CARMAweb users guide version 0.9.1

Johannes Rainer

October 5, 2005

Contents

List of Figures

Chapter 1

Introduction

The web application CARMAweb (Comprehensive R based Microarray Analysis web frontend) is an intuitive graphical user interface for the normalization and analysis of microarray data derived from any microarray platform. The web application is build as a multi tier application based on the Java 2 Platform Enterprise Edition using the Struts framework [1] with Java Server Pages (JSP) and Servlets for the web tier and Enterprise Java Beans (EJB) for the business tier.
Because R [9] and BioConductor [5] are used as analytical backend a wide range of approved normalization and analysis methods for all current microarray platforms are available through the web application. The cluster analysis module *GenesisWeb* uses facilities of the GenesisServer application server [12] to perform cluster analysis using a wide range of clustering algorithms.
A usermanagement guarantees password protected access to the users data and analysis results. Additionally all facilities of CARMAweb can be used without an user account, but in this case the user has to stay connected to the web application until all analysis have been completed; once logged out all data are deleted.
A typical analysis workflow in CARMAweb can be delineated as follows:

- Data upload:raw data files (e.g. Affymetrix **.CEL* files, GenePix **.gpr* files ...) can be uploaded to the web application using the *Upload area* link in the navigation tree (see figure 1.1).

- Preprocessing:using the *New Analysis* button in the button bar of the main page a new analysis can be defined. Depending on the platform used the user has to choose between a Affymetrix analysis, Two color microarray analysis, ABI microarray analysis or, if the analysis should be performed on already normalized data, Normalized data analysis (as the data is already normalized no preprocessing of the data will be performed).

- Replicate handling:the replicate handling step (see section 2.5) is part of the data preprocessing and allows to average the measured and normalized intensity values of replicated microarrays. This step can also be

skipped if no replicates are available in the microarray experiment.

- Detection of differentially expressed genes:differentially expressed genes can be detected by using simple fold change cut–offs in the Fold change analysis, or using statistical tests.

- Cluster analysis:a cluster analysis can be performed independently of the previous steps. As input file any tabulator delimited text file containing expression or regulation values can be used. A new cluster analysis can be started using the *New Cluster analysis* button in the button bar of the main page, or by selecting a file for the cluster analysis in a result folder (see chapter 4).

- GO mapping analysis:the GO analysis can be started by the *New GO mapping* button in the button bar. As input any tabulator delimited text file containing in one column the EntrezGene (formerly LocusLink) identifiers for the genes of interest is accepted.

The main web page of the CARMAweb web application can be split up into the three main parts button bar, navigation tree and center frame (see figure 1.1). The buttons from the button bar can be used to start a new microarray analysis (*New Analysis*), a new GO analysis (*New GO mapping*) or a new Cluster analysis (*New Cluster analysis*). The navigation tree in the left side of the web page can be used to navigate through the web application. It contains links to the *upload area* and to all analysis of the user. The *upload area* link can be used to upload data files to CARMAweb. Each analysis is stored on the server until the user deletes them by clicking on the trashcan button next to the link in the navigation tree. The usermanagement of CARMAweb and the user accounts allow the users to store their data on the server and to access all their analysis. Additionally the user does not have to stay on-line for the time that is needed to perform an analysis.

Figure 1.1: CARMAweb main page.

Chapter 2

Preprocessing

The preprocessing of microarray data includes the background correction, normalization of the expression values within each microarray and between the microarrays (to adjust the expression values accross microarrays). The aim of the data preprocessing step is to remove the technical variance and systematic errors without altering the biological variance within the data. Different adjustment methods have been developed in the last years for the different microarray platforms.

CARMAweb provides preprocessing methods for all different microarray platforms using facilities of the corresponding BioConductor packages. Affymetrix GeneChip data, spotted two color microarrays or ABI microarrays can be analyzed. The following sections give a brief introduction to the preprocessing steps of each of these platforms.

2.1 Preprocessing of Affymetrix GeneChip data

This section describes the pre–processing steps (background correction, normalization) for the Affymetrix GeneChip platform. CARMAweb uses the `affy` and `gcrma` packages from `BioConductor` for this purpose and allows therefore the usage of methods like the Affymetrix standard method MAS5 or more sophisticated methods like MBEI (model based expression index) [8], RMA (robust multi–array average) [7, 6] or GCRMA (robust multi–array average with correction for non specific binding according to the GC content of each probe sequence) to perform the preprocessing. A comparison of the different Affymetrix preprocessing methods is outlined in [3]. Additionally it is possible to define custom preprocessing methods by selecting different algorithms for each one of the preprocessing steps.

As input files the Affymetrix *.CEL* files are required.

The example analysis below has been performed using the Affymetrix test datasets available in CARMAweb. This test dataset consists of 6 Affymetrix GeneChips (hgu133plus2) on which samples from 3 T–ALL (acute

lymphoblastic leukemia) patients before, and after treatment for 6 to 8 hours with GC (glucocorticoids), were hybridized. The first step of the analysis consists in the selection of the files that should be analyzed (see figure 2.1).

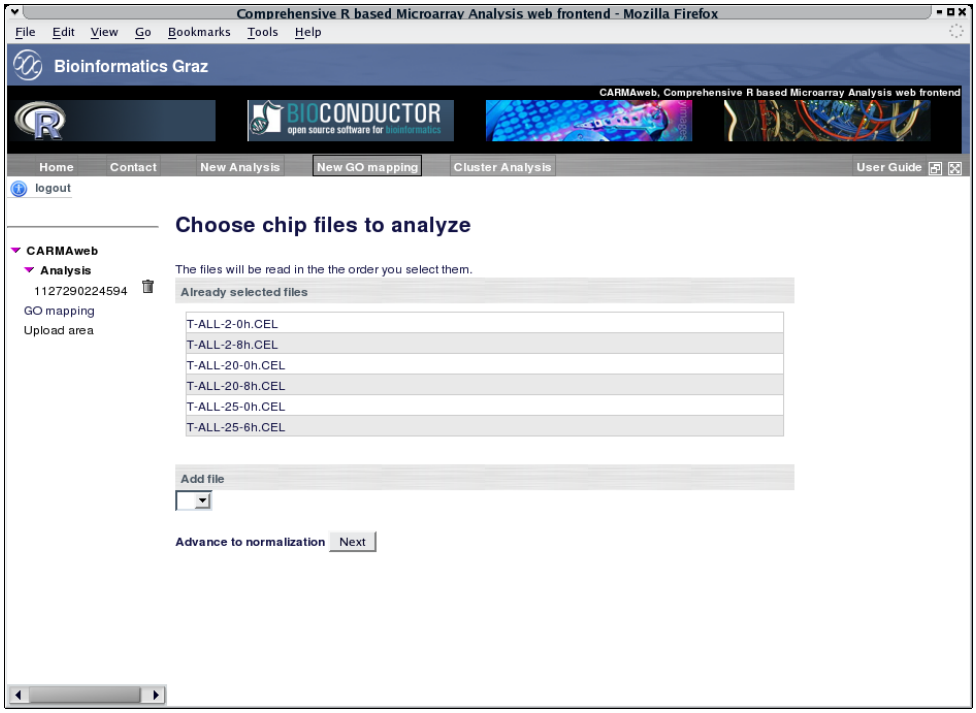

Figure 2.1: Selection for the Affymetrix GeneChip raw data files that should be analyzed.

Next the preprocessing method can be selected (figure 2.2). The Affymetrix preprocessing consists of the steps:

- Background correction
- Normalization
- Probe specific background correction
- Summarization of the probe level data

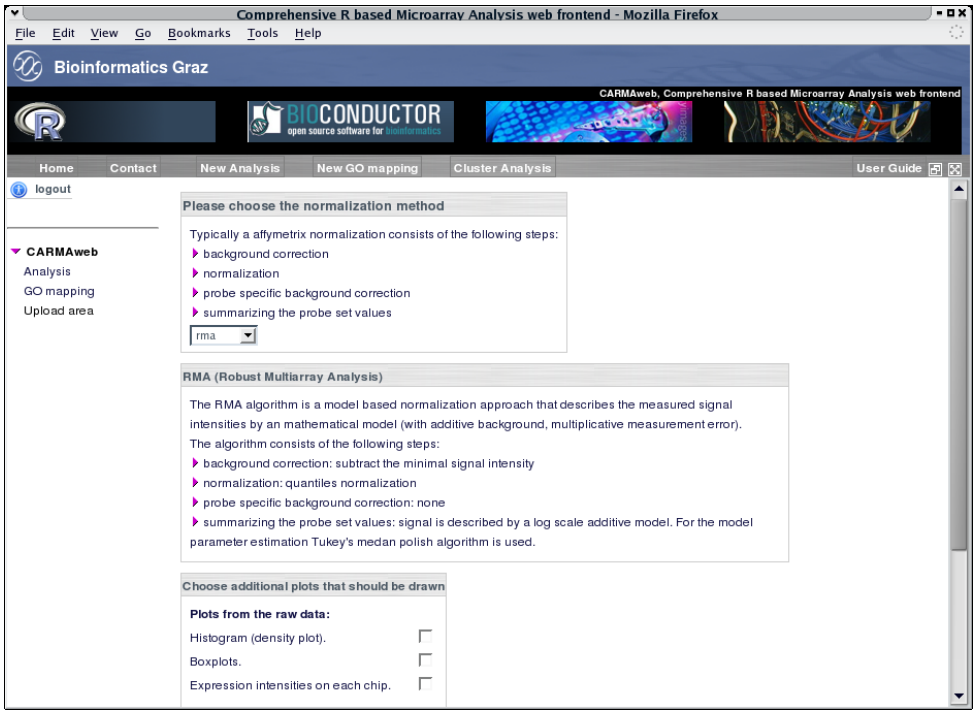

Figure 2.2: Affymetrix preprocessing.

Depending on the method used a short description for the selected processing method is displayed. By clicking on the *Analysis* button the user gets forwarded to the analysis page (figure 2.3), where he can add additional analysis steps, like detection for differentially expressed genes, to the analysis (the detection of differentially expressed genes for this example using test statistics is described in section 3.2). The analysis can be started by clicking on the *Next* button on the page (see figure 2.3).

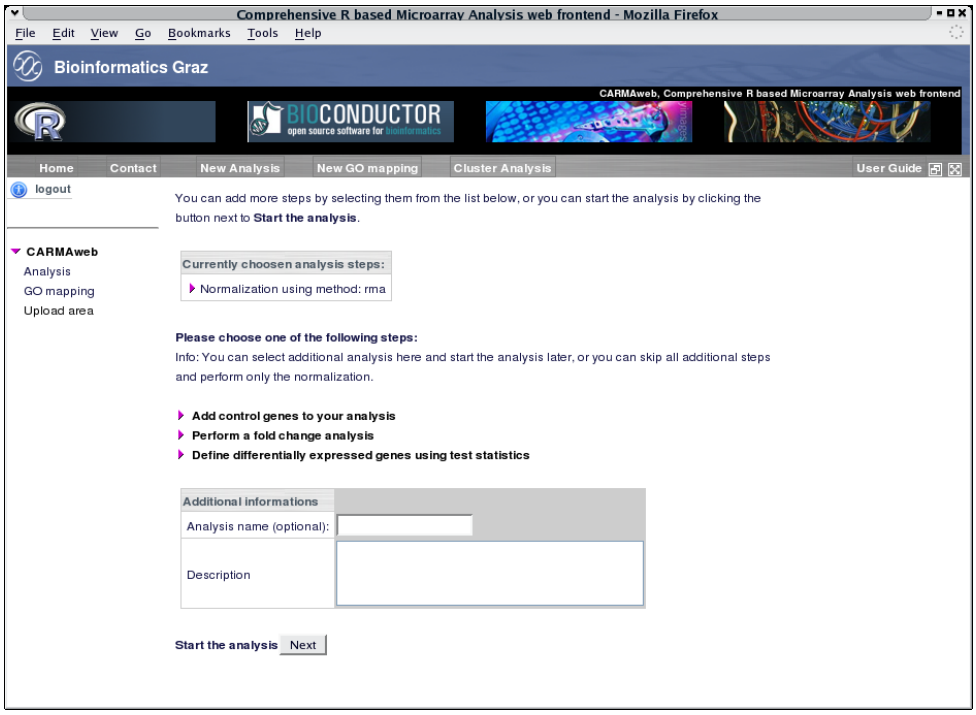

Figure 2.3: Main analysis page for Affymetrix GeneChips. Additional analysis steps can be added by clicking on the appropriate links. The analysis can be started by clicking on the *Next* button.

