

‘maGUI’ – a Graphical User Interface for Analysis and Annotation of DNA Microarray Data

Dhammapal Bharne, Vaibhav Vindal

January 1, 2017

CONTENTS

1. Introduction	1
1.1. Types of micro array data	2
1.2. Why this package	2
2. Usage Guide	2
2.1. Loading	2
2.2. Analysis	3
2.3. Miscellaneous	3
2.4. Others	3
3. What was used to create this document	3
4. REFERENCES	6

1. INTRODUCTION

For the analysis of different types of microarray data, different packages are available in the Bioconductor software[1]. It is very challenging for the users without prior programming experience to use R command line environment[2] and Bioconductor packages. This package allows biologists to perform the most

microarray data analyses and annotations through GUI[3][4] without the need of learning any programming language.

1.1 Types of micro array data

NCBI GEO[5] contains large amount of microarray data produced by various manufacturers like Affymetrix, Agilent, Illumina, Nimblegen and so on. Though, different Bioconductor packages are needed to pre-processes microarray data produced by different manufacturers, it follows the common procedures for analysis like filtering, clustering, classification and so on.

1.2 Why this package

Once the data is loaded successfully, a single click makes pre-processing and analyzing of the microarray data automatically. Therefore, analyzing microarray data produced by Affymetrix, Agilent, Illumina and Nimblegen manufacturers is just few clicks away. Series matrix and platforms files can also be used either online or offline for the microarray data analysis.

2. USAGE GUIDE

To execute maGUI package, type `maGUI:::maGUI()` at the R terminal.

2.1 Loading

Loading of Affymetrix data requires raw files with .CEL extension. Loading of Agilent-one color data requires raw files with foreground mean signal and background median signal values while loading of Agilent-two color data requires raw files with Agilent source. Loading of Nimblegen data require raw files with .xys or .pair extensions. Non-normalized files are used to load both Illumina beadarray and lumi data. Series matrix file along-with platform file is required to load Series Matrix data. GSE number is required to load On-line data.

This package use ReadAffy function from affy package to load CEL files of Affymetrix platform. It uses read.maimages function from limma package to load Agilent one and two color data. Agilent one color data should have foreground mean and background median signal values while Agilent-two color data with Agilent source. Loading of Illumina beadarray and lumi data require non-normalized files while Xys or pair extension files are used to load Nimblegen data. The user can also use series matrix and soft files of an experiment. Further, user can employ online method where only the experiment GSE number is required for the analysis.

2.2 Analysis

In the beginning of the analysis, the loaded microarray data is pre-processed through log-transformation, imputation, normalization and quality check methods. The pre-processed data can be used for differential gene expressions, principal component analysis, clustering and classification of the data.

2.2 Miscellaneous

Miscellaneous applications such as gene set enrichment analysis, test analysis, and identifying gene symbols requires Bioconductor annotation package information which can be retrieved from GEO metadata SQLite database or from database information attached with this GUI. Further they require internet connection if the required database packages are not pre-installed. Other miscellaneous applications include sample size estimation and building co-expression network which do not require the internet connection. Using this GUI, protein-protein associations can also be predicted through correlation of co-expressions in two normalized microarray experimental data.

2.3 Others

The tasks performed on microarray data can be cleared, saved or opened as R file. Figures and tables generated during the analysis can be exported to desired location on the hard disk.

3. WHAT WAS USED TO CREATE THIS DOCUMENT

The following R version and packages are used to generate maGUI:

```
> sessionInfo()
```

```
R version 3.1.1 (2014-07-10)
```

```
Platform: i686-pc-linux-gnu (32-bit)
```

```
locale:
```

```
[1] LC_CTYPE=en_IN    LC_NUMERIC=C      LC_TIME=en_IN
```

```
[4] LC_COLLATE=en_IN  LC_MONETARY=en_IN LC_MESSAGES=en_IN
```

```
[7] LC_PAPER=en_IN    LC_NAME=C         LC_ADDRESS=C
```

```
[10] LC_TELEPHONE=C   LC_MEASUREMENT=en_IN LC_IDENTIFICATION=C
```

```
attached base packages:
```

```
[1] grid    tcltk  stats4  parallel stats  graphics grDevices
```

```
[8] utils  datasets methods base
```

```
other attached packages:
```

```
[1] KEGG.db_3.0.0      mouse4302.db_3.0.0  org.Mm.eg.db_3.0.0
```

```
[4] mouse4302cdf_2.15.0  annotate_1.44.0     ssize_1.40.0
```

