

# SECIM: Galaxy Interface

## User Guide

Southeast Center for Integrated Metabolomics  
University of Florida  
Gainesville, Florida

Updated: 7 July 2015

# Table of Contents

|  |    |
|--|----|
| <a href="#">Create UF Research Computing Account</a>                   | 3  |
| <a href="#">Login to UF Research Computing Account (Access Galaxy)</a> | 4  |
| <a href="#">Access Data on Galaxy</a>                                  | 4  |
| <a href="#">Add Data to Galaxy History</a>                             | 4  |
| <a href="#">Share Data with Others in Galaxy</a>                       | 5  |
| <a href="#">Galaxy Tools Intro</a>                                     | 7  |
| <a href="#">Data Format for SECIM Galaxy Tools</a>                     | 7  |
| <a href="#">Convert Data to TAB Delimited Format</a>                   | 7  |
| <a href="#">Hierarchical Cluster Heatmaps</a>                          | 8  |
| <a href="#">One-Way ANOVA</a>  | 9  |
| <a href="#">Principal Component Analysis</a>                           | 11 |
| <a href="#">Distribution Analysis</a>                                  | 12 |
| <a href="#">Log-Transform Data</a>                                     | 13 |
| <a href="#">Mean Standardization of Data</a>                           | 15 |
| <a href="#">Bland Altman Plots</a>                                     | 16 |
| <a href="#">Count Digits</a>   | 18 |
| <a href="#">Random Forest</a>  | 19 |

## **Create UF Research Computing Account** – Create account needed to access UF Galaxy

Request an account by navigating to <http://www.rc.ufl.edu/help/account-request/>

Enter all information and submit the form. Sponsor Name **MUST** be SECIM and Sponsor Email **MUST** be [amitch@ufl.edu](mailto:amitch@ufl.edu). Be sure to click the box agreeing to the Acceptable Use Policy at the bottom of the page before submitting the form. It may take up to 24 hours for your account to be approved.

|  |   |
|--|---|
| <b>Name *</b>  |   |
| <input type="text" value="YOUR FIRST NAME"/>   | <input type="text" value="YOUR LAST NAME"/> |
| <small>First</small>   | <small>Last</small>                         |
| <b>Organization *</b>  |   |
| <input type="text" value="YOUR UNIVERSITY / ORGANIZATION"/>  |   |
| <b>Email *</b>   |   |
| <input type="text" value="YOUR EMAIL"/>  |   |
| <b>Gatorlink Username *</b>  |   |
| <input type="text" value="PREFERRED ACCOUNT USERNAME"/>  |   |
| <b>Sponsor Name *</b>  |   |
| <input type="text" value="SECIM"/>   | <input type="text" value="SECIM"/>          |
| <small>First</small>   | <small>Last</small>                         |
| <b>Sponsor Email *</b>   |   |
| <input type="text" value="amitch@ufl.edu"/>  |   |
| <b>Type of Research *</b>  |   |
| <input type="text" value="YOUR RESEARCH TYPE"/>  |   |
| <b>Comments</b>  |   |
| <input type="text" value="SECIM client"/>  |   |
| <b>Acceptable Use Policy *</b>   |   |
| <input checked="" type="checkbox"/> I agree to comply with the acceptable use policy   |   |
| <b>ACCEPTABLE USE</b>  |   |
| I acknowledge that the access to the HPC resources operated by UF Research Computing is subject to the UF Acceptable Use Policies and the Research Computing policies at <a href="http://www.hpc.ufl.edu/about/policies">http://www.hpc.ufl.edu/about/policies</a> and that I am responsible for following these policies.   |   |
| <b>RESTRICTED DATA</b>   |   |
| I also certify that using restricted data and software on the HPC resources requires extra steps described at <a href="#">UFRC Policies</a> and at <a href="#">UFRC Export Policies</a> , and that I will notify both my account sponsor and the Office of Research (Research Compliance) and Research Computing at <a href="mailto:help@hpc.ufl.edu">help@hpc.ufl.edu</a> when I am working with such data. |   |
| <input type="button" value="Submit"/>  |   |

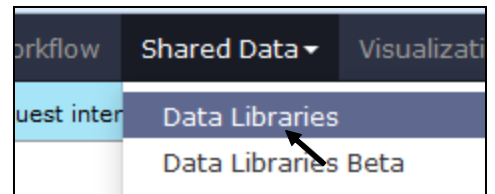
## Log in to UF Research Computing to Access Galaxy

You will receive an email with your username and password for your UF Research Computing account once it has been created. Once you receive this email you can login to Galaxy by navigating to [galaxy.rc.ufl.edu](http://galaxy.rc.ufl.edu) and entering your information



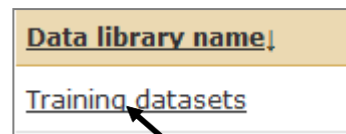
## Accessing Data on Galaxy

Once you have logged on to the UF Research Computing Galaxy website click on the Shared Data icon in the middle at the top of the page and then click on Data Libraries.

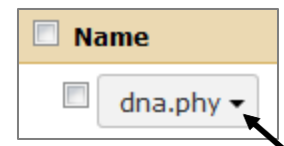


On the Data Libraries page you should see a data library with your study name and study description. Files related to your project will be in the data library. Depending on the number of samples and datasets in your study you may have one or multiple libraries on this page.

