



Gene Extract USER'S MANUAL

Isa Kemal Pakatci

July 24th, 2014

USER'S MANUAL

TABLE OF CONTENTS

	<u>Page #</u>
1.0 GENERAL INFORMATION	1-1
1.1 System Overview.....	1-1
1.2 Authorized Use Permission.....	1-1
1.3 Points of Contact	1-1
1.3.1 Information	1-1
2.0 APPLICATION USAGE.....	2-1
2.1 Users and Login	2-1
2.2 Interface Organization	2-12
2.2.1 Header Panel	2-2
2.2.1 Navigation Panel	2-2
2.2.1 Main Panel.....	2-2
2.3 Extracting New Data	2-33
2.3.1 By KEGG Network Name.....	2-3
2.3.2 By Gene ID	2-4
2.3.4 By Probe ID.....	2-5
2.4 Previously Extracted Date.....	2-5
2.5 Result of Extraction and Analysis	2-6
2.6 Uploading a Dataset	2-9
2.7 Managing Users.....	2-10

1.0 GENERAL INFORMATION

1.0 GENERAL INFORMATION

1.1 System Overview

Gene Extract v1 is a web application intended to assist those people who want to quickly analyze their mouse gene expression data and share the results. Using Gene Extract, users can extract expression values that correspond to the subset of genes they select and immediately see the result of commonly used analysis techniques such as clustering and principal component analysis. As a web application, Gene Extract can serve as a data repository as well as collaboration tool for teams so that team members can easily share analysis results of subset or network of genes that shows interesting patterns in their data.

1.2 Authorized Use Permission

Usage of this software is limited to its owner via the terms of its development. Gene Extract is wholly owned by CS Bio Group at University of North Carolina at Chapel Hill, and may not be used or referenced without their express consent.

1.3 Points of Contact

1.3.1 Information

For additional information, CS Bio group at UNC can be contacted through Group Leader/ Leonard McMillan (mcmillan@cs.unc.edu).

2.0 APPLICATION USAGE

2.0 APPLICATION USAGE

2.1 Users and Login



Username:

Password:

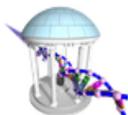
The Gene Extract is a collection of genome-wide transcriptional expression data from mice adipose cells selected by various mechanisms.

This web interface provides access to the current version (**build 01**) of Gene Extract which contains more than 450 expression profiles for more than 22000 probes. It is designed to allow biologists, pharmacologists, chemists and clinical scientists to use basic tools without the need for any specialist ability in the analysis of gene-expression data.

Definition of Gene Extract terms can be found by clicking 'terms and conditions' below this page. If you need help either check the User Manual below or [contact us](#).

The Gene Extract is based at University of North Carolina at Chapel Hill, North Carolina. The Gene Extract team is Daniel Pomp, Leonard McMillan, Wei Wang, Kemal Pakatci and Xiang Zhang