2.2 Preprocessing of two color microarray data

Because R and BioConductor are used as calculation backend a wide range of approved normalization and analysis methods for all current microarray platforms are available through the web application. This methods include for two color microarrays a great set of background correction methods, within array normalization methods like the loess or the print tip loess normalization, as well as a median or mean normalization, and for the between arrays normalization, that adjusts expression values across arrays, methods like the quantile or median scaling method. Most of the available preprocessing methods are outlined in Smyth and Speed [10]. The preprocessing of two color microarray data is carried out using functions from BioConductors limma package [11]. Plots and replicate handling step are performed using functions from the `maDB` package. Technically replicated arrays or replicated spots on each array can be merged by calculating the median or mean expression value across the replicates.
Genes flagged as bad spots by the scanning software can be excluded from the analysis (for Genepix users: choose *give all flagged spots a weight of 0* where *How should spots flagged by the software be handled?* is asked).
The following analysis was performed using the two color microarray test files available in the web application. The files were scanned in a Axon scanner using the GenePix software (version 5). The annotation for each spot is available in the corresponding *Batch08_modUG.gal* file. The two test files were selected for this analysis (see figure 2.4).

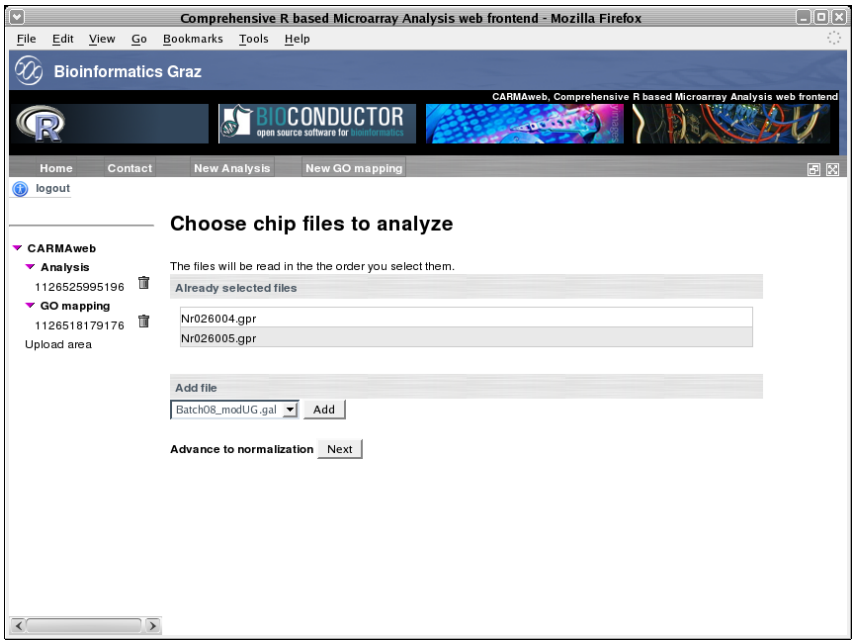

Figure 2.4: Like in the Affymetrix analysis the data files containing the raw data files can be added to the analysis using the shown page.

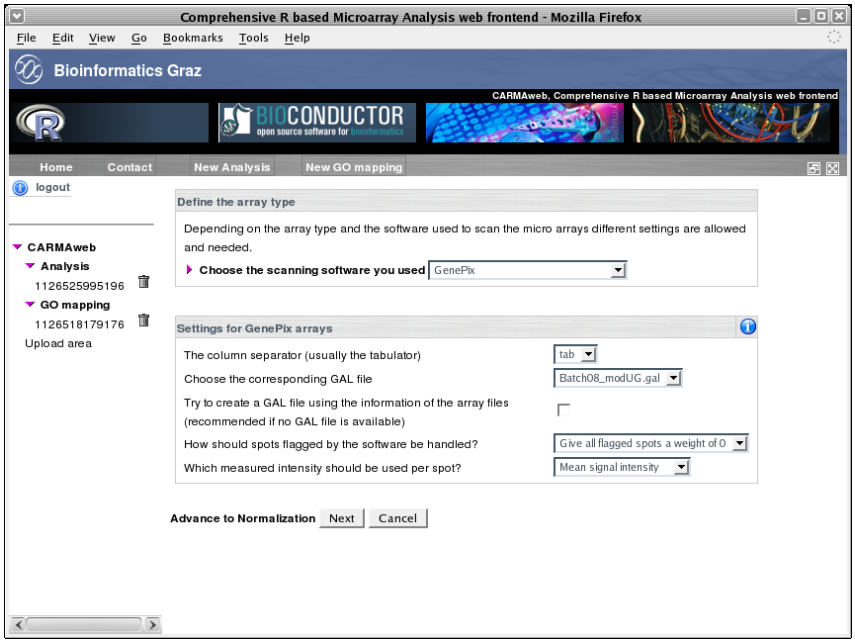

Figure 2.5: The software with which the microarrays were scanned has to be specified to allow the analysis of the raw data files. In the example above the GenePix software was used. Different parameters specific to this scanning software can be defined too on this page.

The user has to specify the software with which the microarrays where scanned (figure 2.5) and can also select different parameters specific for the selected scanning software (for the example analysis with microarrays scanned with the GenePix software a special handling of the flagged genes can be applied (discarding all spots that were flagged bad by the GenePix software). Additionally the user can select the GAL (gene array list) file that contains the annotation of the spots).

The preprocessing for two color microarrays consists of the three steps Background correction, within array normalization and between array normalization. The settings for each step can defined on different tabs of the page shown in figure 2.6. Different methods can be selected for the background correction (figure 2.6).

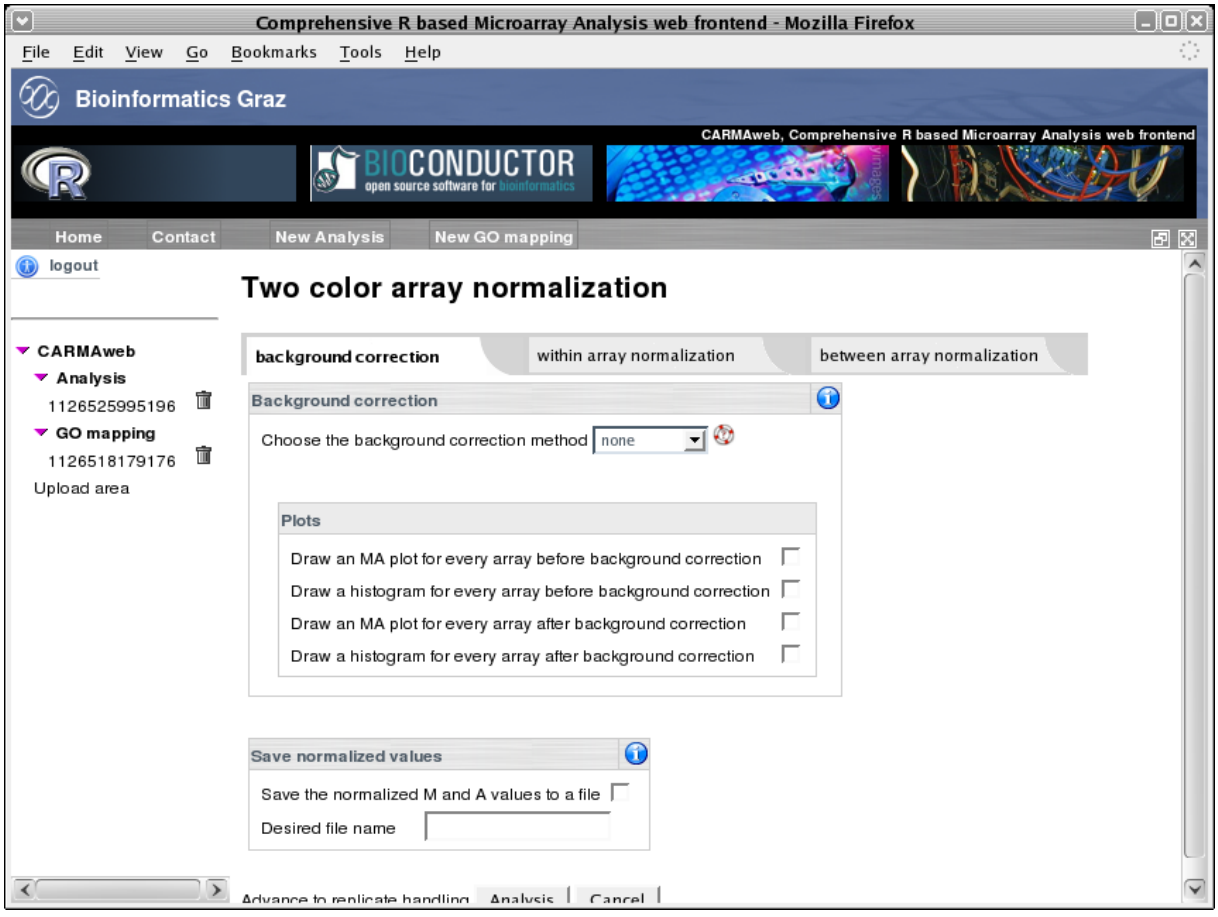

Figure 2.6: Background correction of two color microarrays.

The within array normalization corrects for the dye bias and other systematic errors. The default method in CARMAweb is the print tip loess normalization, which performs a normalization for each print tip group separately to account also for the print tip effect. The within array normalization method can be selected on the *within array normalization* tab of the two color microarray normalization page (see figure 2.7).

