Reviewer 1 Manuscript Review...

*Introduction*

*well written and brings up the main issues. The rationale is well explained.*

*1) What is missing from the list of existing softwares in microarray analysis*

*are those that do something similar (although not at this scale), such as e.g.*

*RACE (Univ. of Lausanne) or Genesifter (commercial in the meantime).*

The manuscript text was modified to include references to relevant publications about similar tools.

*Results*

*Not much to say here, the content is clear and well written.*

*2) Just a small question: whereas the settings of the import wizard can be saved for future*

*upload of similar experiments, it was not clear to me whether an experiment*

*that has been imported can be saved and opened in a subsequent "session". If yes,*

*what would be saved (results, experiment design, raw data, processed data,...)?*

All Robin workflows have been extended to include an import option. When starting a new analysis the user now has the choice between starting a new analysis or opening an existing project. When choosing to import, all settings, the list of imported raw data files and the groups of replicates will be loaded from the imported experiment. The user can freely change all settings, add or remove raw data files and re-run the analysis. To make sure that the imported project will not be overwritten, the results of the modified analysis will be written to a new folder that is named after the imported project plus a user-choosable name extension to distinguish it from the previous analysis. E.g. when importing the project “AFFYTEST”, the new project could be named “AFFYTEST\_NEW”. We think that this extension will be very convenient for many users and also enhance consistency and traceability in situations where the same data set is to be analyzed using different settings and statistical approaches.

*Outlook*

*3) The Robin platform seems to be focused on plants only, although the same*

*problems are faced in other research fields. Is it possible to use Robin for*

*the analysis of e.g. drosophila microarrays? If not, is there any plan to offer*

*such a service?*

Although Robin was developed in a plant-centric institute it is not restricted to plant microarrays only. All Affymetrix GeneChip arrays that are supported by the bioconductor project (for a complete list please see http://www.bioconductor.org/packages/release/data/annotation/).) and all two color and generic single channel array platforms that contain one oligonucleotide probe per target transcript can be analyzed using Robin. The only prerequisite is that the user knows the layout of the chip used. A sentence clearly stating this has been included in the manuscript at the beginning of the RESULTS AND DISCUSSION section.

*4) Although Robin seems to be "biologist-friendly" and may require little*

*support, how will support be provided for users who still have problems? (besides the User Manual)*

A support forum for Robin, that can be used to ask questions and discuss issues not covered by the User’s Manual, has been set up on the MapMan website (<http://mapman.gabipd.org/web/guest/forum>). Users are kindly invited to post their questions there so that the answers can viewed by a larger audience.

*5) Will the software continue to be freely available for academics for a longer*

*period of time? (in recent years, several good tools have become commercial*

*due to shortage of public funding, and some have disappeared).*

Since Robin is containing GNU GPL-licensed software from several projects (e.g. the R project) it is itself distributed under the terms and conditions of the GNU GPL v3. This license ensures that the Robin application and its source code will always be freely available. A passage clarifying this has been added to the manuscript at the end to the “Installation and scope” section.

*Reviewer 2 Manuscript Review...*

*The authors present a stand-alone, all-in-one software that will be of great*

*utility for microarray data analyses by experimental biologists.*

*Although these are several software suites for microarray data analyses,*

*there are not suitable for experimental biologists because of high cost (e.g.*

*GeneSpring) or requirement of statistics and programming skills (e.g. R language). This*

*situation strongly limits proper opportunities for experimental biologists to start*

*microarray experiment or to re-analyze vast amounts of microarray data stored in public*

*repositories. The authors' new software equips a user interface to easily*

*control R language. This software also equips many options for advanced users, so that*

*the target user of this software is broad. These backgrounds, function details and usage with an actual example are concisely written in this manuscript. I couldn't find any problems to*

*improve this manuscript.*

*Reviewer 3 Manuscript Review...*

*The manuscript “Robin: An intuitive wizard application for R-based*

*expression microarray quality assessment and analysis” by Marc Lohse, Adriano*

*Nunes-Nesi, Peter Krüger, Axel Nagel, Jan Hannemann, Federico M. Giorgi, Liam Childs,*

