 'Review Projects', etc., and a 'Review Comparisons' option which is selected and highlighted with a red box. In the center, there's a table titled 'Comparisons (237)' with columns for Actions, Comparison ID, Case Disease State, Comparison Type, Platform Name, and Project Name. A specific row for 'GSE16879.GPL570.test1' is highlighted with a red box and labeled '1. Review comparisons'. To the right of the table is a search bar with the placeholder 'Search: GSE16879' and a red box around it labeled '2. Search for a particular comparison you may have in mind'. Below the table, another red box highlights the 'Comparison ID' column, labeled '3. Click on the Comparison.ID'.

Actions	Comparison ID	Case Disease State	Comparison Type	Platform Name	Project Name
C V P	GSE16879.GPL570.test1	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
C V P	GSE16879.GPL570.test10	ulcerative colitis (UC)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
C V P	GSE16879.GPL570.test11	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
C V P	GSE16879.GPL570.test12	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
C V P	GSE16879.GPL570.test13	ulcerative colitis (UC)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
C V P	GSE16879.GPL570.test14	crohn's disease (CD)	glm	Affymetrix.HG-U133_Plus_2	GSE16879
C V P	GSE16879.GPL570.test15			Affymetrix.HG-U133_Plus_2	GSE16879

When users view details of a comparison, the functional enrichment results are shown. Briefly, for each comparison, we generated the up and down regulated gene lists and use these lists to compare with all genes in the genome to identify functions that are significantly enriched.

Comparison Details ➤ Search All Comparisons

**Comparison ID:**  
GSE16879.GPL570.test14

**Category:**  
Disease vs. Normal

**Contrast:**  
DiseaseState => crohn's disease (CD) vs  
normal control

[View Details](#)

[Comparison Volcano Chart](#)

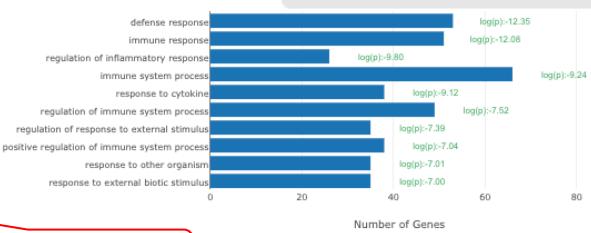
[Pathway View](#)

[Related Samples](#)

[Show Genes](#)

Upregulated Genes
Biological Process
Cellular Component
Molecular Function
KEGG
Molecular Signature
Interpro Protein Domain
Wiki Pathway
Reactome
<a href="#">» Enrichment Report</a>

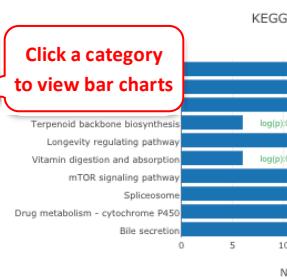
[Download SVG File](#)



Top 10 terms from selected category

Downregulated Genes
Biological Process
Cellular Component
Molecular Function
KEGG
Molecular Signature
Interpro Protein Domain
Wiki Pathway
Reactome
<a href="#">» Enrichment Report</a>

[Download SVG File](#)



Click a category to view bar charts

In the example above, this comparison is between Crohn's disease (CD) vs normal control, and the top up-regulated biological processes are defense and immune response.

Clicking on the left menu will switch the bar charts for different categories (Gene Ontology, KEGG, Molecular signature, Protein domain etc.).

The bar charts here show the top 10 categories. To view complete results, click the Enrichment Report.

## Gene Ontology Enrichment Results

[Text file version of complete results \(i.e., open with Excel\)](#) [Show/Hide](#)

### Enriched Categories

[Toggle columns:](#)  GO Tree  TermID  Term  Enrichment  logP  Genes in Term  Target Genes in Term  Fraction of Targets in Term  Total Target Genes  Total Genes

<a href="#">Column visibility</a>	<a href="#">CSV</a>	<a href="#">Excel</a>	<a href="#">PDF</a>	<a href="#">Print</a>	Show <a href="#">10</a> <a href="#">▼</a> entries	Search: <input type="text"/>
GO Tree	TermID	Term	Enrichment	logP	Genes in Term	
MSigDB	SABATES_COLORECTAL_ADENOMA_UP	SABATES_COLORECTAL_ADENOMA_UP	1.04316187722353e-24	-23.981648297912	128	<a href="#">View</a>
MSigDB	MCLACHLAN_DENTAL_CARIES_UP	MCLACHLAN_DENTAL_CARIES_UP	7.14258507073829e-23	-22.1461445783121	232	<a href="#">View</a>
MSigDB	GSE36888_UNTREATED_VS_IL2_TREATED_STAT5_AB_KNOCKIN_TCELL_2H_UP	GSE36888_UNTREATED_VS_IL2_TREATED_STAT5_AB_KNOCKIN_TCELL_2H_UP	7.73673220280426e-18	-17.1114424354362	178	<a href="#">View</a>
MSigDB	MODULE_5	MODULE_5	3.85418906122208e-17	-16.4140669855208	424	<a href="#">View</a>
MSigDB	GSE36888_UNTREATED_VS_IL2_TREATED_TCELL_17H_DN	GSE36888_UNTREATED_VS_IL2_TREATED_TCELL_17H_DN	2.63186936200225e-15	-14.5797356715815	180	<a href="#">View</a>
Gene Ontology	GO:0005615	extracellular space	1.7444184468598e-13	-12.7583493294011	1411	<a href="#">View</a>
MSigDB	GSE25123_WT_VS_PPARG_KO_MACROPHAGE_UP	GSE25123_WT_VS_PPARG_KO_MACROPHAGE_UP	2.18631393827778e-13	-12.6602874764795	173	<a href="#">View</a>
MSigDB	GSE45365_NK_CELL_VS_CD11B_DC_DN	GSE45365_NK_CELL_VS_CD11B_DC_DN	2.89449591768889e-13	-12.5384270586128	196	<a href="#">View</a>
Gene Ontology	GO:0006952	defense response	4.42239838286997e-13	-12.3543421374495	1127	<a href="#">View</a>

In the enrichment report, the full list of functional terms are shown by order of Enrichment.

## 4.2 View Changed Genes from a Functional Term in Volcano Plot

From the bar chart, click a functional term, and you have the option to view these genes in a volcano plot.

### Comparison Details

[» Search All Comparisons](#)

Comparison ID:  
GSE16879.GPL570.test14

[Click to view volcano plot](#)

Category:  
Disease vs. Normal

[View Details](#)

[Comparison Volcano Chart](#)

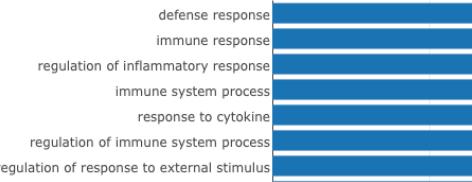
[Pathway View](#)

[Related Samples](#)

[Show Genes](#)

Upregulated Genes
Biological Process
Cellular Component
Molecular Function
KEGG
Molecular Signature
Interpro Protein Domain
Wiki Pathway
Reactome
<a href="#">» Enrichment Report</a>

[Download SVG File](#)



Once you click the link, volcano plot will be generated for the comparison with the changed genes from the selected term highlighted.

## Volcano Plot » View Genes

Comparison of interest

Comparison ID: GSE16879.GPL570.test14

Please enter the comparison id, e.g., GSE44720.GPL10558.test16

P-value  FDR

Chart Name: Volcano Chart

Fold Change Cutoff: 2

Statistic Cutoff: 0.05

Show Gene Name:  Auto (based on cutoff)  Customize

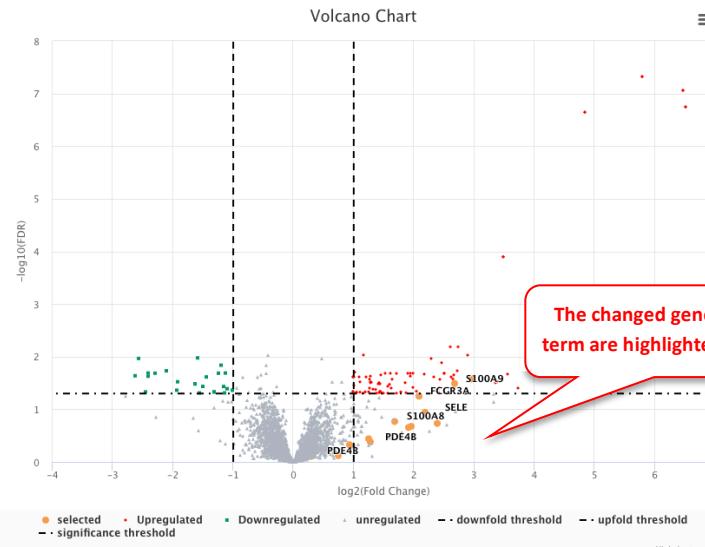
Enter Genes Names:  
  
 ST00AO  
 MNDA  
 FCGR3A  
 FCGR3B  
 SELL  
 SELE  
 PTGS2

» Add A New Chart

Changed genes from the functional term

### Summary

Comparison:	GSE16879.GPL570.test14
Fold Change Cutoff:	2
Log <sub>2</sub> (Fold Change Cutoff):	1.000
Stat Cutoff:	0.05
-Log <sub>10</sub> (Stat Cutoff):	1.301
# of Upregulated Genes:	87
# of Downregulated Genes:	24
All Genes:	54,076



The changed genes from the term are highlighted by orange

Show 10 entries

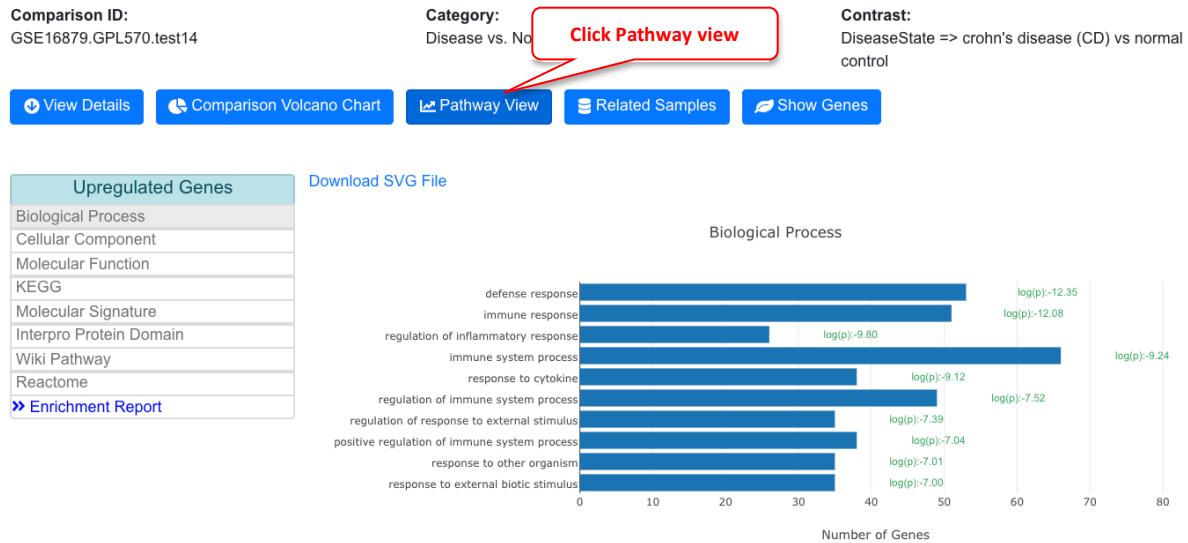
Table of changed genes

Search:

Gene ID	Description	Log2FC	FDR	P-Value
FCGR3A	Fc fragment of IgG receptor IIIa	2.0886	5.619e-002	5.234e-004
FCGR3B	Fc fragment of IgG receptor IIIb	2.0886	5.619e-002	5.234e-004
FCGR3B	Fc fragment of IgG receptor IIIb	2.971	2.677e-002	6.056e-005
MNDA	myeloid cell nuclear differentiation antigen	1.9234	2.219e-001	2.188e-002
PDE4B	phosphodiesterase 4B	1.66	0.0001	0.0001

## 4.3 View Enriched Pathways Directly from Comparison Details

### Comparison Details [» Search All Comparisons](#)



This will automatically open the pathway visualization page and preload the pathway and comparison.

## 4.4 Gene Set Enrichment from Ranked Genes

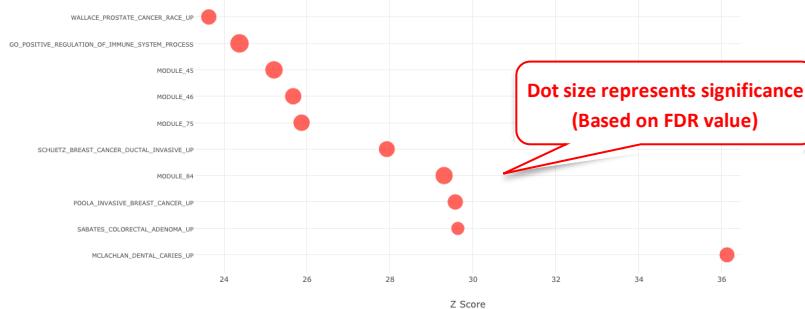
For each comparison, we produce a rank file for all genes using logFC. We use PAGE (Parametric Analysis of Gene Set Enrichment) to identify significant biological changes. PAGE can be more sensitive for comparisons where the logFC is relatively small, but most genes in a functional set show the same direction of change.

The predefined gene sets were from MSigDB.

For each comparison, the top up-regulated and down-regulated gene sets are plotted.

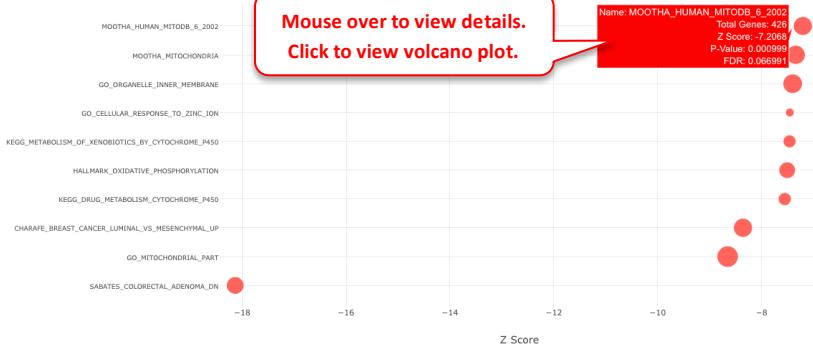
PAGE Report for Upregulated Genes

PAGE Plot - Up-regulated Genes



PAGE Report for Downregulated Genes

PAGE Plot - Down-regulated Genes



To view the genes within a gene set using volcano plot, you can click the dot, and use the link from the pop-up.



## 4.5 Pathway Heatmap From Comparisons

Users can display the enriched pathways from several related comparisons and visualize the top enriched pathways across comparisons. Users can mix public data and inhouse comparisons.

