

IsoGeneGUI: a graphical user interface for analyzing dose-response studies in microarray experiments

Setia Pramana

Interuniversity Institute for Biostatistics and Statistical
Bioinformatics, Universiteit Hasselt, Diepenbeek, Belgium



Outline

- Introduction
 - Dose-response Studies
 - IsoGene Package
 - IsoGeneGUI Package
- Methodology
 - Test Statistics for Trend Test
 - Asymptotic \bar{E}^2
 - Permutation
 - Significance Analysis of Microarray (SAM)
- The IsoGeneGUI Package
 - Menu
 - Data Exploration
 - Permutation
 - Likelihood Ratio Test E^2 Analysis
 - SAM
 - Plots
- Availability

Dose-response studies

- The fundamental objective of drug development.
- Aims: find dose or range of dose that efficacious and safe.

Dose-response studies

To carry out the analysis of dose-response microarray experiments, Lin et al. (2007) discussed the following test statistics which take into account the order restriction of means with respect to the increasing doses:

- Williams (Williams, 1971, 1972)
- Marcus (Marcus, 1976)
- the likelihood ratio test (E^2) (Bartholomew 1961 , Barlow et al. 1972, and Robertson et al. 1988)
- M statistic (Hu et al. 2005)
- Modified M statistic (Lin et al. 2007)

IsoGene Package

- An R package that carries out analysis of dose-response microarray experiments discussed by Lin et.al 2007.
- The inference is based on resampling methods, both permutations (Ge et al., 2003) and the Significance Analysis of Microarrays (SAM, Tusher et al., 2001).
- To control the False Discovery Rate (FDR) the Benjamini Hochberg (BH) procedure is implemented.
- Available at: <http://cran.ii.uib.no/web/packages/IsoGene/>

IsoGeneGUI Package

- The interface of the IsoGene package.
- A graphical user interface based on R-tcl/tk package (Dalgaard, 2001).
- A menu based package and data analysis can be performed simply by selecting options from the menus of the package
- Produces some default graphical displays as well as user-defined graphical output.

Test Statistics for Trend Test

Test statistic	Formula
Likelihood Ratio Test (LRT)	$\bar{E}_{01}^2 = \frac{\sum_{ij}(y_{ij}-\hat{\mu})^2 - \sum_{ij}(y_{ij}-\hat{\mu}_i^*)^2}{\sum_{ij}(y_{ij}-\hat{\mu})^2}$
Williams	$t = (\hat{\mu}_K^* - \bar{y}_0) / \sqrt{2 \times \sum_{i=0}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2 / (n_i(n-K))}$
Marcus	$t = (\hat{\mu}_K^* - \hat{\mu}_0^*) / \sqrt{2 \times \sum_{i=0}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i)^2 / (n_i(n-K))}$
M	$M = (\hat{\mu}_K^* - \hat{\mu}_0^*) / \sqrt{\sum_{i=0}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i^*)^2 / (n-K)}$
Modified M (M')	$M' = (\hat{\mu}_K^* - \hat{\mu}_0^*) \sqrt{\sum_{i=0}^K \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i^*)^2 / (n-l)}$

Analysis

Three analyses can be performed in the package:

- \bar{E}_{01}^2 using its null distribution.
- The five test statistics using permutations.
- The Significance Analysis of Microarrays.

Likelihood ratio test (\bar{E}^2)

The null distribution of \bar{E}^2 test statistic is given by

$$P[\bar{E}_{01}^2(v) \geq c] = \sum_{l=1}^k P(l, k; \mathbf{w}) P[B_{(l-1)/2, (v+k-l)/2} \geq c], \quad (1)$$

where $l = 1, 2, \dots, k$, $v = N - k$, $B_{a,b}$ is a beta variable with parameters a and b , \mathbf{w} is the weight vector through the level probabilities $P(l, k; \mathbf{w})$ as defined in Robertson et al. 1988.

Permutation

By permuting the labels of arrays randomly, the permutation test statistics of m genes are re-calculated and obtain the permutation matrix T :

$$T = \begin{pmatrix} t_{11} & t_{11} & \dots & t_{1B} \\ t_{21} & t_{22} & \dots & t_{2B} \\ \vdots & & \vdots & \\ t_{m1} & t_{m2} & \dots & t_{mB} \end{pmatrix} \quad (2)$$

where B is the number of permutations and each element $t_{i,b}$ of matrix T is the test statistic for the i -th gene in the b -th permutation.

