# Tutorial for CEBS Gene Signature Lookup Tool

**Page Navigation (Quick Links):**

## Function

The Gene Signature Lookup Tool provides the ability to search motif signatures from the CEBS data repository and estimate similarity to a user's gene expression pattern.

This tool is currently designed for use with rat data gathered on Affymetrix and Codelink arrays only. Specifically, data correlations will be made to Affymetrix Rat Genome 230 2.0 array and Codelink UniSet Rat I array. Data for other arrays can be analyzed only if there is a commonality of probe IDs (gene identifiers).

The CEBS Gene Signature Lookup Tool has 4 steps:

- Upload Data
- Review Data and Provide Additional Details
- Perform Data Analysis
- Review Results

From the CEBS Homepage (http://tools.niehs.nih.gov/cebs3/ui/), select the Look Up Gene Signatures Workflow, then follow the instructions below.

*Last Updated 18 July, 2016*

## 1) Upload Data

Upload log 2 normalized data as a tab delimited text file (.txt extension) containing the following minimum information

- Gene identifier (probe ID)
- Fold change data

Currently upload is limited to a file size of 50 kb.

To upload a file:

1. Select the "Choose" button, and then navigate to and select the data file of choice; upload is successful when the correct file name appears in the File Name box
2. Indicate using the checkbox whether column headers / names are contained in the data file
3. Indicate how values are delimited in the data file; currently only tab delimited files are accepted
4. Select 'Continue'



## a) Problems with Uploading Data

Data files that exceed the upload limit (50 kb) will not be saved. If this occurs, the File Name field will not populate and the option to select 'Continue' remains unavailable (see below). The file may be removed by clicking the 'x'.

Data files that are not properly formatted will not be saved and will produce an error message (see below). Please revise the file to the correct format and try again.



## 2) Review Data and Provide Additional Details

Uploaded data are provided for review prior to analysis in table format.



Additional details are required before data analysis is performed; select the following parameters from the pull-down lists provided.

- Identifier Column: indicate the column name for the gene identifiers
- Identifier Type: indicate the type of gene identifier provided
- Fold Change Column: indicate the column name for the fold change data
- Platform: indicate whether the platform used in the microarray experiments was Affymetrix or Codelink
- Array (Chip): indicate the name of the array used for the microarray experiments

After all fields are selected, click the "Submit" button to submit data.



3

## 3) Perform Data Analysis

Review the information being submitted for analysis and initiate data analysis by clicking 'Compute'.

ID column:  ProbeSet ID
ID type:  Probe Set Name
Value Column:  FOLD_CHANGE(LOG2)
Chip/array type:  Affy_Rat230_2
Platform:  Affymetrix

Your data has been submited successfully. You may use this unique number to retrieve the result later: **39**

It may take seconds to complete the data submission. Please click the **Compute** button to start the process and visit this page later.

[ Compute ]       0%

*Note: The Data Submission Number provided on the data analysis page.  This may be used to retrieve analysis results at any time by entering this number on the Upload Data page.*

### *Upload Data Page*

If you have previous submission for the pattern search, please enter you submission ID and retrieve your result if available.

**Retrieve Correlation Result from Previous Submission**

Submission Identifier:  39    [ Retrieve ]

## 4) Review Results

Pearson's correlation coefficient values are provided in table format along with the correlated experiment from the CEBS repository.

- Comparison: correlated experiment name
- Compound: correlated compound name
- Correlation: Pearson's correlation coefficient value

Data can be sorted by clicking the arrows in the header rows of the table.  By default, the data are sorted according to the correlation coefficient.