Pathway Heatmap Tool

Comparison IDs:

GSE57945.GPL11154.DESeq2.test1  
GSE16879.GPL570.test14  
GSE16879.GPL570.test15  
GSE16879.GPL570.test3  
GSE16879.GPL570.test5

Set:

Data Filter:

Display Option:

Upregulated Pathways (20):

- Staphylococcus aureus infection
- Rheumatoid arthritis
- Leishmaniasis
- Cytokine-cytokine receptor interaction
- Complement and coagulation cascades
- Chemokine signaling pathway
- Cell adhesion molecules (CAMs)
- Hematopoietic cell lineage

Downregulated Pathways (20):

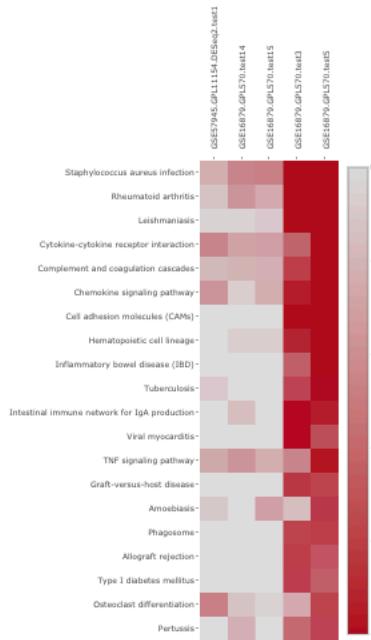
- Chemical carcinogenesis
- Drug metabolism - cytochrome P450
- Retinol metabolism
- Metabolism of xenobiotics by cytochrome P450
- Porphyrin and chlorophyll metabolism
- Ascorbate and aldarate metabolism
- Steroid hormone biosynthesis
- Pentose and glucuronate interconversions

The heatmap shows pathways in rows, comparisons in columns. The statistical significance is color-coded (log P-value, or Z-score). Pathways are sorted by the negative logP values from the highest to the lowest.

### Upregulated Pathways vs. Comparison ID

KEGG: log<sub>10</sub>(p-value)

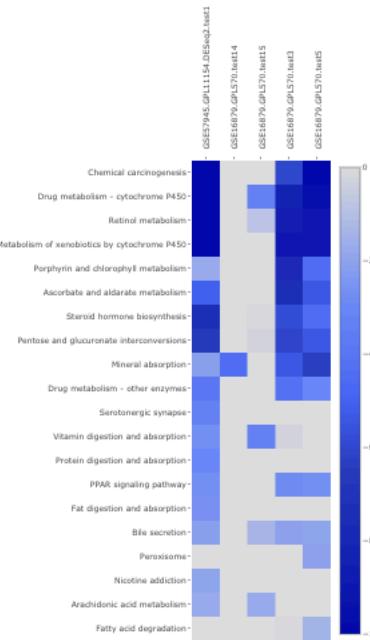
[Download: SVG - Heatmap Data](#)



### Downregulated Pathways vs. Comparison ID

KEGG: log<sub>10</sub>(p-value)

[Download: SVG - Heatmap Data](#)



From the pathway heatmap, users can click any data point to view details.

**Summary**

Gene Set: Staphylococcus aureus infection  
Comparison ID: GSE16879.GPL570.test14  
log(p-value): -4.219  
# of Genes: 9

- [Review Comparison \(GSE16879.GPL570.test14\)](#)
- [View Data in Volcano Plot](#)
- [KEGG Pathway Visualization \(GSE16879.GPL570.test14 only\)](#)
- [KEGG Pathway Visualization \(All Comparisons\)](#)

**Inflammatory bowel disease (IBD) -**

[Close](#)

## 5 Data Upload

OmicsView gives users the option to add their own data or published data that is not on the portal yet. This is achieved through the “Internal Data” option under “My Results”.

The screenshot shows the 'Review Internal Data' page. On the left, a sidebar menu includes 'Gene Expression Plots', 'Comparison Plotting Tools', 'Review', and 'My Results' (with 'Internal Data' selected). The main area displays a table of datasets with columns for Project, Platform Type, # of Comparisons, # of Samples, Permission, Date, Owner, and Actions. Three datasets are listed: '2017\_JM\_LRRK2\_Neuron\_RNASeq\_P4', '2017\_JM\_LRRK2\_Neuron\_Proteomics\_P3', and 'Cx3cr1-deficient microglia\_P2'. A red box surrounds the 'Internal Data' link in the sidebar, and another red box surrounds the 'Import Internal Data' button in the top left of the main area. Red callout bubbles point to these elements with the instructions '1. Click here' and '2. Click to add new dataset'. A third red callout bubble points to the table header with the text 'Datasets entered by the User already'.

To add a new dataset, click on the link “Import Internal Data” that will lead you to a page like below.

The screenshot shows the 'Import Internal Data' form. The sidebar on the left shows 'My Results' with 'Internal Data' selected. The main form has three main sections: 'About Your Data' (with 'Data Type' set to 'All' and 'How would you like to upload' set to 'Multiple Files'), 'Upload Your Data' (with fields for 'Project Info', 'Sample Info', 'Gene Level Expression', 'Gene Count', 'Comparison Info', and 'Comparison Data'), and several 'Choose File' input fields for selecting files. Example files are provided for each category.

### 5.1 Data Type

There are three types of project data. The file requirements for each data type are summarized in the table below.

Data Type	All	Sample and Gene Data Only	Comparison Data Only
Files Required	Project Info file Sample Info file Gene Level Expression file Comparison Info file Comparison Data file Gene Count file (Optional)	Project Info file Sample Info file Gene Level Expression file Gene Count file (Optional)	Project Info file Comparison Info file Comparison Data file

When you select different Data Types, the **Upload Your Data** session in the page changes accordingly.

The figure consists of three side-by-side screenshots of a web form titled 'About Your Data'. Each screenshot shows a different combination of 'Data Type' and 'How would you like to upload?' options. In each screenshot, the first two rows are identical, while the third row varies. Red boxes highlight specific fields in each row.

About Your Data		About Your Data		About Your Data	
Data Type:	<input checked="" type="radio"/> All <input type="radio"/> Sample and Gene Data Only <input type="radio"/> Comparison Data Only	Data Type:	<input type="radio"/> All <input checked="" type="radio"/> Sample and Gene Data Only <input type="radio"/> Comparison Data Only	Data Type:	<input type="radio"/> All <input type="radio"/> Sample and Gene Data Only <input checked="" type="radio"/> Comparison Data Only
How would you like to upload?	<input checked="" type="radio"/> Multiple Files <input type="radio"/> One Zip File	How would you like to upload?	<input checked="" type="radio"/> Multiple Files <input type="radio"/> One Zip File	How would you like to upload?	<input checked="" type="radio"/> Multiple Files <input type="radio"/> One Zip File

Upload Your Data		Upload Your Data		Upload Your Data	
*: Required		*: Required		*: Required	
Project Info*:	<input type="file"/> No file chosen <a href="#">Example file</a>	Project Info*:	<input type="file"/> No file chosen <a href="#">Example file</a>	Project Info*:	<input type="file"/> No file chosen <a href="#">Example file</a>
Sample Info*:	<input type="file"/> No file chosen <a href="#">Example file</a>	Sample Info*:	<input type="file"/> No file chosen <a href="#">Example file</a>	Comparison Info*:	<input type="file"/> No file chosen <a href="#">Example file</a>
Gene Level Expression*:	<input type="file"/> No file chosen <a href="#">Example file</a>	Gene Level Expression*:	<input type="file"/> No file chosen <a href="#">Example file</a>	Comparison Data*:	<input type="file"/> No file chosen <a href="#">Example file</a>
Gene Count:	<input type="file"/> No file chosen <a href="#">Example file</a>	Gene Count:	<input type="file"/> No file chosen <a href="#">Example file</a>		
Comparison Info*:	<input type="file"/> No file chosen <a href="#">Example file</a>				
Comparison Data*:	<input type="file"/> No file chosen <a href="#">Example file</a>				

## 5.1.1 Supported Data File Format

The system support two data formats:

- The csv format (comma separated)
- The tab delimited format

Be sure that your source data files are in the supported format. Usually **Auto Detect** is selected as default and doesn't need to change.

## 5.1.2 Expression Data Format

The Gene Level Expression file can have two formats: Matrix format and Table format. Click on the **Help** link below the selection box to see examples.

The application supports two different formats:

### Matrix Format

This format supports only one kind of data at a time.

Gene	Sample_1	Sample_2	Sample_3
CREB1	6.08	4.92	9.56
TP53	6.27	7.55	6.68
WASH7P	3.08	3.17	8.76

### Table Format:

This format supports only multiple kinds of data.

Gene	Sample_ID	Expression	Count
CREB1	Sample_1	4.55	89
CREB1	Sample_2	2.41	18
CREB1	Sample_3	3.53	76
TP53	Sample_1	2.87	71
TP53	Sample_2	8.72	64
TP53	Sample_3	7.29	37
WASH7P	Sample_1	7.51	31
WASH7P	Sample_2	9.79	30
WASH7P	Sample_3	5.71	39

The **Study** option is useful if you want to include the uploaded project in a specific study. And the **Access** option can be checked if you want to make your project viewable by all the users.

## 5.2 Data Upload options

Instead of uploading the source data file individually, you can compress your data files into a zip file and upload only this zip file. The application will determine the data type based on the file name. Make sure that your zip file contains the following files with the same file names. You can check the requirement of the zipped file for each data type by clicking on the Requirement link under the Choose File box in the **Upload Your Data** session.

Upload Your Data

\*: Required

Zip\_File\*:

Choose File No file chosen

[Requirement](#)

This table below summarizes the file requirement for all data types.

Data Type	File requirement
<p>Data Type:</p> <p><input checked="" type="radio"/> All <input type="radio"/> Sample and Gene Data Only <input type="radio"/> Comparison Data Only</p> <p>How would you like to upload?</p> <p><input type="radio"/> Multiple Files <input checked="" type="radio"/> One Zip File</p>	<p>*: Required</p> <ul style="list-style-type: none"><li>• Project Info*: <a href="#">Project_Info.csv</a></li><li>• Sample Info*: <a href="#">Sample_Info.csv</a></li><li>• Gene Level Expression*: <a href="#">Gene_Expression_Data.csv</a></li><li>• Gene Count: <a href="#">Gene_Count.csv</a></li><li>• Comparison Info*: <a href="#">Comparison_Info.csv</a></li><li>• Comparison Data*: <a href="#">Comparisons_Data.csv</a></li></ul>
<p>Data Type:</p> <p><input type="radio"/> All <input checked="" type="radio"/> Sample and Gene Data Only <input type="radio"/> Comparison Data Only</p> <p>How would you like to upload?</p> <p><input type="radio"/> Multiple Files <input checked="" type="radio"/> One Zip File</p>	<p>*: Required</p> <ul style="list-style-type: none"><li>• Project Info*: <a href="#">Project_Info.csv</a></li><li>• Sample Info*: <a href="#">Sample_Info.csv</a></li><li>• Gene Level Expression*: <a href="#">Gene_Expression_Data.csv</a></li><li>• Gene Count: <a href="#">Gene_Count.csv</a></li></ul>
<p>Data Type:</p> <p><input type="radio"/> All <input type="radio"/> Sample and Gene Data Only <input checked="" type="radio"/> Comparison Data Only</p> <p>How would you like to upload?</p> <p><input type="radio"/> Multiple Files <input checked="" type="radio"/> One Zip File</p>	<p>*: Required</p> <ul style="list-style-type: none"><li>• Project Info*: <a href="#">Project_Info.csv</a></li><li>• Comparison Info*: <a href="#">Comparison_Info.csv</a></li><li>• Comparison Data*: <a href="#">Comparisons_Data.csv</a></li></ul>

## 5.3 Source File format

You can always check the source file format by looking at the example files. They are located below each file selection box.

Upload Your Data

\*: Required

Project Info\*:

Choose File No file chosen

[Example file](#)

Sample Info\*:

Choose File No file chosen

[Example file](#)

Gene Level Expression\*:

Choose File No file chosen

[Example file](#)

Gene Count:

Choose File No file chosen

[Example file](#)

Comparison Info\*:

Choose File No file chosen

[Example file](#)

Comparison Data\*:

Choose File No file chosen

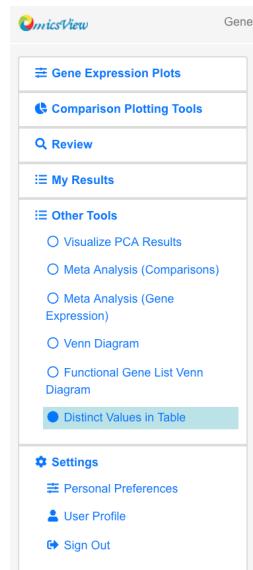
[Example file](#)

### 5.3.1 Project Info file

ProjectID	Description	Disease	Platform	Accession
GSE76161	Background: The bile acid-activated nonalcoholic steatohepatitis (NASH)		GPL11532	GSE76161
GSE57227	We report the genome-wide primary nonalcoholic steatohepatitis (NASH)		GPL11154	GSE57227
GSE67492	Gene expression in the right ventricular pulmonary arterial hypertension (PAH)	HuGene-1_0-st-v1		GSE67492

- Only ProjectID is required and must be unique across the system.
- The other four fields are highly recommended.
- The user can add one or more projects.

Save the file in the supported format and upload to the system. The system has a tool to check all the available fields and existing values for the project info file. Click on **Distinct Values in Table** under the **Other Tools** in the left menu.



And you will see a searching tool.

Distinct Values Projects is for Project Info file

select distinct  from  Submit

Distinct Values Click to view available fields

Select your interested fields to check the existing values for that field. Try your best to use the same value for the same info, which can ensure your imported projects are comparable when doing cross-project analysis.

### 5.3.2 Sample Info file

#### Recommended fields

SampleID	ProjectID	PlatformName	Description	Tissue	DiseaseState	SampleSource	Gender
E-MTAB-2325_10	E-MTAB-2325	Agilent.Expression.Human4x44k	No Info	neocortex	normal control	neocortex	male
E-MTAB-2325_11	E-MTAB-2325	Agilent.Expression.Human4x44k	No Info	neocortex	amyotrophic lateral sclerosis	neocortex	male
E-MTAB-2325_12	E-MTAB-2325	Agilent.Expression.Human4x44k	No Info	neocortex	amyotrophic lateral sclerosis	neocortex	female

- The first three fields (in bold) are required.
- SampleID should be unique in the same project.
- The projectID field should match with the projectID field in the Project Info file.
- The platformName should be one from the available values in the Distinct Values table.
- The other fields listed here are not required but highly recommended. These are used to capture information provided by researchers about the samples.
- The headers must match one of the distinct fields in the Distinct Values tool. If the column header provided by the user does not match any available field, the system will show the unmatched fields in the file and let you manually map it to a known field.

To check the available fields and existing values, you can also check the **Distinct Values** tool by selecting **Samples** in the box.

### 5.3.3 Comparison Info file

#### Recommended fields

ComparisonID	PlatformName	ProjectID	Case.SampleIDs	Control.SampleIDs	ComparisonCategory	ComparisonContrast	Case.DiseaseState	Case.Tissue
E-MTAB-4304.GPL11; NGS.III.illumina.HiSeq	E-MTAB-4304	E-MTAB-4304_0;E-MTA	E-MTAB-4304_1;E-MTA Tissue1 vs. Tissue2			TissueRegion => distal medial condyl knee osteoarthritis	knee cartilage	
E-MTAB-4586.GPL11; NGS.III.illumina.HiSeq	E-MTAB-4586	E-MTAB-4586_0;E-MTA	E-MTAB-4586_14;E-MT Disease vs. Normal			DiseaseState => Parkinson's disease	(Parkinson's disease	skin
E-MTAB-4586.GPL11; NGS.III.illumina.HiSeq	E-MTAB-4586	E-MTAB-4586_0;E-MTA	E-MTAB-4586_14;E-MT Other Comparisons			GeneticModification => LRRK2 G2019S	Parkinson's disease	skin

- The first three fields (in bold) are required.
- ComparisonID should be unique in the same project.
- The projectId field should match with the projectId field in the Project Info file.
- The platformName should be one from the available values for the platformName in the Distinct Values table.
- The next two fields Case.SampleIDs and Control.SampleIDs should be used to list the sampleIDs used as cases or controls for the comparison. Separate sampleIDs by comma.
- The other fields listed here are not required but highly recommended. For example, ComparisonCategory is used in group comparisons in dashboard. Case.DiseaseState and Case.Tissue are used in coloring or grouping bubble plots.
- The headers must match one of the distinct fields in the Distinct Values tool. If the column header provided by the user does not match any available field, the system will show the unmatched fields in the file and let you manually map it to a known field.