The raw p -values:

$$P_i = \frac{\#(b : |t_{i,b}| \geq |t_i|)}{B - 1} \quad (3)$$

where t_i is the observed test statistic for gene i (Ge et al. 2003).

Significance Analysis of Dose-response Microarray Data (SAM, Tusher et al., 2001)

- The generic algorithm of SAM discussed by Chu et al. (2001) is implemented in this package.
- The SAM regularized test statistic:
 - For the t -type test statistics (i.e., Williams, Marcus, the M , and the M'), a fudge factor is added in the standard error of the mean difference. For example:

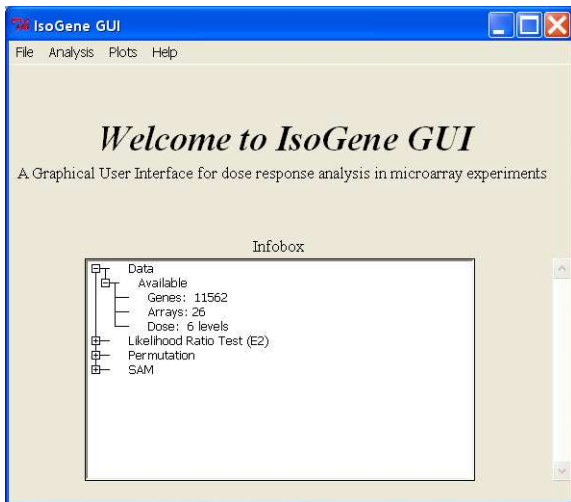
$$M'^{SAM} = \frac{\hat{\mu}_K^* - \hat{\mu}_0^*}{\tilde{S}' + s_0}, \quad (4)$$

where s_0 is the fudge factor.

- For the F -type test statistic, such as \bar{E}_{01}^2 , is defined by

$$\bar{E}_{01}^{2SAM} = \frac{\sqrt{\hat{\sigma}_{H_0}^2 - \hat{\sigma}_{H_1}^2}}{\sqrt{\hat{\sigma}_{H_0}^2 + s_0}}. \quad (5)$$

The IsoGeneGUI



The IsoGeneGUI: Menu

- File

This menu is used for loading and showing the data.

Sub-menus:

- Open Data
 - R workspace (*.RData files)
 - Excel or text file (*.xls or *.txt files)
- Show Data
- Exit

- Analysis

This is the main menu where the analyses are performed.

Sub-menus:

- Set seed
- Likelihood Ratio Test E^2 Analysis
- Permutation Analysis
- Significant Analysis of Microarrays (SAM)
 - SAM Permutation
 - SAM Analysis

The IsoGeneGUI: Menu

- Plots

Besides plots that can be produced in each the analysis dialog box, the package also provides extra plots:

- IsoPlot
- Permutation Plot
- SAM Plot
 - Plot of FDR vs. Delta
 - Plot of number of significant genes vs. Delta
 - Plot of number of False Positive vs. Delta
- User defined scatter plot

- Help

- IsoGene Help
- IsoGeneGUI Manual and Examples
- About

Data Exploration

The IsoPlot dialog box

Isotonic Regression Plot

Insert Gene name/ row name you want to draw:

Gene Name(s), e.g., gene1,gene3

Row Number(s),e.g., 2,3,4

Range of Row Number From: To:

Dose as ordinal

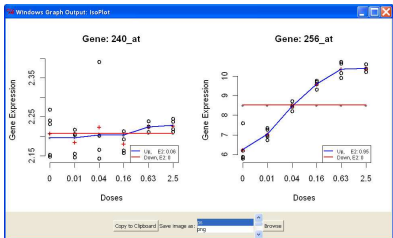
Show isotonic regression curve for both directions

Show summary of the data

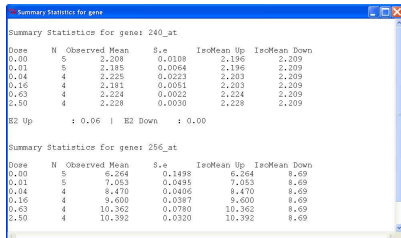
OK Cancel

Data Exploration

The Isotonic regression plots



The summary statistic



Likelihood Ratio Test E^2 Analysis

The dialog box

Global Likelihood Ratio Test (E^2) Analysis

Calculate the p-values for E^2

Select the genes:

All Genes

Genes Range From To

P-value adjustment:

BH BY SidakSS SidakSD

Holm Hochberg Bonferroni

Overall Significant Level:

Display the significant genes

Plots:

Plot E^2 values vs raw p-values

Plot Fold change vs E^2

Ranking Plot (Adjust the p-values first)