To access files in your library click on your study library name (Training datasets in this example).



You should now see any files related to your study. These files can be viewed, downloaded to your computer, or imported into a selected Galaxy history by clicking on the arrow next to the file name and selecting the desired option.

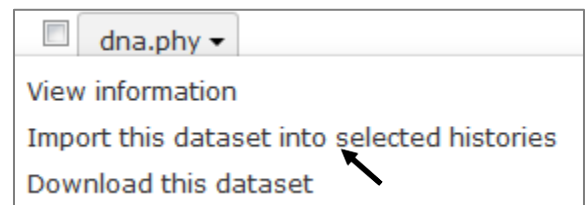


## Adding Data to Your History in Galaxy

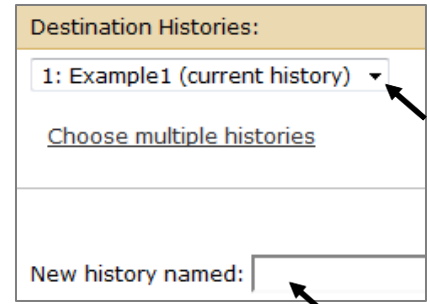
Your history is located on the right side of the Galaxy interface. The history shows the files, tools, and processes that are taking place with the data.

To add data to your history you should first access your data using the methods described in the previous section.

To add data to your history click on the arrow next to the file name and select *Import this dataset into selected histories*.

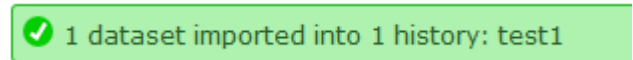


You will then have the option of adding the data to an existing history or creating a new history. To add the data to an existing history select the appropriate history from the dropdown menu. To create a new history type a name into the New History Named box.

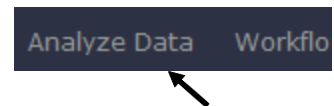


Click on Import library datasets and your data will be added to your history.

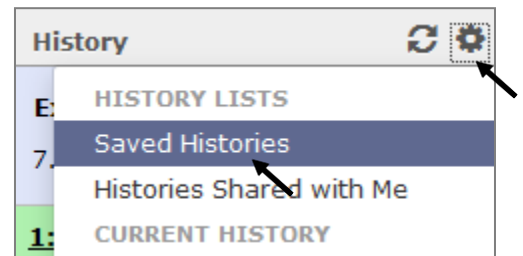
Once data are successfully added to your history you will see a green notification at the top of your screen notifying you that the data were successfully added.



To see your history with the newly added data click on the Analyze Data button in the middle at the top of the screen. Your dataset should be displayed in your history on the right side of the screen.



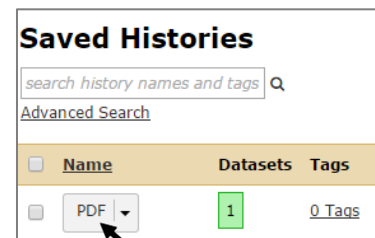
If you cannot see your data click on history options (gear icon) by histories on the right side of the screen.



Click on Saved Histories and a list of your saved histories will be displayed.

Click on the name of the history you previously created (PDF in this example) with the dataset you want to work with.

Your history and dataset should now be displayed on the right side of the screen and you can now click the Analyze Data button at the top of the screen to begin working with your data.

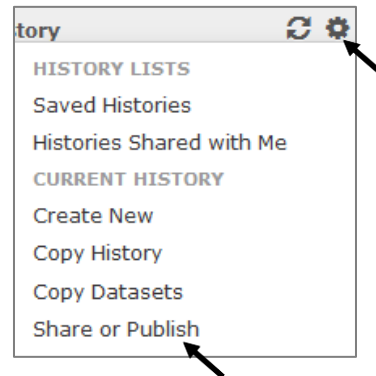


### **Sharing Data with others in Galaxy**

Each person you intend to share data with using Galaxy will need to have a UF Research Computing account and they must have logged into Galaxy at least once before they can receive shared files.

To share data you should first add data to your history using the steps described in the previous section. Once data are in your history you can share you history by clicking on the settings (gear) icon and then clicking on Share or Publish.

On the Share or Publish History page you will have the option to make a web link that you can share with people, make the history accessible and publish to Galaxy's Published Histories where it is then publicly available, or share with specified users. The most common method is to share with specified users.



To share with specified users click on the share with a user button and enter the users email address that is associated with their Galaxy account and click submit. **Each person you intend to share data with using Galaxy will need to have a UF Research Computing account and they must have logged into Galaxy at least once before they can receive shared files.**

## SECIM Galaxy Tools

There are many tools that may be useful for analyzing your data in Galaxy. In addition to the many tools available, SECIM has created tools specifically for analyzing data and preparing data for analyses. Step by step guides for SECIM tools are presented individually below. Although they are combined here, the guides are written as stand-alone documents so that users using one tool do not have to reference other sections to completely understand the tool. As such, some redundancy is common and expected among descriptions.

## SECIM Tools Data Format

You will need two data files to use most SECIM Tools on Galaxy. Data **MUST** be TAB delimited. You can convert datasets using Edit Datasets > Convert Characters.

