Robin: An intuitive wizard application for R-based microarray quality assessment and analysis.

Marc Lohse1, Adriano Nunes-Nesi1, Sonia Osorio1, Peter Krüger1, Liam Childs1, Jan Hannemann2, Axel Nagel1, Dirk Walther1, Joachim Selbig3, Nese Sreenivasulu4, Mark Stitt1, Alisdair R. Fernie1, Björn Usadel1.

1 Max-Planck-Institute of molecular plant physiology

Am Mühlenberg 1

14476 Potsdam-Golm

2 University of Victoria, Centre for Forest Biology

PO Box 3020 STN CSC Victoria

Canada BC V8W 3N5

3 University of Potsdam

Karl-Liebknecht-Strasse 24-25

14476 Potsdam-Golm

4 Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK)

Corrensstraße 3

06466 Gatersleben

**ABSTRACT**

The wide application of high-throughput transcriptomic analysis using microarrays has generated a plethora of different technical platforms, data repositories and sophisticated statistical analysis methods, leaving the individual scientist with the problem of choosing the appropriate approach to successfully address his biological question. Several software applications that provide a rich environment for microarray analysis and data storage are available (e.g. GeneSpring, EMMA2), but these are mostly commercial or require an advanced informatics infrastructure. What has been missing, is a non-commercial, easy-to-use graphical application that aids the lab researcher in finding the proper method to analyze microarray data without requiring expert understanding of the complex underlying statistics. We have developed Robin, a Java-based graphical wizard application that harnesses the power of the advanced statistical analysis functions provided by the R/BioConductor project. To aid inexperienced users, Robin implements a streamlined workflow guiding through all steps of two-color, generic single color and Affymetrix microarray analysis, providing in-line help and documentation. Being more than just a graphical user interface to BioConductor functions, Robin also assesses the data and automatically generates warnings to notify the user of potential outliers, low quality chips and low statistical power.

**INTRODUCTION**

Since the first microarray experiments have been performed in the 1990’s (Schena et al., 1995) a lot of effort has been put into the development of this technique as well as into approaches to properly analyze the resulting data. The wide application and technical improvement of microarrays has on one side resulted in the establishment of large publicly accessible expression data repositories such as GEO, AtGenExpress or GeneVestigator, (Schmid et al., 2005; Barrett et al., 2007) and on the other side the development of an array of sophisticated statistical methods to process and analyze the data to infer new biological insights from it. This wealth of data and methods leaves the individual researcher with the problem of choosing the correct strategy since it is not directly obvious to the inexperienced user which approach is suitable for a given experimental design. While data mining of public collections is facilitated by descriptive meta data that is attached to the expression data (MIAME and MIAME/Plant (Brazma et al., 2001; Zimmermann et al., 2006), XEML, (Hannemann et al., 2009)), choosing the correct approach to statistically (re-)analyze the data inevitably requires expertise in statistics.

One of the most advanced tools to analyze high-throughput experimental data is the open source statistics environment R that is constantly being developed and refined by leading statisticians (R Development Core Team, 2008). Together with the packages provided by the BioConductor project (Gentleman et al., 2004), R represents a powerful yet flexible platform for microarray data analysis and quality assessment. The big downside of using R, though, is its general lack of an intuitive (or even graphical user interface (GUI)). The largest part of the functionality of R can only be accessed via a text console which is constituting a considerable obstacle for many biologists that are inexperienced in the usage of such interfaces. Furthermore, making full usage of the power of R/BioConductor-based data analysis requires basic programming skills.

Several GUI applications have been developed that allow analysis of microarray data generated by different technical platforms, however these are often commercial (GeneSpring, GeneMaths XT etc), not very intuitive (limmaGUI, affylmGUI (Wettenhall and Smyth, 2004; Wettenhall et al., 2006)), not available on all computing platforms (PreP+07, (Martin-Requena et al., 2009)) or are web-based solutions that would either require uploading of potentially sensitive, unpublished data or laborious local installation (CARMAWEB, EMMA 2 Rainer et al., 2006; Dondrup et al., 2009). Packages like the TM4 suite (Saeed et al., 2003) provide a collection of excellent tools for microarray analysis. Due to the multi-program structure, though, it does not offer a consistent, workflow-oriented interface to the user.

To amend this lack of a free, user-friendly and instructive tool for microarray analysis we developed Robin. Robin provides a Java-based GUI to up-to-date R/BioConductor functions for the analysis of both two-color and single channel (Affymetrix GeneChip) microarrays and implements wizard-like workflows that guide the user through all steps of the analysis including quality assessment, evaluation and experiment design. Robin assists the user in the interpretation of the results by automatically issuing warnings if quality check parameters exceed or undercut conservatively chosen threshold values or statistical analysis indicates problems like insufficient input data. During the whole workflow the major attention has been put on simplicity and intuitiveness of the graphical user interface – advanced options to tweak the parameters of the analysis functions are by default hidden from the user. More experienced users of course have the possibility to modify the settings to meet their individual needs and even review and modify the R scripts that were generated from the inputs before they are executed by the embedded R engine by activating the expert mode (Fig. 1).

The generated output comprises complete statistical results on differentially expressed genes that can directly be imported into common spreadsheet applications and meta-analysis tools like MapMan for visualization, informative plots visualizing the quality check and statistical results and the R scripts that have been automatically generated from the users’ input.