## 5.4 Data Formats

There are certain formats for gene expression data and comparison data files. The data format below applies to both RNA-Seq or Array data.

### 5.4.1 Gene Expression Data

For gene expression data (RNA-Seq or microarray), two types of formats are supported.

Matrix format										Table format		
Ensembl_geneID	D14_CH0_E_1	D14_CH0_3	D14_CH0_4	D14_H1_1	D14_H1_2	D14_H1_5	D21_CH0_E_4	D21_CH0_E_5	Gene	SampleID	Expression	
ENSG00000239779	0.6999720549569	0.2199708170379	0.3020909655117	0.3599950951579	0.6530050951579	0.47173077494616	0.3939850914177	0.35458650626006	ARRB2	lesional_trt_2	16.0249	
ENSG00000272520	1.6935770086553	16.0731060504737	13.4048053743489	13.22919320233	14.84650451000	12.949622325051	12.156852195789	12.156852195789	ARRB2	non_lesional_none_2	10.1214	
ENSG00000280551	1.1660500000000	0.60672006274945	0.60672006274945	0.60672006274945	0.60672006274945	0.60672006274945	0.60672006274945	0.60672006274945	ARRB2	non_lesional_trt_2	8.4008	
ENSG00000289053	0.1665000000000	0.00672006274945	0.00672006274945	0.00672006274945	0.00672006274945	0.00672006274945	0.00672006274945	0.00672006274945	ARRB2	lesional_none_3	25.01616	
ENSG00000289748	0.2689510880511951	0.11730714745722	0.64190203318596	0.220134815181299	0.16472005014006	0.13950931155841	0.05098116032798	0.1944796250234	ARRB2	lesional_trt_3	16.5416	
ENSG00000289855	0.6930132049335	7.74320179369565	10.28756688065	7.74320179369565	11.895571578205	10.2792903593191	12.7474151050378	13.415150688987	ARRB2	non_lesional_none_3	9.1222	
ENSG00000290002	0.819623702931881	0.8398736980801402	1.8132407908008	3.42327356484147	3.37032082744007	0.9150206490472	1.9781255713208	2.9374848676234	ARRB2	non_lesional_trt_3	9.501	
ENSG00000290160	0.760021597385181	0.7605737048474715	0.7305408420207	1.21500516191649	1.33803461970374	0.9443273243099	1.00221410404843	0.9571543019649	ARRB2	lesional_none_4	20.2206	
ENSG00000290200	0.10403180745474	0.020000000000000	0.10403180745474	0.10403180745474	0.10403180745474	0.10403180745474	0.10403180745474	0.10403180745474	ARRB2	lesional_trt_4	21.5801	
ENSG00000290270	3.77884695209	0.798170000000000	0.151640400000000	0.84721740200000	0.78477880000000	0.50700000000000	0.17100000000000	0.17100000000000	ARRB2	non_lesional_trt_4	14.3564	
ENSG00000290575	1.151772402000000	0.85745420714805	0.151640400000000	2.3020850031045	1.804320032717988	0.669003031106	1.011730931454	0.9000465237102	ARRB2	lesional_trt_5	8.947	
ENSG00000290685	0.151772402000000	0.109302129000000	0.66465489033201	0.15177105108788	0.15177105108788	0.09545590006199	0.09545590006199	0.09545590006199	ARRB2	non_lesional_trt_5	12.11129	
ENSG00000290693	0.869940503000000	1.970116187474670	1.9381270710714	1.3290505965485	1.6313190059302	0.2208278156667	1.7787327314887	1.4631400365079	ARRB2	lesional_none_5	9.8666	
ENSG00000290747	105.148148320353	85.3491331067867	123.182183381176	78.804583383028	72.030307093374	71.982011153009	85.294483290508	101.137440036874	ARRB2	non_lesional_none_5	11.22102	
ENSG00000290759	153.182183381176	136.32755948464	125.182183381176	120.200742002024	103.182183381176	186.92799854593	127.182183381176	127.182183381176	ARRB2	lesional_trt_5	11.22102	
ENSG00000290779	0.253000000000000	0.261000000000000	0.261000000000000	0.261000000000000	0.261000000000000	0.261000000000000	0.261000000000000	0.261000000000000	ARRB2	non_lesional_trt_5	11.22102	
ENSG00000290844	47.4517319000000	0.041021019132	30.361107032004	23.0545261014029	18.2700052005089	18.2700052005089	18.2700052005089	18.2700052005089	ARRB2	non_lesional_none_5	11.22102	
ENSG00000290949	12.86479791995	7.508439739714989	10.5458177377166	4.119968777166	5.05602640505901	5.7112182470291	4.4315678192294	5.4430015210513	ARRB2	lesional_trt_5	11.22102	
ENSG00000290957	864.34781000000	950.99427171805	526.5900283000000	568.111059091164	543.2840250726676	1154.07709448274	1122.11338900537	1122.11338900537	ARRB2	non_lesional_none_5	6.7286	
ENSG00000290744	121.116646884483	107.068810269004	129.845320132055	83.254482001481	47.5482411328997	57.004351310582	45.474488956285	45.474488956285	ARRB2	non_lesional_none_5	6.7286	

The **Gene** must be listed in the first column. The IDs can be official gene symbols, gene IDs (Entrez Id), or Ensembl IDs. The system will automatically recognize the IDs and map the IDs to the gene annotation table. The SampleIDs must match the sampleIDs in the Sample Info file.

### 5.4.2 Comparison Data

For comparison data, the following template should be used.

Gene	ComparisonID	logFC	P.Value	adj.P.Val
ENSG00000223972	D14_CHD8.vs.D14_H9	-0.480181603380462	0.201172869981358	0.276823114956239
ENSG00000227232	D14_CHD8.vs.D14_H9	-0.0260065294185647	0.782508711249897	0.834365881843091
ENSG00000243485	D14_CHD8.vs.D14_H9	0.614759995164825	0.12397256363312	0.183494049719199
ENSG00000238009	D14_CHD8.vs.D14_H9	-0.489128093633942	0.325707592057561	0.414162918866014
ENSG00000233750	D14_CHD8.vs.D14_H9	-1.10014257817111	0.0192044755497128	0.0369617315251983
ENSG00000237683	D14_CHD8.vs.D14_H9	-0.605292542573319	3.41124756968239E-06	2.19219164320563E-05
ENSG00000239906	D14_CHD8.vs.D14_H9	-1.0010125449076	0.0636863585580712	0.103931911218667
ENSG00000241860	D14_CHD8.vs.D14_H9	-0.714487254252856	0.000262496694401327	0.000917136064345589
ENSG00000228463	D14_CHD8.vs.D14_H9	-1.18951263566867	0.0140347276118906	0.0282154235613193
ENSG00000237094	D14_CHD8.vs.D14_H9	-0.691939730736816	3.54212236297494E-07	3.17551702158323E-06
ENSG00000250575	D14_CHD8.vs.D14_H9	-1.28179600460206	0.000183559753891093	0.000670021987202936
ENSG00000233653	D14_CHD8.vs.D14_H9	-1.56406411915058	0.0289389868237879	0.0526241552771901
ENSG00000236679	D14_CHD8.vs.D14_H9	0.0814532190594377	0.835473327403157	0.877408636003058
ENSG00000225972	D14_CHD8.vs.D14_H9	0.545641008555407	0.000139660955234846	0.000526980960540253

List one or more comparisons in the table. The **Gene** must be listed in the first column. The IDs can be official gene symbols, gene IDs (Entrez Id), or Ensembl gene IDs. The system will automatically recognize the IDs and map the IDs to the gene annotation table. The second column must be **ComparisonID**, which must match the comparisonIDs for the Comparison Info file. For each row, three values should be entered, **logFC** is the log2 Fold Change, **P.Value**, and **adj.P.Val** (FDR). Most statistical packages output these three values. If some values are missing, enter NAs.

## 6 Advanced Analyses

### 6.1 Correlation Tools

Once the user has identified a gene of interest, the user can use correlation tools to find other genes that share similar (or opposite) profiles in terms of gene expression or fold change. First, enter the gene of interest, and samples to be used for correlation. In the example below, we entered OSM gene, and 69 samples involved in Crohn's disease.

Correlation Tools Using Gene Expression

Source Gene Names:  Load Saved Genes

Sample IDs:  Load Saved Sample IDs

1. Start here

2. Enter gene of interests

3. Enter or load saved samples

Advanced Options

How do you like to compare the genes?

Correlation Method:  Pearson Correlation  Spearman Correlation (Pearson Correlation Coefficient Between Ranked Variables)

4. Set options

Enable Log<sub>2</sub> Transform

Direction of Correlation: Both

Cut-off of Correlation Coefficient: 0.80

Maximum Number of Top Matched Genes: 100

5. Submit

Please click [here](#) for the search summary.

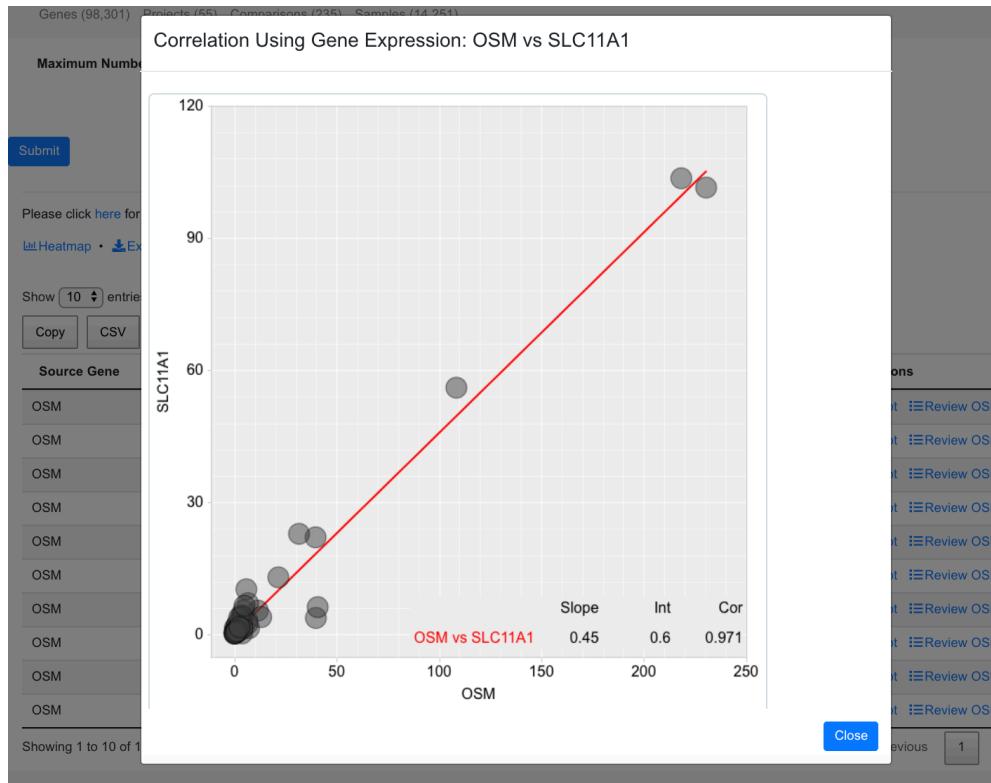
The result shows a table of correlated genes, ranked by R<sup>2</sup>

Source Gene	Matched Gene	Correlation Coefficient	R <sup>2</sup>	# of Data Point	Actions
OSM	SLC11A1	0.9855	0.97121	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review SLC11A1</a>
OSM	VNN3	0.91343	0.83435	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review VNN3</a>
OSM	BCL6	0.90543	0.8198	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review BCL6</a>
OSM	ADM	0.90802	0.8245	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review ADM</a>
OSM	MEFV	0.90972	0.82759	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review MEFV</a>
OSM	HP	0.90977	0.82768	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review HP</a>
OSM	LOC729737	0.91066	0.8293	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review LOC729737</a>
OSM	OSCAR	0.91094	0.82981	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review OSCAR</a>
OSM	SIGLEC9	0.9115	0.83083	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review SIGLEC9</a>
OSM	C5AR1	0.91282	0.83324	68	<a href="#">Plot</a> <a href="#">Review OSM</a> <a href="#">Review C5AR1</a>

Showing 1 to 10 of 100 entries

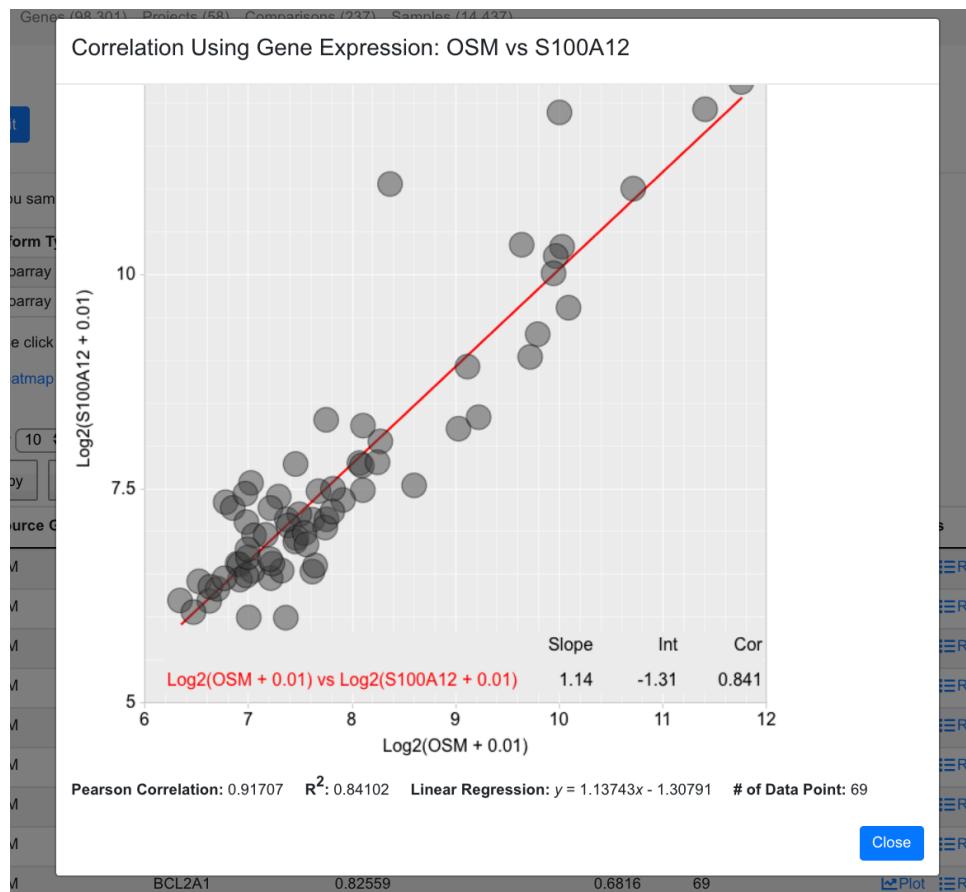
Previous 1 2 3 4 5 ... 10 Next

Click the plot icon will show scatter plot of the target and the correlated gene.



An additional way to plotting the correlation with other genes, is to enable the log2 transformation of the expression.

The screenshot shows the "Correlation Tools Using Gene Expression" interface. On the left, a sidebar menu is open, showing various tools and options. The "Correlation Tools Using Gene Expression" option is selected and highlighted with a red box. The main panel displays "Source Gene Names" (OSM) and "Sample IDs" (GSM155564, GSM155565, GSM155566, GSM155567, GSM155568, GSM155569, GSM155570, GSM155571). Below this, there are sections for "Advanced Options" and "Correlation Method". A red box highlights the "Enable Log<sub>2</sub> Transform" checkbox, which is checked. Other settings shown include "Value to Be Added for Log Transformation: 0.01", "Direction of Correlation: Both", "Cut-off of Correlation Coefficient: 0.80", and "Maximum Number of Top Matched Genes: 100".