A cutoff value can be applied to the correlation to limit the results if desired.  Type the value into the 'Correlation Cutoff' box and select 'Apply Cutoff'.

| Correlation Match : Total = 20. (Correlation Cutoff: 0.5) | | |
|---|---|---|
| **Comparison** ⬍ | **Compound** ⬍ | **Correlation** ⬍ |
| MIFEPRISTONE-5d-300mg/kg-LI-RATM-RG230-2 | MIFEPRISTONE | 0.596796 |
| CLOTRIMAZOLE-3d-89mg/kg-LI-RATM-RG230-2 | CLOTRIMAZOLE | 0.587295 |
| CARBAMAZEPIN-5d-490mg/kg-LI-RATM-RG230-2 | CARBAMAZEPINE | 0.57171 |
| FLUCONAZOLE-5d-394mg/kg-LI-RATM-RG230-2 | FLUCONAZOLE | 0.501315 |
| PHENOBARBITA-3d-54mg/kg-LI-RATM-RG230-2 | PHENOBARBITAL | 0.501869 |
| ECONAZOLE-1d-334mg/kg-LI-RATM-RG230-2 | ECONAZOLE | 0.50745 |
| MICONAZOLE-5d-200mg/kg-LI-RATM-RG230-2 | MICONAZOLE | 0.512779 |
| DIAZEPAM-3d-710mg/kg-LI-RATM-RG230-2 | DIAZEPAM | 0.513182 |
| RABEPRAZOLE-3d-1024mg/kg-LI-RATM-RG230-2 | RABEPRAZOLE | 0.513681 |
| CARVEDILOL-5d-2000mg/kg-LI-RATM-RG230-2 | CARVEDILOL | 0.520394 |
| SAFROLE-3d-488mg/kg-LI-RATM-RG230-2 | SAFROLE | 0.528027 |
| CLOTRIMAZOLE-3d-52mg/kg-LI-RATM-RG230-2 | CLOTRIMAZOLE | 0.53041 |
| CLONAZEPAM-5d-2500mg/kg-LI-RATM-RG230-2 | CLONAZEPAM | 0.547022 |
| ANASTROZOLE-1d-400mg/kg-LI-RATM-RG230-2 | ANASTROZOLE | 0.549945 |
| MICONAZOLE-1d-920mg/kg-LI-RATM-RG230-2 | MICONAZOLE | 0.551599 |

Correlation Cutoff: 0.5  [Apply Cutoff]  [Remove Cutoff]

Select a correlated experiment row to review associated for the study

- Characteristics:  experimental details and compound structure

| Characteristics | Histopathology | Signatures |
|---|---|---|

| Attribute Name | Attribute Value |
|---|---|
| ADMINISTRATION_FRE( | DAILY |
| ADMINISTRATION_ROU | ORAL GAVAGE |
| ARRAY_TECHNOLOGY | GENECHIP |
| CHIP_NAME | Affy_Rat230_2 |
| COMPOUND_NAME | CLOTRIMAZOLE |
| DOSE | 89 |
| DOSE_LEVEL | M1 |
| DOSE_UNIT | mg/kg |
| EXPERIMENT_DATE | 20011127 |
| EXPERIMENT_NAME | CLOTRIMAZOLE-3d-89mg/kg-LI-RATM-RG230-2-20011127 |
| ICXNUMBER | ICX5600186 |
| ORGANISM_DESCRIPTI | Rat, Male, CD-IGS, Sprague Dawley |
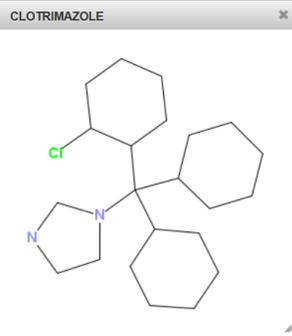
CLOTRIMAZOLE