Robin installer packages including an embedded minimal R engine (plus the required packages) are available for the most abundant operating systems Microsoft Windows (XP or higher) and Mac OS X (version 10.5 or higher) from http://bioinformatics.mpimp-golm.mpg.de/projects/own/robin. Installing these packages will leave an existing installation of R on the target system untouched. For all other systems a lightweight package that can incorporate and configure an existing R installation for usage with Robin is available.

**RESULTS AND DISCUSSION**

Robin implements standardized workflows for the analysis of common microarray experiment designs such as common reference and direct design two-color experiments and simple multifactorial designs in which more than one experimental condition is being varied. The user can choose between three separate workflows specialized for Affymetrix GeneChip, generic single channel (e.g. Agilent etc) and two-color microarray data normalization and analysis respectively.

*Importing raw data*

Due to the various file formats in use for non-Affymetrix microarray data, special care has been taken to provide a versatile import wizard that assists the user in the import of arbitrary tabular single- and two-color data. The only restriction imposed is that the data has to be in tabular text format. During the import, Robin tries to automatically separate header information from the tabular data section in the input file and asks the user to specify which columns contain the fields required for analysis (i.e. red channel foreground and background, green channel foreground and background intensities and a unique identifier for each measured signal). When importing single- and two-color data, Robin tries to determine whether the chip layout comprises probes spotted in duplicates. For convenience, the layouts of many common plant microarrays are bundled with Robin as layout presets. Detection of spot replication relies on the spot identifiers and is based on the assumption that if the gene spots are not duplicated (but the controls are, which is most probably true) the number of unique identifiers will be greater than 50% of the total number of spots. This should be true for all array types that have more gene spots than control spots, which might not be the case for “boutique” arrays that only contain few probes (e.g. custom arrays designed for small organellar genomes). If replicate spots are detected, Robin sorts the input data by identifier to make sure that replicates are consecutive, sets the number of duplicates to two and the spacing between duplicates to one. Obviously, this is incorrect in cases, where more than 2 replicates are spotted on the array. When analyzing arrays on which the spacing of replicate spots is not uniform, this approach might lead to overestimation of significance and underestimation of correlation for replicate spots that are close together on the array. To account for this possible bias, Robin generates a warning when replicates are detected and informs the user of the assumptions made. In addition to this essential information on the input data, the user either has to choose the chip grid layout from the list of predefined layouts or enter a custom layout. All settings of the import wizard interface can be saved as an input data preset to speed up loading of similar data. After importing the data, the user is asked to define the ‘targets table’ by entering the different RNA samples and specifying which sample has been hybridized to which color channel on each chip. For subsequent analysis, a reference sample must be specified. In very simple experiments that only comprise replicate chips of two different treatments (possibly including dye swaps), Robin uses the first entered sample as reference by default. If data conforming to a common reference design was entered, Robin automatically detects the common reference sample and prompts the user in case this sample was not set as reference. Already at this step, Robin analyses the input and tries to make sure that the data is consistent e.g. by verifying that the samples are not disconnected. Import of Affymetrix single channel data does not cause such problems, since the data format is uniform and it is not necessary to define a targets table.

*Quality assessment*

After importing the chip data a variety of quality assessment methods (Fig. 1) that can be freely chosen and combined can be run to allow the user to get an overview of the quality of input data and subsequently exclude chips that show strong technical artifacts individually. Additionally, the user can already at this step choose which normalization, p-value correction and statistical analysis approach (linear model-or rank product-based) to use, although these more advanced settings are not displayed by default and robust standards, that yield reliable results in most cases, are preselected. Advanced users can of course take control of analysis parameters and modify them according to their needs. To support the user in the evaluation of quality assessment results, warnings are issued automatically if quality measures of individual chips exceed conservatively chosen threshold values (see Materials and Methods section for details). Specifically, methods available for quality assessment of single channel data are (I) RNA degradation analysis, (II) box plots and (III) density plots of raw probe signal intensities, (IV) pseudo-images of probe level model (PLM) residuals, (V) scatter plots of the average probe intensity (A) against the logarithmic fold change in expression (M; MA plots), (VI) scatter plots comparing all possible combinations of two individual chips, (VII) visualization of principal component analysis and hierarchical clustering of the normalized expression values (VIII) box plots showing the normalized unscaled standard errors (NUSE) and relative logarithmic expression (RLE) of the probe level models and (IX) false color images of the background signal intensity for non-Affymetrix arrays. All PLM-based methods are exclusively available for Affymetrix arrays only, while the other function can also be run on generic single channel chips. Methods available for two–color chip quality assessment are (I) image plots visualizing the chip background signal intensities, (II) density plots of the probe intensity distribution before and after normalization, (III) MA plots of raw and normalized data for each chip and (IV) image plots showing the M value for each probe color coded on a pseudo chip.

All of the above mentioned quality checks have been implemented in R using functions provided by the Bioconductor packages affy, affyPLM, simpleaffy, gcrma, plier, limma, marray and RankProd (Wang et al., 2002; Bolstad, 2004; Gautier et al., 2004; Smyth, 2004; Wu et al., 2004; Affymetrix, 2005; Wilson and Miller, 2005; Hong et al., 2006). Some functions were modified to enhance the visual output. Depending on the type of input data the user can choose between different analysis approaches: In case of single channel data, linear model based (limma) or rank product based (RankProd) analysis is available. Two color data will always be analyzed using limma functions. Quality analysis (QA) results will by summarized in a scrollable list showing clickable thumbnail images of the QA plots. Individual chips showing warnings may be manually excluded from the analysis to prevent them from introducing technical bias in the subsequent assessment of differential gene expression.