Now the expression is log transformed.

The examples above are from Correlation Tools using Expression. The interface and usage for Correlation Tools using Comparison is similar.

**Gene Expression Plots**

- Comparison Plotting Tools**
- Bubble Plot (Single Gene)
- Bubble Plot (Multiple Genes)
- Volcano Plot
- WikiPathways Visualization
- KEGG Visualization
- Reactome Visualization
- Pathway Heatmap Tool
- Correlation Tools Using Comparisons**
- Export Genes and Comparisons
- Similar Comparisons (GO)
- Similar Comparisons (PAGE)
  - Comparisons Venn Diagram (GO)
  - Comparisons Venn Diagram (PAGE)
- Review
- Review Genes
- Review Projects

**Correlation Tools Using Comparisons**

Source Gene Names:

Please enter one or more gene names, separated by line break.

Comparison IDs:

Please enter two or more comparison IDs, separated by line break.

**Advanced Options**

How do you like to compare the genes?

Calculate the correlations against all available genes in database  
 Calculate the correlations among the entered genes only

Correlation Method:

Pearson Correlation  
 Spearman Correlation (Pearson Correlation Coefficient Between Ranked Variables)

Direction of Correlation:

Cut-off of Correlation Coefficient:

Maximum Number of Top Matched Genes:

## 6.2 PCA Analysis

Users can select a few samples and use PCA plot to visualize the sample relationships.

**PCA Tool for Genes & Samples**

Gene Names:  Load Saved Genes

Sample Names:  Load Saved Samples

Gene Expression Plots

Comparison Plotting Tools

Review

My Results

1. Start here

Other Tools

Visualize PCA Results (selected)

Meta Analysis (Comparisons)

Meta Analysis (Gene Expression)

Venn Diagram

Functional Gene List Venn Diagram

Distinct Values in Table

Settings

Personal Preferences

User Profile

Sign Out

Basic PCA Tool

FactoMineR Analysis

1. Start here

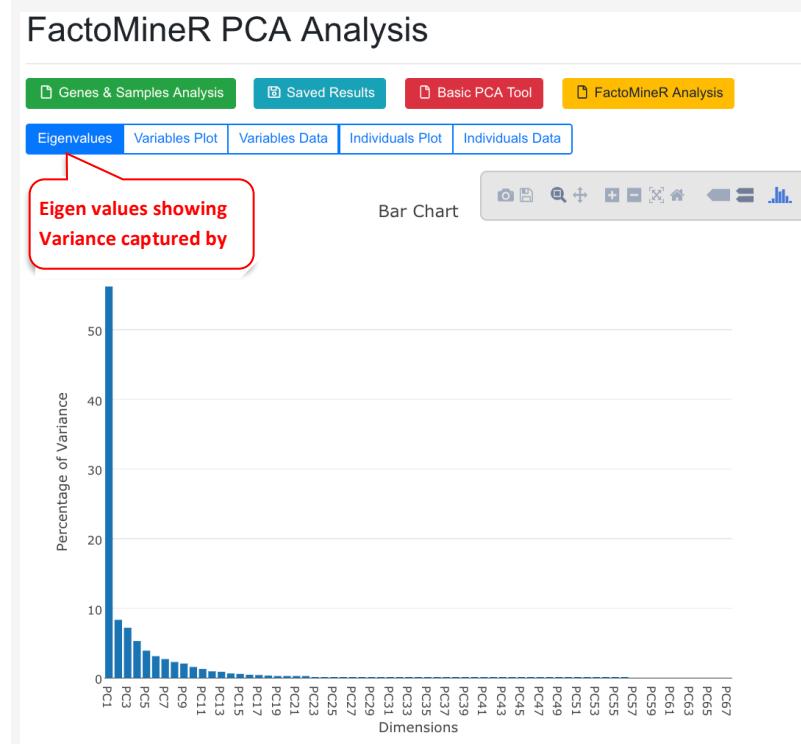
2. Enter gene of interests

3. Enter or load saved samples

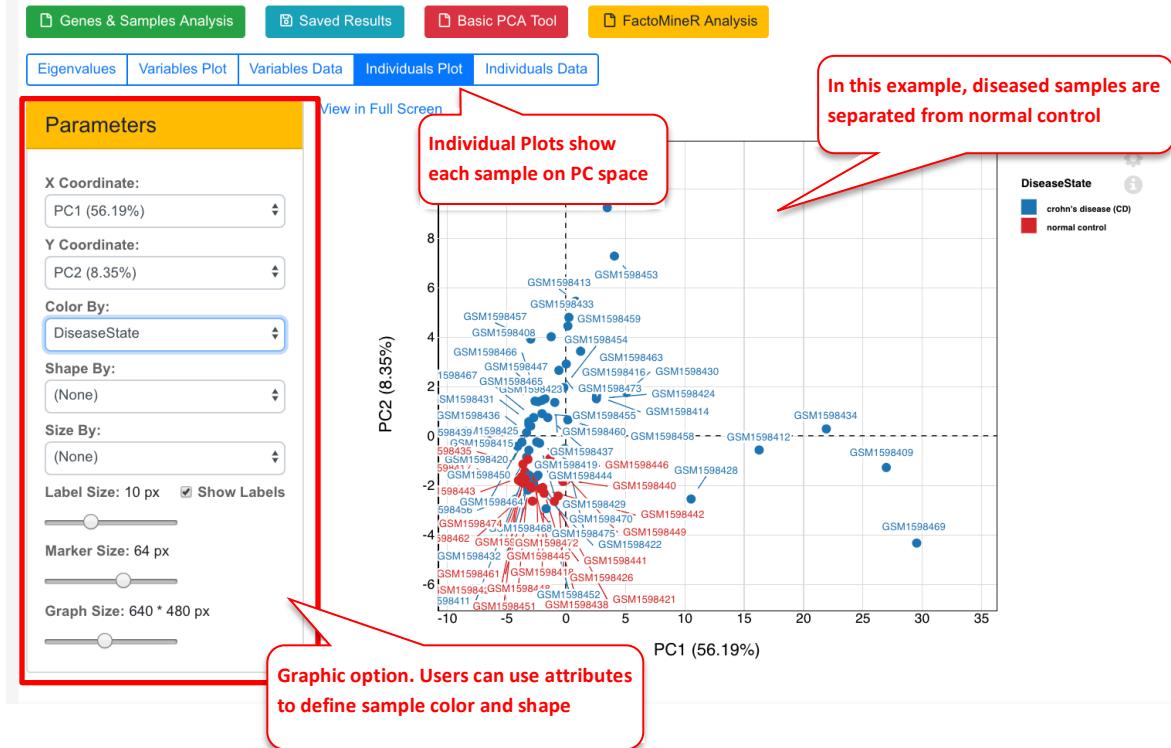
4. Set options

5. Submit

The system will use FactorMineR package to run PCA analysis and display the results.



## FactoMineR PCA Analysis



The PCA tool can also display results from pre-calculated data (basic PCA, or FactorMineR results).

## PCA Scatter Plot Tool

The screenshot shows the PCA Scatter Plot Tool interface. At the top, there are four tabs: Genes & Samples Analysis (green), Saved Results (blue), Basic PCA Tool (red, selected), and FactoMineR Analysis (yellow). Below these are four upload fields: Data File (Choose File, No file chosen), Attributes File (Choose File, No file chosen), Variance File (Choose File, No file chosen), and Format (radio buttons for csv and txt / tsv, with csv selected). A red box highlights the 'Basic PCA Tool' tab.

## FactoMineR PCA Analysis

Required Files

- PCA\_barchart.csv
- PCA\_ind.contrib.csv
- PCA\_ind.coord.csv
- PCA\_ind.cos2.csv
- PCA\_var.contrib.csv
- PCA\_var.coord.csv
- PCA\_var.cor.csv
- PCA\_var.cos2.csv

Optional Files

- PCA\_attributes.csv
- PCA\_quali.sup.coord.csv
- PCA\_quali.sup.cos2.csv
- PCA\_quanti.sup.coord.csv
- PCA\_quanti.sup.cor.csv
- PCA\_quanti.sup.cos2.csv

Data File:  Choose File No file chosen  
[Download Example Zip](#)

### 6.3 Meta-Analysis

Meta-Analysis can be used to identify genes that are changed consistently across multiple projects. This can be either implemented by doing Meta-analysis using comparisons or meta-analysis using gene expression.

#### 6.3.1 Meta-Analysis (Comparisons)

In the example below, we are looking for the most significant DEGs in Disease vs. Normal comparisons of Crohn's Disease.

1. Start here

2. Leave empty to analyze all genes or enter gene lists for focused analysis

3. Enter comparison list

4. Set options

5. Click and wait for results, may take a few minutes

The system will use the comparison data (logFC, p-value) to compute combined p-value and rank product. It also uses a simple cutoff to get counts of up and down regulated genes. This method is fast and can be applied to any comparison data. However, it does not use the individual sample data or consider number of samples in each comparison.

Please click [here](#) for the search summary.

[Save to a Study](#) [Create a Gene List](#) [Bubble Plot](#) [Export Genes and Comparisons](#) [Download](#)

Show 100 entries [Copy](#) [CSV](#) [Excel](#) [PDF](#)

**Simple count results**

**Rank product results**

**MetaDE results**

Actions	Gene Name	Entrez ID	Description	# of Data Points	Upregulated (%)	Downregulated (%)	Rank Products	Rank Products (log2 Fold Change)	Rank Products (p-value)	Rank Products (FDR)	Combined p-value (Fisher)	Combined p-value (Maximum p-value)	Combined FDR (Fisher)	Combined FDR (Maximum p-value)
	hsa-mir-151a			0										
	AL157440			8	0	25	305.8	-1.11	0	0	0	0	0	0
	hsa-mir-1302-7			0										
	LIPN	643418	lipase family member N	2	0	0	930.6	0.87	0	0.00015				
	ANKR022	118932	ankyrin repeat domain 22	7	85.71	0	247	1.65	0	0	0	0	0	0
	hsa-mir-4472-1			0										
	FBXO6	26270	F-box protein 6	7	42.86	0	730.1	1.06	0	4.0E-5	0	0	0	1.0E-5
	DG588968			0										
	LINC00864	728218	long intergenic non-protein coding RNA 864	1	0	100	210	-1.09	0	0				
	hsa-mir-1302-11			0										

We can enable filter by percentage up-regulated genes to focus on genes that are consistently up-regulated.

From the output table, the user can select the genes most interesting to them. Some key columns are explained below.

- # of data points: this shows the number of comparisons with valid data for this gene. Although 8 comparisons were entered here, not all genes show up in all experiments. We used minimal 3 data points when we set up display options.
- Simple count results. Using the cutoff values that we entered (logFC 1, FDR 0.05), the system checks if a gene passes the cutoff for each comparison, and computes percentage up and down regulation.
- Rank Product results (recommended results for meta-analysis). This is done with the RankProd package. The system ranks genes in each comparison using logFC and computes a combined rank and statistical values. The Rank Products columns show the rank (smaller is more consistent change), log2Fold Change and FDR.
- The Combined p-value/FDR (Fisher or Maximum) are computed using the individual p-values with MetaDE package. Note unlike RankProd, these values didn't consider the direction of the change, so the data need to be interpreted carefully, and we recommend using these p-values together with the simple count values to have a better understanding.

The table can be sorted by any column. In the above view, we sorted the data by logFC. The table can also be sorted by # of data points to view the genes with most data. In the example below, we chose 10 up-regulated genes with most data points from the table and used the bubble plot button to create bubble plot. Indeed, all these genes show consistent up regulated in the selected comparisons.

## Genes & Comparisons Bubble Plot

[Load Example Data](#)

Gene Names: [Load Saved Genes](#)

[Select Gene Set](#)

Comparison IDs: [Load Saved Comparison IDs](#)

S100A8  
MMP1  
MMP3  
HCAR3  
DUOX2  
REG1B

GSE57945.GPL11154.DESeq2.test1  
GSE16879.GPL570.test14  
GSE16879.GPL570.test15  
GSE16879.GPL570.test3  
GSE16879.GPL570.test5  
GSE52746.GPL17996.test1

Source of the Comparison IDs:

Omicsoft Data  Internal Data

Chart Height Scale Factor:

1

Chart Left Margin Scale Factor:

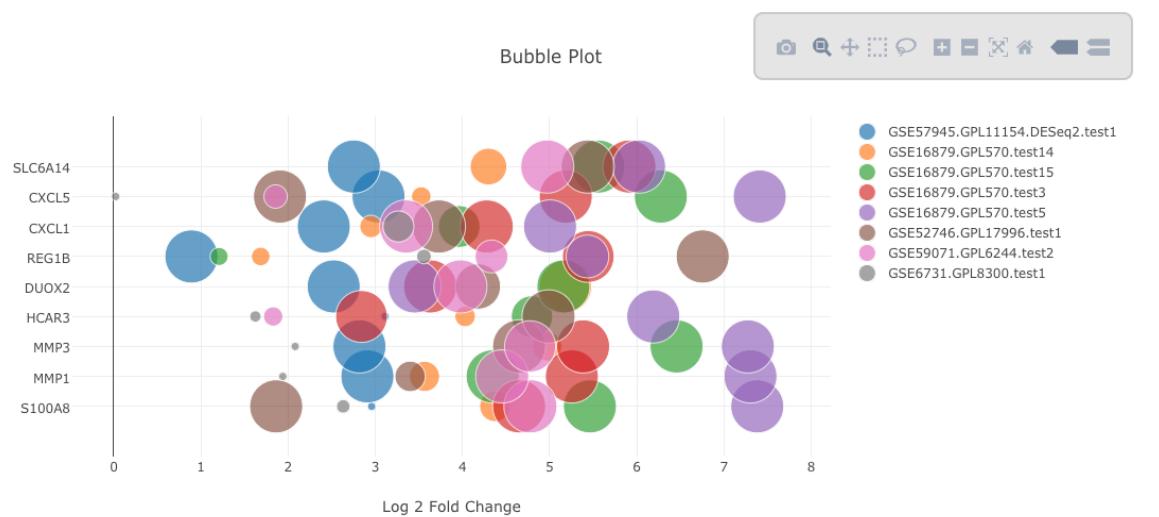
1

Display Columns:

Log<sub>2</sub> Fold Change  p-value  FDR

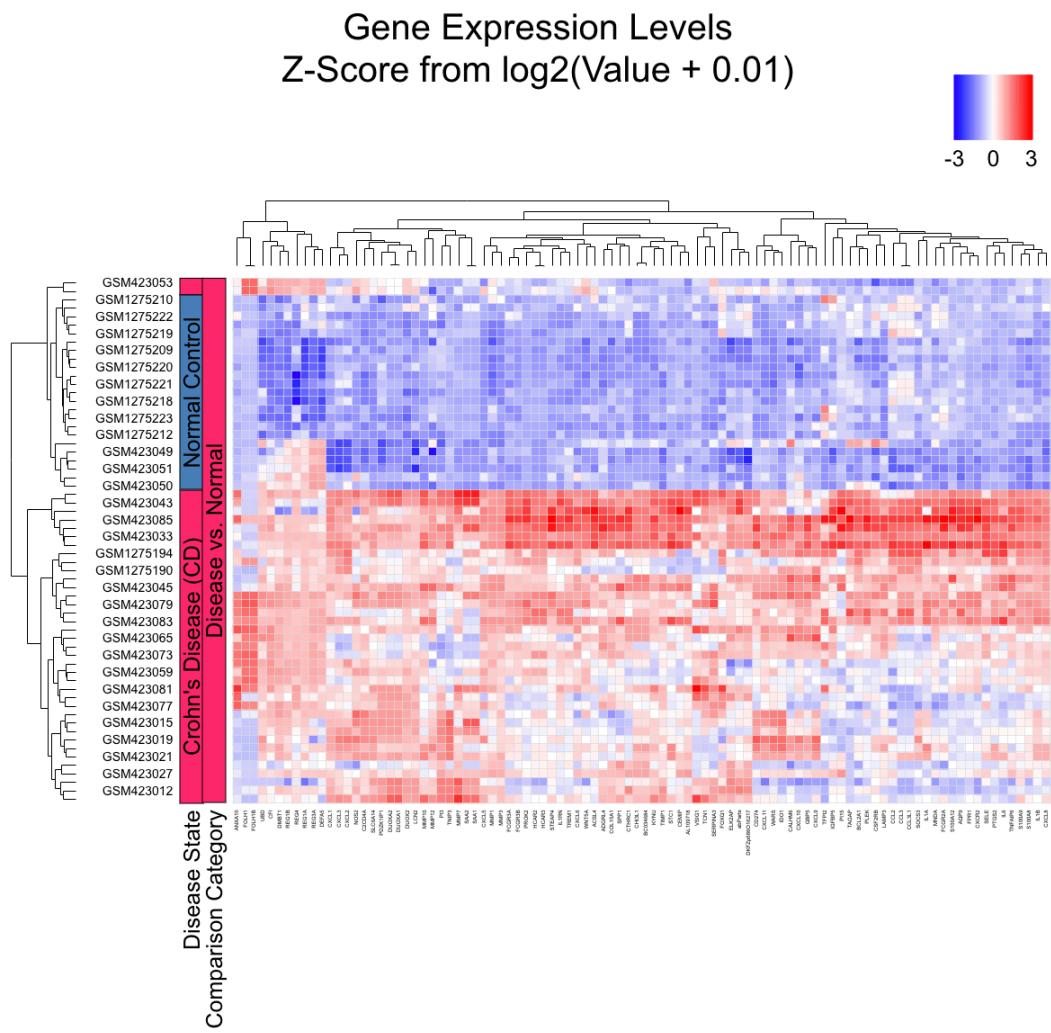
[Plot](#)

The resulting bubble plot will show all 8 comparisons for each gene.