### Wide Format Dataset

A wide formatted dataset will contain measurements for each sample. Each data variable (e.g. compound, sample1, sample2) will be stored in a separate column.

| Compound  | sample1 | sample2 | sample3 | ... |
|-----------|---------|---------|---------|-----|
| compound1 | 10      | 20      | 10      | ... |
| compound2 | 5       | 22      | 30      | ... |
| compound3 | 30      | 27      | 2       | ... |
| compound4 | 32      | 17      | 8       | ... |
| ...       | ...     | ...     | ...     | ... |

### Design Dataset

The design dataset is used to relate samples to various groups or treatments. Columns will contain information for each of your samples (sampleID) and link them to groups of interest (e.g. treatments, male/female).

| sampleID | group1 | group2 |
|----------|--------|--------|
| sample1  | g1     | t1     |
| sample2  | g1     | t1     |
| sample3  | g1     | t1     |
| sample4  | g2     | t1     |
| sample5  | g2     | t1     |
| sample6  | g2     | t1     |
| ...      | ...    | ...    |

**NOTE:** *The design file **MUST** have a column named “sampleID” and the values of the column **MUST** match the column names of the wide formatted dataset*

## Convert Data to TAB delimited file (if necessary)

If your data are not TAB delimited they will not appear in the dropdown menus for the SECIM Tools. Files can be converted in Galaxy by selecting the *TEXT MANIPULATION* menu on the left side of the screen.

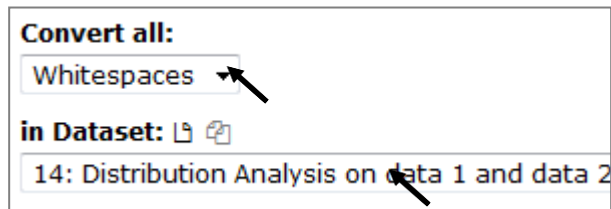
**Text Manipulation**

Select the *Convert* tool which will convert delimiters to TAB

**Convert delimiters to TAB**

Select the delimiter to convert (e.g. whitespaces, columns, pipes) and the dataset you wish to convert

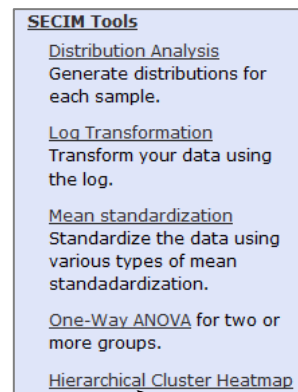
Press Execute and a new TAB delimited dataset will be created



**Hierarchical Cluster Heatmap** – This tool generates a Hierarchical Cluster Heatmap from a wide-dataset and a design file and an optional Annotation file.

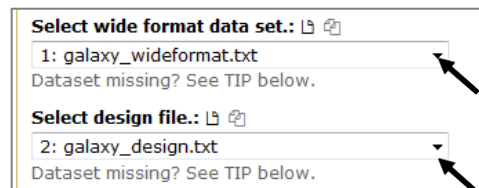
Click on the *Hierarchical Cluster Heatmap* tool in the *SECIM Tools* menu on the left side of the screen.

*You will need a wide dataset and a design dataset to create Hierarchical Cluster Heatmaps in Galaxy. See above for details on creating necessary files.*



Select your wide format data set from the drop down menu

*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the *Edit Datasets > Convert characters* tool to convert your data to a TAB delimited file. See above for more details.*



Select your design file in the second drop down menu

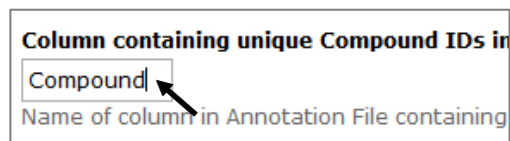
(Optional) Enter any groups or treatments if you want to create a column Color Bar for treatment groups

(Optional) Select the Annotation File if you wish to relate groups to pathways

*NOTE: Compound names must match compound names in the Wide Format Dataset*

**The next steps are only completed if an annotation file was selected. If not, skip below to Execute.**

Type in the column name containing unique compound names (or sample IDs). (column name is Compound in this example)





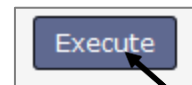
Type in the column name containing unique annotations for each compound (column name is Groups in this example). This is used to generate a color bar.



**Column containing annotations to color**  
Groups  
Name of the column in the annotation

### Execute

Once all of the information is correctly input, click on the “Execute” button to create Hierarchical Cluster Heatmaps.



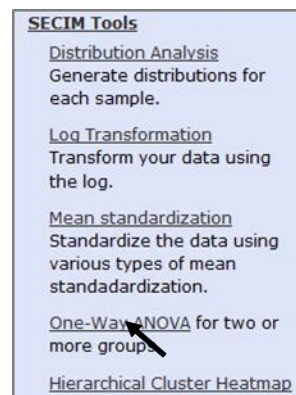
Execute

After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

**One-Way ANOVA** – This tool does a row-by-row analysis calculating a One-way ANOVA on the selected groups. Two input datasets are needed.

Click on the *One-Way ANOVA* tool in the *SECIM Tools* menu on the left side of the screen.