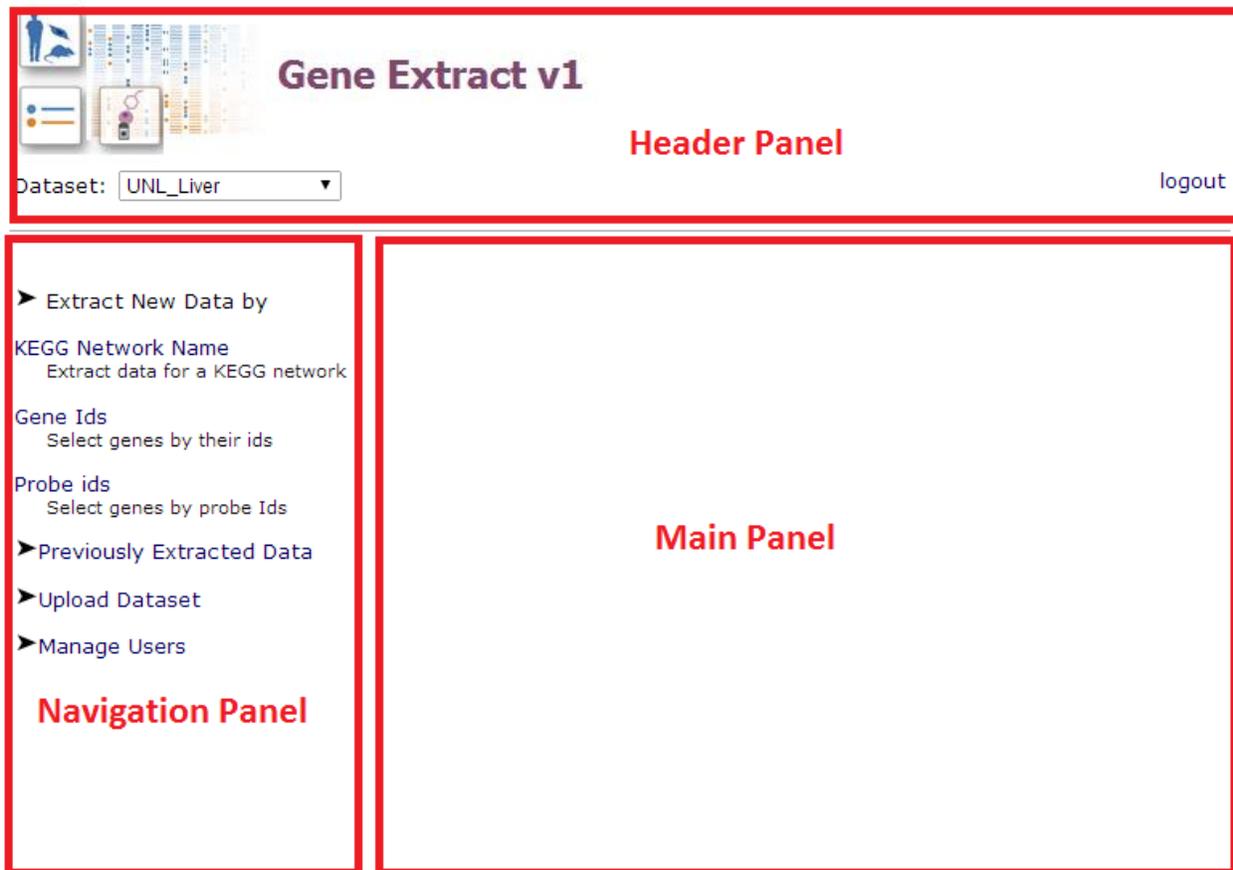
[User Manual](#) | [terms and conditions](#)



© 2009 UNC Computational Genetics Group

In this login screen users need to enter their username and password. Future users need to contact Leonard McMillan to have a new username for them so that they can access the application. Login session keeps open until user logs out by using the logout link on the top left of the screen. There are 2 levels of user permissions: Administrator and user. Administrator users can control which users can access to specific datasets that exists in the data repository.

2.1 Interface Organization



Header Panel

There are two actions in the header panel: Logout link and dataset selection. By clicking logout link users can log out of their session. On the left side of this panel, user can select the dataset that they want to extract gene expressions from. The set of datasets are automatically populated as the list of datasets that the current user have access to.

Navigation Panel

This panel allows user to navigate through the application. It shows 4 basic functionalities of the application: Extract New Data, Previously Extracted Data, Upload Dataset and Manage Users. Only administrators can see and use Manage Users functionality. The detailed description of these functionalities are described later in this document.

Main Panel

This panel changes each time user clicks on a link on the navigation pane whereas Navigation Panel content stay the same. Main panel shows commands and results based on the link clicked.

2.2 Extracting New Data

By Kegg Network Name

► Extract New Data by

Include gene names in output csv files

KEGG Network Name
Extract data for a KEGG network

Include strain ids in output csv files

Gene Ids
Select genes by their ids

Network name:
1- and 2-Methylnaphthalene degradation - mmu00624 - (15 genes) ▼

Probe ids
Select genes by probe Ids

► Previously Extracted Data

► Upload Dataset

► Manage Users

This page allows users to extract the subset of genes based on predefined mouse gene networks. The networks of genes are taken from KEGG database on January 12th, 2009. As of this date, 203 networks are defined in this application. By using the drop down box, user can select set of genes, by choosing appropriate network which is described by network name, KEGG network ID which starts with mmu and number of genes in this network. This way, users don't need to know the names of the genes in a particular network of genes which are generally grouped by biological relevance.