*Experiment design*

When working with Affymetrix data, depending on the statistical analysis strategy chosen, the user can define two (when using rank product) to any number (using limma) of groups of replicates and sort the imported data files accordingly. Unique labels identifying the groups have to be chosen – these labels will be used later on when defining the contrasts of interest. Robin will generate a warning if groups contain less than three replicates accounting for the lower reliability of the results if too few data points are available for the analysis of differential expression. At present, all replicate experiments are treated as true biological replicates. Providing data that is technically replicated will lead to an overestimation of significance when analyzing differential gene expression.

Subsequently, the groups are depicted as draggable boxes on the graphical designer panel that allows the user to visually lay out comparisons of interests between the groups. To achieve this, one simply has to draw an arrow by (shift-) control-click-dragging from one box to the other, e. g. from ‘wildtype’ to ‘mutant’ as shown in (Fig. 1). Robin interprets this as the contrast ‘wildtype minus mutant’. If more than one experimental condition is being varied, so called ‘interaction terms’ can be defined by creating ‘meta groups’ and drawing arrows between them (see Fig. 1). Specifically, the contrast of meta groups shown in figure 1 will be interpreted as the interaction term ‘(wildtype minus wiltype stressed) minus (mutant minus mutant stressed)’ which will extract those genes that respond to stress differently in mutant and wild type. The expert settings box included on the experiment designer panel again allows advanced users to tweak all relevant parameters of the statistical analysis, like p-value- and minimal log2-fold change cutoff, correction method for multiple testing, normalization (although it is not recommended to use different normalization methods for quality analysis and main analysis) and the statistical strategy for multiple testing across contrasts. Additionally, expert users can choose to review the R script that is generated from the inputs before it is sent to the R engine and include custom code or use Robin to quickly and comfortably generate skeletons of analysis scripts that can then be used as starting points for more sophisticated customized analyses.

*Analysis and Results*

The statistical methods Robin employs to identify differentially expressed genes are based on two different approaches: Linear modeling (limma, (Smyth, 2004)) and rank product-based analysis (RankProd, (Breitling et al., 2004; Hong et al., 2006)). When analyzing Affymetrix data, the user can choose between these two options with the restriction that rank product-based inference of differential expression is only available when two groups are to be compared. When working with two-color microarrays rank product-based analysis is not available yet. The two methods differ in that they take two completely different approaches to the detection of differentially expressed genes. While the linear model-based method relies on advanced statistical modeling and bayesian inference, the rank product approach more resembles biological reasoning on the data.

Since rank product-based analysis is limited to comparing two experimental conditions, the linear model based analysis offers far more options and flexibility with respect to the available settings and design of the experiment (e.g. if two factors, like genotype and treatment, are being varied in an experiment and the user is interested in the interaction effect).

After collecting all necessary information from the user, Robin generates an R script that is subsequently executed by the embedded R engine. The script produces a comprehensive set output files that are organized in a folder structure.

The results comprise several informative plots summarizing the statistical analysis: MA plots are created for each contrast showing the genes that are called significantly differentially expressed highlighted in red (see supp. Fig. 2). If less than five contrasts have been defined, Robin generates Venn diagrams visualizing the number of genes responding differentially and the overlap of response between contrasts (see Fig. 2). In addition to this, dendrograms showing the hierarchical clustering of the data based on Pearson correlation of expression and scatter plots of principal components (PCA) give an overview of the internal structure of the data. Robin automatically saves several tables containing the complete statistics results for all the genes and the top 100 differentially expressed genes for each comparison made. Please note that when a minimal log2-fold change of 1 cut-off is chosen, genes showing a weaker response than a two-fold up- or down regulation will not be listed in the full tables. Summary tables formatted for direct import and visualization in the meta analysis tools MapMan and PageMan (Usadel et al., 2005; Usadel et al., 2006) allow for a simple integration of Robin with downstream analyses. These files contain the log2 fold change in expression for each gene in each contrast plus a flag denoting the results of the statistical testing (0 = not significantly regulated, 1 = significantly up regulated, -1 = significantly down regulated). These flags can be used for convenient filtering in MapMan (see (Usadel et al., 2009) for further details). For Affymetrix data, present and absent calls are calculated using the mas5calls implementation provided by the affy BioConductor package (Gautier et al., 2004). All plots generated in the quality analyses, processed input files, the generated R source code and a short text file summarizing the analysis are written to the output folder to completely document the analysis workflow and ensure reproducibility of the results.