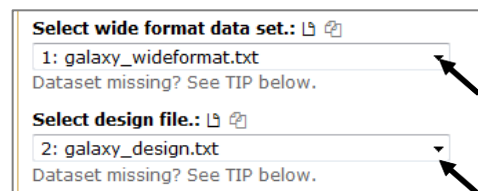
*You will need a wide dataset and a design dataset to execute one-way ANOVA in Galaxy. See above for details on creating necessary files.*





**SECIM Tools**  
[Distribution Analysis](#)  
Generate distributions for each sample.  
[Log Transformation](#)  
Transform your data using the log.  
[Mean standardization](#)  
Standardize the data using various types of mean standardization.  
[One-Way ANOVA for two or more groups](#)  
[Hierarchical Cluster Heatmap](#)

Select your wide format data set from the drop down menu

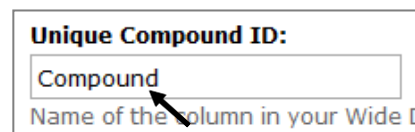
*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the [Edit Datasets > Convert characters](#) tool to convert your data to a **TAB delimited file**. See above for more details.*



Select wide format data set.:   
1: galaxy\_wideformat.txt  
Dataset missing? See TIP below.  
Select design file.:   
2: galaxy\_design.txt  
Dataset missing? See TIP below.


Select your design file in the second drop down menu

Type the name of the column (Compound in this example) in your wide dataset that has unique compound or sample IDs into the Unique Compound ID box.



**Unique Compound ID:**  
Compound  
Name of the column in your Wide D

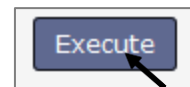
Type in the name of the column in your design file that identifies your different groups (column name is Group1 in this example).



**Group/Treatment:**  
Group1  
Name of the column in your Design File t

## Execute

Once all of the information is correctly input, click on the “Execute” button to submit the One-Way ANOVA Job to Galaxy.



After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

## Output

You will get three different outputs from using the one-way ANOVA tool: a results table, qq plots, and volcano plots.

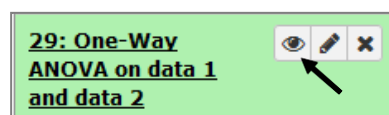
The results table is a TSV table that contains the one-way ANOVA results and analysis of means.

The qq plots output is a PDF that contains quantile – quantile plots (i.e. normal quantile plots) from ANOVA. This plot is helpful for visualizing the extent to which data are normally distributed.

The volcano plots output is a PDF that contains volcano plots comparing differences between group means. Volcano plots give an overview of potentially interesting samples. The log change is on the x-axis and the negative log<sub>10</sub> p-value is on the y-axis.

## Viewing Output

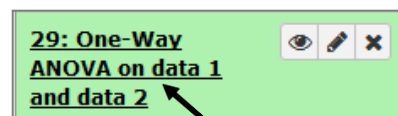
To view each of the output files click on the “eye” symbol on the green box in your history on the right side of the screen.



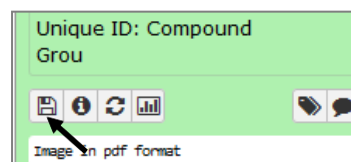
Click on the “eye” symbol for each of the outputs to examine each of the output files.

## Saving Output to Computer

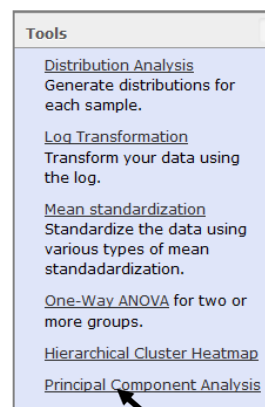
To download the output to your computer first click on the title of the output in your history. This will expand the box and give more information about the tool and data.



After expanding the green box in your history, click on the *download* button on the bottom of the box to save the file to your computer.



**Principal Component Analysis** – This tool performs Principal Component Analysis on the given numeric input data using functions from R statistical package – ‘princomp’ function (for Eigenvector based solution) and ‘prcomp’ function (for Singular Value Decomposition based solution).

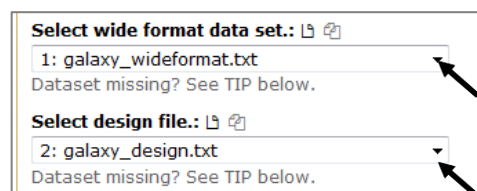


Click on the *Principal Component Analysis* tool in the *SECIM Tools* menu on the left side of the screen.

*You will need a wide dataset and a design dataset to execute Principal Component Analysis in Galaxy. See above for details on creating necessary files.*

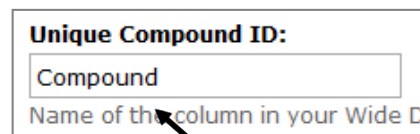
Select your wide format data set from the drop down menu

*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the Edit Datasets > Convert characters tool to convert your data to a TAB delimited file. See above for more details.*



Select your design file in the second drop down menu

Type the name of the column (Compound in this example) in your wide dataset that has unique compound or sample IDs into the Unique Compound ID box.



Select the method that you would like to use to execute your Principal Component Analysis with. Eigenvectors of Correlation (princomp) will create a PCA using the correlations among variables. Eigenvectors of Covariance (princomp) will create a PCA using the covariances among variables. Singular Value Decomposition (prcomp) will create a PCA using SVD.

*(Optional) If you select Singular Value Decomposition you will need to select a method to center and scale your variables.*

### Execute

Once all of the information is correctly input, click on the “Execute” button to run Principal Component Analysis.