Likelihood Ratio Test E^2 Analysis

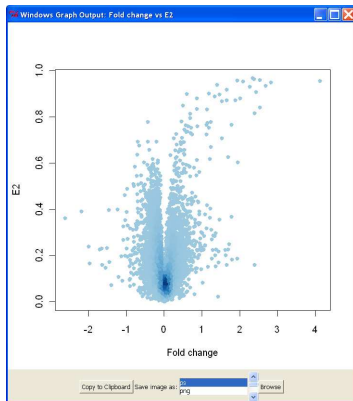
Output: List significant genes

	ProbeID	row.num	Mu, Diff	Direction	E2Val	pval.Asymp	BH.p.val	EY.p.val
1	6415_at	6215	2.34951303	u	0.9654130	2.378309e-16	0.000000	0.000000
2	1803_at	1603	2.38403091	u	0.9628712	4.870726e-16	0.000000	0.000000
3	4242_at	4042	2.53672430	u	0.9584245	1.623438e-15	0.000000	0.000000
4	4833_at	4633	1.93891924	u	0.9548713	3.818261e-15	0.000000	0.000000
5	256_at	56	4.12847733	u	0.9528423	6.030189e-15	0.000000	0.000000
6	380_at	180	2.09541507	u	0.9507085	9.554572e-15	0.000000	0.000000
7	1360_at	1160	2.82725604	u	0.9468426	2.105868e-14	0.000000	0.000000
8	3941_at	3741	1.40675989	u	0.9381753	1.018548e-13	0.000000	0.000000
9	924_at	724	2.69715769	u	0.9339285	2.039269e-13	0.000000	0.000000
10	4848_at	4648	2.27824021	u	0.9285061	4.653657e-13	0.000000	0.000000
11	3746_at	3546	1.65736950	u	0.9163942	2.397639e-12	0.000000	0.000000
12	896_at	696	2.44329292	u	0.9064505	7.796396e-12	0.000000	0.000000
13	7541_at	7341	1.19362811	u	0.8988077	1.778998e-11	0.000000	0.000000
14	5099_at	4899	0.60531374	u	0.8968981	2.165298e-11	0.000000	0.000000
15	1359_at	1159	1.47442793	u	0.8956025	2.469078e-11	0.000000	0.000000
16	1568_at	1368	1.34435519	u	0.8878167	5.261434e-11	0.000000	0.000000
17	1269_at	1069	0.86672056	u	0.8798123	1.086907e-10	0.000000	0.000001
18	9835_at	9635	2.01765002	u	0.8787428	1.193201e-10	0.000000	0.000001
19	8754_at	8554	1.88526389	u	0.8713131	2.232660e-10	0.000000	0.000001
20	9931_at	9731	1.66185945	u	0.8711961	2.254151e-10	0.000000	0.000001
21	11124_at	10924	1.53998549	u	0.8665769	3.268060e-10	0.000000	0.000002
22	11448_at	11248	0.50159972	u	0.8486610	1.234955e-09	0.000001	0.000006
23	6992_at	6792	2.54487370	u	0.8386370	2.431316e-09	0.000001	0.000012
24	6922_at	6722	0.86925494	u	0.8334940	3.387152e-09	0.000002	0.000016
25	9421_at	9221	1.13472792	u	0.8305570	4.074771e-09	0.000002	0.000019
26	7917_at	7717	2.40088293	u	0.8139082	1.098002e-08	0.000005	0.000048
27	283_at	83	0.69466377	u	0.8115807	1.252357e-08	0.000005	0.000053
28	5684_at	5484	0.69700485	u	0.7937990	3.253788e-08	0.000013	0.000133
29	10058_at	9858	0.46128782	u	0.7795599	6.601657e-08	0.000026	0.000261
30	477_at	277	1.24433243	u	0.7778047	7.180537e-08	0.000028	0.000275

Can be saved as R object or/and an excel file.

Likelihood Ratio Test E^2 Analysis

Output: Fold change vs. E^2 plot



Can be copied into clipboard and be saved as different image types.

Permutation

Permutation Analysis - IsoGene GUI

New Permutation

Select the genes:

All Genes

Genes Range From To

Number of Permutation:

Save the permutations result:

Obtain the raw p-value from a file:

Statistic:

M E2 Marcus Williams M'

Controlling FDR Procedure:

BH BY

Controlling FWER Procedure:

SidakSS SidakSD Holm Hochberg Bonferroni

Overall Significant Level:

Display the significant genes

Plots:

Specify the statistic that you want to plot:

M E2 Marcus Williams M'

Specify the plot(s):

Plot of the specified statistic vs its raw p-values

Plot of fold change vs the specified statistic

Rank plot (Adjust the p-value first)

Permutation

The output:

- Result for all genes.
- list of significant genes using the selected statistic(s).
- Graphical displays.