CLOTRIMAZOLE

- Histopathology:  results of histopathology evaluation with statistical significance determined using a ridit ("relative to an identified distribution") score calculation

| Characteristics | Histopathology | Signatures |
| --- | --- | --- |

| Histopathology ⇕ | pValue ⇕ |
| --- | --- |
| AUTOLYSIS | 0.504787 |
| BILE DUCT DILATATION | 0.504787 |
| BILE DUCT HYPERPLASIA | 0.509574 |
| BILE DUCT, NECROSIS, ONCOCYTIC | 0.504787 |
| CAPSULE, INFLAMMATORY CELL INFILTRA | 0.516751 |
| CAPSULE, MESOTHELIAL CELL, HYPERPL | 0.504787 |
| CAPSULE, THROMBUS | 0.504787 |
| CENTRILOBULAR FIBROSIS | 0.509574 |
| CENTRILOBULAR, INFLAMMATORY CELL | 0.559618 |
| CENTRILOBULAR, INFLAMMATORY CELL | 0.692869 |
| CLEAR CELL FOCUS | 0.504787 |
| HEPATOCYTE, CENTRILOBULAR, ATROPH | 0.504787 |
| HEPATOCYTE, CENTRILOBULAR, CYTOPL | 0.504787 |
| HEPATOCYTE, CENTRILOBULAR, DEGENE | 0.504787 |
| HEPATOCYTE, CENTRILOBULAR, GLYCOG | 0.504787 |

- Signatures: classifies the pathological and pharmacological effects of the correlated compound using gene expression data
  - Select a tab to review data
  - Click on a the arrows to expand the data group
  - Select individual findings to view additional details
- Classification algorithms used include:
  - ASPLP (Adjusted Sparse Linear Programming)
  - SPLP (Sparse Linear Programming)
  - T-rank

| Characteristics | Histopathology | Signatures |
| --- | --- | --- |

| ASPLP | SPLP | TRANK |
| --- | --- | --- |

▸ Body and Organ Weights
▸ Clinical Pathology
▸ Histopathology
▸ Literature
▾ Structure Activity Class
  ▾ LIVER
    ▯ Thyroperoxidase inhibitor_LIVER_TRANK
    ▯ Toxicant, DNA alkylator_LIVER_TRANK
▸ Structure Activity Class and Pharmacology

## 5) Example Section

There is an example data set on the Tutorial page, which is located on the menu on the left side of the page. This example will detail how to complete the gene signature lookup with that sample data set, and the type of question that can be answered using the Gene Signature Lookup. To download this file, click the link and follow the instructions in your browser.

The question we will answer in this example is: what is the most highly correlated comparison to the sample data, and what were the significant histopathology results?

### a) Upload Data

Once you have downloaded and saved the tutorial file, navigate back to the Gene Signature Lookup page. Then use the + Choose button to upload the sample file. The file name should appear in the box. The other settings can remain at their default values. Click Continue.

### b) Review Data

The next page will display the newly uploaded file. The Additional information section needs filling in. Select the following options:

1) Identifier Column: ProbeSet ID
2) Identifier Type: Probe Set Name
3) Fold Change Column: FOLD_CHANGE(LOG2)
4) Platform: Affymetrix
5) Array (Chip): Affy_Rat230_2

When finished, click the Submit button

### c) Perform Data Analysis

Ensure the input criteria on the next page match the list above. Also make note of the unique Gene Signature Lookup number. If everything is correct, click the Compute button. This process may take some time.

***Note:*** *To retrieve your results at a later time return to the Gene Signature Lookup homepage, enter the submission identifier under the heading Retrieve Correlation Result from Previous Submission, and select Retrieve.*

### d) Review Results

To find the closest correlation, click on the Correlation header until it is sorted in decreasing order (click twice). The highest correlation should be 'Mifepristone…' Click to select that row. Because we are interested in Histopathology results, choose the

7

Histopathology tab. To sort the pValue column in ascending order, click the column header. There is one significant result: hepatocyte, periportal, lipid accumulation, macrovesicular. Clicking on this row will open a window with more details about this result, including images.

## 6) Additional Instructions/Information

1) Flash movie running flex code for CEBS indicates a software platform is being used to create graphics or texts for CEBS

2) CEBS data may be cited by navigating to the CEBS Support page, and selecting the [Citing CEBS](#) document under FAQs.  Here, instructions can be found on citing NTP Data as well as non-NTP data in CEBS.