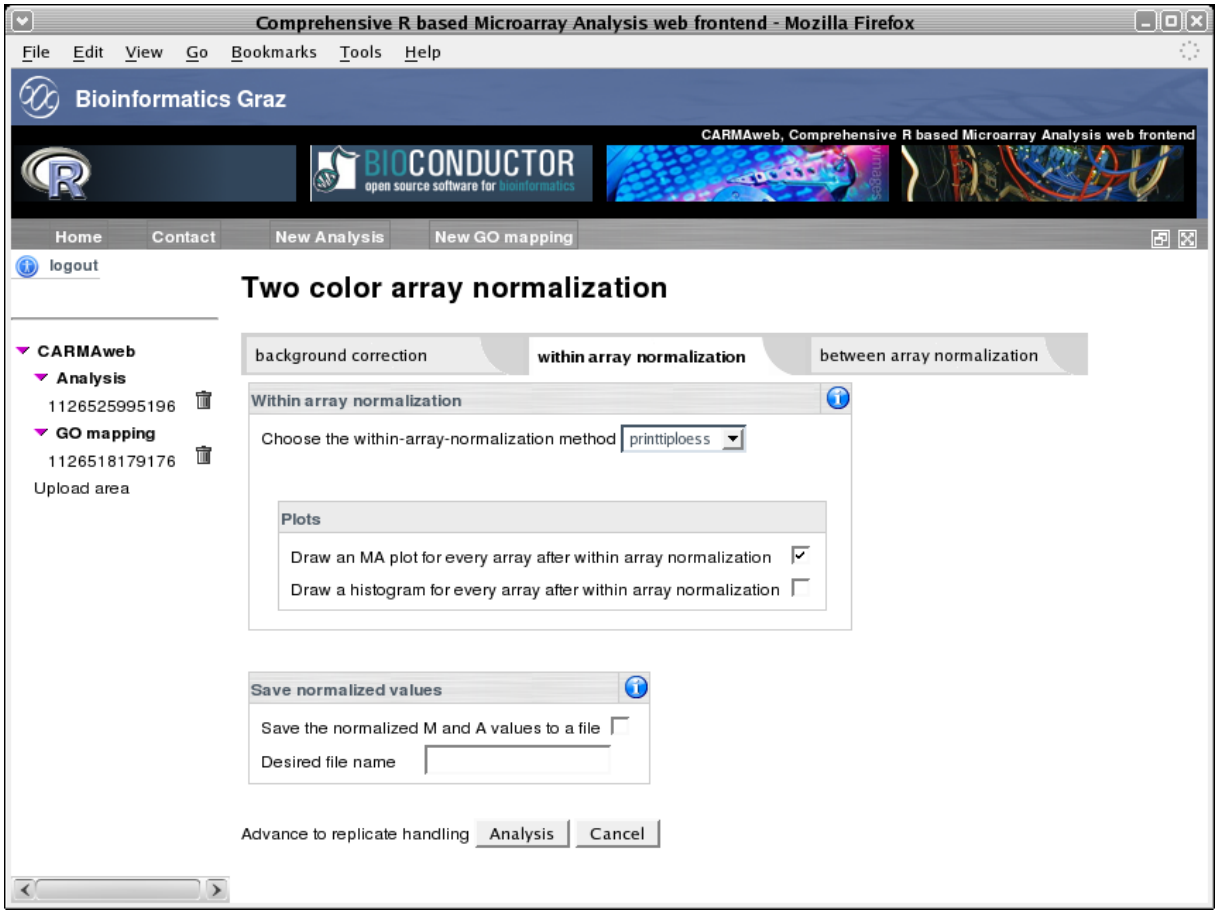

Figure 2.7: Within array normalization of two color microarrays.

The between array normalization (figure 2.8) adjusts the expression values for each spot across the microarrays of a microarray experiment. Available methods are a simple scaling method or the quantile mothod (known from the Affymetrix platform).

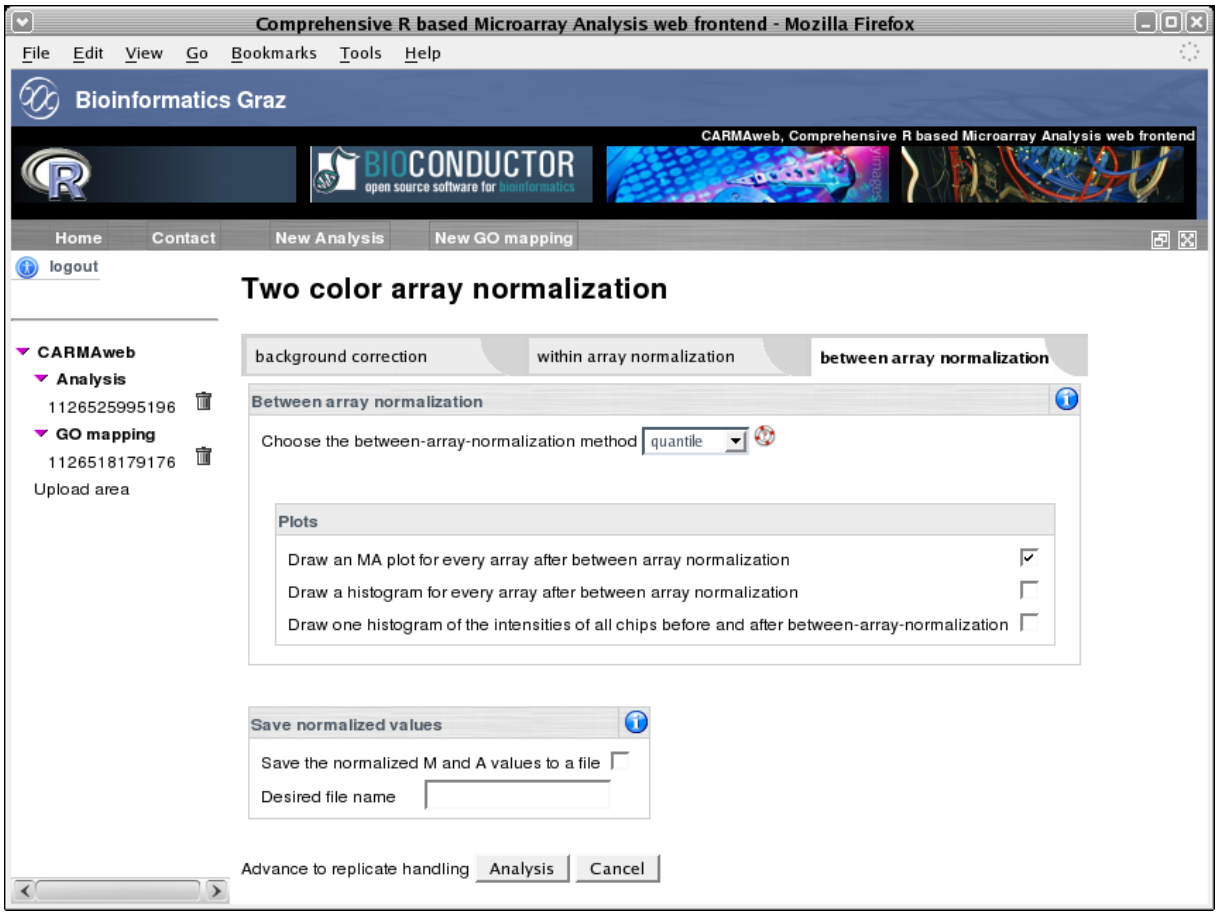

Figure 2.8: Between array normalization of two color microarrays.

By clicking on the *Analysis* button at the bottom of the page the user gets to the replicate handling page (figure 2.9), where he can define how replicated spots on an array or technically replicated microarrays should be handled. In the example all the replicated spots on each array should be averaged. Therefore the number of arrays to which the microarrays should be merged was set to 2, and each microarray was assigned to one of the so called *merged arrays*. If replicated arrays are available in the microarray experiment (technical replicates), these arrays can be merged, by assigning the replicated microarrays to one *merged array*.

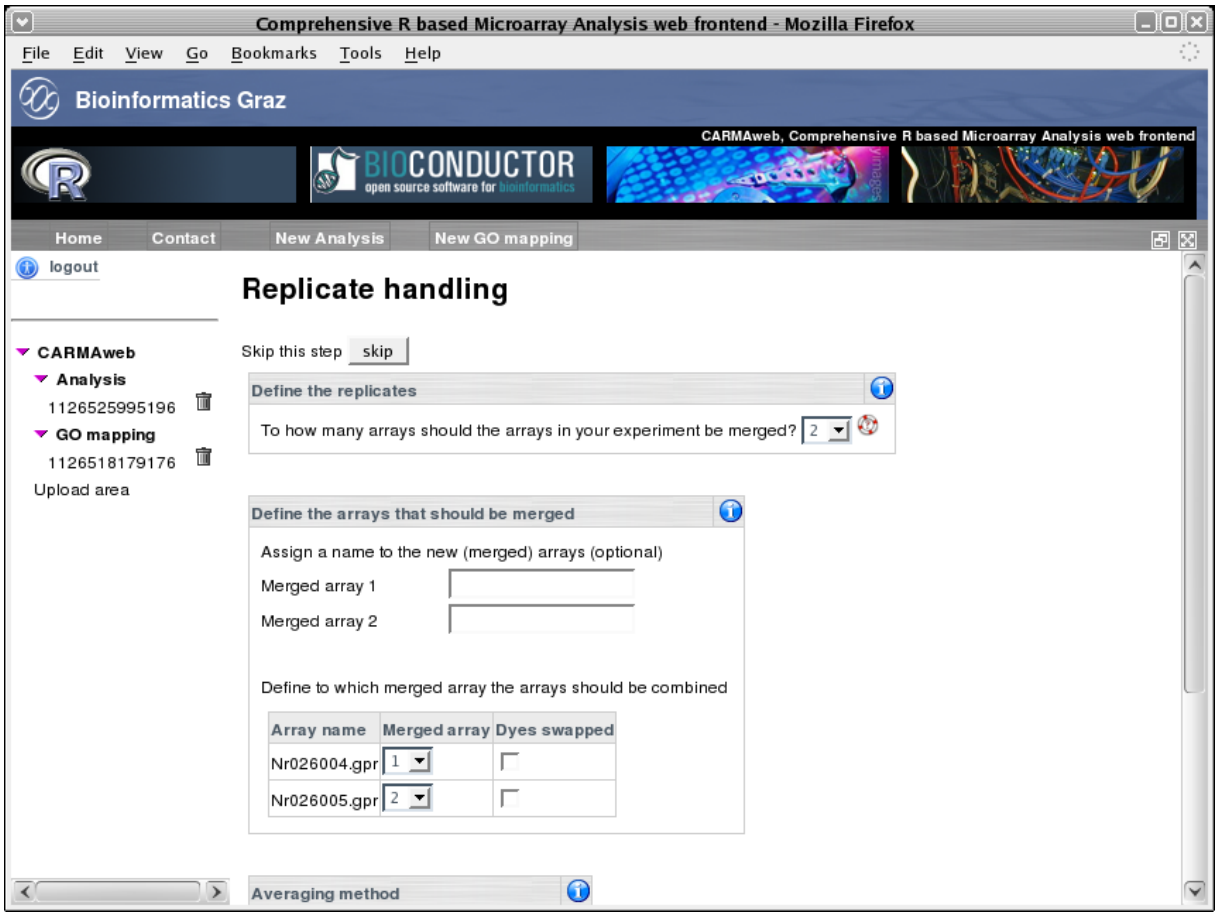

Figure 2.9: Replicate handling. On this page replicated spots within a microarray or technical replicated microarrays can be averaged (merged).

After the replicate handling step the analysis can be started or additional steps can be added to the experiment (like the detection of differentially expressed genes). The detection of differentially expressed genes using fold change analysis for this example is described in chapter 3.1.

2.3 ABI microarray normalization

The output from ABIs microarray scanning and analysis software consists of expression values that are already background corrected using internal controls. Unlike all other platforms ABI stores these expression values from different microarrays into one single file. CARMAweb can handle these files and allows the user also to select microarrays from different input files for the further analysis. Raw (background corrected) expression values can be adjusted across all microarrays within one experiment using the quantile normalization. Alternatively between–array–normalized expression values calculated using ABIs internal scaling method can be used for further processings.

Unfortunately no test files are available for the ABI microarray platform.

2.4 Analysis using already normalized data

The detection of differentially expressed genes can also be performed directly on already normalized expression values. This can be achieved by selecting *Perform an analysis on already normalized data* after clicking the *New Analysis* button. On the next page (see figure 2.10) the file containing the normalized expression values can be selected. The user has to select the columns that contain the normalized expression values, additionally it is possible to choose the decimal point (either . or ,) that is used or to set the data to be log2 transformed. All columns that are not selected to contain expression values are used as *annotation* columns.