*Case study – Comparison of tomato tissues*

Robin was used to analyse differential gene expression in tomato flowers, roots and leaves. Tomato plants were grown and samples prepared as described and subsequently hybridized to TOM2 microarrays in a two color microarray experiment setup. Quality assessment showed that there were no obvious and severe technical artifacts visible on the chips when investigating the background intensity images and the signal intensity distributions plots (supp. Fig. 1). However, warnings were generated for all MA plots of the individual chips because of a slightly elevated percentage (between 10.141% and 13.43%) of genes that showed a more than two fold change in expression. These warnings are based on the assumption that most of the genes will not show differential expression in any given experiment and are issued if the percentage exceeds 5%. This value might be too strictly chosen when comparing very different tissues as it is the case in the experiment described in this study. Another reason might be an overestimation of expression values due to an elevated signal to noise ratio. As often observed in two color microarray experiments, the normalized signal intensities differ in the red and green channel (see supp. Fig. 1). This technical bias could largely be eliminated by using the standard background subtraction and scaling normalization approach in Robin as shown on supplementary figure 1. Since none of the chips showed strongly outlying behavior in the quality assessment step, all were included in the statistical analysis of differential gene expression. The three tomato tissues were compared against each other using a direct design with three biological replicats and dye swaps. In total, 418 genes were found to be specifically significantly differentially regulated between leaves and roots, 200 when comparing leaves to flowers and 234 in the comparison of flowers to roots. As indicated on the Venn diagram (Fig. 2) there was also a substantial numbers of genes showing differential expression levels in more than one comparison.

The results obtained in Robin were subsequently analyzed using MapMan (Usadel et al., 2009) to gain insights into the biological context of relevant differences in gene expression. Using the biological pathway visualization capabilities of MapMan, general differences could be observed when comparing the aboveground organs with roots. The most prominent changes were, as could be expected, present in genes related to photosynthesis. The corresponding MapMan bins (1.1 PS.light reaction, 1.2 PS.photorespiration, 1.3 PS.calvin cycle, 19 tetrapyrrole synthesis) were strongly and very consistently upregulated in leaf and flower tissue (Fig. 3, supp. Table 2 and supp. Fig. 3) when compared to roots while the difference was much less pronounced although still significant between leaves and flowers. This result can clearly be attributed to the fact that leaves as the primary sites of photosynthesis supply sink organs like roots and flowers with assimilates and hence need to maintain the photosynthetic machinery in a functional state.

In addition to the visual inspection of pathways provided by MapMan, the built-in wilcoxon rank sum test function was used on all three comparisons to identify significantly changed MapMan bins (see supp. Table 2). Other general processes that were found to be significantly upregulated in leaves include starch synthesis (2.1.2) and degradation (2.2.2) while sucrose breakdown-related genes (2.2.1) showed increased expression in roots. Since sucrose is the carbohydrate species that is primarily transported via the phloem, this observation again supports the sink-source relation between leaves and roots. Sucrose synthase is presumably involved in sucrose breakdown to provide for carbon supply in sink organs (Sun et al., 1992; Zrenner et al., 1995). Surprisingly, invertases, that are required for normal root growth in *Arabidopsis* (Barratt et al., 2009), showed slightly stronger expression in leaves.

YABBY transcription factors (27.3.10), that have previously been shown to be involved in the regulation of lateral organ development (Street et al., 2008; Stahle et al., 2009), were found to be significantly upregulated in leaf (SGN-U603003) and flower tissue (SGN-U591723, SGN-U577176, SGN-U603003, see supp. Fig. 3). The expression of YABBY proteins was strongest in flowers supporting their well described prominent role in flower development (Fourquin et al., 2007; Ishikawa et al., 2009; Orashakova et al., 2009). Investigation of the development-specific expression pattern of *Arabidopsis* YABBY proteins using the Genevestigator (Zimmermann et al., 2004) tool revealed a similar expression pattern with the CRC (crabs claw) protein showing highest expression in mature flowers (supp. Fig. 4). Similarly, the MADS-box transcription factors homologous to SEPALLATA (SEP1/2) and AGAMOUS-like (AGL8/12) from *Arabidopsis*, that are known to regulate flower and seed development ((Mizukami et al., 1996; Pelaz et al., 2000) also see (Robles and Pelaz, 2005) for a review) , show strongest expression in flower tissues (see supp. Fig. 3), confirming the sanity of the results generated using Robin.

MapMan bins that were primarily upregulated in root tissue comprise lignin biosynthesis (16.2.1), plasma membrane intrinsic proteins like aquaporins (34.19), genes related to flavonoid synthesis and metabolism of phenolic compounds. Although the latter two were not significantly responding according to the wilcoxon rank sum, individual genes showed significant responses. Since expression of flavonoid biosynthesis genes in root tissue is induced in the light (Hemm et al., 2004) the upregulation of SGN-U565166, SGN-U565164 (similar to flanonol synthase) and SGN-U563058 (similar to flavonone-3-hydroxylase) might indicate an artifact due to exposure of the root to light during sample harvesting.

Flower tissue displayed a strong expression of cell wall degrading enzymes like pectin esterase (10.8), pectate lyases and polygalacturonases (10.6.3) in comparison to both leaves and roots. Pectin methyl esterases (PME) catalyze the demethylation of pectin prior to cleavage by pectate lyases and polygalacturonases. Previous screens of cDNA libraries derived from maize pollen have shown high expression levels of pectin degradation related genes (Wakeley et al., 1998) that are believed to play a role in pollen tube elongation. Interestingly, two putative PMEs (SGN-U585819 and SGN-U585823) exhibited deviating behavior with low expression in flowers. Further investigations using the tomato genome browser provided by the sol genomics network (http://solgenomics.net/gbrowse/gbrowse/ITAG\_devel\_genomic/) revealed that both genes are located on the same chromosome in direct vicinity of each other possibly indicating that they originate from a duplication event. As the annotation of these genes as PME is solely based on sequence similarity and has no direct experimental support, it may be suspected that they are not involved in the same process. The observations reported above were highly significant both on the pathway level, as tested by the wilcoxon rank sum test, and on the level of individual genes as confirmed by the statistical analysis of differential gene expression (please see supp. Table 1 for full details).

