09-Apr-2012  
NAR-00599-Web-B-2012  
<i>R</i>obi<i>NA</i>: A user-friendly, integrated software solution for RNA-Seq based transcriptomics  
  
Dear Dr. Lohse  
  
Thank you for giving us the opportunity to consider your manuscript.  
  
The referees have raised substantial criticisms, which are detailed below. We will consider publishing your manuscript only if you can accommodate their suggestions in a revised version or explain satisfactorily why their comments are invalid.  
  
Detailed instructions for submitting your revised manuscript are provided BELOW the referees' reports. When you submit your revised manuscript, you should provide a concise point-by-point response to the referees’ comments. Any text in the manuscript that you change or add in response to referee or Editor comments should be marked in red. You should also upload high-resolution figures which conform to the minimum resolution and special requirements outlined below.  
  
The revised version must be uploaded within 20 days of the date of this letter.  
  
NAR is fully Open Access, and ALL papers published are made freely available at the NAR website and also at PubMed Central immediately upon publication.  
For papers published in 2012 the Open Access charge for a corresponding author based at an institution with NAR membership will be $1385/£710/€€1065, while the charge for a corresponding author at an institution which does not have NAR membership will be $2770/£1420/€€2130.  
Papers published that occupy more than 9 pages will incur an additional charge of $195/£100/€150 for each page in excess of 9. There are no excess page charges for papers that occupy 9 pages or less. Waivers or discounts will be considered sympathetically for corresponding authors from developing countries and those in genuine hardship. Further information is available from our website ([http://nar.oupjournals.org/openaccess](https://outlook.mpimp-golm.mpg.de/owa/redir.aspx?C=905289e2b14d405e856ffc72fd95f2ff&URL=http%3a%2f%2fnar.oupjournals.org%2fopenaccess" \t "_blank)).  
  
We look forward to receiving your revised manuscript.  
  
Yours sincerely,  
  
Gary Benson  
  
Executive Editor  
Nucleic Acids Research

We thank you and the two referees for the rapid review of our manuscript: “RobiNA: A user-friendly, integrated software solution for RNA-Seq based transcriptomics”. We were very pleased to read that both reviewers found RobiNA to be a useful and easy to use tool, and appreciate their constructive suggestions on how we might improve both the tool and the manuscript.

We have now revised our manuscript and we should be grateful if you would consider the revised version for publication in the special issue of *Nucleic Acids Research*. We have listed our responses to the reviewers’ comments below.

Reviewers' Comments to Author  
  
(Line numbers mentioned in a report may not coincide with the original line numbers.)  
  
Referee: 1  
Comments for the Author  
The authors have developed a good user-friendly tool to analyzed RNA data. However, maybe some features can be improved:  
  
1-The demo only include files to be mapped to the transcriptome, it would be nice to have the genome and the gff file to test this function also

We have added the current *Arabidopsis thaliana* TAIR10 genome release sequences and GFF3 annotation file to the test set available from our web site.

2-It would be great that when you open a project that exists you can add new samples without analyze everything again (If the tools already does this, please put in the tutorial how to do it). Moreover, I had some problem when open a project that exists, the tool did not recognize any sample after the read filtering, when the user has to design the experiment.

When importing an existing project, the quality checking and trimming steps can now??? be skipped, since the data imported have already been checked/trimmed in the analysis that is imported. However, when adding additional input files, the mapping and analysis steps have to be repeated. We implemented this behaviour based on the assumption that a user who imports an existing project does this most likely because they want to modify the settings of the mapping and statistical analysis. However, when the analysis is to be repeated with identical settings, but with additional input data files, RobiNA does not yet provide the option to run the mapping step just for these new files (to save time). We are very grateful for pointing out this shortcoming and will include this feature in the next release of the tool.

We have corrected a bug causing the samples to disappear upon import of an existing project so that project import now works properly on all operating system available to us for testing.

3-It would be good to have a quick start showing how to use it in 5 steps

Thank you for this very good suggestion. We have created a screenshot-based quick start guide that can be downloaded from the RobiNA website.

4-Can the program detect isoforms?

No – in its current state, RobiNA works on the level of genes when using a genome as the reference and on the level of transcripts if a transcriptome is provided. If the provided transcriptome contains isoform transcripts, only those reads that map into the unique regions of the isoforms will be considered for differential gene expression analysis. Proper handling of reads that map ambiguously e.g. to several transcript isoforms is a non-trivial task, however, we are currently working on including this feature in the next version of the software.

We have clarified this behaviour in the accompanying software manual.

5-If reads are longer than 36 (72) it is more difficult to bowtie to map them. Those this tool fragment the reads in this case? and how longer reads would affect the analysis?