Significance Analysis of Microarrays (SAM)

The SAM permutation dialog box

Significance Analysis of Microarrays (SAM) - IsoGene GUI

New Permutation

Select the genes:

All Genes

Genes Range From To

Specify Fudge Factor (S0):

No fudge factor

Automatic

Specify percentile for S0: e.g. 5,45,100

Fudge plots:

Statistic: E2 Williams Marcus M M'

Standar error plot

CV plot

Number of Permutation:

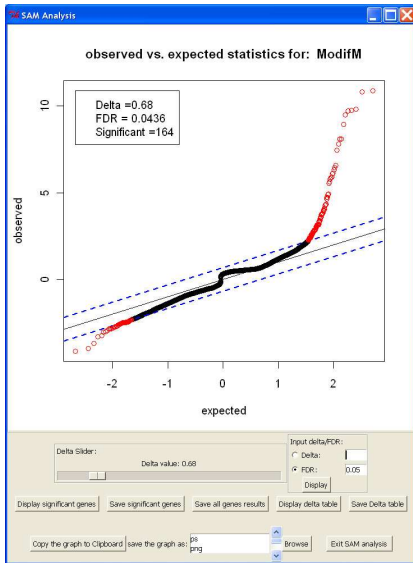
Save the permutations result: browse

Run Permutation Please wait... Performing 63 / 100 permutations

Obtain the permutation result from a file: browse

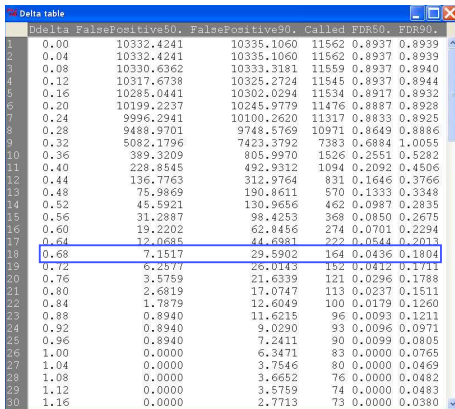
Ok Cancel

Significance Analysis of Microarrays (SAM)



Significance Analysis of Microarrays (SAM)

Output: Delta table



	Ddelta	FalsePositive50	FalsePositive90	Called	FDR50	FDR90
1	0.00	10332.4241	10335.1060	11562	0.8937	0.8939
2	0.04	10332.4241	10335.1060	11562	0.8937	0.8939
3	0.08	10330.6362	10333.3181	11559	0.8937	0.8940
4	0.12	10317.6738	10325.2724	11545	0.8937	0.8944
5	0.16	10285.0441	10302.0294	11534	0.8917	0.8932
6	0.20	10199.2237	10245.9779	11476	0.8887	0.8928
7	0.24	9996.2941	10100.2620	11317	0.8833	0.8925
8	0.28	9488.9701	9748.5769	10971	0.8649	0.8886
9	0.32	5082.1796	7423.3792	7383	0.6884	1.0055
10	0.36	389.3209	805.9970	1526	0.2551	0.5282
11	0.40	228.8545	492.9312	1094	0.2092	0.4506
12	0.44	136.7763	312.9764	831	0.1646	0.3766
13	0.48	75.9869	190.8611	570	0.1333	0.3348
14	0.52	45.5921	130.9656	462	0.0987	0.2835
15	0.56	31.2887	98.4253	368	0.0850	0.2675
16	0.60	19.2202	62.8456	274	0.0701	0.2294
17	0.64	12.0685	44.6981	222	0.0544	0.2013
18	0.68	7.1517	29.5902	164	0.0436	0.1804
19	0.72	6.2577	26.0143	152	0.0412	0.1711
20	0.76	3.5759	21.6339	121	0.0296	0.1788
21	0.80	2.6819	17.0747	113	0.0237	0.1511
22	0.84	1.7879	12.6049	100	0.0179	0.1260
23	0.88	0.8940	11.6215	96	0.0093	0.1211
24	0.92	0.8940	9.0290	93	0.0096	0.0971
25	0.96	0.8940	7.2411	90	0.0099	0.0805
26	1.00	0.0000	6.3471	83	0.0000	0.0765
27	1.04	0.0000	3.7546	80	0.0000	0.0469
28	1.08	0.0000	3.6652	76	0.0000	0.0482
29	1.12	0.0000	3.5759	74	0.0000	0.0483
30	1.16	0.0000	2.7713	73	0.0000	0.0380