After user clicks submit expressions that correspond to selected set of genes are extracted from the dataset that is selected on the left side of the header panel. The results can be downloaded as cvs file. The checkbox "Include gene names in output csv files" allows users to control whether the resulting cvs contain gene names in the first row. If user wants to view results in a spreadsheet editor, including gene names is generally preferred. But if users is going to load it into a program like MATLAB or Stata, not including the gene names is more useful. The same functionality goes for the strain ids. But strain ids will be added as first column if the appropriate checkbox is selected.

By Gene ID

▶ Extract New Data by Data name: (for future reference)

KEGG Network Name
 Extract data for a KEGG network Include gene names in output csv files

Gene Ids
 Select genes by their ids Include strain ids in output csv files

Probe ids
 Select genes by probe Ids

▶ Previously Extracted Data

▶ Upload Dataset

▶ Manage Users

Gene names:

Strains: (Comma delimited list, leave empty to select all strains)

Another way to extract new data is to provide individual gene names. If the user clicks on “Gene Ids” link on the navigation panel, user sees a screen similar to above figure. User can enter multiple genes by typing the gene name in the “Gene names” text box. A helper list shows the list of gene names base that includes the letters user have typed. User can select a gene by using mouse or arrow keys and hitting enter. Entered genes will be highlighted in blue and they can be deleted either by clicking the ‘x’ sign right after the gene name or using backspace button.

“Data name” textbox can be entered optionally to give a name to the subset of data that is extracted and analyzed. This name will be seen on the “Previously Extracted Data” page. If no data name is entered, a name that includes current date and time will be assigned automatically. User can also select subset of mouse strains that exist in the data.

figure above. As can be seen some of the datasets are named by network name (mmuxxxxxxx) followed by dataset name. These are extracted using Kegg network name. Some data names show date and time which means that the user who created that data didn't enter any name so creating date and time is used as data name. The other data names are given as the user entered.

2.5 Result of Extraction and Analysis

➤ Extract New Data by
 KEGG Network Name
 Extract data for a KEGG network

Gene Ids
 Select genes by their ids

Probe ids
 Select genes by probe Ids

➤ Previously Extracted Data
 ➤ Upload Dataset
 ➤ Manage Users

Dataset: mmu00791_UNL_Liver
Number of genes: 4
Number of experiments: 471

Information about the data

Maximum Likelihood ratio matrix
 Maximum Likelihood average matrix
 Standard deviation matrix
 P value matrix
 Correlation matrix

Link to download useful matrices

Result of gene clustering (dendrogram)