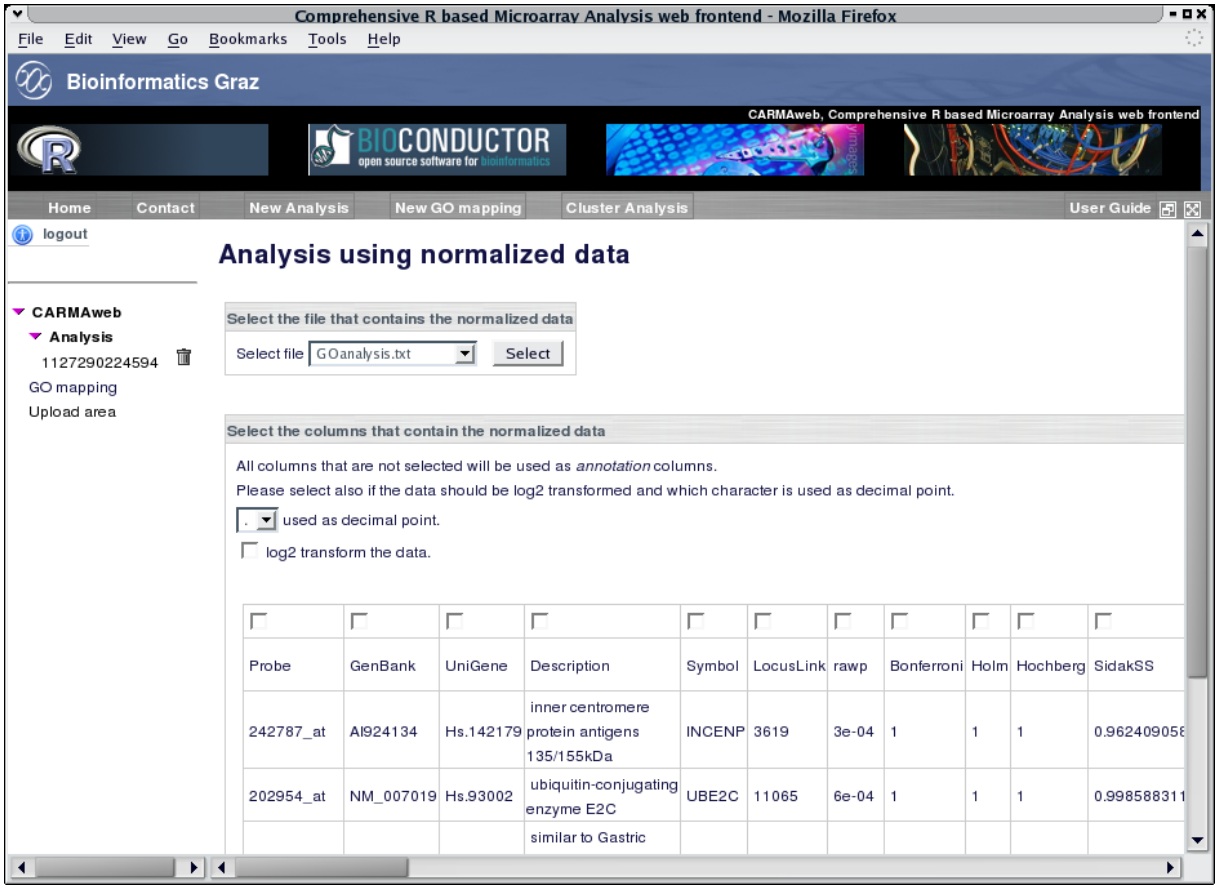

Figure 2.10: Analysis using already normalized data. After selecting the file, the first 4 rows are displayed, so that the user can select the columns containing the normalized expression values.

2.5 Replicate handling

The replicate handling step allows to average the normalized intensity (expression) values of replicated microarrays within a microarray experiment. Additionally replicated spots (genes or probe sets for Affymetrix GeneChips) can be averaged within each microarray. All spots that are flagged by the scanning software as bad spots are excluded from the replicate handling step (provided that this information was available in the input data and that the user has selected the appropriate settings to read this information from the input files (for Genepix users: choose *give all flagged spots a weight of 0* where *How should spots flagged by the software be handled?* is asked; ABI users can use the flags from the input file)).

The replicate handling step for two color microarrays is shown in figure 2.9, in figure 2.11 the settings for the replicate handling of all other platforms are shown. The user can select the number of microarrays to which the microarrays of the experiments should be merged. In the example below the expression values of the Affymetrix GeneChip *T-ALL-2-0h.CEL* are averaged with those of *T-ALL-20-0h.CEL* and the same for the to chips with the patient samples after 8 hours GC treatment. As number of *merged arrays* 2 was selected, optional it is possible to give a (unique!) name to the merged arrays. Next the microarrays of the experiment have to be assigned to the merged arrays. Additionally it is possible to average the values of replicated spots (genes) within each microarray, or to log2 transform the data before averaging them and the method to calculate the average can be selected (mean or median value).

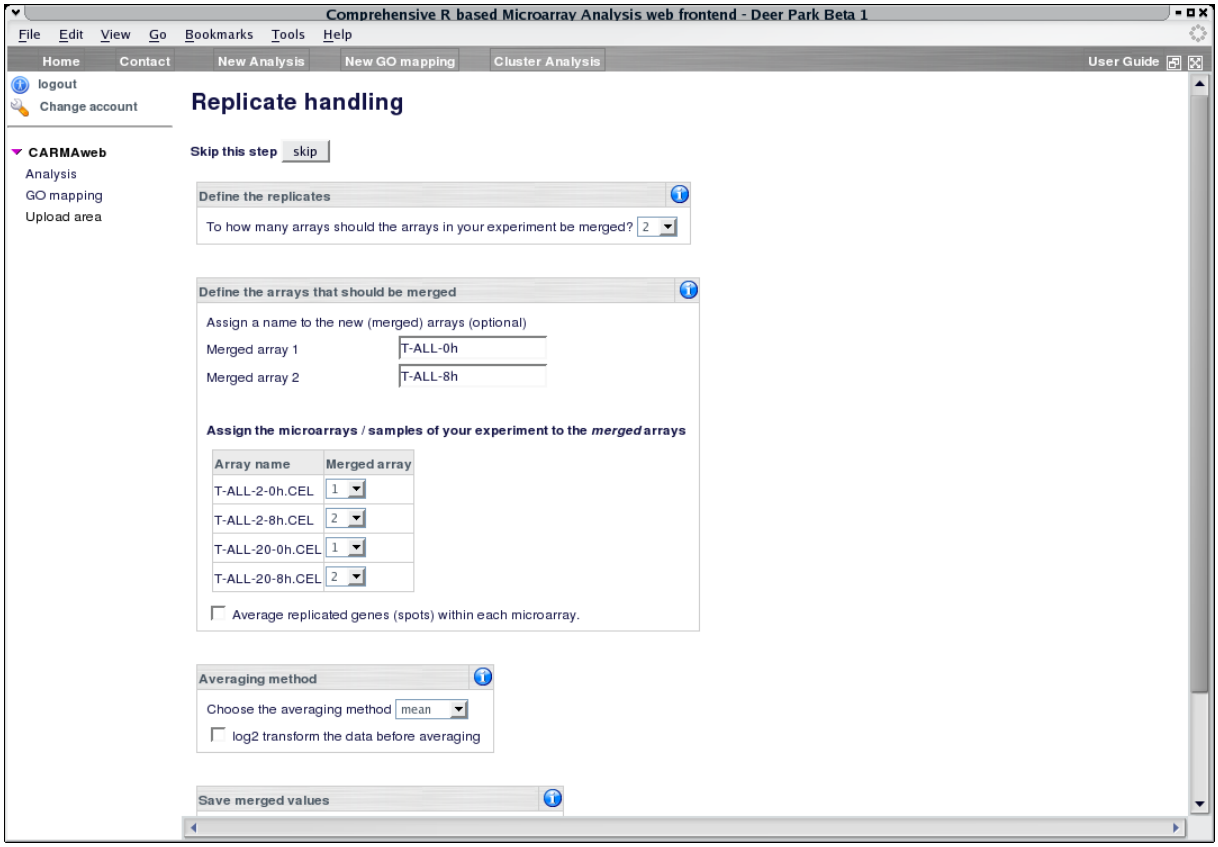

Figure 2.11: Replicated microarrays / samples can be averaged (merged) in this page.

Chapter 3

Detection of differentially expressed genes

The two methods for the detection of differentially expressed genes described in the following sections can be applied to the normalized expression values of all three supported platforms (Affymetrix, two color microarrays, ABI microarrays).

3.1 Fold change analysis

For microarray experiments with a small number of arrays and thus samples, the detection for differentially expressed genes using a simple fold change analysis is possible. Differentially expressed genes are defined in this context as those genes that show a bigger increase or decrease in their expression, comparing the expression values of the sample to the control sample, than a user defined value. CARMAweb allows also to rank genes according to the number of comparisons in which they were defined to be differentially expressed.
CARMAweb allows to set cut–off values for both the M (regulation) and the A (average expression) value. In the example a M cut–of value of 1 was selected. Therefore all genes that show a two fold up– or down–regulation are defined to be differentially expressed.
The web page where comparisons can be defined is shown in figure 3.1. The number of comparisons for the microarray analysis has to be defined, then the sample (signal channel) and the control sample can be selected for each comparison. Genes that have a bigger expression value in the sample compared to the control sample lead to positive M values whereas negative M values derive from genes that are higher expressed in the control compared to the sample. Other parameters for a comparisons are the M and A value cut–off and the filename for the data table containing the regulated genes from the comparisons. Additionally a MA plot can be drawn for every comparison.

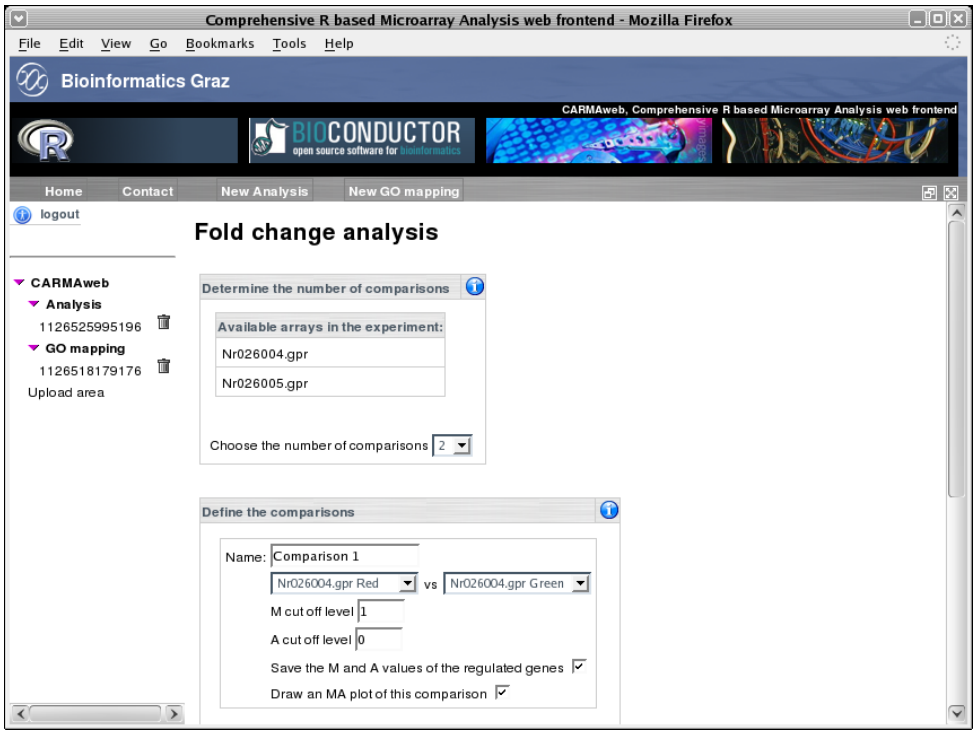

Figure 3.1: Foldchange analysis.

