

Analysis of Dose-response Studies in Microarrays Experiments Using The IsoGeneGUI Package

Setia Pramana, Dan Lin, Ziv Shkedy

April 2, 2010

1 Introduction

The IsoGene Graphical User Interface (IsoGeneGUI) is a user friendly interface of the IsoGene package which is aimed to perform analysis of dose-response studies in microarray experiments. The IsoGeneGUI is developed for the user with no or limited knowledge about R programming so he/she can perform the analysis of dose-response in microarray setting easily. This GUI was developed using tcl/tk package. The statistical methodologies (test statistics, etc.) used in this package are discussed by Lin et.al (2007, 2008, 2010).

2 Usage

To run the package

```
> library(IsoGeneGUI)
> IsoGeneGUI()
```

3 Menus

The main window of the IsoGene-UI is presented in Figure 1. The package has four main menus: File, Analysis, Plots and Help. In the middle of the main window there is an info box which provides information about the data (availability and summary) and the result summary of the last performed analysis.

1. File:
 - (a) Open Data
 - i. R workspace (*.RData files)
 - ii. Excel or text file (*.xls or *.txt files)
 - (b) Show Data
 - (c) Exit
2. Analysis:
 - (a) Set seed
 - (b) Likelihood Ratio Test E2 Analysis

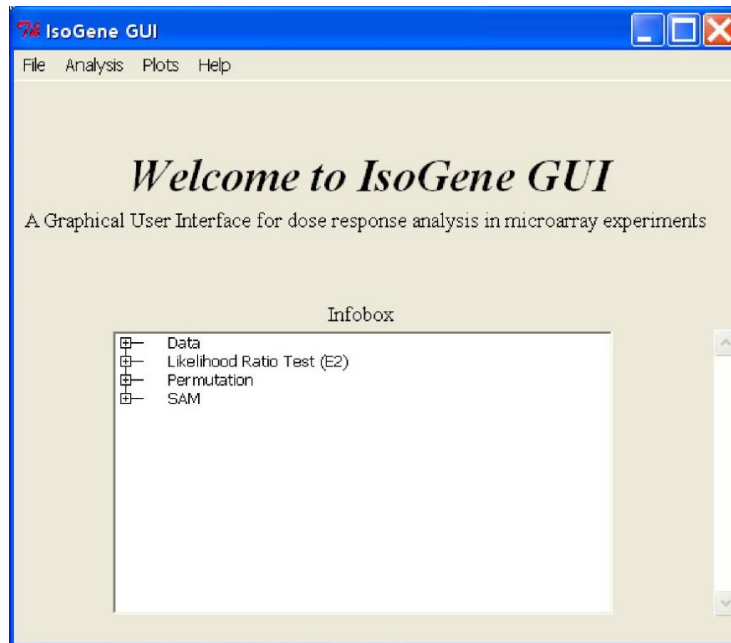


Figure 1: The main window of the ISoGeneGUI.

- (c) Permutation Analysis
- (d) Significant Analysis of Microarrays (SAM)
 - i. SAM Permutation
 - ii. SAM Analysis
- 3. Plot:
 - (a) IsoPlot
 - (b) Permutation Plot
 - (c) SAM Plot
 - i. Plot of FDR vs. Delta
 - ii. Plot of number of significant genes vs. Delta
 - iii. Plot of number of False Positive vs. Delta
 - (d) User defined scatter plot
- 4. Help:
 - (a) IsoGene Help
 - (b) IsoGeneGUI Help
 - (c) About

4 Reading the Data

In the first step of the analysis, the data is uploaded to the package using the File menu. The package can read data in R workspace file (*.RData), excel and text files.

4.1 R workspace files

The format of the gene expression data should be a matrix or table where the columns are the arrays and the rows are the gene names. The dose information should be a vector or table contain the dose levels which corresponds to the arrays in the expression matrix/table.