We can also create a gene list from the table for up or down gene lists from meta-analysis (using the green button at top of meta-analysis results) and use the list for subsequent analysis like functional enrichment or generating heatmap. Below is a heatmap using 100 up-regulated genes from meta-

analysis, and it is very clear that the diseased samples have an upregulation of these genes.



### 6.3.2 Meta-Analysis (Gene Expression)

In the example below, we use the same comparisons from Crohn's disease.

**Gene Expression Plots**

**Comparison Plotting Tools**

**Review**

**My Results**

**Other Tools**

- Visualize PCA Results
- Meta Analysis (Comparisons)
- Meta Analysis (Gene Expression) 1
- Venn Diagram
- Functional Gene List Venn Diagram
- Distinct Values in Table

**Settings**

**Personal Preferences**

**User Profile**

**Sign Out**

**Meta Analysis Using Gene Expression Data**

**Select Comparisons** 2

**Edit Samples**

**Review Results**

**1. Select Comparisons**

**Meta Analysis Name:** Crohns disease from expression 2

**Comparison IDs:** Load Saved Comparison IDs 3

GSE57945.GPL11154.DESeq2.test1  
GSE16879.GPL570.test14  
GSE16879.GPL570.test15  
GSE16879.GPL570.test3  
GSE16879.GPL570.test5  
GSE52746.GPL17996.test1  
GSE59071.GPL6244.test2  
GSE6731.GPL8300.test1

**Source of the Comparison IDs:**  
 Omicssoft Data  Internal Data Select Internal Project: 3

**Options:**  
 Perform Rank Product Analysis 4

**Continue**

By default, the Rank Product Analysis is checked. Uncheck this if the data is too big for Rank product analysis, or you just need a quick run using limma for each individual comparison.

On the next page, system will display the Case and Control samples for each comparison. At this step, users can edit samples. Users can use this page to remove outliers or add other samples they want to include in the meta-analysis.

## Meta Analysis Using Gene Expression Data



### 2. Edit Samples

Comparison #1:  
GSE57945.GPL11154.DESeq2.test1

Case Sample IDs:

- GSM1598408
- GSM1598409
- GSM1598410
- GSM1598411
- GSM1598412
- GSM1598413
- GSM1598414
- GSM1598415

Control Sample IDs:

- GSM1598495
- GSM1598496
- GSM1598497
- GSM1598524
- GSM1598525
- GSM1598527
- GSM1598528
- GSM1598587

Comparison #2:  
GSE16879.GPL570.test14

Case Sample IDs:

- GSM423053
- GSM423055
- GSM423057
- GSM423059

Control Sample IDs:

- GSM423047
- GSM423048
- GSM423049
- GSM423050

After users review the samples for each comparison, they can go to the next step to run the meta-analysis. This step can take a while if the number of samples is large. For the meta-analysis, the system will do the following:

- 1) Run limma to get logFC, p-value and confidence interval for each individual comparison
- 2) Run MetaDE to get combined p-values
- 3) Run Rank Product package for all the expression data to get statistical significance and combined logFC.

After the analysis is done, the system will show a summary and the result table.

## Meta Analysis Using Gene Expression Data

[Create New Meta Analysis](#)

[Summary](#) [Meta Analysis Results](#)

Meta Analysis Results for Crohns disease meta

Comparisons used in the analysis:

Comparison Name	Comparison Number	Average Number of Genes for Meta Analysis	Average Number of Not Available (NA) Genes	# of Case Samples	# of Control Samples
GSE57945.GPL11154.DESeq2.test1	1	30,373	0	4	4
GSE16879.GPL570.test14	2	21,893	8,480	4	4
GSE52746.GPL17996.test1	3	21,893	8,480	3	4
GSE59071.GPL6244.test2	4	21,562	8,811	4	4

[View samples in each comparison](#) ([Comparison\\_List.csv](#)) | [View number of genes in each comparison](#) ([Sample\\_geneCount.csv](#))

From **30,373** genes listed in the comparisons, **19,214** genes are present in all comparisons and thus produce statistical results from meta-analysis.

Note: The statistical values from meta analysis are only available for genes that are present in all comparisons. If the number of genes in one comparison is much lower than others, consider running another meta-analysis without this comparison to get statistical values for more genes.

Meta Analysis Results

[Download](#)

Show **100** entries

[Copy](#) [CSV](#) [Excel](#) [PDF](#)

Search:

Actions	Symbol	Description	logFC_RP	P.Val_RP	FDR_RP	RankProd	ES_pval	ES_FDR	logFC_Ave	logFC_1	SE_1	CIL_1	CLR_1	PValue_1	adj.PVal_1	logFC_2	SE_2	CIL_2	CLR_2	PValue_2	adj.PVa
<a href="#">View</a>	SS_rRNA	NA	-0.00345	0.00377	-0.00908	0.00908	1	1													
<a href="#">View</a>	SS_rRNA	NA	-0.00408	0	0.00377	-0.00908	0.00908	1	1												
<a href="#">View</a>	SS_rRNA	NA	-0.00052	0.00377	-0.00908	0.00908	1	1													
<a href="#">View</a>	SS_rRNA	NA	0	0	0.00377	-0.00908	0.00908	1	1												
<a href="#">View</a>	SS_rRNA	NA	0.00314	0.00377	-0.00908	0.00908	1	1													
<a href="#">View</a>	SS_rRNA	NA	-0.46437	-1.24584	0.40456	-2.22073	-0.27095	0.01993	0.30861												
<a href="#">View</a>	SS_rRNA	NA	0	0	0.00377	-0.00908	0.00908	1	1												
<a href="#">View</a>	SS_rRNA	NA	0.00445	0	0.00377	-0.00908	0.00908	1	1												
<a href="#">View</a>	SS_rRNA	NA	0.00029	0	0.00377	-0.00908	0.00908	1	1												
<a href="#">View</a>	SS_rRNA	NA	-0.00108	0	0.00377	-0.00908	0.00908	1	1												
<a href="#">View</a>	SS_rRNA	NA	0	0	0.00377	-0.00908	0.00908	1	1												

In the summary, the four comparisons are shown, along with number of genes used for meta-analysis, and number of case and control samples. Because different platforms (especially arrays) have different number of genes present, and some genes may have too low signals to be detected in certain projects, not all genes are present in all the comparisons. In this example, 19,214 genes are present in all comparisons. In some cases, it may be beneficial to remove some comparisons with small number of genes.

In the meta-analysis result table, results from Rank Product are shown (from RankProd package, preferred results to use), followed by effective size method (MetaDE.ES from MetaDE package). The logFC, SE, confidence interval, p-value and FDR for each comparison are shown next. The comparison numbers are the same as those listed in the summary table above. The gene list can be sorted by RankProd (most significant changes) or by logFC\_RP (up or down regulated).

From this analysis, we can create a Forest plot for each gene by clicking the icon near the gene symbol.

Show 100 entries

**Sort by Rank Product**

Search:

Actions	Symbol	Description	logFC_RP	P.Val_RP	FDR_RP	RankProd	ES_pval	ES_FDR	logFC_Ave	logFC_1	SE_1	CIL_1	CIU_1	P.Value_1	adj.PVal_1	logFC_2	SE_2	CIL_2	CIU_2	P.Value_2	adj.PVal_2	
<a href="#">Copy</a>	<a href="#">CSV</a>	<a href="#">Excel</a>	<a href="#">PDF</a>																			
	DUOX2	dual oxidase 2	4.06492	0	0	34	0	0	4.39622	4.70365	1.00363	2.28514	7.12215	0.00284	0.131	5.22679	0.49383	4.12697	6.32661	0	0.00287	4.
	SLC6A14	solute carrier family 6 (amino acid transporter), member 14	4.12759	0	0	37.07	0	0.00011	4.32919	1.57219	0.88038	-0.5493	3.69369	0.12123	0.59147	4.69476	0.47514	3.63656	5.75295	0	0.00373	5.
	S100A8	S100 calcium binding protein A8	4.50373	0	0	39.74	0	0	4.23725	3.3074	1.11241	0.62676	5.98803	0.02298	0.31688	3.97632	0.63784	2.55577	5.39687	0.0001	0.02651	4.
	MMP3	matrix metallopeptidase 3	3.98462	0	0	44.82	0	0	4.46926	4.74546	1.18491	1.89011	7.60081	0.00619	0.1807	5.22652	0.63144	3.82023	6.63281	1.0E-5	0.00779	3.
	CXCL8	chemokine (C-X-C motif) ligand 8	3.38594	0	0	54.88	0	0	4.09838	3.58179	0.7145	1.86003	5.30356	0.002	0.11173	4.51577	0.754	2.83651	6.19502	0.00013	0.03024	5.
	REG1B	regenerating islet-derived 1 beta	2.82828	0	0	76.76	0.00601	0.30687	4.36029	4.59136	0.74814	2.78854	6.39419	0.00067	0.06874	1.4676	0.35593	0.6749	2.2603	0.00205	0.10255	7.
	CLDN8	claudin 8	-3.43029	0	0	90.02	8.0E-5	0.01266	-2.49827	-0.99252	0.87849	-3.10945	1.12441	0.29907	0.81485	-2.81377	0.70135	-4.39376	-1.26977	0.00235	0.10904	-2.
	MMP1	matrix metallopeptidase 1	3.09665	0	0	105.9	0	0	3.39109	4.13223	1.17003	1.31274	6.95172	0.01107	0.23571	3.4349	0.49049	2.34253	4.52728	4.0E-5	0.01578	2.

**Click to view Forest Plot**

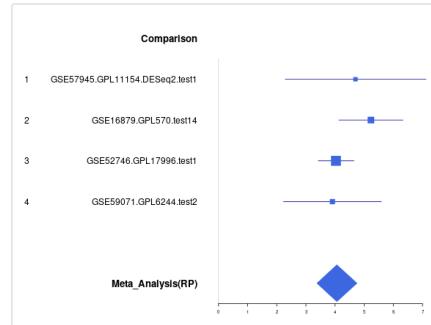
In the Forest plot, results from individual comparison and combined data are shown. The confidence interval for individual comparison is based on limma output. The confidence interval for combined results is based on p-value from RankProd (or ES method when RankProd is not run).

### Forest Plot for Gene DUOX2

[Return to Meta Analysis Results](#) [Review Gene Details](#)

Meta Analysis: Crohns disease meta  
 Gene Name: DUOX2  
 EntrezID: 50506  
 Description: dual oxidase 2

Download Plot: [SVG](#) • [PNG](#) • [PDF](#)



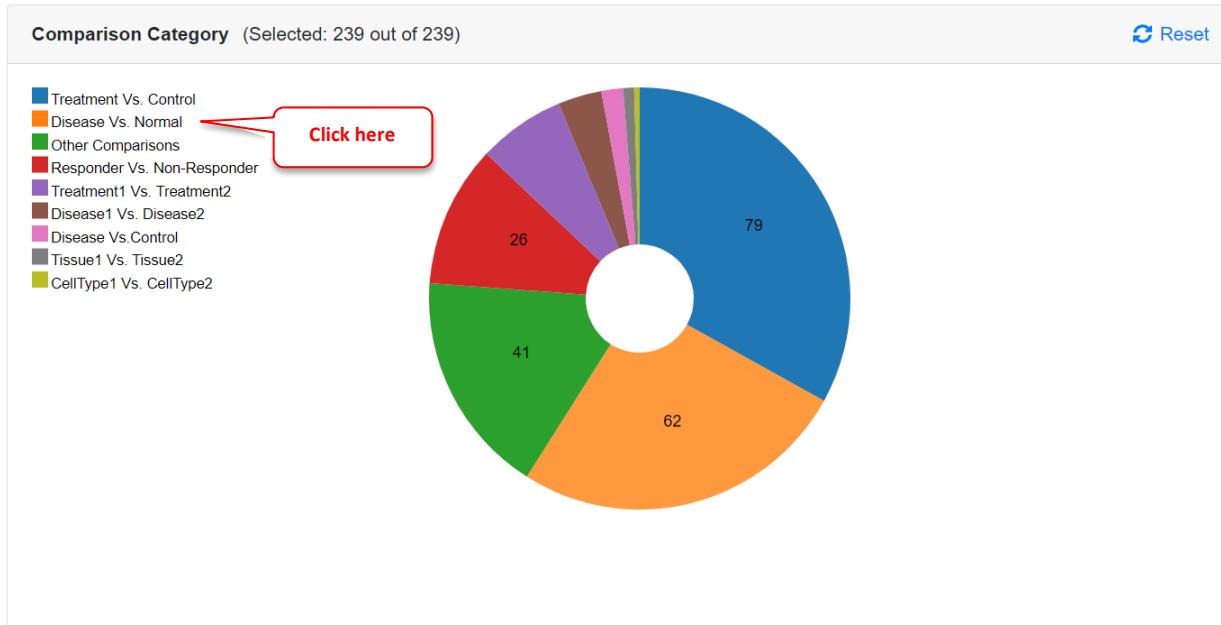
The plot can be saved as SVG or PDF. The summary table is shown below the plot.

## 7 Step-by-step guide of running the Crohn's disease meta-analysis

To provide more guidance on how to carry out an analysis, in this section an end-to-end example is shown to demonstrate how to draw meaningful conclusions by data exploration.