Plots

Permutation p -values plot

Permutation p -values Plot

Insert Gene name/ row name you want to draw:

Gene Name(s), e.g., gene1, gene3

Row Number(s), e.g., 2,3,4

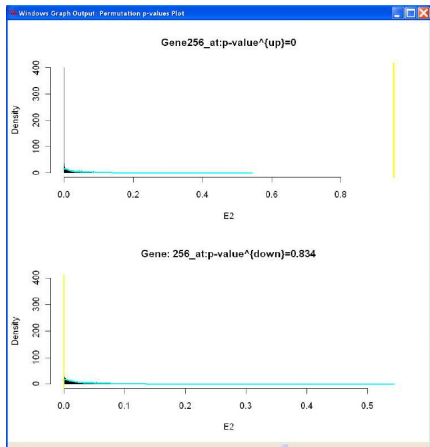
Range of Row Number From: To:

Number of Permutation:

Statistic:

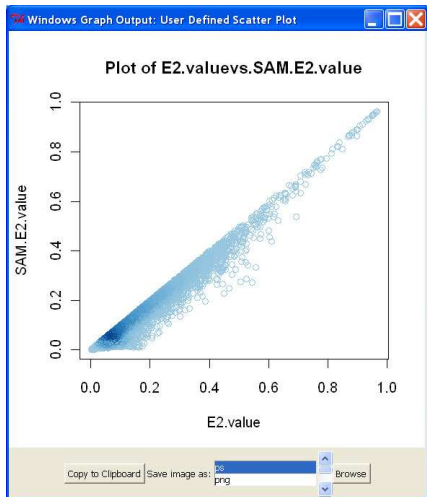
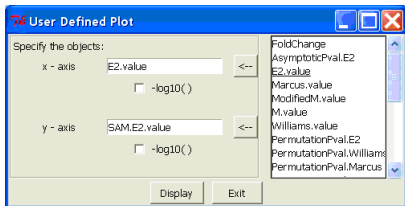
M E2 Marcus Williams M'

OK Cancel



Plots

User Defined Scatter Plot



Help

Html help with screenshots



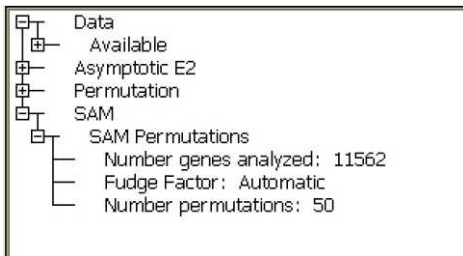
IsoGeneGUI: A Graphical User Interface for the IsoGene

User Guides

1. [About the IsoGeneGUI](#)
2. [The statistics for testing for monotonic trend](#)
3. [IsoGeneGUI input files](#)
4. [Data exploration](#)
5. Analysis:
 - a. [Likelihood ratio test statistic \(E2\)](#)
 - b. [Permutation](#)
 - c. [Significance Analysis of Microarrays \(SAM\)](#)
6. [Example Data: Dopamine Study](#)

Infobox

Provides information about the data (availability and summary) and the result summary of the last performed analyses.



Availability

- R-forge site:
<https://r-forge.r-project.org/projects/isogenegui/>
- To install: `install.packages("IsoGeneGUI",
repos="http://R-Forge.R-project.org")`
- The full user manual and example data will be soon
available at: <http://www.censtat.uhasselt.be/software/>
- Plan to be submitted to Bioconductor

Selected References

- Barlow, R.E., Bartholomew, D.J., Bremner, M.J. and Brunk, H.D. (1972) *Statistical Inference Under Order Restriction*, New York: Wiley.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing, *J. R. Statist. Soc. B*, **57**, 289-300.
- Lin, Dan, Shkedy, Ziv, Yekutieli, Dani, Burzykowski, Tomasz, Göhlmann, Hinrich, De Bondt, An, Perera, Tim, Geerts, Tamara and Bijens, Luc.(2007) Testing for Trends in Dose-Response Microarray Experiments: A Comparison of Several Testing Procedures, Multiplicity and Resampling-Based Inference, *Statistical Applications in Genetics and Molecular Biology: Vol. 6 : Iss. 1, Article 26*.
- Robertson, T., Wright, F.T. and Dykstra, R.L. (1988), *Order Restricted Statistical Inference*, Wiley.

Thank you for your attention...

Demo??