3.2 Test statistics

For microarray experiments with an appropriate number of arrays differentially expressed genes can be detected using statistical tests like the Mann Withney U test, the Students t test or the permutation (randomization) test (Figure 1). Microarray experiments generate large multiplicity problems in which thousands of hypothesis are tested simultaneously within one experiment [4] (e.g. is gene x differentially expressed between a treated and untreated sample). BioConductors multtest package provides suitable methods to adjust p values according to this multiple hypothesis testing problem. Available adjustment methods are for example the procedure introduced by Benjamini and Hochberg [2] for strong control of the FDR (false discovery rate, expected proportion of false positives among the rejected hypotheses) or the procedure by Westfall and Young [13] to control the FWER (family–wise error rate, probability of at least one false positive). To alleviate the loss of power from the formidable multiplicity of gene-by-gene hypothesis testing within a microarray experiment, a non–specific prefiltering of the data can also be performed in CARMAweb. Non–specific means without reference to the group the samples are into. The aim of this prefiltering step is to remove from consideration that set of genes that is not differentially expressed under any comparison. The web application allows the prefiltering of the data based on the variance within the data.

The example described in this section uses the normalized data from 6 Affymetrix GeneChips (two groups, treated and untreated with each 3 samples per group (usually bigger sample sizes are required)). First the sample

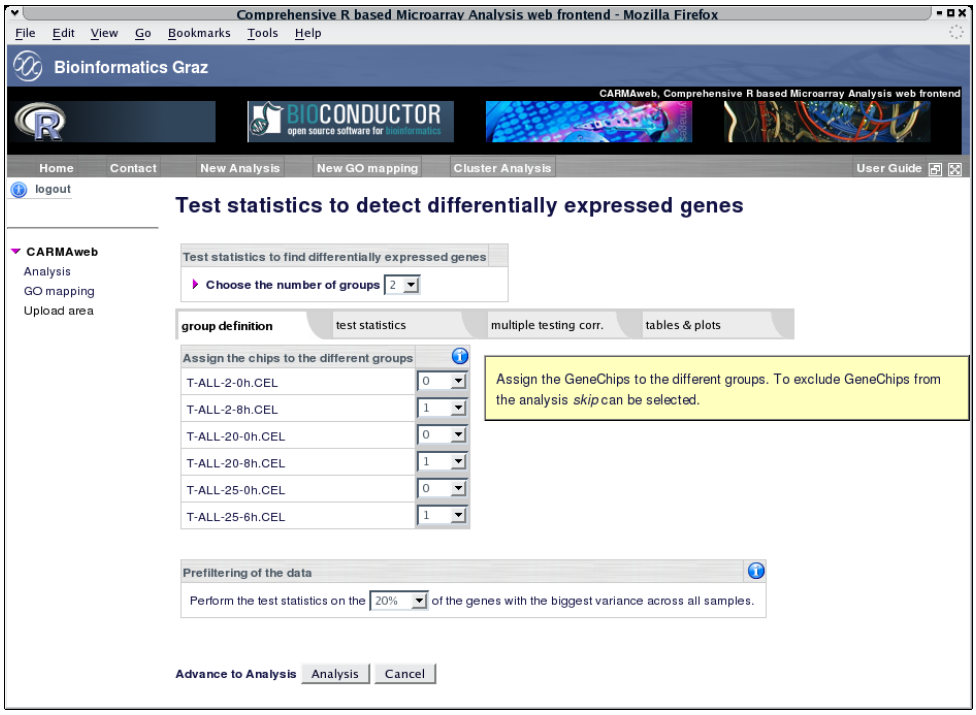

Figure 3.2: Main page for the `multtest` package. Samples can be assigned to the different groups and the data set can be restricted to the x% of genes with the biggest variance. Additionally a information pop–up is shown. Such information pop–up windows are displayed when the user moves the mouse over an information button.

groups have to be defined (figure 3.2). To alleviate the loss of power from the formidable multiplicity of gene-by-gene hypothesis testing a non-specific prefiltering should be carried out. This can be done by restricting the analysis for example to those genes that have 20% of the biggest variance over all samples. The different settings for the `multtest` package can be chosen in the different tabs of the main `multtest` page (figures 3.2 to 3.5). The test statistic can be selected on the page displayed in figure 3.3 (*test statistics* tab). By default a permutation based t test is selected. Other tests that can be applied are the t–test and the wilcoxon test.

The method (or methods) that should be used to adjust for the multiple testing within the experiment can be selected on the page shown in figure 3.4.

Finally the plots and tables that should be created can be defined on the 4th tab (*tables & plots*) shown in figure 3.5. Each analysis creates additionally to the selected tables and plots a tab delimited text file containing the p values calculated for each gene in the data set.

The detection of differentially expressed genes analysis can be added to the analysis by clicking on the *Analysis* button at the bottom of the page. CARMAweb allows to add more *detection–of–differentially–expressed–genes–steps* to one microarray analsis.

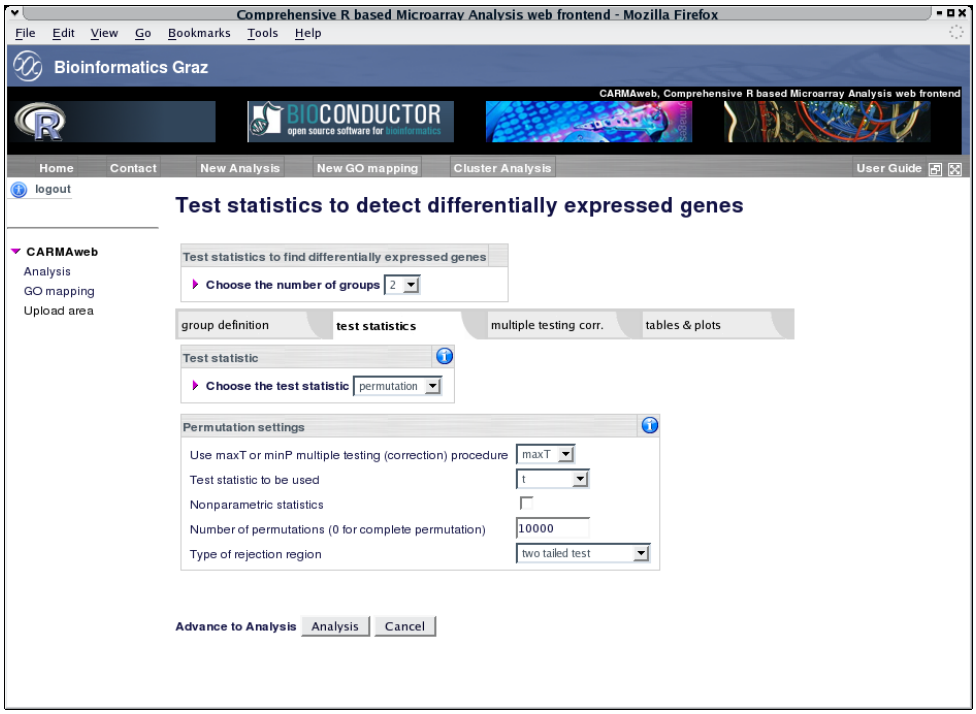

Figure 3.3: *test statistics* tab. The test statistic that should be used to detect the differentially expressed genes can be selected on this page.

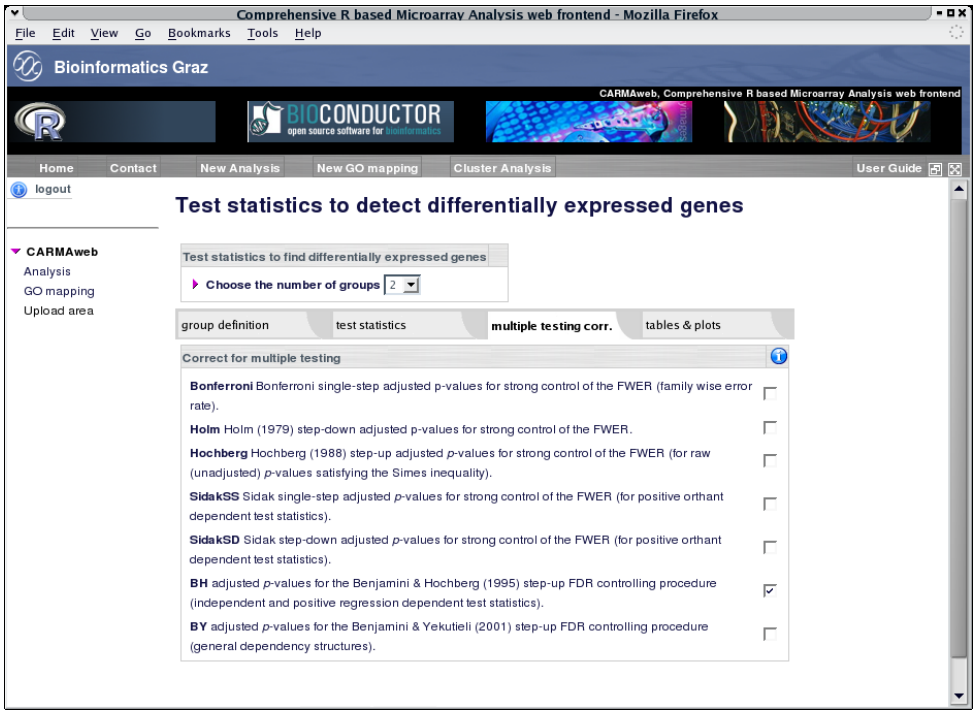

Figure 3.4: *multiple testing correction* tab. The method (or methods) that should be used to adjust for the multiple testing problem can be selected on this page.

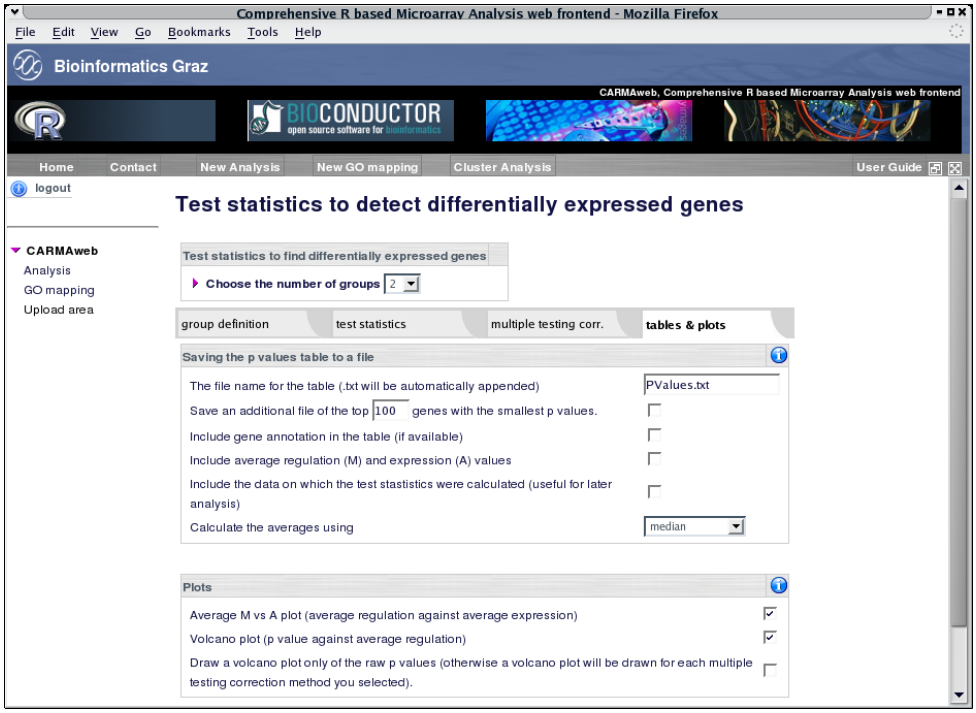

Figure 3.5: *plots & tables* tab. The plots and tables that should be created from the data can be defined using this page.