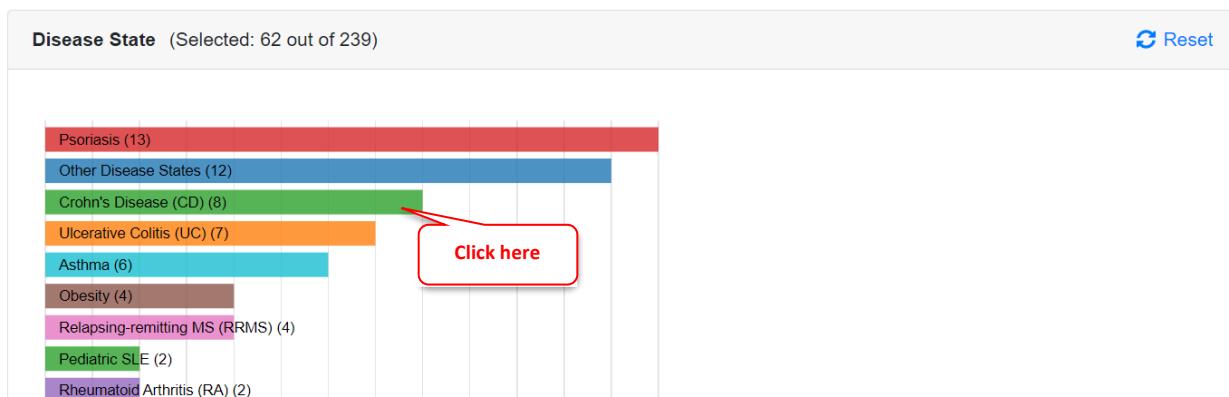
### Step 1: Select Disease Vs. Normal from the Comparison Category

Click on the OmicsView logo on the top-left corner or select “Comparison Dashboard” under “Review” found in the left panel. Then click on “Disease Vs. Normal” under Comparison Category.



### Step 2: Select Crohn's Disease comparisons

Click on the “Crohn's Disease” bar in the Disease State subpanel of the Comparison Dashboard.



### Step 3: Add the comparisons to a Comparison List

On the same Comparisons Dashboard page, scroll down to the table at the bottom.

Click on the top checkbox to select all 8 comparisons – You will see 8 check marks as in screenshot below. Then click on “Create a Comparison List”.

The screenshot shows a table titled "Comparisons (8)". At the top left, there is a "Actions" column with a checkbox. A red box labeled "1. Click here" points to this checkbox. Another red box labeled "2. Then click here" points to the "Create a Comparison List" button at the top left of the table area.

Comparisons (8)								
	Actions	Comparison ID	Case Cell Type	Case Disease State	Case Treatment	Comparison Category	Comparison Contrast	Comparison Type
<input checked="" type="checkbox"/>		GSE16879.GPL570.test14	crohn's disease (CD)	infliximab	Disease vs. Normal	DiseaseState => crohn's disease (CD) vs normal control	glm	Affymetrix.HG-U133_Plus_2 GSE16879
<input checked="" type="checkbox"/>		GSE16879.GPL570.test15	crohn's disease (CD)	infliximab	Disease vs. Normal	DiseaseState => crohn's disease (CD) vs normal control	glm	Affymetrix.HG-U133_Plus_2 GSE16879
<input checked="" type="checkbox"/>		GSE16879.GPL570.test3	crohn's disease (CD)	infliximab	Disease vs. Normal	DiseaseState => crohn's disease (CD) vs normal control	glm	Affymetrix.HG-U133_Plus_2 GSE16879
<input checked="" type="checkbox"/>		GSE16879.GPL570.test5	crohn's disease (CD)	infliximab	Disease vs. Normal	DiseaseState => crohn's disease (CD) vs normal control	glm	Affymetrix.HG-U133_Plus_2 GSE16879
<input checked="" type="checkbox"/>		GSE52746.GPL17996.test1	crohn's disease (CD)	NA	Disease vs. Normal	DiseaseState => crohn's disease (CD) vs normal control	glm	Affymetrix.HG-U133_Plus_2 GSE52746
<input checked="" type="checkbox"/>		GSE57945.GPL11154.DESeq2.test1	crohn's disease (CD)	NA	Disease vs. Normal	DiseaseState => crohn's disease (CD) vs normal control	DESeq2.v1.10.1.os.v101316	NGS.Illumina.HiSeq2000 GSE57945
<input checked="" type="checkbox"/>		GSE59071.GPL6244.test2	crohn's disease (CD)	NA	Disease vs. Normal	DiseaseState => crohn's disease (CD) vs normal control	glm	Affymetrix.HuGene-1_0-st-v1 GSE59071
<input checked="" type="checkbox"/>		GSE6731.GPL8300.test1	crohn's disease (CD)	NA	Disease vs. Normal	DiseaseState => crohn's disease (CD) vs normal control	glm	Affymetrix.HG_U95Av2 GSE6731

Showing 1 to 8 of 8 entries

#### Step 4: Name the comparison list

Fill in the “Name” field, then click on “Create Comparison List”. Providing a description is optional.

The screenshot shows a form titled "Create Comparison List". A red box labeled "1. Put name here" points to the "Name" input field, which contains "Crohns demo". Another red box labeled "2. Then click here" points to the "Create Comparison List" button at the bottom left.

Name:	Crohns demo
A unique name of the list	
Comparison IDs:	GSE57945.GPL11154.DESeq2.test1 GSE16879.GPL570.test14 GSE16879.GPL570.test15 GSE16879.GPL570.test3 GSE16879.GPL570.test5 GSE52746.GPL17996.test1 GSE59071.GPL6244.test2 GSE6731.GPL8300.test1
✓ You have entered 8 comparison IDs, all of them are available in the database.	
Description:	
<input type="button" value="Create Comparison List"/>	

## Step 5: Create Pathway Heatmap

Use the menu on the left and click on “Pathway Heatmap Tool” under “Comparison Plotting Tools”.

Then click “Load Saved Comparisons”.

Then select the Comparison List you just created, “Crohns demo”.

Then mark the checkbox “Use comparison info instead of comparison ID” (to provide more information) and click on “Get Gene Sets” button.

The screenshot shows the software's sidebar on the left and the main tool window on the right.

**Left Sidebar:**

- Gene Expression Plots**
  - Single Gene (RNA-Seq)
  - Multiple Genes (RNA-Seq)
  - Single Gene (Microarray)
  - Multiple Genes (Microarray)
  - Heatmap
  - Correlation Tools Using Gene Expression
  - Export Genes and Samples
- Comparison Plotting Tools**
  - Bubble Plot (Single Gene)
  - Bubble Plot (Multiple Genes)
  - Volcano Plot
  - WikiPathways Visualization
  - KEGG Visualization
  - Reactome Visualization
  - Pathway Heatmap Tool
  - Correlation Tools Using Comparisons

**Main Window - Pathway Heatmap Tool:**

- Comparison IDs:** [Load Saved Comparisons](#) (highlighted with a red box and arrow labeled "1. Click here")
- Data Source:**  Omicsoft Data  Internal Data [Select Internal Project: 2](#)
- Set:** PAGE List
- Data Filter:**  Enable z-score lower limit
- Display Option:**  Use comparison info instead of comparison ID
- Display top:** 20 gene sets with highest |z-score|
- Get Gene Sets** button (highlighted with a red box and arrow labeled "2. Then click here")

Please select a comparison list you like to load:

X

3. Select and  
close window

Comparison IDs	Count	Actions
<input type="radio"/> all_comparisons	235	
<input type="radio"/> Crohn's disease versus normal	5	
<input type="radio"/> Crohn8_DvsN	8	
<input type="radio"/> Crohn8_DvsN_and_1_TvsC	9	
<input checked="" type="radio"/> Crohns demo	8	
<input type="radio"/> Disease_vs_Control_D35_D56	2	
<input type="radio"/> GSE52746_Crohns_disease_TNFalpha	3	
<input type="radio"/> GSE62832	4	
<input type="radio"/> psoriasis Test3	3	
<input type="radio"/> res nres	6	

**Close**

Comparison IDs:

GSE57945.GPL11154.DESeq2.test1  
GSE16879.GPL570.test14  
GSE16879.GPL570.test15  
GSE16879.GPL570.test3  
GSE16879.GPL570.test5  
GSE52746.GPL17996.test1  
GSE59071.GPL6244.test2  
GSE6731.GPL8300.test1

Data Source:

Omicsoft Data  Internal Data Select Internal Project: 2

Set:

PAGE List

Data Filter:

Enable z-score lower limit

4. Click this

Display Option:

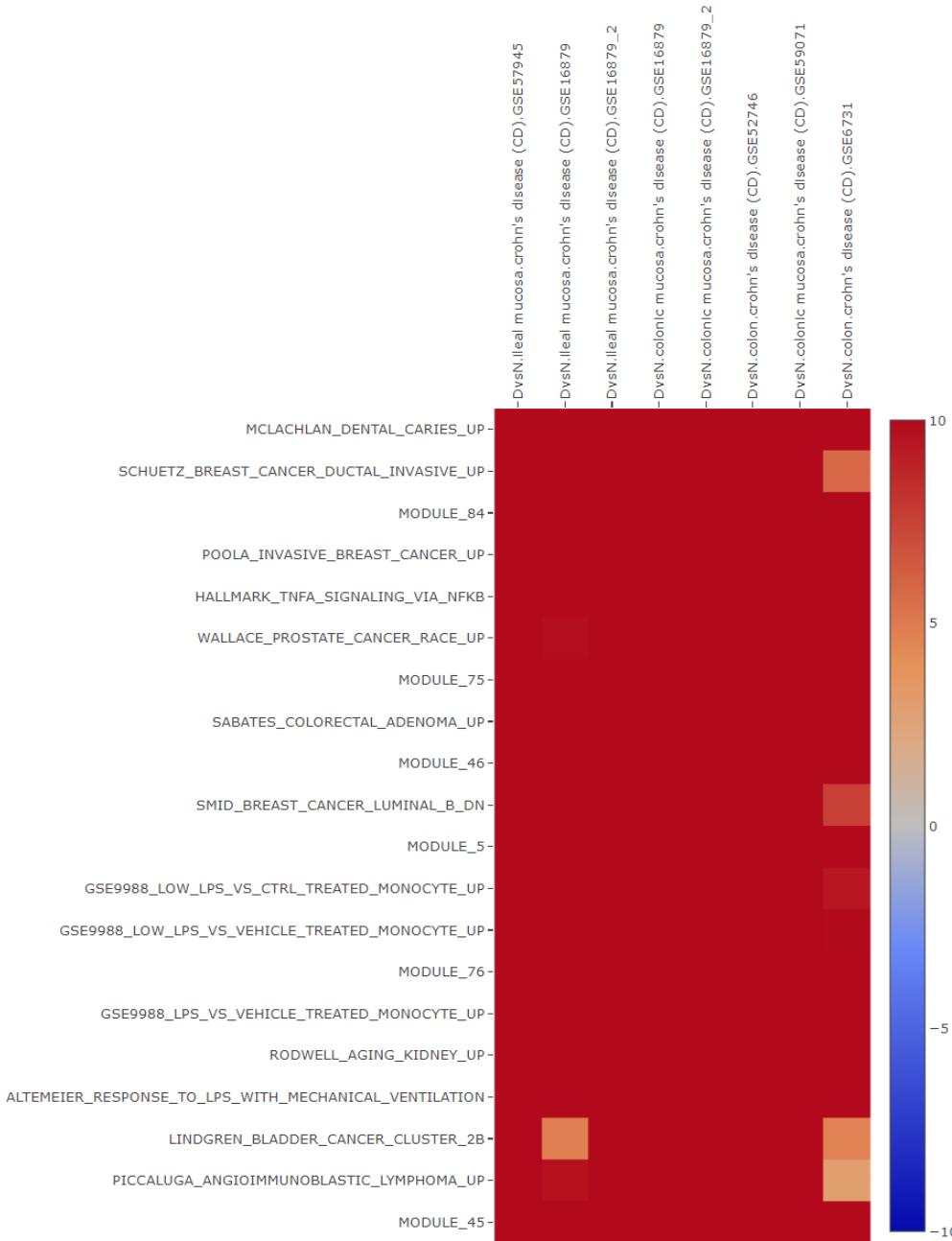
Use comparison info instead of comparison ID

Display top  gene sets with highest |z-score|

Get Gene Sets

5. Then this

Pathway heatmap is then displayed which provides insight into highly activated immune related signature responses including TNF alpha:



## Step 6: Create Pathway Level visualization for TNF alpha pathway

On the left side menu, pick “KEGG Visualization” under “Comparison Plotting Tools”. Alternatively, you could pick WikiPathways or Reactome, and search for TNF alpha pathway).

Then click “Select Pathway”.

Then type TNF in the search box and hit “Select” on the TNF signaling pathway that comes up.

The screenshot shows the OmicsView software interface. On the left, there is a sidebar with two main sections: "Gene Expression Plots" and "Comparison Plotting Tools". Under "Comparison Plotting Tools", the "KEGG Visualization" option is highlighted with a red box and labeled "First". Below it are other options: "Reactome Visualization", "Pathway Heatmap Tool", "Correlation Tools Using Comparisons", and "Export Genes and Comparisons". The main area is titled "KEGG Pathway Visualization" and shows a search bar with "Search" and a dropdown for "Show [10] entries". It also has fields for "Enter comparison names (one per row)" and "Or, upload your comparison files: Choose File". A "Submit" button is at the bottom. To the right, a modal window titled "Select Pathway" is open, listing various KEGG pathways with their codes and names. A red box highlights the search bar in the modal window, with the text "Type in TNF in search" above it. Another red box highlights the "Select" column in the modal table, with the text "Select" next to it. The modal also includes a "Close" button at the bottom right.

Code	Pathway Name	Action
hsa00010	Glycolysis / Gluconeogenesis	>> Select
hsa00020	Citrate cycle (TCA cycle)	>> Select
hsa00030	Pentose phosphate pathway	>> Select
hsa00040	Pentose and glucuronate interconversions	>> Select
hsa00051	Fructose and mannose metabolism	>> Select
hsa00052	Galactose metabolism	>> Select
hsa00053	Ascorbate and aldarate metabolism	>> Select
hsa00061	Fatty acid biosynthesis	>> Select
hsa00062	Fatty acid elongation	>> Select
hsa00071	Fatty acid degradation	>> Select

Next load the saved comparisons by clicking on “Load from saved lists”.

Select “Crohns demo” from My Saved List and Close the window.

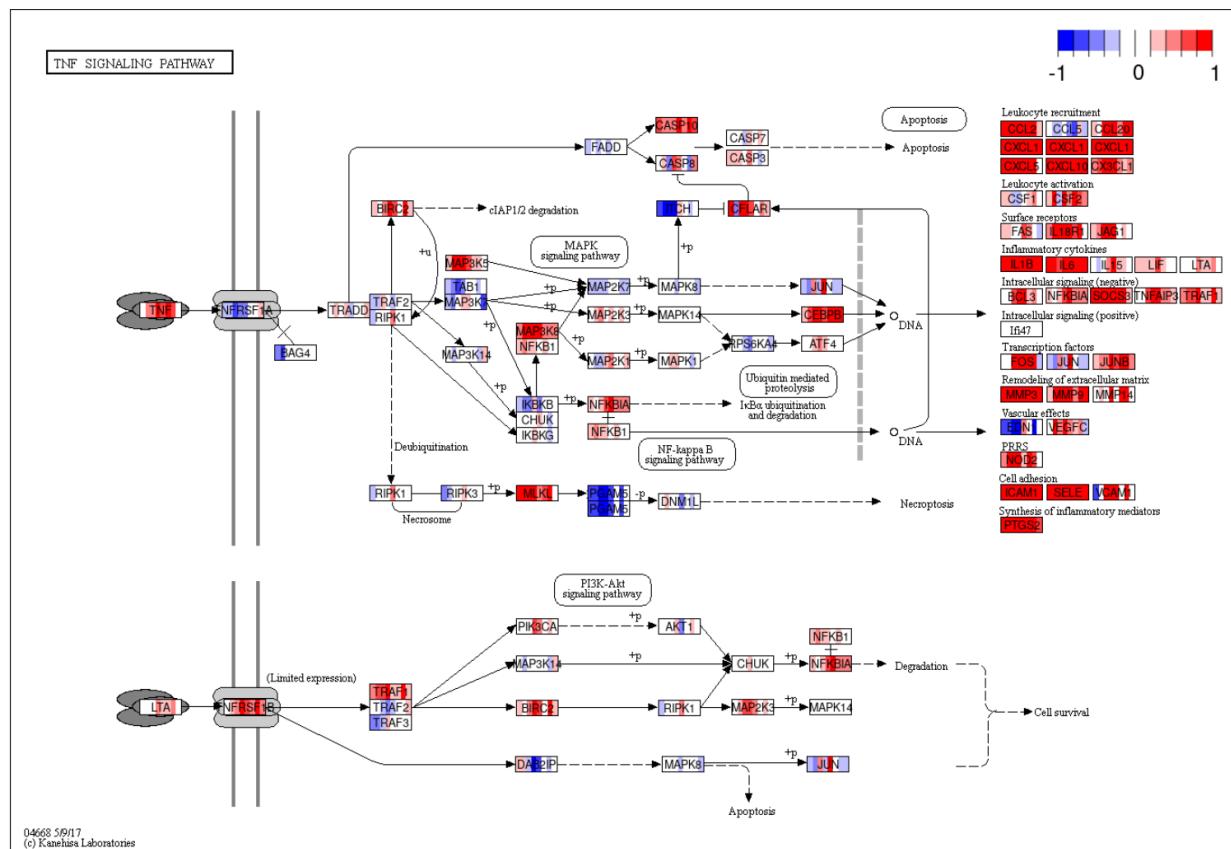
Then click on “Submit”.