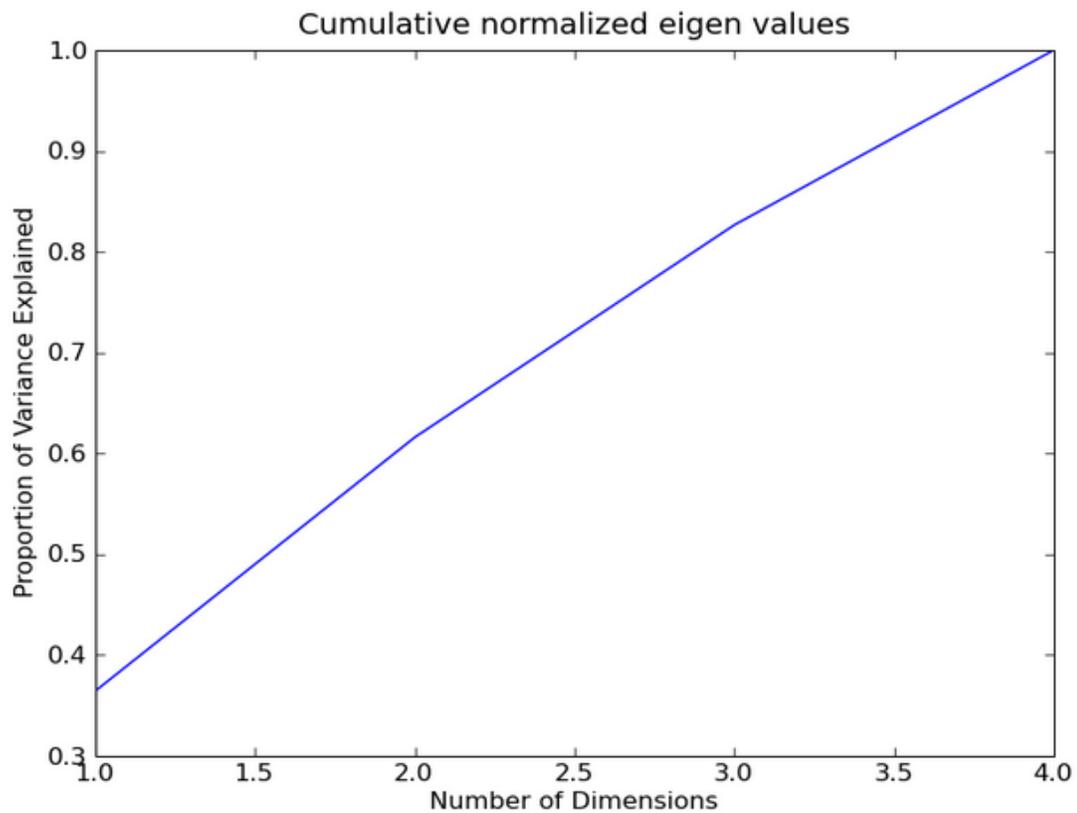
Visualization of correlation matrix

	xobec2	xobec3	Adar	xobec1
xobec2	1.0	0.8	0.2	0.4
xobec3	0.8	1.0	-0.2	0.2
Adar	0.2	-0.2	1.0	0.4
xobec1	0.4	0.2	0.4	1.0

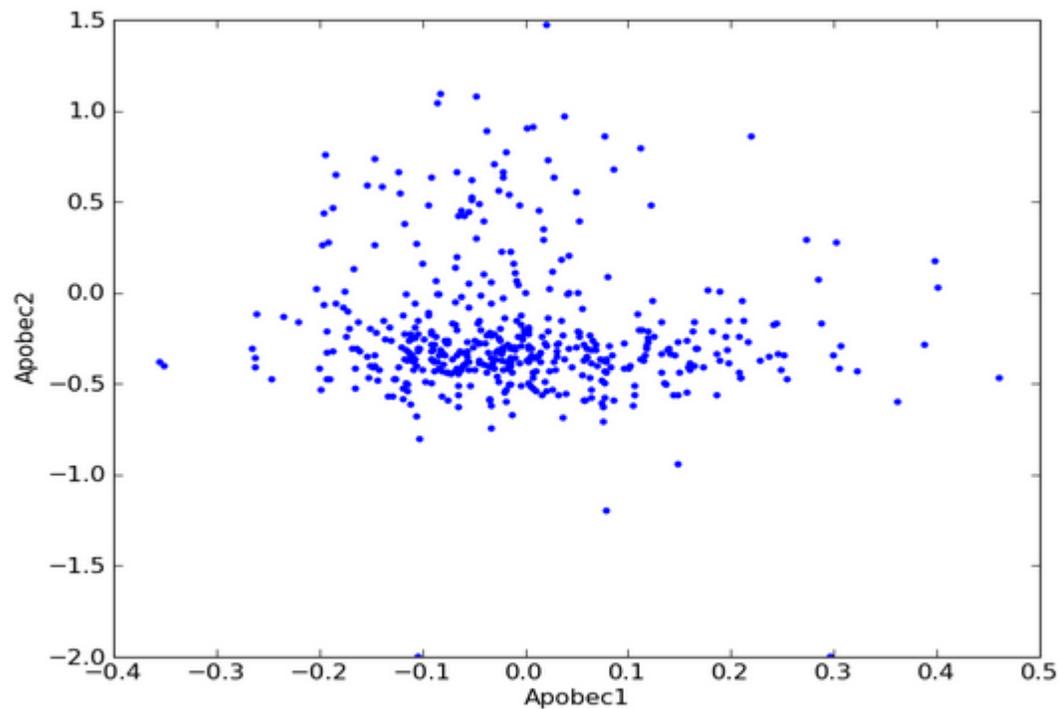
When user extracts new data or view previously extracted data, he/she sees a screen similar to above figure. The first 3 lines show information about the extracted data such as number of genes and experiments. Next 5 lines are the links to download useful matrices as csv files, so these are the result of data extraction operation. The figures that follow these lines are result of the analysis of the data. First analysis that is applied on the data is correlation and cluster analysis. A visualization of correlation matrix

is shown to the user. The order of genes is based on the cluster analysis of the genes. Dendrogram that shows how genes are clustered are also shown