Chapter 4

Analysis result folder

Once the calculation of an analysis is finished its results are accessible through the *R Analysis workspace* page (figure 4.1). The content of this folder can be downloaded as a zip archive, or each single file can be downloaded separately (by clicking on the  icon next to the file). Additionally single files can be copied to the *upload area* using the  icon or be deleted by clicking on the  icon or directly sent to the cluster analysis module GenesisWeb using the icon.

For each analysis a report file is generated, in which all performed analysis steps and the descriptions for the various methods used in the analysis are delineated, thus guarantee a maximal transparency and reproducibility of the analysis and allowing the user to re–perform the whole analysis on a local workstation by typing the R commands from the analysis report into a R console. Depending on the users choices different tables with expression or regulation values are created. This may range from tables containing the normalized expression values of all genes in all arrays, to tables with regulation or expression values of the subsets of differentially expressed genes or tables containing the raw p values and p values adjusted using the various adjustment methods.

All plots are embedded in the analysis report file and are additionally available as single pdf (end eps) files. In figure 4.1 such a file is opened showing an average MA plot.

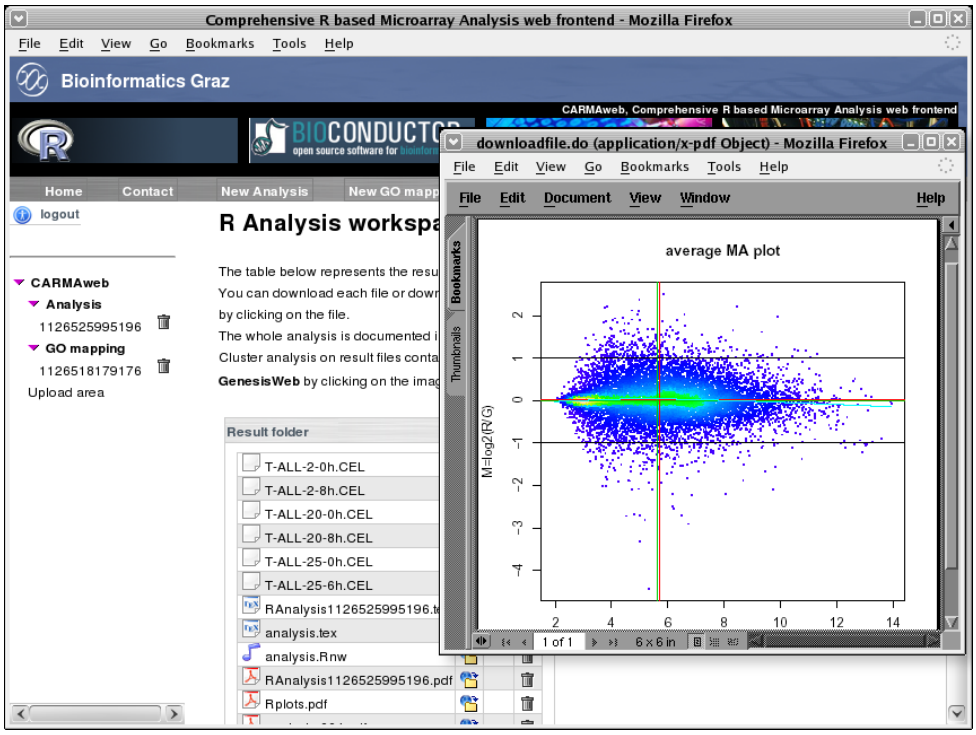

Figure 4.1: All files used and created during the analysis are accessible through this page. The MA plot shown represent an average MA plot comparing the expression levels of the genes in sample group 1 against those of sample group 0 of a detection of differentially expressed genes using test statistics. The points in the plot are colored according to the local point density.

Chapter 5

GO analysis

Mapping genes to Geneontology (GO) terms should assist in the biological interpretation of microarray results. The GO analysis in `CARMAweb` is performed by `BioConductors` `GOstats` and `GO` packages.
As input file any tab delimited file can be used, that contains in one column the Entrez Gene (formerly LocusLink) identifiers of the genes of interest. A new GO analysis can be started using the link *New GO mapping* in the button bar of the main page. The following GO mapping analysis is performed using the test file called *GOanalysis.txt*. To perform a GO analysis on own data, the file has to be uploaded first to the web application (using the *upload area* link on the left side of the main page). The file containing the corresponding Entrez Gene IDs can be selected and once the file has been selected the column can be chosen that contains the needed Entrez Gene identifiers (see figure 5.1). A preview of the first 4 lines of the file should help in the selection of the correct column.

To calculate hypergeometric p–values that give information about the over–representation of the genes of interest in the GO terms compared to all genes that can be mapped to the corresponding terms, a file containing all Entrez Gene identifier of all genes that can be detected with the microarray used has to be selected (figure 5.2). Affymetrix users can choose the Affymetrix GeneChip used in the analysis and do not have to upload or select a file. By skipping this step no (correct) p values will be calculated and only the GO graph is generated.
The genes of the test file were detected using Affymetrix GeneChip *hgu133plus2*, therefore this chip has been selected (figure 5.2).

The ontologies for which the GO analysis should be performed can be selected on the page shown in figure 5.3. Once the wanted parameters and settings are chosen the analysis can be started by clicking on the *analysis* button.

The analysis results are available through the results page (figure 5.4). For each ontology a GO graph is produced and a table containing all the GO terms to which the genes of interest can be mapped (together with the number of genes of interest that can be mapped to the GO term, the number of genes on the array that can be mapped to the GO term and the total number of genes that correspond to the GO term (from every species)). The row numbers of the GO terms in this table are used as the labels of the GO terms in the corresponding GO graph.

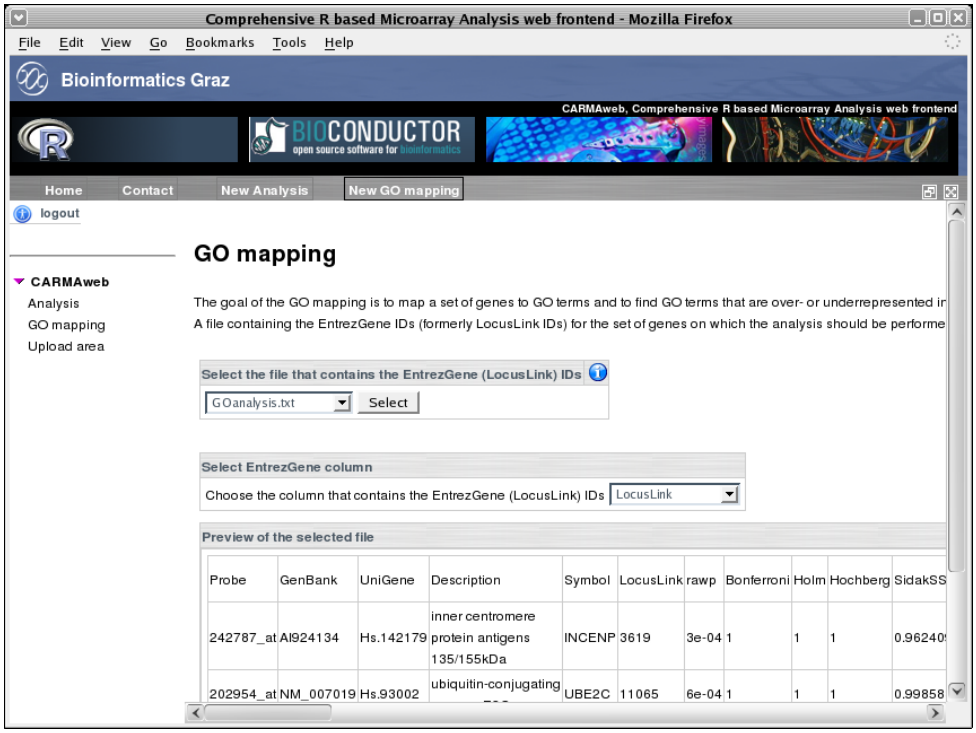

Figure 5.1: GO analysis. Selecting the file containing the Entrez Gene identifiers of the genes of interest, and the column in the selected file where they can be found.

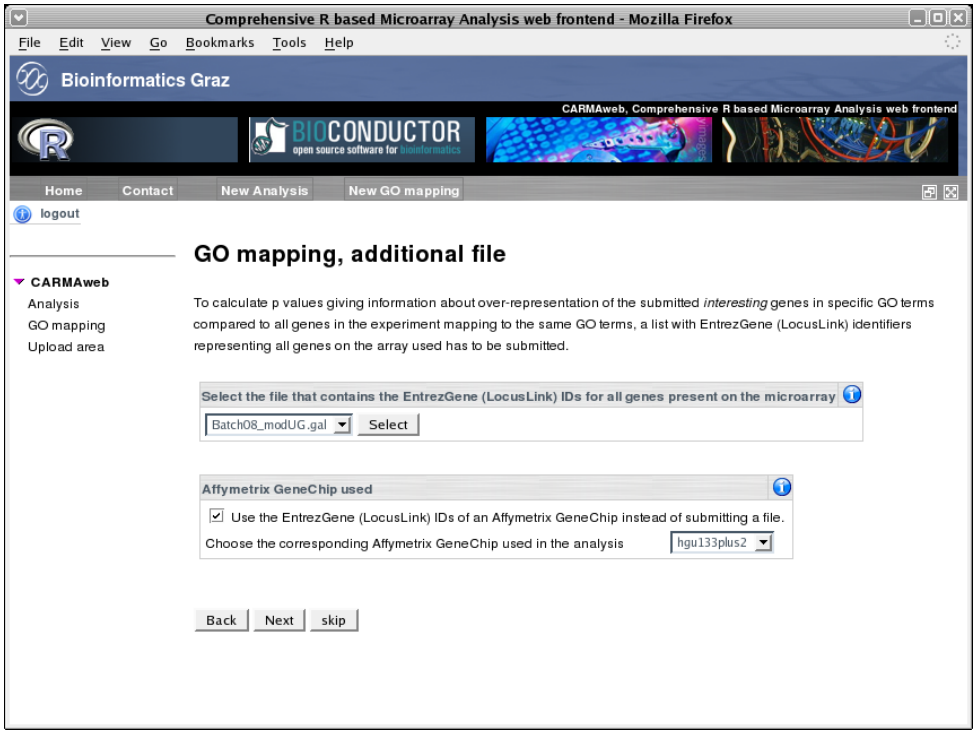

Figure 5.2: GO analysis. Selecting the file that contains the Entrez Gene identifiers of all genes on the used microarray. Affymetrix users can select the GeneChip used.