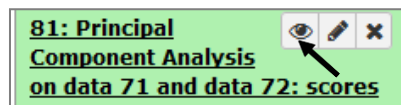
After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

## Output

You will get two different outputs from using the Principal Component Analysis tool: a TSV file containing eigenvectors/variable loadings and a TSV file containing scores of input data on principal components.

## Viewing Output

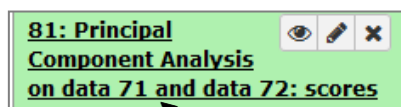
To view each of the output files click on the “eye” symbol on the green box in your history on the right side of the screen.



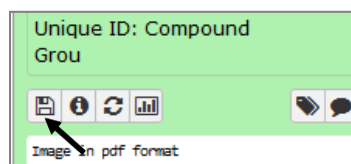
Click on the “eye” symbol for each of the outputs to examine each of the output files.

## Saving Output to Computer

To download the output to your computer first click on the title of the output in your history. This will expand the box and give more information about the tool and data.



After expanding the green box in your history, click on the *download* button on the bottom of the box to save the file to your computer.



## Distribution Analysis – Generate distributions for each sample.

Click on the *Distribution Analysis* tool in the *SECIM Tools* menu on the left side of the screen.

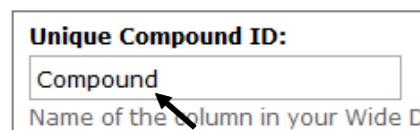
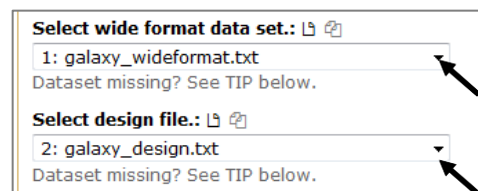
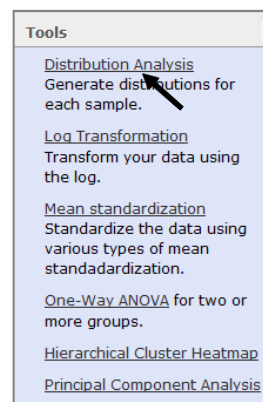
*You will need a wide dataset and a design dataset to execute Distribution Analysis in Galaxy. See above for details on creating necessary files.*

Select your wide format data set from the drop down menu

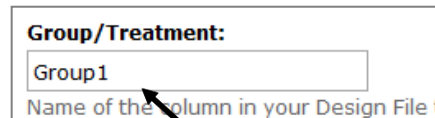
*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the *Edit Datasets > Convert characters* tool to convert your data to a *TAB delimited* file. See above for more details.*

Select your design file in the second drop down menu

Type the name of the column (Compound in this example) in your wide dataset that has unique compound or sample IDs into the Unique Compound ID box.

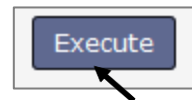


Type in the name of the column in your design file that identifies your different groups (column name is Group1 in this example).



### Execute

Once all of the information is correctly input, click on the “Execute” button to run Distribution Analysis.



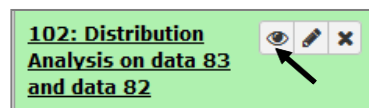
After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

### Output

You will get two different outputs from running the Distribution Analysis tool: A PDF file containing distributions and an identical HTML distribution plot with interactive zoom.

### Viewing Output

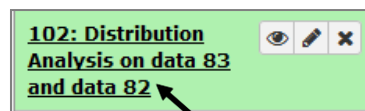
To view each of the output files click on the “eye” symbol on the green box in your history on the right side of the screen.



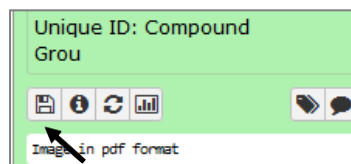
Click on the “eye” symbol for each of the outputs to examine each of the output files.

### Saving Output to Computer

To download the output to your computer first click on the title of the output in your history. This will expand the box and give more information about the tool and data.

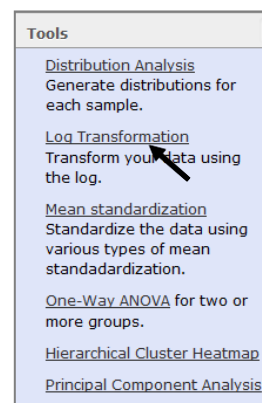


After expanding the green box in your history, click on the *download* button on the bottom of the box to save the file to your computer.



### Log Transformation – Transform data using the log

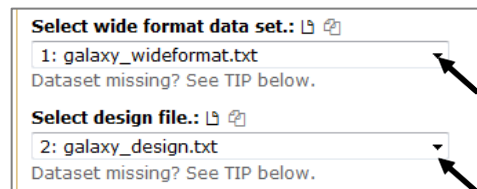
Click on the *Log Distribution* tool in the *SECIM Tools* menu on the left side of the screen.



You will need a wide dataset and a design dataset to perform Log Distribution Analysis in Galaxy. See above for details on creating necessary files.