*Sonia Osorio-Algar, Dirk Walther, Joachim Selbig, Nese Sreenivasulu, Mark Stitt,*

*Alisdair R. Fernie and Björn Usadel describes a new application meant to*

*facilitate microarray data analysis for biologists without bioinformatics*

*training. They built a Java application that taps into the power of the R*

*language and the Bioconductor project to process microarray data. The*

*authors demonstrate the usefulness of the software by presenting a case study where*

*they applied Robin to analyze a two color microarray experiment comparing gene*

*expression in tomato leaves, flowers and roots.*

*General comments:*

*The authors succeed at creating a very user friendly application that it can*

*be attractive to both beginners and advanced users. I was able to use the*

*software to analyze two sets of microarray experiments from my own laboratory without*

*a problem the first time I tried it (aside from some installation oddities,*

*indicated below). The application is excellent as it provides many tools in*

*a simple way and it is very well documented. I will definitely be using this*

*software to analyze my own microarray data.*

*I think one great advantage of Robin is to provide R scripts on the fly.*

*This makes the tool really useful for more advanced users, something where “user*

*friendly” applications typically fail. This would also be very useful for*

*teaching purposes.*

*One aspect that I think must be addressed and improved is the follow up to*

*the initial analysis of microarray data. Most of the results discussed in their*

*case study were NOT obtained with Robin but with the tools used after Robin. In*

*this sense, most biological insight cannot be gained by Robin alone. For many*

*people, moving from the normalized data or lists of differentially expressed*

*genes to an application such as Mapman will not be trivial. This may be a*

*hurdle similar to analyzing the data from scratch. The authors could really make*

*their tool a “must have” by making a more direct connection/transition to Mapman*

*(or other available data analysis tool) for the next step in data analysis. For*

*example, in the last step the user may choose between saving the data*

*(current version) or launching Mapman with their data.*

We implemented a function that allows users to directly open the results obtained in the Robin analysis in MapMan and PageMan by simply clicking a “view results in MapMan”-button at the end of the workflow. This functionality will be available with Robin version at least 1.1.x and the next release of MapMan that will also contain PageMan as an integrated module. However, since the implementation has not yet undergone rigorous testing, we do still regard it as potentially unstable and experimental and it is not yet available in the current release of the MapMan application.

*Minor comments:*

*1. I think the warnings the software provides during quality checking are*

*very useful. However, it is not clear to me how the authors decided the*

*default values to issue said warnings. I think this is an important aspect to*

*clarify in the manuscript since many users may proceed (or not) based on the default*

*warning calls of the software. For example, why use 10% for RNA degradation*

*plots? I could not find a justification for this (and other) cutoffs. The*

*reference provided does not justify this cutoff either.*

The thresholds for automatic warnings generation were chosen based on a combination of generally accepted values and values that we set up based on our own experiences. Specifically, the RNA degradation curve slope threshold value of three was used, because it is the quality cut-off applied to the 5’ to 3’ signal intensity ratios of control probe sets like the glycerinaldehyde-3-phosphate dehydrogenase (GAPDH) by most facilities providing Affymetrix hybridization services. Since the RNA degradation analysis performed by Robin uses information from all probes in the data set and it is more important that the global RNA degradation profiles of all chips that are analyzed together are in agreement, we chose to add the 10% and 30% deviation from median cut offs to notify the user of individual chips deviating from the global profile. The values were chosen after reviewing the body of microarray data available in house. The same approach was taken to determine suitable cut offs for the values calculated in the MA plots. In general we preferred to use conservative values that are more likely to produce a warning message in cases where the quality is only moderately reduced to make sure that the user is aware of possible bias in the results. We want to stress that the warnings Robin generates are to be understood as notifications of possible sources of bias rather than strict indications of unusable or extremely bad quality data.

*2. I analyzed an Affymetrix experiment and I wanted to then analyze two-*

*color arrays. However, I was not able to go back all the way to the*

*beginning to make the appropriate selection. I had to restart the software. Although not*

*essential, I think it would be useful to be able to go back all the way to the*

*beginning to start the analysis of a new data set. Specially, considering*

*the software takes a few seconds to load. Just in my tests, I analyzed 5*

*different data sets.*

This issue has been corrected. Clicking the “Restart”-button at the end of any workflow will now restart Robin and lead the user back to the entry screen.