**First**

**Second**

**Third**

A detailed map of the TNF signaling pathway is then displayed with coloring by fold change of up or down regulation for each of the 8 disease-versus-control comparisons. It allows the user to home in on key nodes that show consistent dysregulation across datasets.



### Step 7 : Meta-analysis of differentially expressed genes (Comparisons) across studies

Click on “Meta Analysis (Comparisons)” under “Other Tools” in the left-hand navigation panel.

- ≡ Other Tools
- Visualize PCA Results
  - Meta Analysis (Comparisons) Click here
  - Meta Analysis (Gene Expression)
  - Significantly Changed Genes
  - Venn Diagram
  - Functional Gene List Venn Diagram
  - Distinct Values in Table

Click on “Load Saved Comparison IDs” (i.e., saved Comparison List).

## Meta Analysis Using Comparison Data

Gene Names:

Please use more than two genes or leave empty to explore all available genes.

Comparison IDs:  Click here

Please enter two or more comparison IDs, separated by line break.

**Gene Attributes (Selected: 3):**  
 Gene Name  Entrez ID  Description ([More Attributes](#))

**Source of the Sample IDs:**  
 Omicsoft Data  Internal Data [Select Internal Project: 2](#)

[Meta Analysis Options](#)

[Data Display Options](#)

Select the previously created “Crohns demo” list and click on “Select”.

Please select a comparison list you like to load:

	Comparison IDs	Count	Actions
<input type="radio"/>	all_comparisons	235	Review
<input type="radio"/>	Crohn's disease versus normal	5	Review
<input type="radio"/>	Crohn8_DvsN	8	Review
<input type="radio"/>	Crohn8_DvsN_and_1_TvsC	9	Review
<input checked="" type="radio"/>	Crohns demo	8	Review
<input type="radio"/>	Disease_vs_Control_D35_D56	2	Review
<input type="radio"/>	Genes_TNFalpha	3	Review
<input type="radio"/>	Genes	4	Review
<input type="radio"/>	psoriasis Test3	3	Review
<input type="radio"/>	res nres	3	Review

First

Second

Select

Then click on “Submit”.

## Meta Analysis Using Comparison Data

Gene Names: Load Saved Genes Search by Gene Set

Please use more than two genes or leave empty to explore all available genes.

Comparison IDs: Load Saved Comparison IDs

GSE57945.GPL11154.DESeq2.test1  
GSE16879.GPL570.test14  
GSE16879.GPL570.test15  
GSE16879.GPL570.test3  
GSE16879.GPL570.test5  
GSE52746.GPL17996.test1  
GSE59071.GPL6244.test2  
GSE6731.GPL8300.test1

Gene Attributes (Selected: 3):  
 Gene Name  Entrez ID  Description ([More Attributes](#))

Source of the Sample IDs:  
 Omicsoft Data  Internal Data [Select Internal Project: 2](#)

Meta Analysis Options

Data [Click here](#)

Submit Reset

The results table shows a list of genes that have been identified as differentially expressed in any one of the previously selected datasets. The table can be sorted by column headers to get a top set of genes ranked by a variety of metrics. This allows the user to select a subset of genes that have consistent up and down regulation across all conditions/studies.

	Actions	Gene Name	Entrez ID	Description	# of Data Points	Upregulated (%)	Downregulated (%)	Rank Products	Rank Products ( $\log_2$ Fold Change)	Rank Products (p-value)
<input type="checkbox"/>		SLC6A14	11254	solute carrier family 6 member 14	7	100	0	4.01	5.05	0
<input type="checkbox"/>		MMP3	4314	matrix metallopeptidase 3	8	87.5	0	6.18	4.8	0
<input type="checkbox"/>		CXCL8	3576	C-X-C motif chemokine ligand 8	8	75	0	8.23	4.44	0
<input type="checkbox"/>		S100A8	6279	S100 calcium binding protein A8	8	75	0	9.68	4.26	0
<input type="checkbox"/>		MMP1	4312	matrix metallopeptidase 1	8	87.5	0	11.55	4.15	0
<input type="checkbox"/>		DUOX2	50506	dual oxidase 2	7	100	0	13.27	4.01	0
<input type="checkbox"/>		HCAR3	8843	hydroxycarboxylic acid receptor 3	8	50	0	18.17	3.68	0
<input type="checkbox"/>		CXCL1	2919	C-X-C motif chemokine ligand 1	8	87.5	0	19.59	3.62	0
<input type="checkbox"/>		CHI3L1	1116	chitinase 3 like 1	8	75	0	19.96	3.6	0

## Step 8 Create bubble plot from meta-analysis results

Select the top genes from the Meta Analysis table.

Click on “Bubble Plot” above the table.

The screenshot shows the Gene Comparison tool interface. At the top, there are several buttons: "Save to a Study", "Create a Gene List", "Bubble Plot" (which is highlighted with a red box and a red arrow pointing to it), "Export Genes and Comparisons", and "Download". Below these are buttons for "Copy", "CSV", "Excel", and "PDF". A search bar is also present. The main area displays a table of gene information, with the first five rows highlighted in grey. Red boxes numbered 1 through 5 are overlaid on the interface:

- 1. Select: A red box surrounds the first row of the table.
- 2. Select: A red box surrounds the second row of the table.
- 3. Select: A red box surrounds the third row of the table.
- 4. Select: A red box surrounds the fourth row of the table.
- 5. Then click: A red box surrounds the "Bubble Plot" button at the top of the page.

Show 100 entries	Gene Name	Entrez ID	Description	# of Data Points	Upregulated (%)	Downregulated (%)	Rank Products	Rank Products ( $\log_2$ Fold Change)
<input checked="" type="checkbox"/>	SLC6A14	11254	solute carrier family 6 member 14	7	100	0	4.01	5.05
<input checked="" type="checkbox"/>	MMP3	4314	matrix metallopeptidase 3	8	87.5	0	6.18	4.8
<input checked="" type="checkbox"/>	CXCL8	3576	C-X-C motif chemokine ligand 8	8	75	0	8.23	4.44
<input checked="" type="checkbox"/>	S100A8	6279	S100 calcium binding protein A8	8	75	0	9.68	4.26

Load the previously saved comparisons by selecting “Load Saved Comparison IDs” and selecting “Crohns demo”. Then click on “Plot” on the bottom of the new page.

# Genes & Comparisons Bubble Plot

[Load Example Data](#)

Gene Names: [Load Saved Genes](#) [Search by Gene Set](#)

S100A8  
MMP3  
CXCL8  
SLC6A14

Comparison IDs: [Load Saved Comparison IDs](#)

GSE57945.GPL11154.DESeq2.test1  
GSE16879.GPL570.test14  
GSE16879.GPL570.test15  
GSE16879.GPL570.test3  
GSE16879.GPL570.test5  
GSE52746.GPL17996.test1  
GSE59071.GPL6244.test2

Source of the Comparison IDs:

Omicsoft Data  Internal Data

Chart Height Scale Factor:

1

Chart Left Margin Scale Factor:

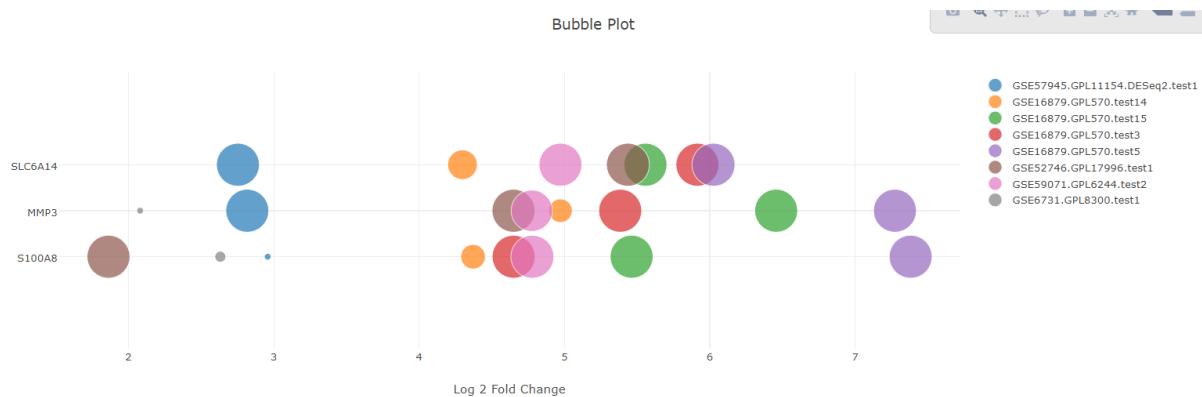
1

Display Colu  
 Log<sub>2</sub> Fold Change  p-value  FDR

[Plot](#)

Click

The bubble plot gives an intuitive view of the extent of change of each gene (log<sub>2</sub> Fold Change on x axis) and the significance (size of the bubble) across the studies. It also allows user to see if any of the studies/comparisons are outliers that display a different pattern from the others with respect to gene expression changes.



## 8 Methods

### 8.1 Data processing

**Public datasets.** The public datasets that are accessible through OmicsView are a subset of the DiseaseLand <https://www.qiagenbioinformatics.com/diseaseland/> database. The DiseaseLand data service uses common analysis pipelines to quantify and normalize publicly available microarray and RNA-seq expression data from raw data. For each project, and each sample, metadata are curated to apply controlled vocabularies and ensure consistent formatting of terms. Information on the data processing pipeline and the DiseaseLand product is available at the following links ([http://www.arrayserver.com/wiki/index.php?title=DiseaseLand\\_Curation\\_Pipeline](http://www.arrayserver.com/wiki/index.php?title=DiseaseLand_Curation_Pipeline), [http://www.arrayserver.com/wiki/index.php?title=Omicsoft\\_Affymetrix\\_Microarray\\_Preprocessing](http://www.arrayserver.com/wiki/index.php?title=Omicsoft_Affymetrix_Microarray_Preprocessing), [http://www.arrayserver.com/wiki/index.php?title=RNA-Seq\\_Normalized\\_FPKM\\_Values\\_in\\_Land](http://www.arrayserver.com/wiki/index.php?title=RNA-Seq_Normalized_FPKM_Values_in_Land)). The data in OmicsView was exported from DiseaseLand prior to 8/27/2019. No OmicsView data can be exported from the portal, distributed with a software package, or otherwise redistributed without express permission from QIAGEN.

### 8.2 Functional and Pathway Enrichment

The function enrichment algorithms implemented in OmicsView come in two flavors.

The first is based on hypergeometric enrichment of differentially expressed genes (DEGs) against several knowledge databases, based on the HOMER package (<http://homer.salk.edu/homer/microarray/go.html>). This requires a significance cutoff to be applied to the two group comparison expression data such that we can select a set of genes that are upregulated and downregulated, i.e., differentially expressed genes. The HOMER package comes with some collections of gene knowledge databases that we use for enrichment.

The second approach is based on a variant of the Gene Set Enrichment Analysis (GSEA) approach called PAGE (Parametric Analysis of Gene Set Enrichment) which is faster than the original GSEA approach and more sensitive [1]. We use an implementation of PAGE that is available in the R piano package (<https://www.bioconductor.org/packages/release/bioc/html/piano.html>). GSEA based approaches just require a ranked list of genes, i.e., genes with a numeric quantity for ranking them that shows how different they are between groups in the comparison, usually log Fold Change (logFC) values. HOMER and PAGE are typically more sensitive than traditional GSEA in terms of revealing pathway enrichment. Both packages come pre-built with a set of known functions/pathways in the form of gene sets, against which we look for enrichment in the comparison of interest.

#### 8.2.1 HOMER workflow

##### 8.2.1.1 Selection of Differentially Expressed Genes (DEGs) for enrichment analysis

For each comparison, we generated the up- and down-regulated gene lists as input for functional enrichment analysis. We used a dynamic cutoff of logFC and AdjustedPValue / PValue to aim for 200-2000 genes in each list. To achieve this, we start with a stringent cutoff of adjusted p-value of 0.05, and 2-

fold change up or down. Getting a list of greater than 200 genes is ideal for the downstream analysis. If there are fewer, we drop the nominal p-value to 0.01 (with 2-fold up or down), and subsequently if 200 is not reached we drop the fold change to 1.2 in either direction or increase the adjusted p-value to 0.1. If the most relaxing cutoff doesn't generate a list of at least 50 dysregulated genes, we select the top 50 genes ranked by absolute logFC values to generate the lists for enrichment analysis.

#### *8.2.1.2 Functional Enrichment of the DEG list against the genome*

The findGo.pl program from Homer package ( <http://homer.salk.edu/homer/microarray/go.html> ) is used to analyze the functional enrichment for each DEG list.

There are several different "ontologies", or libraries of gene groupings that came with the Homer package. We used Homer version v4.8.3, human-o v5.8 and mouse-o v5.8 libraries (current as of Dec 6, 2016). These include the following categories:

Gene Ontology: Biological Process, Molecular Function, Cellular Component

Chromosome Location: Genes with similar chromosome localization (NCBI Entrez Gene)

KEGG Pathways: Groups of proteins in the same pathways (From KEGG)

Protein-Protein Interactions: Groups of proteins interacting with the same protein (From NCBI Entrez Gene)

Interpro: Proteins with similar domains and features (Interpro)

Pfam: Proteins with similar domains and features (Pfam)

SMART: Proteins with similar domains and features (SMART)

Gene3D: Proteins with similar domains and features (Gene3D Database)

Prosite: Proteins with similar domains and features (Prosite Database)

PRINTS: Proteins with similar domains and features (PRINTS Database)

MSigDB: Lists of genes maintained by the Molecular Signature Database (includes many different categories of genes (MSigDB)

BIOCYC: Groups of proteins in the same pathway (NCBI Biosystems/BIOCYC)

COSMIC: Human proteins that are mutated in the same cancers (COSMIC)

GWAS Catalog: Human genes with risk SNPs identified in their vicinity for the same disease (GWAS Catalog)

Lipid Maps: Mouse proteins found in the same lipid processing pathways (NCBI Biosystems/LIPID MAPS)

Pathway Interaction Database: Proteins in the same pathway (NCBI Biosystems/PID)

REACTOME: Proteins in the same biochemical pathways (NCBI Biosystems/REACTOME)

SMPDB: Proteins in the same pathway (SMPDB)

WikiPathways: Protein in the same pathway (WikiPathways)

The HTML output file of HOMER was further enhanced by a custom php script to make it more user friendly.