RobiNA does not fragment reads that are longer than 36 nucleotides although it has been discussed that with increasing read length bowtie’s inability to allow for gapped alignments and more than 3 mismatches in the seed region becomes problematic. The upcoming new version of bowtie (bowtie2) will allow for gapped alignments and also offers a range of other improvements and options. However, we decided against including it in the current release of RobiNA since it is still in beta testing stage, despite its superior performance (Langmead and Salzberg, 2012 Nature Methods) . Once having left the beta stage we will include bowtie2 after extensive testing as the standard read mapper.

6-If the tool is installed in linux without root permissions gives an error because it has no permission to install R packages or similar.  Is there a way to install first all the dependencies, so it would not be necessary for the \*jar to do anything else?

This depends on the individual configuration of R on the host computer and hence is difficult to tackle by the RobiNA installer. By default, R tries to install new packages for all users, however, the user can choose to install packages in a user-writable directory. This would eliminate the problem and allow installation of new packages to proceed without problems. We have included a paragraph specifically dealing with setting up and running RobiNA on linux in the manual.

6-After differential expression analysis, it would be nice to show which R package the user used to calculated (in the title, or beginning of the output pdf). It helps when you compare the output using one or another.

This feature is in fact present in the submitted version of RobiNA... The packages used and also the settings are given and the references are included the summary that is shown at the end of the analysis. A complete R session information is given at the end of the summary. Summaries of finished projects can be viewed by opening the corresponding project in RobiNA and choosing the “browse” option. Several analysis summaries can be opened in parallel allowing for comparison of the results. However, as this information appears to be easily overlooked, it is now provided as an additional output (PDF) file that is automatically generated and saved in the project folder.

7-It has been demonstrated that there is a bias in the CG content of RNA-seq. And there is package trying to solve that. It would be good to add this to the tool.

This is a very good suggestion, we have therefore added the GC content bias correction methods described by Risso et al., 2011 (package EDASeq) as an option in the statistics settings panel. When this option is not activated, the normalization will be performed according to the methods suggested in the edgeR and DESeq packages.

Referee: 2  
Comments for the Author  
In this paper the authors present a bioinformatic tool for analyzing RNAseq data. The application covers all required steps (i.e., quality control, mapping, assessment for differential expression, ...) except for normalization. I assume that this step is performed using the methods described in DESeq and edgeR packages. However, this can be a limitation since these methods do not consider, for instance, GC-content information.   
  
The program also considers the use of raw data as well as mapped data obtained from other tools (sam/bam files). I found the interface to be intuitive and user friendly. Furthermore, most of the options are automatized so that non-specialized users can use it without any problem. The Quality Control (QC) plots are well documented and it considers several QC filters. The application is well documented and easy to be installed.   
  
I believe that both the topic and application are of great interest for research community. However, I have found several problems when analysing real data. Herein, I provide a list of them:  
  
  
Major comments  
--------------------------------------------  
  
A. RNAseq Normalization  
  
There is a lot of literature about RNAseq normalization procedures that are also available at Bioconductor (e.g., EDAseq, cqn, ...). As I previously mentioned, I assume that this step is performed using the own methods proposed by DESeq and edgeR packages (I could not check it because the program crashes before doing it). These other normalization methods should be implemented since it has been shown that results may be biased when, for example, GC-content is not considered in the normalization step.

This is indeed an excellent suggestion and as already mentioned in the response to issue no. 7 of reviewer 1, we have included the GC content bias correction methods described by Risso et al., 2011 and implemented in the EDASeq Bioconductor package. Users can select this GC content normalization of the statistics settings panel and specify which combination of within- and between lane normalization is to be applied. If this option is not selected, the normalization methods described in the selected DE analysis package will be used.  
  
B. General practical issues   
  
1. When selecting as input a SAM/BAM file, if you try to go backwards it forgets about your selection and takes you to the menu of read mapping

Quality checking of reads taken from BAM/SAM files will be integrated into the RobiNA workflow in the next release. To make sure the interface stays consistent in the current release, we have disabled the possibility to go backwards from the experiment layout step when importing SAM/BAM files.

2. When trying to recover a previous project, the program freezes and does not answer

The project import functionality, available via the “Open existing project” option, has been revisited and tested on all operating systems available to us (Windows XP, Windows 7, Mac OSX 10.6, Ubuntu Linux 11.10). Several bugs causing inconsistent behaviour have been fixed Based on our testing of the updated program, we believe this problem has now been eliminated, at least on the above operating systems.

However, due to the heterogeneous nature of the possible operating systems and Java JVMs RobiNA could be installed on, we cannot guarantee that problems occurring only under certain circumstances (e.g. combinations of OS and JVM and R engine) are impossible. In such situations, detailed information on the system used and further configuration details are crucial to finding and eliminating the problem. Such information can be posted on our forum, and we encourage all users to provide feedback in this way to enable us to improve the capabilities and usefulness of the tool.

Regarding the specific case of the crashes reported by Reviewer 2, we are unfortunately lacking this information and are hence not able to determine whether the changes we made eliminate this particular error. However, as we have also posted RobiNA as a software on the SEQanswers Forum (<http://seqanswers.com/>) (besides its publication in the seqanswersWIKI some time ago) we hope that the very active user base will find errors caused by an unusual combination of operating system allowing us to respond to these problems. Feedback gained from the seqanswers forum (and all other) users will also be used to improve the functionality to make RobiNA a useful tool for the RNA-Seq community.