**MATERIAL AND METHODS**

*Implementation of Robin*

Robin was implemented in Java and R using free extension libraries developed by several software projects. Specifically, the NetBeans visual API (http://graph.netbeans.org/) was used to develop the visual experiment designer, the AffxFusion (http://www.affymetrix.com/partners\_programs/programs/developer/index.affx) library was employed for the extraction of detailed information from Affymetrix chips, Apache commons (<http://commons.apache.org/>) facilitates generic string operations. To achieve an improved user experience and better integration into the Mac OS X platform, we used the AppleJavaExtensions provided by Apple, Inc. and the QuaQua (http://www.randelshofer.ch/quaqua/) look and feel. We provide a stand alone “slim-line” R engine that was embedded in the Robin package and is independent of user installed versions of R. All required BioConductor packages have been included to provide an all-in-one package that works directly after installation. Installer packages for different operating systems were created using the free IzPack installer generator (http://izpack.org/). We also provide a lightweight package without R that can be deployed on any Java-enabled platform. At the first start, Robin will ask the user for a path to a working R installation, check this installation and automatically download all required packages (if not already present), provided the computer has a working internet connection.

*Automatic input assessment and generation of warnings*

Robin tries to aid the user in assessing the quality of the microarray data by automatically generating warnings if diagnostic measures are exceeding preset threshold values. The assessment of global RNA degradation effects as implemented by the AffyRNAdeg function ((Gautier et al., 2004)) yields slopes for each of the degradation curves. If the slopes of individual RNA degradation curves are exceeding a value of three or are deviating by more than 10% from the median slope of all curves, a warning message indicating the affected chips is diplayed in the quality check result list. MA plots visualizing the log2 fold change in expression of gene G under condition C vs. condition D (M = logGC - logGD) plotted against the average log2 probe intensity (A = ½ \* (logAC + logAD)) are generated for each individual chip. In the case of two color microarrays the red channel signal intensity is compared against the green channel signal intensity. To display MA plots for affymetrix arrays, the normalized expression values of each chip are compared against a synthetic chip created using the median expression values of all probesets across all chips in the experiment. Based on the assumption that most genes will not respond differentially to a given treatment Robin automatically warns the user if more than 5% of the probesets on an individual chip are more than two fold up- or down regulated. This threshold might be too restrictive in certain experiments where e.g. very different developmental stages of an organism are compared or a drastic treatment is applied. Generally, though, a high percentage of differentially responding probesets might indicate technical artifacts caused e.g. by a low signal-to-noise ratio or large differences in probe signal intensity that could not be eliminated by normlization. Again based on the aforementioned assumption, the M values plotted on a MA plots should be centered around M=0. A lowess fit (Cleveland, 1979) is calculated for the MA plots. In the ideal case the lowess fit curve would be identical to the M=0 line. As an estimate for a strong deviation of the lowess fit from the M=0 line, the absolute numerical integral of the lowess curve over the M=0 line is calculated. If the integral exceeds a value of 1, a warning will be issued to notify the user of possible artifacts that might be caused by e.g. a bimodal probe signal intensity distribution. Probe signal intensity oversaturation is estimated by calculating the percentage of probes whose raw signal intensity is at the maximal intensity value measured within each chip. Usually only one or a few probes display maximal intensity (in the case of Affymetrix GeneChips the theoretically possible maximal dynamic range of probe signal intensity is 0 to 216 due to the 16 bit data precision of Affymetrix GeneChip scanning devices). If more than 0.25 % of the probes have maximal intensity, the chip is considered oversaturated and a warning is generated, informing the user of the possible information loss. Since the rank product-based analysis does not accept duplicated spots on one array, Robin checks the input data and collapses replicated values identified by the same identifier to the median value within each array. If replication is detected a file containing the replicated spot identifiers and values will be written to disk. In addition to the warnings issued during the quality assessment, Robin will also inform the user of problems that occurred during the statistical analysis of differential expression, like low or imbalanced numbers of biological replicates and low significance of the results (e.g. none of the probes tested is called significantly differentially expressed given the chosen thresholds). At the end of the analysis workflow, Robin will present a summary list of all generated warnings to ensure that the user is made aware of possible shortcomings of the data.

*Plant material*

*Solanum lycopersicum* plants cultivar M82 seeds were allowed to germinate directly on soil and were then transferred to a vermiculte-based groth substrate and further cultivated as described in (van der Merwe et al., 2009). Plant materials for microarray analysis were harvested from 6 week-old plants. Specifically, leaf samples were taken from the third to fourth node from the top, roots were washed in tap water to remove growth substrate and all fully expanded flowers were collected. In order to minimize circadian effects, samples were taken on two consecutive days at the same time of day within 1 ½ hours. Tissue samples were immediately shock frozen in liquid nitrogen and stored at -80°C.