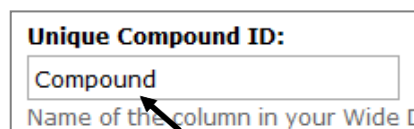
Select your wide format data set from the drop down menu

*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the Edit Datasets > Convert characters tool to convert your data to a TAB delimited file. See above for more details.*

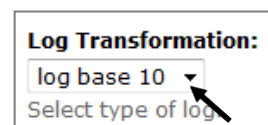


Select your design file in the second drop down menu

Type the name of the column (Compound in this example) in your wide dataset that has unique compound or sample IDs into the Unique Compound ID box.

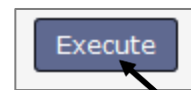


Select the Log Transformation type that you would like to perform. You can conduct a log base 10 transformation, log base 2 transformation, or Natural Log transformation.



### Execute

Once all of the information is correctly input, click on the “Execute” button to run Distribution Analysis.



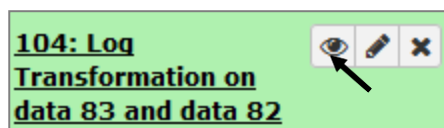
After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

### Output

A wide format dataset will be output with log transformed values.

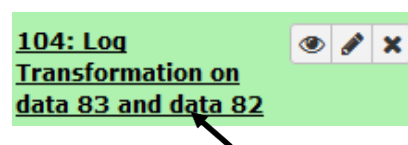
### Viewing Output

To view the output file click on the “eye” symbol on the green box in your history on the right side of the screen.

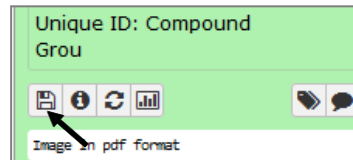


### Saving Output to Computer

To download the output to your computer first click on the title of the output in your history. This will expand the box and give more information about the tool and data.

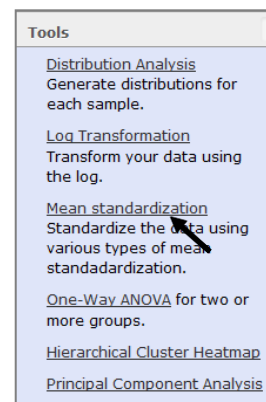


After expanding the green box in your history, click on the *download* button on the bottom of the box to save the file to your computer.



### **Mean Standardization** – Standardize data to mean or standard deviation.

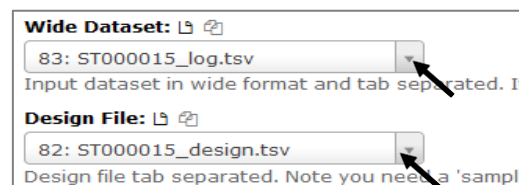
Click on the *Mean Standardization* tool in the *SECIM Tools* menu on the left side of the screen.



*You will need a wide dataset and a design dataset to execute Mean Standardization. See above for details on creating necessary files.*

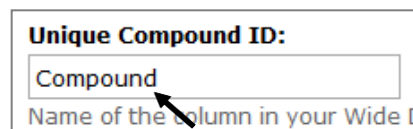
Select your wide format data set from the drop down menu

*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the *Edit Datasets > Convert characters* tool to convert your data to a *TAB delimited* file. See above for more details.*



Select your design file in the second drop down menu

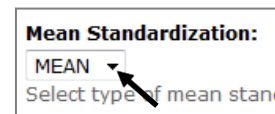
Type the name of the column (Compound in this example) in your wide dataset that has unique compound or sample IDs into the Unique Compound ID box.



Type in the name of the column in your design file that identifies your different groups (column name is Group1 in this example).



Choose the type of standardization that you would like to perform – standardize based on the mean or standardize based on the standard deviation.



### **Execute**

Once all of the information is correctly input, click on the “Execute” button to run Mean Standardization.



After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as

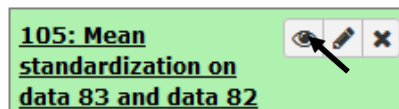
space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

## Output

A wide format dataset will be output with standardized values.

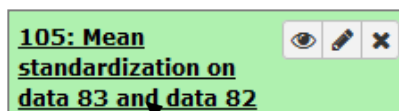
## Viewing Output

To view each of the output file click on the “eye” symbol on the green box in your history on the right side of the screen.

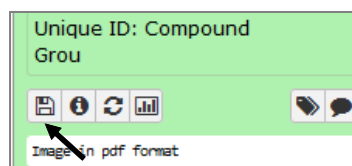


## Saving Output to Computer

To download the output to your computer first click on the title of the output in your history. This will expand the box and give more information about the tool and data.



After expanding the green box in your history, click on the *download* button on the bottom of the box to save the file to your computer.

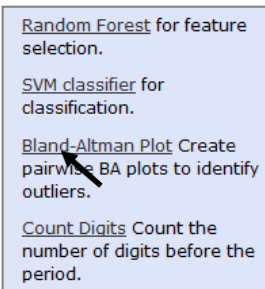


**Bland-Altman Plot** – The Bland-Altman plot is commonly used to look at concordance of data between samples. It is especially useful for looking at variability between replicates. This script will generate BA-plots for all pairwise combinations of samples, or if group information is provided it will only report pairwise combinations within the group.

A linear regression is performed on the BA-plots to identify samples whose residuals are beyond a cutoff. For each compound (row) in the dataset, a sample is flagged as an outlier if the Pearson normalized residuals are greater than a cutoff.