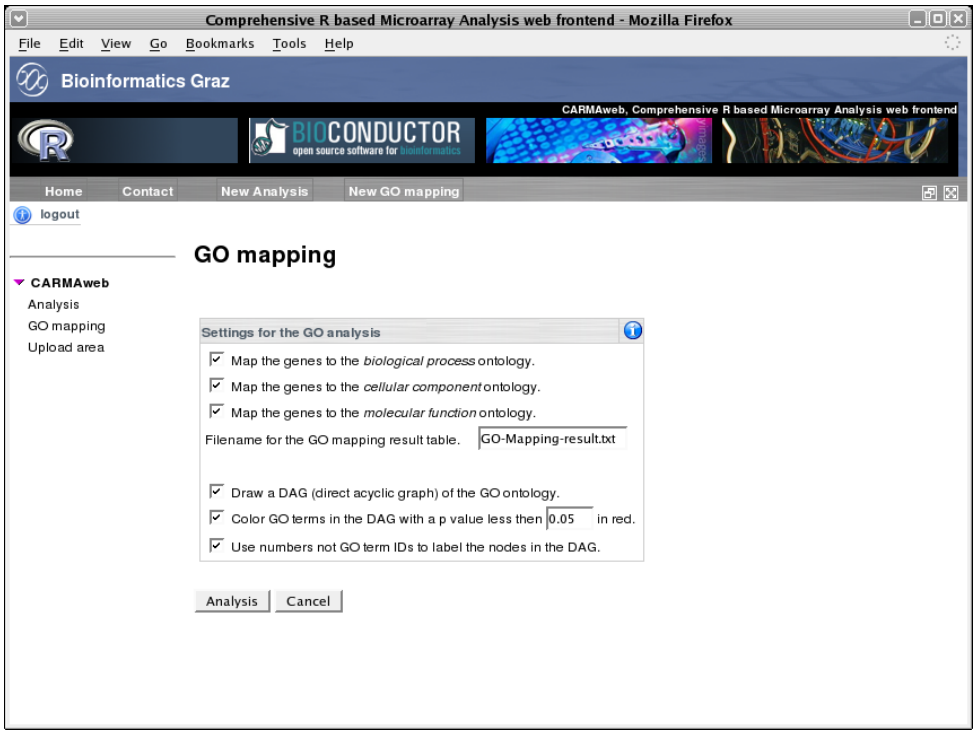

Figure 5.3: GO analysis. Additional parameters for the GO analysis.

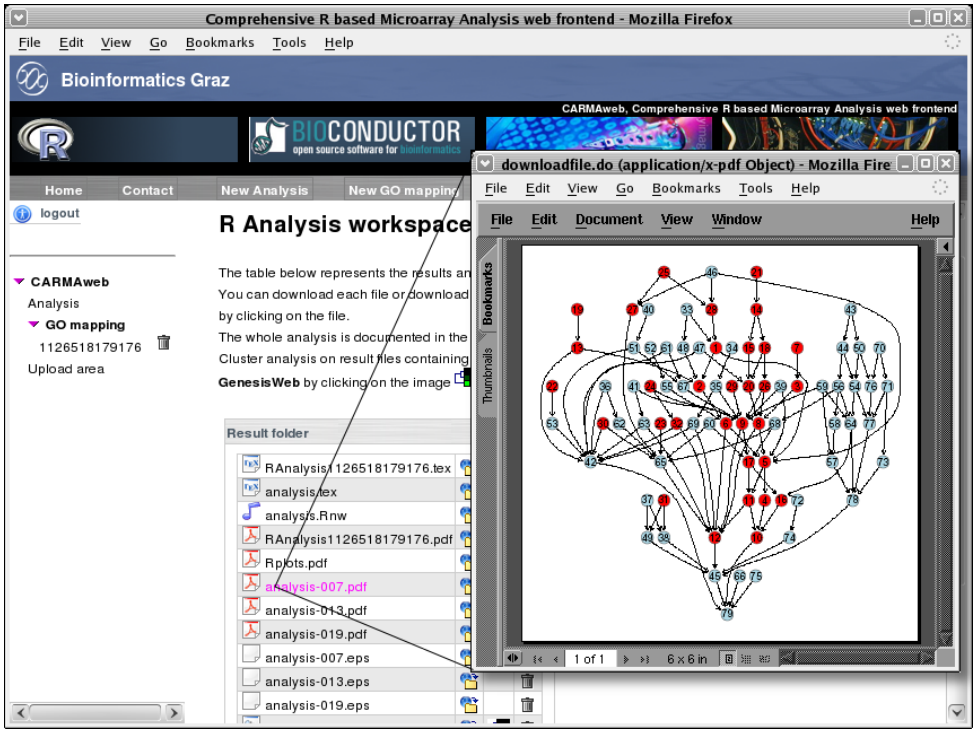

Figure 5.4: GO analysis. GO analysis result. All files used and created in the analysis are available through the presented web page. A GO graph (cellular component) has been opened.

Chapter 6

Cluster analysis

The cluster analysis can be performed using CARMAwebs cluster analysis module `GenesisWeb` which can be used as a standalone web application or in combination with CARMAweb. GenesisWeb uses facilities of the GenesisServer application server [12] to perform the cluster analysis calculations.

The layout of the GenesisWeb main page can be splitted up into the three parts button bar, navigation tree and center frame (see figure 6.2). The button bar can be used to manipulate the opened data set (for example using the *Normalize Data* menu, that allows to normalize the data (for example mean centering of the data)), or by starting a new cluster analysis by selecting one of the cluster analysis methods available. Cluster analysis algorithms implemented in GenesisWeb include the unsupervised methods HCL (hierarchical clustering), SOM (self organizing maps), KMC (k–means clustering), PCA (principle component analysis), CA (correspondence analysis) and the supervised method SVM (support vector machines). Algorithm specific parameters as well as the distance (similarity) measurement method can be selected by the user.

The navigation tree in the left part of the web page allows to open or upload new data files and to access cluster analysis results.

GenesisWeb takes as input files any tabulator delimited table can be used. The columns should correspond to the samples and the rows to the genes.

A cluster analysis using GenesisWeb can also be started from CARMAweb directly by clicking on the  icon next to a file (tab delimited txt file) in a analysis result page (see figure 4.1). In the example below the input file for the GO analysis used in chapter 5 was used. This file contains also the normalized expression values of the 100 genes over a big number of samples (the columns containing the normalized expression values have column names beginning with *B–ALL*). By clicking on the file with the name *GOanalysis.txt* shown in the result folder in figure 5.4, a page is opened (figure 6.1) where the user can select the columns of the data file that contain the data that should be clustered (in the example those columns with column names beginning with *B–ALL*). Additionally the user should select the column that contains a unique identifier for each gene (in the example the *ProbeID* column).

Based on the users selections the data file is generated and opened directly in GenesisWeb (see figure 6.2) by clicking on the next button at the bottom of the page.

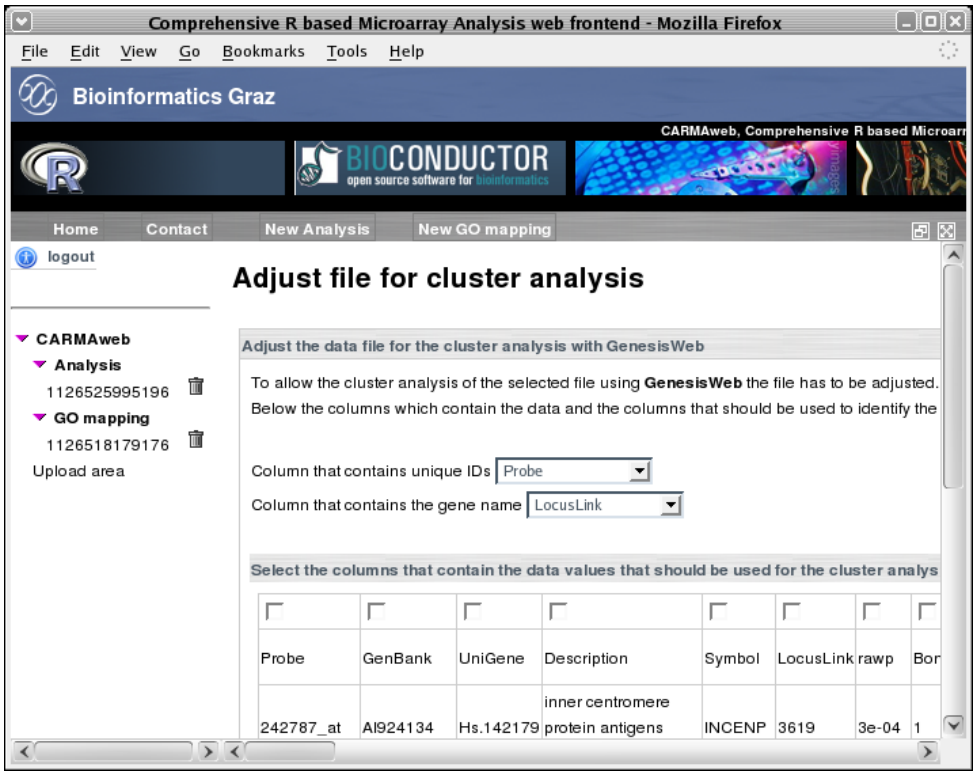

Figure 6.1: Cluster analysis. The columns containing the identifiers for each gene, as well as the columns that contain the data that should be used for the cluster analysis have to be defined.

In the example this dataset was *mean centered* using the *Normalize Data* menu in the button bar of GenesisWeb and a hierarchical cluster analysis (euclidean distance measurement and clustering over the samples and the genes) is started by clicking on the HCL link in the button bar.
The result of a cluster analysis result using the HCL algorithm is shown in figure 6.3.

Gene and / or sample clusters can be selected and colored as well as exported to text files. By clicking on the appropriate node in the hierarchical tree in the result shown in figure 6.3 a web page is opened (shown in figure 6.4) where the cluster color can be selected or the cluster can be exported to a text file.

By making a cluster *public* (using the web page shown in figure 6.4) all the genes (or samples) of the corresponding cluster are also labeled with the same color in other cluster results (for example using a different clustering algorithm) of the same data set. This is shown in figure 6.6, where the genes from the green marked cluster (see figure 6.5) are also labeled green in a PCA clustering result.

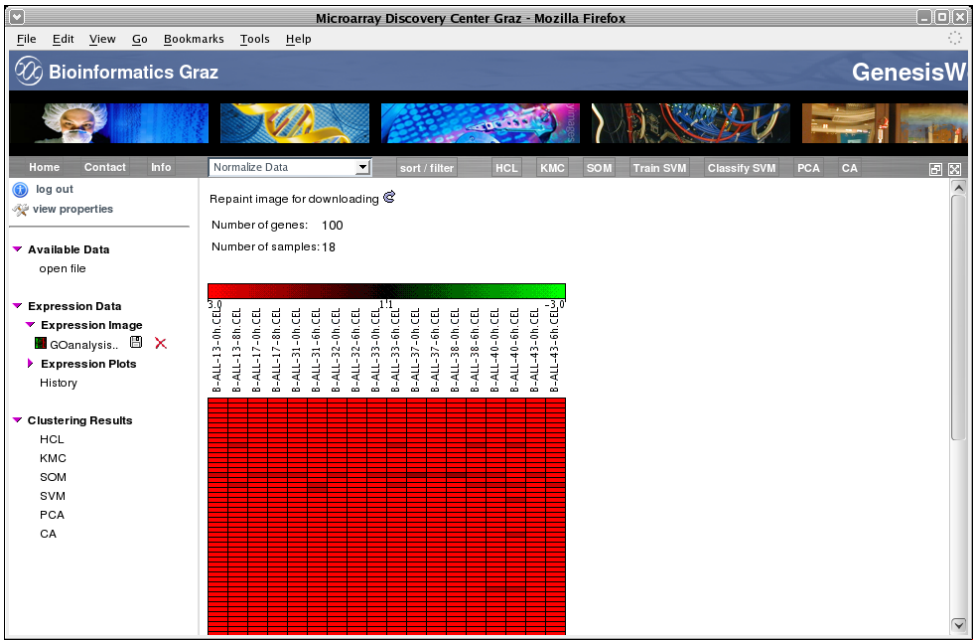

Figure 6.2: Cluster analysis. The cluster analysis module GenesisWeb. The button bar in the top of the page can be used to apply the different cluster analysis algorithms to the data. Analysis results as well as data files are accessible through the navigation tree in the left part of the web page.