*Sample preparation*

Tomato RNA extraction was performed using a modification of the standard TRIzol (Invitrogen GmbH, Karlsruhe) extraction protocol. Briefly, 500 mg of frozen material was finely ground in a mortar and subsequently mixed with 5 ml of TRIzol solution by vortexing. After addition of 3-5 ml chloroform and centrifugation for 20 minutes at 4000xg, the aqueous phase containing the RNA was transferred to a fresh tube. RNA was precipitated over night following addition 0.5 volumes of precipitation solution (0.8 M sodium citrate, 1.2 M sodium chloride) and 0.5 volumes of 2-propanol. Precipitated RNA was recovered by centrifugation for 20 minutes at 4000xg and subsequently washed twice by adding 5 ml of 70% ethanol and centrifuging for 5 minutes at 4000xg. After complete removal of 70% ethanol, the RNA pellets were air-dried and finally dissolved in 40 µl of sterile water. cDNA synthesis and labeling was carried out as described in (Degenkolbe et al., 2005) using Dynabeads Oligo(dT)25 (Dynal, Oslo, Norway) to extract mRNA from the whole RNA samples.

*Chip hybridization and data processing*

The TOM2 microarrays were obtained from the Boyce Thompson Institute (Ithaca, NY, USA). Each microarray contains 11890 oligonucleotide probes designed based on gene transcript sequences from the Lycopersicon Combined Build # 3 unigene database ([http://www.sgn.cornell.edu](http://www.sgn.cornell.edu/)). Chip hybridization was performed as described in (Degenkolbe et al., 2005) with the following modifications: ADRIANO. All three possible comparisons between the three tissues were performed in three biological replicates resulting in nine microarray hybridizations. Raw signal intensity values were computed from the scanned array images using the image analysis software GeneSpotter version 2.3 (MicroDiscovery, Berlin, Germany). The raw intensity values were normalized using Robin’s default settings for two color microarray analysis. Specifically, background intensities estimated by GeneSpotter were subtracted from the foreground values and subsequently a printtip-wise loess normalization (Yang et al., 2002) was performed within each array. To reduce technical variation between chips, the logarithmized red and green channel intensity ratios on each chip were subsequently scaled across all arrays (Yang et al., 2002; Smyth and Speed, 2003) to have the same median absolute deviation. Statistical analysis of differential gene expression was carried out using the linear model-based approach developed by (Smyth, 2004). The obtained p-values were corrected for multiple testing using the strategy described by (Benjamini and Hochberg, 1995) separately for each of the comparisons made. Genes that showed an absolute log2-fold change value of at least 1 and a p-value lower than 0.05 were considered significantly differentially expressed. The log2-fold change cutoff value was imposed to account for noise in the experiment and make sure that only genes that show a marked reaction are recorded. The TOM2 chip oligonucleotide annotation was updated based on BLAST (Altschul et al., 1990) searches against the newest version of the SGN tomato unigene set (Tomato 200607 build2, <http://solgenomics.net/>) and MapMan bins were assigned to each oligonucleotide on the chip based on the SGN tomato unigene mapping. Wilcoxon rank sum tests were performed to test whether there were bins that were significantly and consistently behaving different than the other bins in the MapMan ontology using the built-in function in MapMan.

**FIGURE LEGENDS**

Figure 1: Left panel: Screenshot of the quality assessment functions available for Affymetrix (R) chips. All methods can be freely combined to obtain an overview of the input data quality. Short inline explanations for each method are displayed in the info field on the left side upon clicking the question marks. The expert panel at the bottom of the user interface is providing more option for customizing the analysis settings. By default, robust analysis methods are predefined and panel is hidden to provide a less cluttered interface to inexperienced users. Right panel: Screenshot of the graphical experiment designer panel. Comparisons between the previously defined groups of biological replicate chips can be configured by dragging visual connections between them. The arrowhead defines the direction of the comparison. E.g. the arrow between the ‘wildtype’ group and the ‘wildtype stress’ group is interpreted as the ‘wildtype - wildtype stress’ contrast, meaning that genes showing a higher expression level in the ‘wildtype stress’ group will have a negative log2 fold change value in the output and vice versa. Interaction terms can be defined via ‘metagroups’ shown as orange boxes.

Figure 2: Venn diagram showing the numbers of genes called significantly differentially expressed when comparing tomato leaf, flower and root tissue. The numbers include both up- and downregulated genes. Genes that are differentially regulated in more than one comparison are depicted in the overlapping areas. As indicated by the number in the lower right corner, 10531 genes were not significantly affected.

Figure 3: PageMan analysis of the tomato case study. A wilcoxon test was performed, analogous to the test implemented in MapMan, to identify significantly differentially regulated MapMan bins. Individual bins that show distinct responses are highlighted. The plot shows the color coded Z scores of the p-values computed in the test.

Supplementary figure 1: Summary of all quality check plots generated for the tomato case study experiment. (A) Image plots of the background signals measured on each chip. (B) Chipwise MA plots; (C) False-color images of the log2 ratios of raw red and green channel signal intensities; (D) Overview plots showing the raw and normalized signal intensity distributions on all chips. The upper panel shows density plots and the lower panel shows boxplots of the same values.

Supplementary figure 2: MA plots of the three comparisons made in the tomato case study experiment. The plots show the average signal intensity (A) and the average normalized log2-fold change (M) individually for each comparison. Genes showing significant differential regulation are highlighted by red circles.

Supplementary figure 3: Exemplary visualization of the most strongly reacting bins using the MapMan. Genes that are not significantly regulated are greyed out using the built-in filter function. The comparisons shown are (A) Leaf – Root, (B) Flower – Root and (C) Leaf – Flower.