Click on the *Bland-Altman Plot* tool in the *SECIM Tools* menu on the left side of the screen.

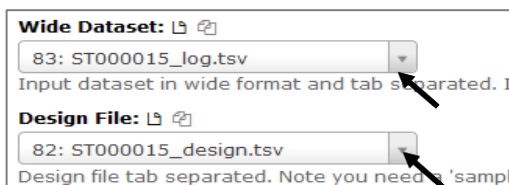
*You will need a wide dataset and a design dataset to create Bland-Altman Plots. See above for details on creating necessary files.*







Select your wide format data set from the drop down menu

*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the [Edit Datasets > Convert characters tool](#) to convert your data to a TAB delimited file. See above for more details.*

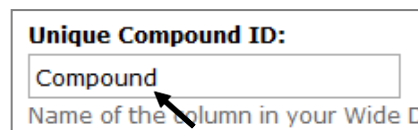


**Wide Dataset:**  83: ST000015\_log.tsv  
Input dataset in wide format and tab separated. I

**Design File:**  82: ST000015\_design.tsv  
Design file tab separated. Note you need a 'sampl

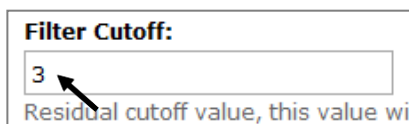
Select your design file in the second drop down menu

Type the name of the column (Compound in this example) in your wide dataset that has unique compound or sample IDs into the Unique Compound ID box.



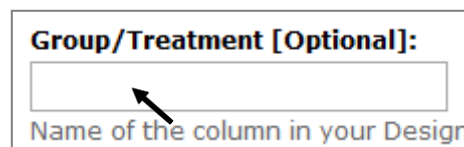
**Unique Compound ID:**  
Compound  
Name of the column in your Wide D

Select the cutoff value for residuals. Any samples with values  $\geq$  the cutoff value will be flagged. (Default value is 3).



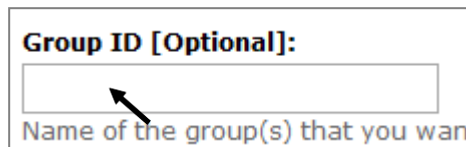
**Filter Cutoff:**  
3  
Residual cutoff value, this value wi

(Optional) If you want to run the analysis on select groups (e.g. qc samples) you will need to input the name of the column that identifies your groups. If you want to run the analysis on all samples leave the box blank.



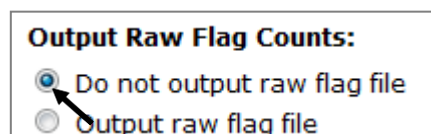
**Group/Treatment [Optional]:**  
Name of the column in your Design

(Optional) If you are performing analyses on groups, enter the Group ID of the group you want to process in the Group ID box. Leave blank if you want to process all groups.



**Group ID [Optional]:**  
Name of the group(s) that you wan

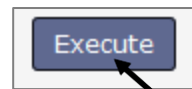
Check the box indicating if you want to output raw flag files. A summary of flags is always output but raw files will only be output if selected.



**Output Raw Flag Counts:**  
 Do not output raw flag file  
 Output raw flag file

## Execute

Once all of the information is correctly input, click on the “Execute” button to create Bland-Altman Plots.



Execute

After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

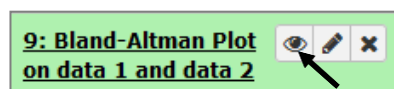
## Output

This tool will always output 3 different files. A PDF of pairwise scatter plots and BA-plots; a PDF of bar graphs for samples showing the number of data points flagged as outliers, and a TSV file containing the sum of pairwise flags, summarized to sample level.

If all options are selected 2 additional files will also be created. A TSV file containing outlier flags for each pairwise comparison, and a TSV design file relating outlier flags back to individual samples.

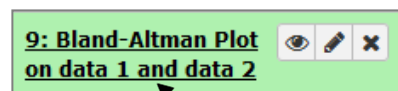
## Viewing Output

To view each of the output file click on the “eye” symbol on the green box in your history on the right side of the screen.

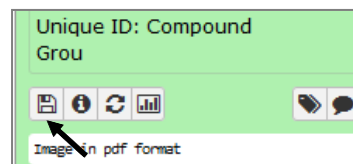


## Saving Output to Computer

To download the output to your computer first click on the title of the output in your history. This will expand the box and give more information about the tool and data.



After expanding the green box in your history, click on the *download* button on the bottom of the box to save the file to your computer.

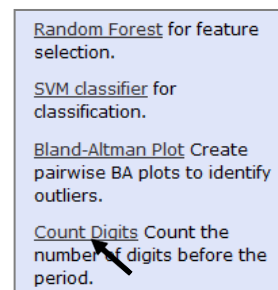


**Count Digits** – For a quick QC, this script allows you to look and see if values are not drastically different orders of magnitude. A constant difference in magnitude may indicate there is something wrong with a sample. For example, if you have a sample that is routinely in the three digit range (100) while all other samples are in the size digit range (10000).

This script counts the number of digits before the decimal place.

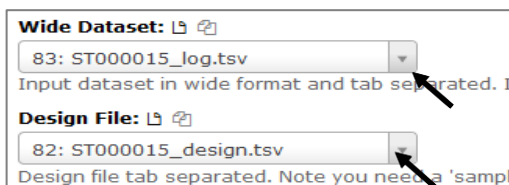
Click on the *Count Digits* tool in the *SECIM Tools* menu on the left side of the screen.