#Dims	Normed Eigen Value	Cum Normed Eigen Value
1	0.364522141615	0.364522141615
2	0.25155825018	0.616080391795
3	0.210515673197	0.826596064992
4	0.173403935008	1.0



Following the results of cluster analysis, the results of Principal Component Analysis (PCA) are shown as a table as well as a line graph. The table shows values of eigen values in PCA which represents the variation in the calculated eigen vectors. Cumulative Normed Eigen Value show the proportion of variation explained by first n eigenvectors, to total variation. Line graph is a visualization of first and third column of the PCA result table.



Following the PCA results, scatterplots for each pair of genes are shown. The purpose of these plots is to allow user to see how pair of genes interact with each other. This is complementary to correlation and clustering results because if we see a high correlation value in the correlation plot we expect to see the points close to diagonal of the scatterplot. Moreover, scatter plot allows user to see that can be missed by correlation plot. For instance there might be small number of outliers in the data which is causing the correlation value to be small whereas other data points suggest strong correlation between the genes. Scatterplot allow user to capture such events.

2.6 Uploading a Dataset

► Extract New Data by Dataset name:

KEGG Network Name
Extract data for a KEGG network Gene expression data file in GEODB SOFT file format: No file chosen

Gene Ids
Select genes by their ids

Probe ids
Select genes by probe Ids

► Previously Extracted Data

► Upload Dataset

► Manage Users

User can upload any dataset that he/she has as long as the data is in GEODB Soft file format. The data should correspond to mouse gene expression data because the gene extraction by KEGG network or Gene name relies on the mouse gene names. User can name the dataset any way he/she wants. The name given by user will be added to the list of datasets that is shown on the left of header panel. By default administrators and the user that uploads the data have access to the data. If the user wants other specific users to access to the data he/she uploaded, the user may ask administrator to grant access.

2.7 Managing Users

▶ Extract New Data by **Administrators**
 KEGG Network Name **mcmillan**
 Extract data for a KEGG network **weiwang**
kemal

Gene Ids
 Select genes by their ids

Probe ids
 Select genes by probe Ids

▶ Previously Extracted Data

▶ Upload Dataset

▶ Manage Users

		dpomp	timw
	BXD	<input type="checkbox"/>	<input checked="" type="checkbox"/>
	UNL_Muscle	<input type="checkbox"/>	<input type="checkbox"/>
	timw	<input type="checkbox"/>	<input type="checkbox"/>
	timw2	<input type="checkbox"/>	<input type="checkbox"/>
	bxd2	<input type="checkbox"/>	<input type="checkbox"/>
	gds_try	<input type="checkbox"/>	<input type="checkbox"/>
	GDS3254	<input type="checkbox"/>	<input type="checkbox"/>
	UNL_Liver	<input checked="" type="checkbox"/>	<input type="checkbox"/>
	compgen_gds	<input type="checkbox"/>	<input type="checkbox"/>
	try	<input type="checkbox"/>	<input type="checkbox"/>
	bxd3	<input type="checkbox"/>	<input type="checkbox"/>
	bxd5	<input type="checkbox"/>	<input type="checkbox"/>
	UNL_Hypothalamus	<input type="checkbox"/>	<input type="checkbox"/>
	bxd4	<input type="checkbox"/>	<input type="checkbox"/>
	GSE8028	<input type="checkbox"/>	<input type="checkbox"/>
	UNL_Adipose	<input type="checkbox"/>	<input type="checkbox"/>

This feature can only be used by administrators. “Manage Users” link will be shown on the navigation panel only if the user have administrative privileges. Using this page administrator can grant or revoke users’ access to specific datasets. List of datasets are shown on the right of the table and each column represents a user. Username is given as the header row of the table in the main panel. A check checkbox means that the corresponding user have access to the data referenced by the name in the first column of the table.