Supplementary figure 4: Expression patterns of three YABBY transcription factor homologs from *Arabidopsis* created using the Genevestigator web application. The Affymetrix probe set identifiers correspond to the following YABBY genes: 245029\_at: YABBY family protein At2g26580; 260355\_at:Crabs claw (CRC) protein At1g69180; 262989\_at: Inner no outer (INO) protein At1g23420.

Supplemental figure 5: Genomic locations of two putative pectin methyl esterases from tomato (SGN-U585819 and SGN-U585823) as shown by the Gbrowse genome browser (<http://solgenomics.net/gbrowse/gbrowse/ITAG_devel_genomic/>). The genes are located on the same chromosome within a range of less than 10kb possibly indicating that they originate from a genetic duplication event.

Supplementary table 1: Detailed statistical results tables as produced by Robin. For convenience, the individual tables have been combined into one MS Excel file containing the original tables on separate worksheets. A second set of work sheets has been included that also contains the MapMan bins associated with each of the oligonucleotides on the TOM2 chip and the annotation of the target transcripts taken from the latest tomato unigene release (Tomato 200607 build2). The columns contain from left to right: (Feature.ID) A unique identifier for the oligonucleotide probes or probe sets on the chips; (logFC) the log2-fold change in expression; (AveExpr) average normalized expression value; (t) t-statistic; (P.Value, adj.P.Val) raw and Benjamini-Hochberg-corrected p-values for differential expression; (B) the log-odds for differential expression.

Supplementary table 2: Wilcoxon rank sum test results generated by MapMan. The ‘Elements’ column refer to the total number of genes classified into the respective MapMan bin. P-values denote the probability that the corresponding bin was incorrectly classified as significantly regulated.

**Literature**

**Affymetrix** (2005) Guide to probe logarithmic intensity error (plier) estimation. Technical Report, Affymetrix, Inc., [www.affymetrix.com/support/technical/technotesmain.a.x](http://www.affymetrix.com/support/technical/technotesmain.a.x)

**Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ** (1990) Basic local alignment search tool. J Mol Biol **215:** 403-410

**Barratt DH, Derbyshire P, Findlay K, Pike M, Wellner N, Lunn J, Feil R, Simpson C, Maule AJ, Smith AM** (2009) Normal growth of Arabidopsis requires cytosolic invertase but not sucrose synthase. Proc Natl Acad Sci U S A **106:** 13124-13129

**Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Edgar R** (2007) NCBI GEO: mining tens of millions of expression profiles--database and tools update. Nucleic Acids Res **35:** D760-765

**Benjamini Y, Hochberg Y** (1995) Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B **57:** 289-300

**Bolstad BM** (2004) Low Level Analysis of High-density Oligonucleotide Array Data: Background, Normalization and Summarization. Ph.D. thesis

**Brazma A, Hingamp P, Quackenbush J, Sherlock G, Spellman P, Stoeckert C, Aach J, Ansorge W, Ball CA, Causton HC, Gaasterland T, Glenisson P, Holstege FC, Kim IF, Markowitz V, Matese JC, Parkinson H, Robinson A, Sarkans U, Schulze-Kremer S, Stewart J, Taylor R, Vilo J, Vingron M** (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet **29:** 365-371

**Breitling R, Armengaud P, Amtmann A, Herzyk P** (2004) Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett **573:** 83-92

**Cleveland WS** (1979) Robust locally weighted regression and smoothing scatterplots. Amer. Statist. Assoc **74:** 829-836

**Degenkolbe T, Hannah MA, Freund S, Hincha DK, Heyer AG, Kohl KI** (2005) A quality-controlled microarray method for gene expression profiling. Anal Biochem **346:** 217-224

**Dondrup M, Albaum S, Griebel T, Henckel K, Junemann S, Kahlke T, Kleindt C, Kuster H, Linke B, Mertens D, Mittard-Runte V, Neuweger H, Runte K, Tauch A, Tille F, Puhler A, Goesmann A** (2009) EMMA 2 - A MAGE-compliant system for the collaborative analysis and integration of microarray data. BMC Bioinformatics **10:** 50

**Fourquin C, Vinauger-Douard M, Chambrier P, Berne-Dedieu A, Scutt CP** (2007) Functional conservation between CRABS CLAW orthologues from widely diverged angiosperms. Ann Bot **100:** 651-657

**Gautier L, Cope L, Bolstad BM, Irizarry RA** (2004) affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics **20:** 307-315

**Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J** (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol **5:** R80

**Hannemann J, Poorter H, Usadel B, Blasing OE, Finck A, Tardieu F, Atkin OK, Pons T, Stitt M, Gibon Y** (2009) Xeml Lab: a tool that supports the design of experiments at a graphical interface and generates computer-readable metadata files, which capture information about genotypes, growth conditions, environmental perturbations and sampling strategy. Plant Cell Environ

**Hemm MR, Rider SD, Ogas J, Murry DJ, Chapple C** (2004) Light induces phenylpropanoid metabolism in Arabidopsis roots. Plant J **38:** 765-778

**Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J** (2006) RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. Bioinformatics **22:** 2825-2827

**Ishikawa M, Ohmori Y, Tanaka W, Hirabayashi C, Murai K, Ogihara Y, Yamaguchi T, Hirano HY** (2009) The spatial expression patterns of DROOPING LEAF orthologs suggest a conserved function in grasses. Genes Genet Syst **84:** 137-146