### 8.2.2 Gene Set Enrichment Analysis (GSEA) workflow (PAGE)

For our second enrichment tool, we used a variation of the GSEA method, PAGE (Parametric Analysis of Gene Set Enrichment) [1] to process the comparison data as PAGE is much faster and more sensitive than GSEA.

For each comparison, we produced a rank file with gene symbols and corresponding logFC values. If a gene symbol appeared multiple times in the same comparison, the average logFC was used. Human gene sets were downloaded from MsigDB (<http://software.broadinstitute.org/gsea/msigdb>) version 5.2 (msigdb.v5.2.symbols.gmt).

Mouse gene sets were downloaded from the Bader Lab website from Univ. of Toronto (<http://baderlab.org/GeneSets>). This version is labeled as December\_01\_2016 (Mouse\_GO\_AllPathways\_with\_GO\_iea\_December\_01\_2016\_symbol.gmt). This gmt file contains special characters that cannot be used by R piano package, therefore we manually replaced special characters to “/” or “\_”. In addition, some mouse gene sets have the same name, so we added suffix “\_altSetX” to make all the names unique.

#### 8.2.2.1 PAGE Analysis using the R piano package

The R package piano (<https://bioconductor.org/packages/release/bioc/html/piano.html>) was used to run the PAGE analysis for all the rank files. To simplify the piano output, we combined down-regulated and up-regulated gene sets into a single table for each gene set. From the piano results, we extracted out the p-value, FDR, and Z-score.

## 8.3 Meta-Analysis

The meta-analysis functions allow a user to combine expression data across multiple studies to find changes that are robust. DiseaseAtlas offers two ways to perform meta-analysis:

The first method works on comparison data. The system uses the comparison data (logFC, p-value) to compute combined p-value and rank product. This method is fast and can be applied to any type of comparison data. However, it does not use the individual sample data, nor does it consider the number of samples in each comparison.

The second method uses per sample Gene Expression data. The user is expected to input a list of factors that indicate comparisons across or within studies, and then gene level significant changes are recomputed by extracting expression data from all samples for each comparison. The RankProd [2] and/or MetaDE [3] packages are then applied to perform the meta-analysis. Limma is applied to get the statistics for each individual comparison. This analysis takes much longer (10 minutes to an hour for a typical analysis, even longer if number of samples are very large), and it has more strict sample requirements (e.g., no samples can occur in two different comparisons).

Statistically, the second method is more robust. However, both methods should detect consistently changed genes.

### 8.3.1 Meta-analysis statistics

For comparison data, the combined p-value is computed using Fisher's method, that is  $-2*(\text{sum of } \ln(p\text{-value}))$  is compared against a Chi-squared distribution with N degrees of freedom, where N is the number of p-values being combined. This is carried out for every gene and yields a combined p-value that is reported. Another simpler approach is to report the maximum p-value for the gene across all the comparisons – a much more stringent measure of overall significance. This approach to combining p-values is implemented in the MetaDE R package. Note MetaDE will not produce results if more than 30% of the comparisons for a gene have missing p-values. In addition, with this approach, the p-value combination does not account for the direction of the fold change (up or down), so the up regulated and down regulated percentage summaries need to be referred to for interpretation.

The RankProd method (<https://bioconductor.org/packages/release/bioc/html/RankProd.html>) converts log<sub>2</sub> fold changes across all genes in a comparison to ranks and then computes a meta-statistic per gene which is the geometric mean of the ranks across comparisons. It is a non-parametric approach, and computes statistical significance based on a permutation approach which also addresses the multiple testing aspect of looking for significance within the set of all genes in the transcriptome.

## References

1. Kim, S.Y., & Volsky, D.J. (2005). PAGE: parametric analysis of gene set enrichment. *BMC bioinformatics*, **6**, 144 (2005).
2. Del Carratore, D et al. RankProd 2.0: a refactored Bioconductor package for detecting differentially expressed features in molecular profiling datasets. *Bioinformatics* **33 (17)**, 2774-2775 (2017).
3. Ma, T. et al. Metaomics: analysis pipeline and browser-based software suite for transcriptomic meta-analysis. *Bioinformatics* **35 (9)**, 1597-1599 (2019).

## 9 Supplementary Tables

### 9.1 Table S1. Feature comparison of tools

Raw file: <https://github.com/interactivereport/OmicsView/blob/gh-pages/TableS1.xlsx>

	Tools	OmicsView	ARCHS4	Bges	BIOMEX	DEE2	Gemma	GEPA	GXD	OASIS	PairOomics 3	PulmonDB	recount3
Year	2022	2018	2020	2020	2018	2021	2017	2021	2016	2018	2020	2021	
<b>Main function</b>													
Data storage	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Data visualization	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
<b>Basic information</b>													
RNA-Seq Species	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Human, Mouse, Rat	Human, Mouse	Human, Mouse	29 Species	6 Species	9 Species	7 Species	Human	Mouse	Human	73 Species	Human	Human, Mouse	
Other datatypes	Microarray, Proteomics	Microarray, EST, in situ hybridization		Proteomic, mass cytometry, single cell RNA-Seq and mass cytometry)		Microarray	Microarray, RNA in situ hybridization, immunohistochemistry, RT-PCR, northern blot, and western blot	Microarray, genomics	Epigentics, proteomics, metabolomics	Microarray, genomics	Single cell RNA-Seq, small/Micro RNA-Seq, etc.		
<b>Key features</b>				A platform with harmonized datasets across many disease areas and equipped with user-friendly data visualization, annotation, and biological interpretation tool.	An RNA-Seq database for gene and transcript expression data and transcriptome analysis, such as species gene expression profile enrichment analysis.	A curated database of four expression RNA-seq types for cross-species gene expression profile enrichment analysis.	A curated RNA-seq repository for multiomics data analysis, such as metabolomics, proteomics, and single cell RNA-seq (all RNA-Seq and mass cytometry).	A manually curated database for transcriptomic datasets with gene and transcript interference quantification.	A web-based tool for RNA-Seq data mining using data from TCGA and GTEx, focusing on cancer and normal tissue expression profiling.	A mouse developmental gene expression database, curating data from RNA-Seq, microarray, RNA in situ hybridization, and RT-PCR, northern blot, and western blot.	A cancer and normal tissue database of microarray and RNA-Seq data, providing gene expression analysis and visualization.	A web-based, integrated multiomics data analysis platform for two primary diseases, cancer and non-cancer, using CORD and iPf.	A large collection of expression data by standardized processing pipeline, and a package for expression query.
<b>Online/Standalone</b>	Both	Online	Online	Standalone	Online	Online	Online	Online	Online	Online	Online	Online	Online
<b>Website</b>	<a href="https://omicsview.org/">https://omicsview.org/</a>	<a href="https://ab.cloudarchitects.com/#/">https://ab.cloudarchitects.com/#/</a>	<a href="https://bgee.org/">https://bgee.org/</a>	<a href="https://camillelab.sil.es/wb.be/en/biomec/">https://camillelab.sil.es/wb.be/en/biomec/</a>	<a href="https://dee2.io/">https://dee2.io/</a>	<a href="https://gemma.ms/">https://gemma.ms/</a>	<a href="https://genie.ca/home.html">https://genie.ca/home.html</a>	<a href="https://genie.ca/help.html">https://genie.ca/help.html</a>	<a href="http://www.informatics.jax.org/GXDbeta/">http://www.informatics.jax.org/GXDbeta/</a>	<a href="http://www.informatics.jax.org/expression.html">http://www.informatics.jax.org/expression.html</a>	<a href="http://www.ncbi.nlm.nih.gov/gene/">http://www.ncbi.nlm.nih.gov/gene/</a>	<a href="http://pulmo.ncbi.nlm.nih.gov/">http://pulmo.ncbi.nlm.nih.gov/</a>	<a href="http://pulmo.ncbi.nlm.nih.gov/recount3/">http://pulmo.ncbi.nlm.nih.gov/recount3/</a>
<b>Source Code Link</b>	<a href="https://github.com/inteligensoft/Omics_recound3">https://github.com/inteligensoft/Omics_recound3</a>	<a href="https://github.com/mMauroLabb/rob4">https://github.com/mMauroLabb/rob4</a>	<a href="https://github.com/mBeepeDB">https://github.com/mBeepeDB</a>	<a href="https://github.com/mMarkoKommuni/dee2">https://github.com/mMarkoKommuni/dee2</a>	<a href="https://github.com/mPavdilab/CeRNA">https://github.com/mPavdilab/CeRNA</a>	<a href="https://github.com/mPavdilab/GXDbeta">https://github.com/mPavdilab/GXDbeta</a>	<a href="https://github.com/nlbcu/home/GXDbeta">https://github.com/nlbcu/home/GXDbeta</a>	<a href="https://github.com/nlbcu/help.html">https://github.com/nlbcu/help.html</a>	<a href="https://github.com/nlbcu/home/GXDbeta/FirstTimeUsers.shtml">https://github.com/nlbcu/home/GXDbeta/FirstTimeUsers.shtml</a>	<a href="http://www.informatics.jax.org/expression.html">http://www.informatics.jax.org/expression.html</a>	<a href="https://pulmo.ncbi.nlm.nih.gov/recount3/">https://pulmo.ncbi.nlm.nih.gov/recount3/</a>	<a href="https://pulmo.ncbi.nlm.nih.gov/recount3/">https://pulmo.ncbi.nlm.nih.gov/recount3/</a>	
<b>Tutorial</b>	<a href="https://interactivegenomics.github.io/OmicsView/tutorial/doc/">https://interactivegenomics.github.io/OmicsView/tutorial/doc/</a>	<a href="https://mavani.ab.cloudarchitects.com/#/help.html">https://mavani.ab.cloudarchitects.com/#/help.html</a>	<a href="https://oebe.org/?page_id=PL4SbInnp7YYzK8tV9_WKmTS02wps7">https://oebe.org/?page_id=PL4SbInnp7YYzK8tV9_WKmTS02wps7</a>	<a href="https://www.youtube.com/watch?v=omnivida2list-PL4SbInnp7YYzK8tV9_WKmTS02wps7">https://www.youtube.com/watch?v=omnivida2list-PL4SbInnp7YYzK8tV9_WKmTS02wps7</a>	<a href="https://dee2.io/help.html">https://dee2.io/help.html</a>	<a href="https://pavdilab.github.io/Gemma/dee2.html">https://pavdilab.github.io/Gemma/dee2.html</a>	<a href="https://genie.ca/help.html">https://genie.ca/help.html</a>	<a href="http://www.ncbi.nlm.nih.gov/GXDbeta/help.html">http://www.ncbi.nlm.nih.gov/GXDbeta/help.html</a>	<a href="http://www.ncbi.nlm.nih.gov/GXDbeta/help.html">http://www.ncbi.nlm.nih.gov/GXDbeta/help.html</a>	<a href="http://www.ncbi.nlm.nih.gov/GXDbeta/help.html">http://www.ncbi.nlm.nih.gov/GXDbeta/help.html</a>	<a href="https://pulmo.ncbi.nlm.nih.gov/recount3/">https://pulmo.ncbi.nlm.nih.gov/recount3/</a>	<a href="https://pulmo.ncbi.nlm.nih.gov/recount3/">https://pulmo.ncbi.nlm.nih.gov/recount3/</a>	
<b>Application Programming Interface (API)</b>	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
<b>Licensing</b>	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free	Free
<b>Analysis software</b>													
<b>Analysis Power Tools (AP)</b>													
Bioconductor	Y	Y	Y	Y			Y	Y		Y	Y	Y	Y
BLAST				Y									
Custadapt							Y						
Enrichr		Y											
FastQC		Y				Y							
featureCounts													
GISTIC2.0													
gcRMA				Y									
HOMER		Y											
IQRay		Y											
Kallisto		Y	Y				Y						
Kraken							Y						
libBioHGS													Y
Megadepth													Y
MetaDE	Y												
Omicsoft OS Aligner	Y												
PAGE	Y												
recount2													
RSEM	Y		Y										
RMA	Using an in-house method similar to RMA												
<b>SCORPIUS</b>					Y								
seqtk													
SIFT													
Skewer						Y							
STAR	Y					Y							
<b>UCSC Xena pipeline</b>													
<b>EBI Expression Atlas iRAP pipeline</b>													
<b>Data source (full or partial)</b>													
ArrayExpress	Y		Y				Y		Y				
dbGAP	Y		Y										
CGPS													
EMBL-EBI							Y		Y				
GTEX	Y		Y					Y	Y				
GEO	Y	Y	Y	Y			Y	Y	Y		Y	Y	
SRA	Y	Y	Y	Y			Y	Y	Y		Y	Y	
TCGA								Y					
QIAGEN DiseaseLand	Y												
<b>Disease area</b>													
<b>Not disease area specific</b>	Y	Y					Y	Y	Y	Y			
<b>Brain-related datasets</b>	Y	Y	Y				Y	Y	Y	Y			
<b>Cancer-related datasets</b>	Y	Y	Y				Y	Y	Y	Major focus			
<b>Healthy weight datasets</b>	Y	Y	Y	Y			Y	Y	Y	Y			
<b>Lung disease-related datasets</b>	Y	Y	Y				Y	Y	Y	Y	Major focus	Y	
<b>Visualization tool</b>													
Bubble plot	Y												
Chromosomal distribution plot													
Dendrogram (hierarchical clustering)	Y						Y						
Expression boxplot	Y	Y					Y		Y				
Expression heatmap	Y						Y	Y	Y				
Expression image									Y				
Pathway enrichment heatmap	Y												
Pathway interaction network													
Sample correlation heatmap							Y						
Pan-cancer heatmap													
Forest plot	Y												
Gene set enrichment plot	Y												
KEGG pathway plot	Y												
WikiPathways plot	Y												
<b>Multi-datasets pathway mapping</b>	Y												
PCA plot	Y	Y					Y	Y	Y				
Survival analysis plot							Y						
Sankey diagram													
Venn diagram	Y												
Volcano plot	Y												
Violin plot	Y												
<b>Table summary</b>													
Anatomical information										Y			
Anatomical enrichment result													
BigWig coverage file													
Drugability score													
Gene expression	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Differential expression	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Cross-species expression comparison	Y												
Co-expression analysis	Y	Y					Y	Y					
GeneOntology Gene set enrichment test	Y	Y	Y				Y						
KEGG pathway	Y	Y											
<b>Phenotype or predicted phenotype</b>													
Predicted upstream transcription factors (CfSEA)													
Predicted kinase interaction													
Mutation, copy number variation													
QC metric													
<b>Database management</b>													
Project/Sample/Comparison	Y	Y					Y	Y	Y	Y	Y	Y	
User							Y						
Session	Y						Y						
Result sharing	Y												

## 9.2 Table S2. Mapping between use cases and OmicsView menu items

Raw file: <https://github.com/interactivereport/OmicsView/blob/gh-pages/TableS2.xlsx>

Use case	OmicsView menu items
Sample clustering/distribution and outlier detection	Visualize PCA Results, Heatmap
Make/review DEG lists	Significantly Changed Genes, Volcano Plot
Inspect/evaluate genes across samples	Single Gene, Multiple Gene, Heatmap
Pathway analysis (single DEG list)	Review Comparisons. Search and click on individual comparisons for pre-computed pathway enrichment results, WikiPathways Visualization, KEGG Visualization
Meta-analysis of multiple DEG lists, integration of DEGs across studies	Pathway heatmap Tool, Meta Analysis (Comparisons), Bubble Plot (Comparisons), Comparisons Venn Diagram
DEG profile similarity searching	Similar Comparisons (GO), Similar Comparisons (PAGE)
Gene-gene similarity across samples, comparisons	Correlation Tools Using Gene Expression, Correlation Tools Using Comparisons