The package provides an R workspace example dataset called `dopamine2`. For the dopamine2 data, the dose levels are given by

```
> dose
[1] 0.00 0.00 0.01 0.01 0.04 0.04 0.16 0.16 0.63 0.63
[11] 2.50 2.50 0.00 0.00 0.00 0.01 0.01 0.01 0.04 0.04
[21] 0.16 0.16 0.63 0.63 2.50 2.50
```

while the expression matrix has the following structure:

```
> dopamine[1:5, 1:6]
      X1      X2      X3      X4      X5
201_at 2.579138 2.318749 2.496895 2.456772 2.479480
202_at 2.140561 2.061804 2.131749 2.107638 2.086722
203_at 6.988566 6.620562 5.764725 6.326178 7.020716
204_at 11.081855 9.974999 10.790689 10.702516 10.544664
205_at 12.104545 12.076975 11.989770 12.151120 12.118520
```

Part of the `dopamine2` can be obtained in folder "exampleData" inside the package.

The full example data in R workspace, text and excel files can be obtained from: <http://www.ibiostat.be/software/IsoGeneGUI/index.html>

4.2 Open the data set

In order to upload the R workspace we choose in the file menu in Figure ??a the following sequence:

File > Open dataset > R workspace

The package will automatically refer to the example data `dopamine2` if the option open R workspace is choosen.

Text files and Excel files can be uploaded to IsoGene-GUI as well using following sequence in the dialogue box:

File > Open dataset > Excel or Text files

Once the the data is uploaded, information about the data is presented automatically in the info box of the main window.

5 Exploratory Data Analysis

To explore the expression of the genes, the IsoGeneGUI provides the submenu **IsoPlot** from **Plots** menu. This feature displays the data points, sample means at each dose and an isotonic regression line.

There are three input options to draw the isotonic regression plot, using gene name(s), row number(s) or using a range of row numbers. There are also three check boxes:

1. Dose as ordinal. This option will draw the plot and treat dose as ordinal scale. The default will plot with dose as a continuous variable.
2. Show isotonic regression curve for both direction. The default plot just display the isotonic trend/curve which is more likely fit the data. By checking this option, the isotonic trend for both directions will be displayed.
3. Show summary of the data. This option is to provide a short summary of each selected gene.

6 Data Analysis

The IsoGene-GUI package provides three options for analysis: (a) analysis with the likelihood ratio test statistic (LRT) using its exact p-values, (2) resampling based analysis and (3) significance analysis of microarrays (SAM).

6.1 LRT Using exact p-values

To perform the exact LRT we choose the following sequence in the analysis menu:

```
Analysis > Likelihood Ratio Test (E2)
```

Then the main dialog box for analysis based on the LRT using exact p-values is shown in Figure 2.

Note that the users can choose to perform the analysis for all the genes on the array or on pre defined subset. In addition, in order to adjust for multiplicity, we need to select from the menu the adjustment method to be used and the overall error rate.

6.2 Resampling Based Methods

Resampling based analysis can be performed using the main menu for resampling method shown in Figure 3. We need to specify the number of permutations, the test statistic(s) to be used and the multiplicity adjustment(s). Similar to the E^2 menu, here we can specify default graphical displays.

6.3 Significance Analysis of Microarrays (SAM)

The **IsoGene-GUI** also provides testing for the dose-response relationship under order restricted alternatives using the Significance Analysis of Microarrays procedure (SAM). To perform this analysis there are two main steps: (1) calculating the SAM regularized test statistic using permutation and (2) the SAM analysis. These two steps are performed by two sub menus in the SAM menu. Then in the first step several options of the use fudge factor are provided. User can specify to use **no Fudge Factor**, **Automatic Fudge factor** (fudge factor will be calculated using the methods described in the SAM manual), or the **fudge factor based on a specific percentile**.

After the statistics are computed or loaded from a file, we now can perform the SAM analysis by following the sequence:

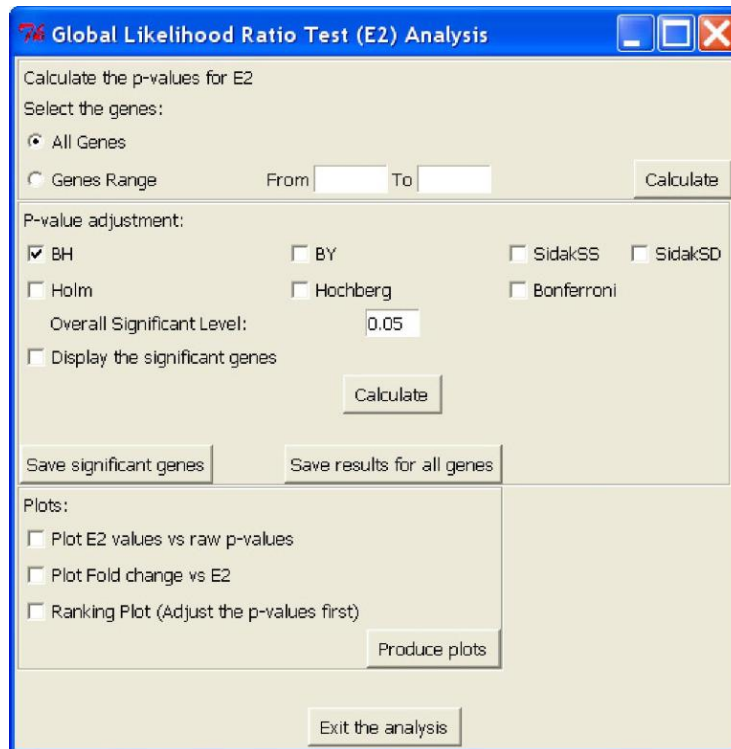


Figure 2: The main dialog box of Likelihood Ratio Test E^2 analysis.

Analysis > Significance Analysis of Microarrays > SAM Analysis

After specifying the test statistic, the main dialog box of SAM analysis, which shows the plot of the observed versus the expected test statistics and other options, will be shown (see Figure 4).

Note that, in this plot we can change the delta value (automatically using delta slider or manually using delta input box) and the FDR level.

7 Saving IsoGeneGUI Outputs

Each plot produced from the analysis in IsoGeneGUI can be copied into clipboard which can be easily pasted into MS.Office documents, such as Ms.Word. It can be done by clicking the **Copy to Clipboard** button located in the bottom of the graph and then paste it directly to the document (or **Ctrl-V**). Furthermore, the plots can also be saved into several image formats (*.ps, *.png, *.jpeg, *.bmp, *.tiff). After the format is selected/highlighted from the list, the user can click **browse** button to specify the name of the image file and also the file location.

Note that results of all genes and lists of significant genes from the three analysis above can be shown and also saved into an **R-workspace** and/or an excel file by clicking **Save** button .

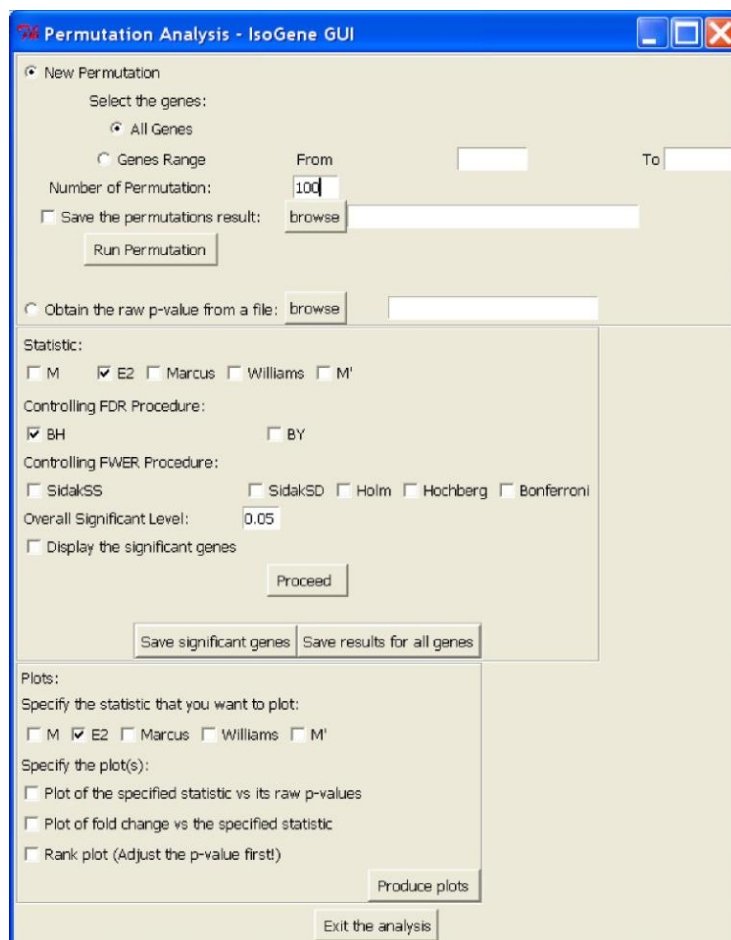


Figure 3: The main dialog box for resampling-based monotone trend test.