**Martin-Requena V, Munoz-Merida A, Claros MG, Trelles O** (2009) PreP+07: improvements of a user friendly tool to pre-process and analyse microarray data. BMC Bioinformatics **10:** 16

**Mizukami Y, Huang H, Tudor M, Hu Y, Ma H** (1996) Functional domains of the floral regulator AGAMOUS: characterization of the DNA binding domain and analysis of dominant negative mutations. Plant Cell **8:** 831-845

**Orashakova S, Lange M, Lange S, Wege S, Becker A** (2009) The CRABS CLAW ortholog from California poppy (Eschscholzia californica, Papaveraceae), EcCRC, is involved in floral meristem termination, gynoecium differentiation and ovule initiation. Plant J **58:** 682-693

**Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF** (2000) B and C floral organ identity functions require SEPALLATA MADS-box genes. Nature **405:** 200-203

**Rainer J, Sanchez-Cabo F, Stocker G, Sturn A, Trajanoski Z** (2006) CARMAweb: comprehensive R- and bioconductor-based web service for microarray data analysis. Nucleic Acids Res **34:** W498-503

**Robles P, Pelaz S** (2005) Flower and fruit development in Arabidopsis thaliana. Int J Dev Biol **49:** 633-643

**Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J** (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques **34:** 374-378

**Schena M, Shalon D, Davis RW, Brown PO** (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science **270:** 467-470

**Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU** (2005) A gene expression map of Arabidopsis thaliana development. Nat Genet **37:** 501-506

**Smyth GK** (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Statistical applications in genetics and molecular biology **3:** Article3

**Smyth GK, Speed T** (2003) Normalization of cDNA microarray data. Methods **31:** 265-273

**Stahle MI, Kuehlich J, Staron L, von Arnim AG, Golz JF** (2009) YABBYs and the Transcriptional Corepressors LEUNIG and LEUNIG\_HOMOLOG Maintain Leaf Polarity and Meristem Activity in Arabidopsis. Plant Cell

**Street NR, Sjodin A, Bylesjo M, Gustafsson P, Trygg J, Jansson S** (2008) A cross-species transcriptomics approach to identify genes involved in leaf development. BMC Genomics **9:** 589

**Sun J, Loboda T, Sung SJ, Black CC** (1992) Sucrose Synthase in Wild Tomato, Lycopersicon chmielewskii, and Tomato Fruit Sink Strength. Plant Physiol **98:** 1163-1169

**R Development Core Team** (2009) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0

**Usadel B, Nagel A, Steinhauser D, Gibon Y, Bläsing OE, Redestig H, Sreenivasulu N, Krall L, Hannah MA, Poree F, Fernie AR, Stitt M** (2006) PageMan: an interactive ontology tool to generate, display, and annotate overview graphs for profiling experiments. BMC Bioinformatics **7:** 535

**Usadel B, Nagel A, Thimm O, Redestig H, Blaesing OE, Palacios-Rojas N, Selbig J, Hannemann J, Piques MC, Steinhauser D, Scheible WR, Gibon Y, Morcuende R, Weicht D, Meyer S, Stitt M** (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of corresponding genes, and comparison with known responses. Plant Physiol **138:** 1195-1204

**Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M** (2009) A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. Plant Cell Environ **32:** 1211-1229

**van der Merwe MJ, Osorio S, Moritz T, Nunes-Nesi A, Fernie AR** (2009) Decreased mitochondrial activities of malate dehydrogenase and fumarase in tomato lead to altered root growth and architecture via diverse mechanisms. Plant Physiol **149:** 653-669

**Wakeley PR, Rogers HJ, Rozycka M, Greenland AJ, Hussey PJ** (1998) A maize pectin methylesterase-like gene, ZmC5, specifically expressed in pollen. Plant Mol Biol **37:** 187-192

**Wang J, Nygaard V, Smith-Sørensen B, Hovig E, Myklebost O** (2002) MArray: analysing single, replicated or reversed microarray experiments. Bioinformatics **18:** 1139-1140

**Wettenhall JM, Simpson KM, Satterley K, Smyth GK** (2006) affylmGUI: a graphical user interface for linear modeling of single channel microarray data. Bioinformatics **22:** 897-899

**Wettenhall JM, Smyth GK** (2004) limmaGUI: a graphical user interface for linear modeling of microarray data. Bioinformatics **20:** 3705-3706

**Wilson CL, Miller CJ** (2005) Simpleaffy: a BioConductor package for Affymetrix Quality Control and data analysis. Bioinformatics **21:** 3683-3685

**Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F** (2004) A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of the American Statistical Association **99:** 909-917

**Yang YH, Dudoit S, Luu P, Lin DM, Peng V, Ngai J, Speed TP** (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res **30:** e15

**Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W** (2004) GENEVESTIGATOR. Arabidopsis microarray database and analysis toolbox. Plant Physiol **136:** 2621-2632

**Zimmermann P, Schildknecht B, Craigon D, Garcia-Hernandez M, Gruissem W, May S, Mukherjee G, Parkinson H, Rhee S, Wagner U, Hennig L** (2006) MIAME/Plant - adding value to plant microarrray experiments. Plant Methods **2:** 1

**Zrenner R, Salanoubat M, Willmitzer L, Sonnewald U** (1995) Evidence of the crucial role of sucrose synthase for sink strength using transgenic potato plants (Solanum tuberosum L.). Plant J **7:** 97-107