*You will need a wide dataset and a design dataset to use the Count Digits tool. See above for details on creating necessary files.*



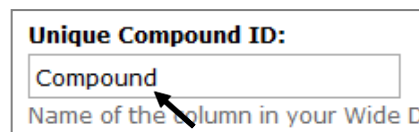
Select your wide format data set from the drop down menu

*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the [Edit Datasets > Convert characters tool](#) to convert your data to a TAB delimited file. See above for more details.*

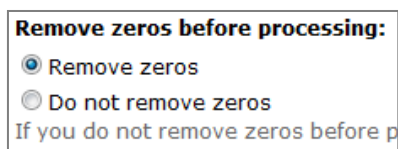


Select your design file in the second drop down menu

Type the name of the column (Compound in this example) in your wide dataset that has unique compound or sample IDs into the Unique Compound ID box.

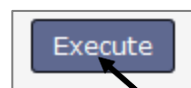


Choose whether you want to remove zeros before processing. If you have zeros in your data and do not remove them they may skew your results.



### Execute

Once all of the information is correctly input, click on the “Execute” run the Count Digits tool.



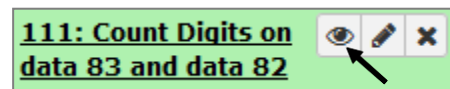
After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

### Output

This tool will always output 2 different files. A TSV with compound as row and sample as column with values indicating the number of digits that a sample had for a given compound. You will also get a PDF of the distribution of digit counts

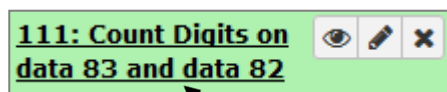
### Viewing Output

To view each of the output file click on the “eye” symbol on the green box in your history on the right side of the screen.

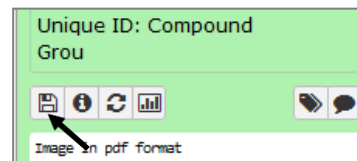


### Saving Output to Computer

To download the output to your computer first click on the title of the output in your history. This will expand the box and give more information about the tool and data.

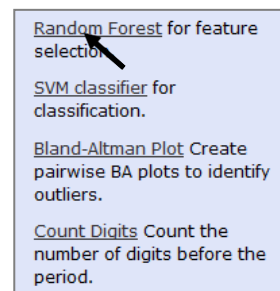


After expanding the green box in your history, click on the *download* button on the bottom of the box to save the file to your computer.



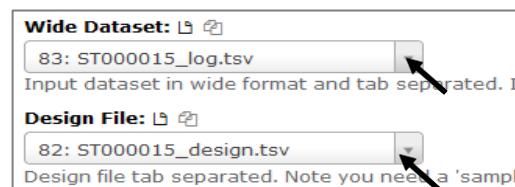
**Random Forest** - This tool uses the Random forest algorithm to select the important features that the groups/classes differentiate the most based on. Click on the *Random Forest* tool in the *SECIM Tools* menu on the left side of the screen.

*You will need a wide dataset and a design dataset to use the Random Forest tool. See above for details on creating necessary files.*



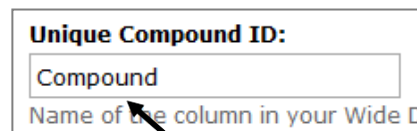
Select your wide format data set from the drop down menu

*If you do not see your dataset on the drop down menu it is likely in the wrong format. Use the *Edit Datasets* > *Convert characters* tool to convert your data to a *TAB* delimited file. See above for more details.*

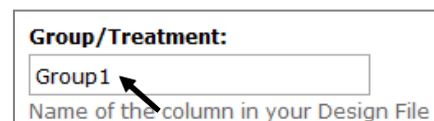


Select your design file in the second drop down menu

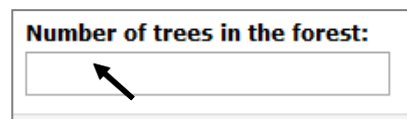
Type the name of the column (Compound in this example) in your wide dataset that has unique compound or sample IDs into the Unique Compound ID box.



Enter the name of the column in your design file that has your unique group IDs.



Enter the number of trees that you want to have in your forest.



## Execute

Once all of the information is correctly input, click on the “Execute” run the Count Digits tool.



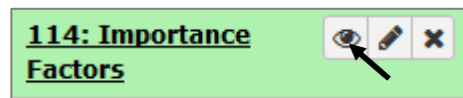
After hitting execute, the job will be added to your history on the right side of the screen. The job will appear gray which means it was added to the line of jobs and will be executed as soon as space is available. Once the job is running, the box will turn yellow, if the job executes successfully the box will turn green, if the job is unsuccessful, it will turn red.

## Output

This tool will always output 2 different files. A transformed dataset and a rand-order list of features and their relative importance.

## Viewing Output

To view each of the output file click on the “eye” symbol on the green box in your history on the right side of the screen.



## Saving Output to Computer

To download the output to your computer first click on the title of the output in your history. This will expand the box and give more information about the tool and data.



After expanding the green box in your history, click on the *download* button on the bottom of the box to save the file to your computer.