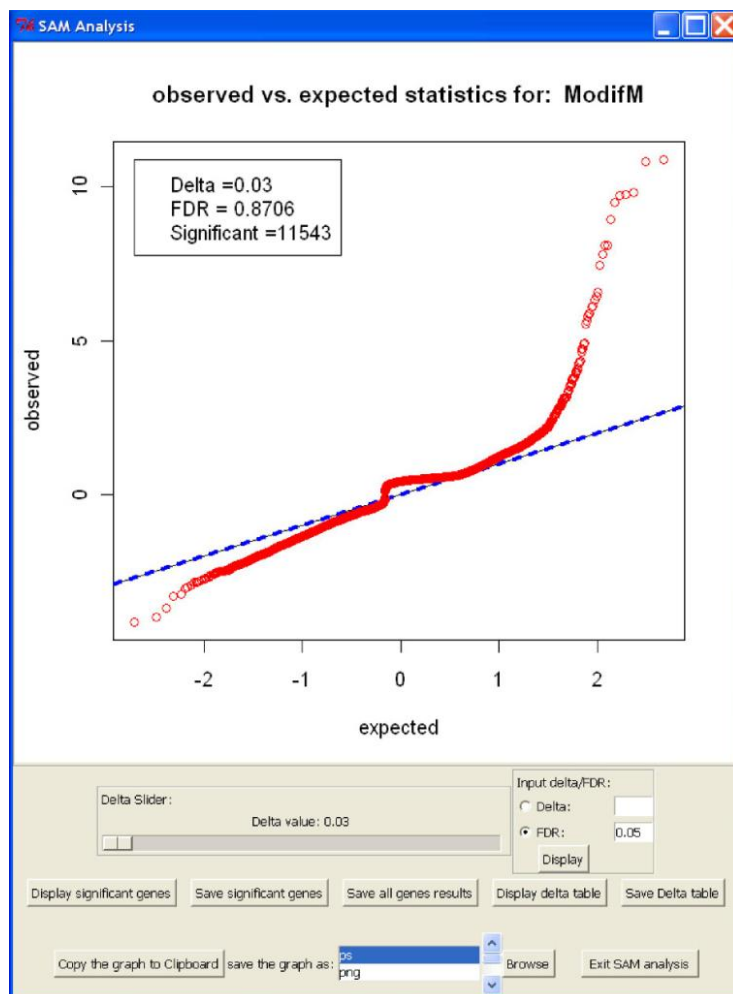


Figure 4: The main dialog box for SAM analysis.

8 Complete users' manual

Users can access the IsoGeneGUI documentation online or by installing the IsoGeneGUI package locally. Then the following code can be typed at R prompt:

```
> if (interactive()) {  
+   browseURL("http://www.ibiostat.be/software/IsoGeneGUI/index.html")  
+ }
```

or alternatively:

```
> if (interactive()) {  
+   library(IsoGeneGUI)  
+   IsoGeneGUIHelp()  
+ }
```

Users can download example dataset and a complete Users' Manual from the site as well.

References

1. Lin, D., Shkedy, Z., Yekutieli, D., Burzykowski, T., Göhlmann, H., De Bondt, A., Perera, T., Geerts, T. and Bijmens, L. (2007) Testing for trends in dose-response microarray experiments: A comparison of several testing procedures, multiplicity and resamplingbased inference. *Statistical Applications in Genetics and Molecular Biology*, **6(1)**, Article 26.
2. Lin, D., Shkedy, Z., Burzykowki, R., Ion, T., Göhlmann, H.W.H., De Bondt, A., Perera, T., Geerts, T. and Bijmens, L. (2008) An investigation on performance of Significance Analysis of Microarray (SAM) for the comparisons of several treatments with one control in the presence of small variance genes. *Biometrical Journal*, Multiple Comparison Problem, Special Issue, **50(5)**, 801–823.
3. Lin, D., Shkedy, Z., Yekutieli, D., Amaratunga, D. and Bijmens, L., editors. *Modeling Dose-response Microarray Data in Early Drug Development Experiments Using R*. Springer, (to be published in 2010).