[7]	xtable_1.8-2	gdata_2.17.0	globaltest_5.20.0
[10]	GOstats_2.32.0	Category_2.32.0	GO.db_3.0.0
[13]	Matrix_1.2-6	lumi_2.18.0	beadarray_2.16.0
[16]	ggplot2_2.1.0	impute_1.40.0	simpleaffy_2.42.0
[19]	gcrma_2.38.0	genefilter_1.48.1	affy_1.44.0
[22]	KEGGgraph_1.24.0	XML_3.98-1.1	KEGGREST_1.6.4
[25]	Rgraphviz_2.10.0	WGCNA_1.51	AnnotationDbi_1.28.2
[28]	GenomeInfoDb_1.2.5	fastcluster_1.1.20	dynamicTreeCut_1.63-1
[31]	RBGL_1.42.0	graph_1.44.1	gWidgetsRGtk2_0.0-83
[34]	cairoDevice_2.23	gWidgets_0.0-54	RGtk2_2.20.31
[37]	amap_0.8-14	GEOmetadb_1.26.1	GEOquery_2.32.0
[40]	convert_1.42.0	marray_1.44.0	limma_3.22.7
[43]	pdInfoBuilder_1.30.6	oligo_1.30.0	oligoClasses_1.28.0
[46]	affxparser_1.38.0	RSQLite_1.0.0	DBI_0.5
[49]	BiocInstaller_1.16.5	Biostrings_2.34.1	XVector_0.6.0
[52]	IRanges_2.0.1	S4Vectors_0.4.0	Biobase_2.26.0
[55]	BiocGenerics_0.12.1		

loaded via a namespace (and not attached):

[1]	acepack_1.3-3.3	affyio_1.34.0	AnnotationForge_1.8.2
[4]	backports_1.0.3	base64_2.0	base64enc_0.1-3

[7]	BatchJobs_1.6	BBmisc_1.10	BeadDataPackR_1.18.0
[10]	beanplot_1.2	BiocParallel_1.0.3	biomaRt_2.22.0
[13]	bit_1.1-12	bitops_1.0-6	brew_1.0-6
[16]	bumphunter_1.6.0	checkmate_1.8.1	chron_2.3-47
[19]	cluster_2.0.4	codetools_0.2-14	colorspace_1.2-6
[22]	data.table_1.9.6	digest_0.6.10	doParallel_1.0.10
[25]	doRNG_1.6	fail_1.3	ff_2.2-13
[28]	foreach_1.4.3	foreign_0.8-66	Formula_1.2-1
[31]	GenomicAlignments_1.2.2	GenomicFeatures_1.18.7	GenomicRanges_1.18.4
[34]	gridExtra_2.2.1	GSEABase_1.28.0	gtable_0.2.0
[37]	gtools_3.5.0	Hmisc_3.17-4	httr_1.2.1
[40]	illuminaio_0.8.0	iterators_1.0.8	KernSmooth_2.23-15
[43]	lattice_0.20-33	latticeExtra_0.6-28	locfit_1.5-9.1
[46]	magrittr_1.5	MASS_7.3-45	matrixStats_0.50.2
[49]	mclust_5.2	methyumi_2.12.0	mgcv_1.8-13
[52]	minfi_1.12.0	multtest_2.22.0	munsell_0.4.3
[55]	nleqslv_3.0.3	nlme_3.1-128	nnet_7.3-12
[58]	nor1mix_1.2-1	openssl_0.9.4	pkgmaker_0.22
[61]	plyr_1.8.4	png_0.1-7	preprocessCore_1.28.0
[64]	quadprog_1.5-5	R6_2.2.0	RColorBrewer_1.1-2

[67]	Rcpp_0.12.3	RCurl_1.95-4.8	registry_0.3
[70]	reshape_0.8.5	reshape2_1.4.1	rngtools_1.2.4
[73]	rpart_4.1-10	Rsamtools_1.18.3	rtracklayer_1.26.3
[76]	scales_0.4.0	sendmailR_1.2-1	siggenes_1.40.0
[79]	splines_3.1.1	stringi_1.1.1	stringr_1.0.0
[82]	survival_2.39-5	tools_3.1.1	zlibbioc_1.12.0

4. REFERENCES

1. Robert C Gentleman, Vincent J. Carey, Douglas M. Bates, et al. Bioconductor: Open soft-ware development for computational biology and bioinformatics. *Genome Biology*, 5:R80, 2004.
2. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013.
3. John Verzani. gWidgets: gWidgets API for building toolkit-independent, interactive GUIs. cran.r-project.org, 2015. URL: https://r-forge.r-project.org/R/?group_id=761
4. Michael Lawrence, John Verzani. Programming Graphical User Interfaces in R. Chapman and Hall/CRC The R Series, 2012.
5. NCBI GEO. URL: <http://www.ncbi.nlm.nih.gov/geo/>