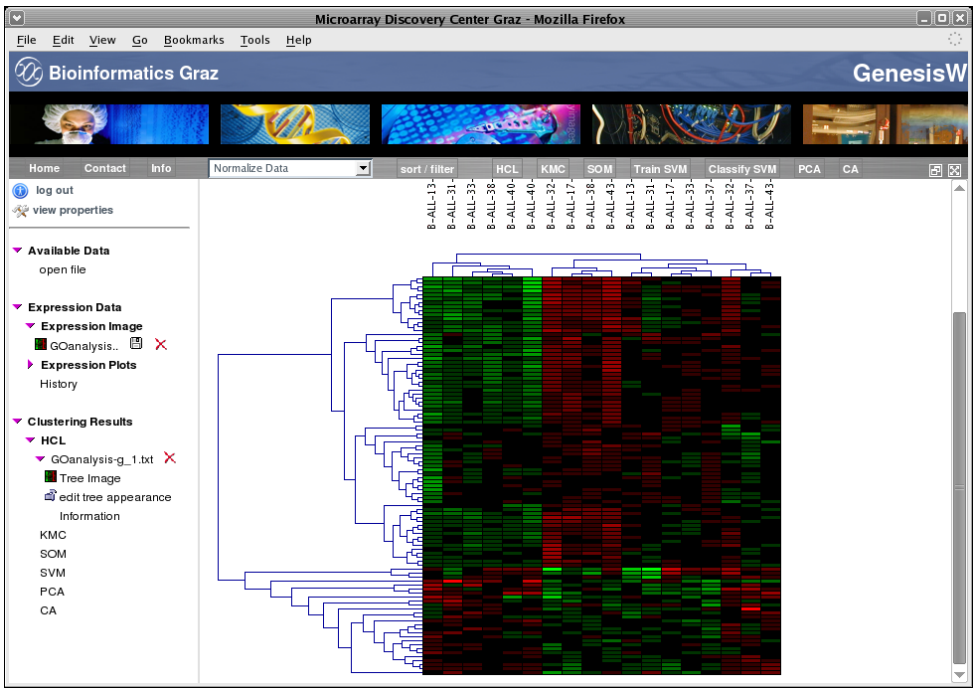

Figure 6.3: Result of a HCL cluster analysis of the test file. Gene and Sample clusters can be selected, colored and exported.

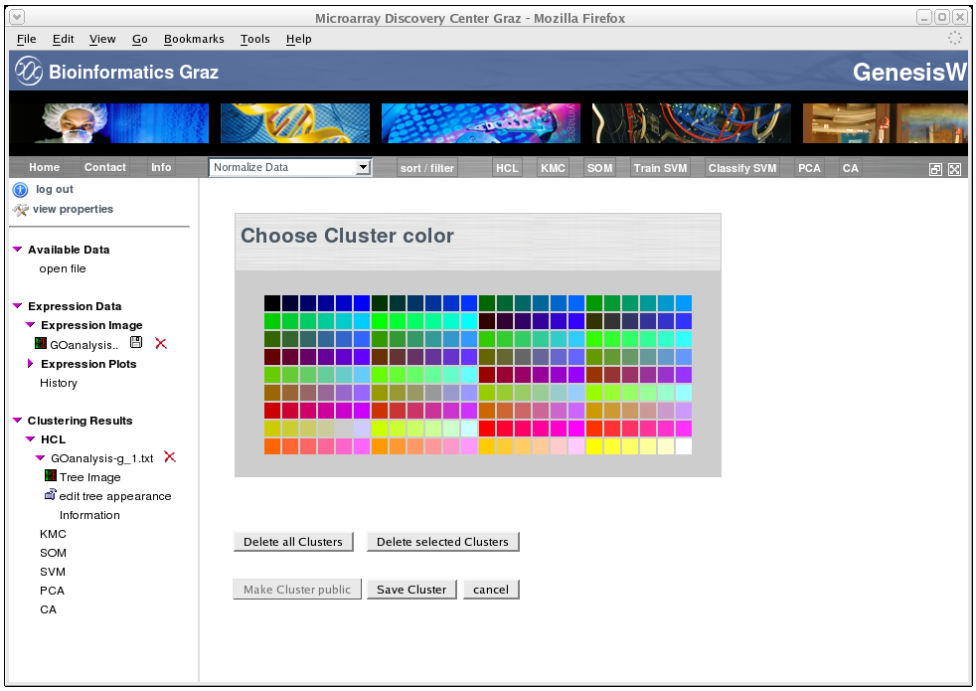

Figure 6.4: Once a cluster has been selected the color can be chosen that should be used to mark this cluster.

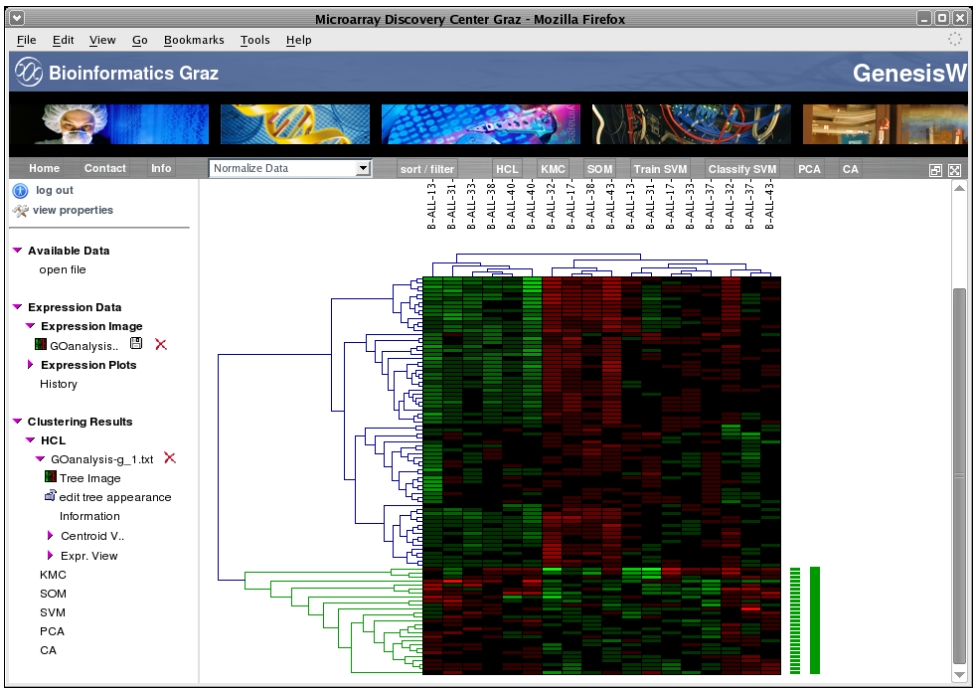

Figure 6.5: HCL cluster result of the test dataset. One cluster has been labeled with a green color.

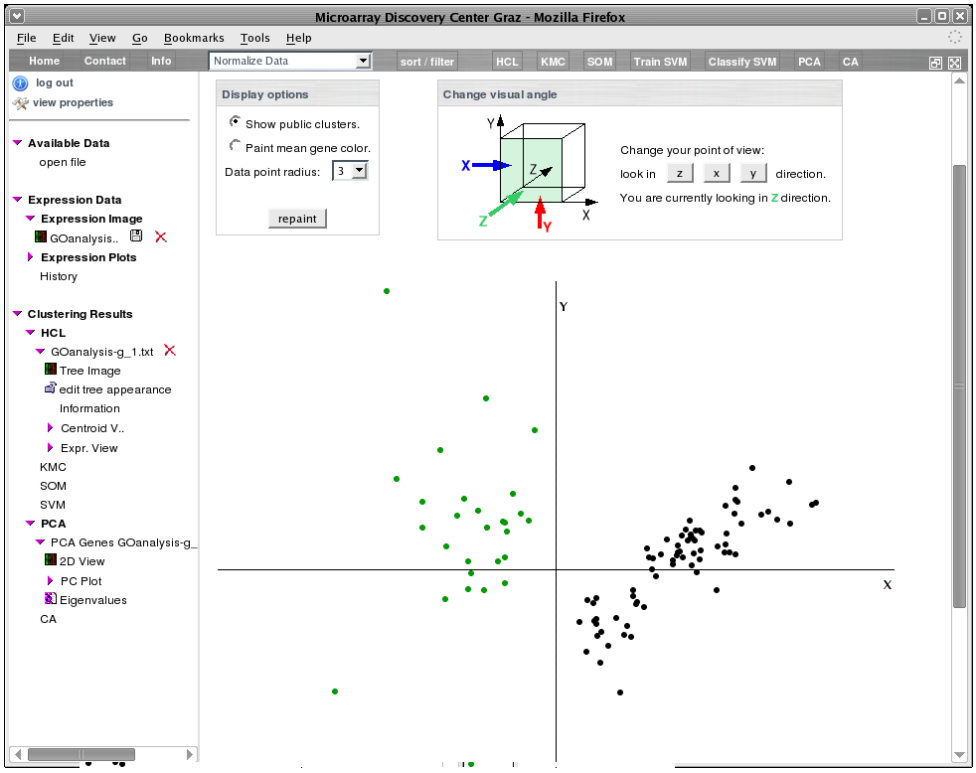

Figure 6.6: PCA cluster analysis of the test data set. The first two principle components are shown. Using the menu in top of the page it is possible to rotate the image. The genes labeled with the green color are those genes that have been marked in the HCL result in figure 6.5

Bibliography

[1] The apache struts web application framework.

[2] Y Benjamini and Y Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser.*, 57:289–300, 1995.

[3] B M Bolstad, R A Irizarry, M Astrand, and T P Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, 19(2):185–193, Jan 2003. Evaluation Studies.

[4] S Dudoit, J P Shaffer, and Boldrick J C. Multiple hypothesis testing in microarray experiments. *Statistical Science*, 18(1):71–103, 2003.

[5] Robert C Gentleman, Vincent J Carey, Douglas M Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, Laurent Gautier, Yongchao Ge, Jeff Gentry, Kurt Hornik, Torsten Hothorn, Wolfgang Huber, Stefano Iacus, Rafael Irizarry, Friedrich Leisch, Cheng Li, Martin Maechler, Anthony J Rossini, Gunther Sawitzki, Colin Smith, Gordon Smyth, Luke Tierney, Jean Y H Yang, and Jianhua Zhang. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*, 5(10):R80, 2004.

[6] Rafael A Irizarry, Benjamin M Bolstad, Francois Collin, Leslie M Cope, Bridget Hobbs, and Terence P Speed. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*, 31(4):e15, Feb 2003.

[7] Rafael A Irizarry, Bridget Hobbs, Francois Collin, Yasmin D Beazer-Barclay, Kristen J Antonellis, Uwe Scherf, and Terence P Speed. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264, Apr 2003.

[8] C Li and W H Wong. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc Natl Acad Sci U S A*, 98(1):31–36, Jan 2001.

[9] R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2005. ISBN 3-900051-07-0.

[10] G K Smyth and T P Speed. Normalization of cDNA microarray data. *Methods*, 31:265–273, 2003.

[11] Gordon K Smyth. Limma: linear models for microarray data. In R C Gentleman, V J Carey, S Dudoit, R A Irizarry, and W Huber, editors, *Bioinformatics and Computational Biology Solutions using R and Bioconductor*, page Chapter 23. Springer, New York, 2005.

[12] Alexander Sturn, Bernhard Mlecnik, Roland Pieler, Johannes Rainer, Thomas Truskaller, and Zlatko Trajanoski. Client-server environment for high-performance gene expression data analysis. *Bioinformatics*, 19(6):772–773, Apr 2003.

[13] P H Westfall and SS Young. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley, New York, 1993.