*3. It was not obvious to this reviewer how the software handles designs*

*with more than two experimental factors in Affymetrix data. Are they always*

*pairwise comparisons? If this is the case, how would the user evaluate*

*interactions between experimental factors? The user’s manual indicates both*

*limma and RanProd are used, but I could not find how to select limma to*

*analyze my data with two experimental factors.*

To analyze experiments with more than two experimental factors, we chose to adopt the approach outlined in the limma user’s guide. To define such comparisons, the user can combine simple comparisons that were defined in the graphical experiment designer (please see manuscript figure 1) to “metagroups” by selecting the two groups of interest and clicking the “create metagroup” button. In the subsequent limma analysis such comparisons will be interpreted as e.g. the term (groupA - groupB) – (groupC - groupB). This allows user to extract the “difference of differences” (interaction effect) in cases where two experimental factors are being varied. Please do also see the corresponding paragraph in the manuscript (last paragraph in page 11 starting “Subsequently, the replicate groups are depicted as draggable boxes...“) and the user’s manual (section 3.1.2) .

*4. It would be really useful to be able to save what has been done. I*

*accidentally closed the software a couple of times in the middle of an*

*analysis and I had to upload data, create groups, define the comparisons, etc all*

*over again. This also happened when trying to go back to step 2 to run some*

*quality checks. I accidently hit the button one more time than necessary and*

*everything was reset at step 1.*

Please see the answer to issue 2 raised by reviewer 1.

*5. I tried the software in a PC computer running Windows 7. I have two*

*comments: a. The installer should indicate when it is done installing the software.*

*b. It would also be very useful to users to have icons and/or shortcuts*

*automatically added to the Desktop or Start Menu by the installer. Otherwise*

*the user has to dig into the installation folder, something many people may not*

*find easy to do.*

We could reproduce this undesired behavior of the installer package using the windows 32bit installer to try to install Robin on a 64bit Windows 7 operating system. It does not occur, though, when using this installer package in 32bit windows systems like Windows XP. To solve the problem, we are now providing an installer package specifically adapted to Windows 7 64bit operating systems. This package does not show these errors any more. In our tests on Windows 7 (using an appropriate 64bit Java installation), shortcuts were correctly placed on the desktop and in the start menu if the user chose this option. However, users of 64bit Windows 7 systems that are using a 32bit Java installation might still experience this problem when trying to install Robin from the 64bit installer package.

*6. I could not interpret the Venn diagrams as the names of the groups*

*overlapped in the PNG file. This should be corrected.*

The Venn diagram function has been modified to display the legend, color coded, on the right side and outside of the actual diagram to make sure that the text labels cannot overlap any more.

*7. Robin does not provide the probe to locus identifier translation.*

*Again, the authors could help the users greatly by providing locus*

*identifiers in the output tables in a separate column.*

At the end of the analysis workflows, Robin does now offer to annotate the results using the MapMan functional classification ontology available via the MapManStore website (URL). The MapManStore provides mapping files that contain assignments of gene and probe identifiers to functional categories (“Bins”) of the MapMan ontology. In the case of Affymetrix chips, these mapping files do also contain the identifiers of the genes that are hit by the respective probesets. Mappings for several frequently used plant microarrays platforms have been packaged with Robin for user convenience. In case the user is working on a platform for which a is not included, additional mapping files can be obtained from the MapManStore (by clicking “Download more mappings”). If an appropriate mapping is also not available via the MapManStore, users can contact us directly and we will try to provide a custom mapping using our automated annotation pipeline.

*8. I have not seen the name “GeneVestigator” with capital V. It is*

*probably more appropriate to use the standard name “Genevestigator”.*

The manuscript text has been modified to show the correct name of the Genevestigator service.

*9. The authors may consider opening the software for other programmers*

*to contribute new tools.*

Please see the corresponding comment on issue 5 of reviewer 1 concerning this topic.