3. Errors during fastq inputting process do not give too many details

This was indeed the case, and we thank the reviewer for pointing out this problem. The fastq import routines have been revised and should now stably read all current fastq versions. In case of input file corruption or when RobiNA encounters fastq files with an unexpected non-standard format, detailed error dialogs point out why the data could not be imported.

4. I tested for several different databases (even for the one available at the RobiNA homepage) and, when the process got to the differential expression (DE) assessment, the program shut down giving random R errors (such as error in scan... or error in exactTest...). This is a big problem because the user must repeat the whole process again (unable to recover a previous project!)

We are sorry for this unfortunate behaviour and having caused the reviewer trouble in assessing our software. We could unfortunately not reproduce this bug on the operating systems available to us. Errors raised by the scan function suggest that a file could not be read properly which might hint at file permission problems or maybe the inaccessibility of a file on a network resource. To track down how and why this specific scan error could occur would unfortunately require more detailed.

R errors in the exactTest function might be related the experiment setup e.g. regarding conditions without replicate samples. We have identified several experiment layout configurations under which crashes of the R analysis process occurred and revised the scripts to handle these situations without terminating the whole analysis. However, especially when using RobiNA on operating systems other than Windows and Mac OS X with an external R installation, we cannot guarantee that the scripts will function with versions of R / Bioconductor other than the one they were designed for. Development of the Bioconductor packages for RNA-Seq analysis is very lively and the functionality and function signatures may change significantly from version to version. The current version of RobiNA is delivered with R 2.15.0 and Bioconductor 2.10.

Since the problems described were not observed by reviewer 1 they might be related to the specific environment and or experiment setup used by reviewer 2. As laid out above (in the response to issue no.2) we hope to identify and eliminate possible errors that occur on specific platforms by drawing on feedback from the active user community.

C. Graphical issues  
  
1. Some text (e.g. in the raw read pre-processing step) gets hidden under other menu and it is not possible to read it

We are sorry for this and have adapted the visualization proecedures, the new graphical user interface has been thoroughly tested under Linux, Windows and Mac OS X since it turned out that the glitches described are not showing on all platforms. The current version 1.2.2 of RobiNA should hopefully not show this poor behaviour any more.

2. When selecting conditions for each sample, the menu appears and disappears randomly. Sometimes it only appears part of it

We are sorry for this behaviour but think that it was caused by the some problem that caused C1 and hope to have abolished the problem. Please see response to issue C.1 above.

Minor comments / Recommendations to improve the application  
------------------------------------------------------------  
  
I would recommend to add these features to improve the application  
  
  
1. The program only accepts fastq or SAM/BAM formats. It would be interesting that the program also allows the user to incorporate count tables such as those available at the Re-Count repository ([http://bowtie-bio.sourceforge.net/recount/](https://outlook.mpimp-golm.mpg.de/owa/redir.aspx?C=905289e2b14d405e856ffc72fd95f2ff&URL=http%3a%2f%2fbowtie-bio.sourceforge.net%2frecount%2f" \t "_blank)). Most of the final users (biologists, medical doctors, ...) may only have access to the table of counts since raw data can be processed by sequencing core facilities using their own pipelines

We were already thinking along these lines, and the reviewer’s helpful suggestion confirms that this would be a useful feature. Therefore, we included it in the current version (1.2.2) of RobiNA.  
  
2. The program only considers a limited number of configurations when using edgeR and DESeq. Current versions of these packages also includes other more sophisticated methods for dealing with overdispersion and variance estimation. Furthermore, there exists other BioConductor packages that could also be considered for DE testing

This is indeed the case, we updated the embedded R and corresponding bioconductor versions to R-2.15.0 and bioconductor 2.10 respectively. New options for estimation of overdispersion and variance implemented in the latest edgeR and DESeq packages are now available on the statistics settings panel of the latest version of RobiNA (v1.2.2).

We are also currently working on the integration of further DE analysis packages (e.g baySeq, BitSeq) into the RobiNA workflow to provide a wider range of methods. These will be included in the upcoming releases of the tool.

Instructions for submitting your revised manuscripts:  
  
\*Submission  
Log on to the online submission web site (http://mc.manuscriptcentral.com/nar ) as before and, in your 'Author Centre', click on 'Manuscripts with Decisions'. At the bottom of the screen you will see those manuscripts that require a revision (or that have been revised). Create a revision of this manuscript by clicking on 'create a revision' under Actions. You will now be able to see the editor and reviewer comments and will be able to respond to these. The ‘Upload files’ screen will automatically be populated with the files that you uploaded at initial submission. You should delete all files that have been changed during revision and upload your revised files in their place, by the procedure used during initial submission. If you click on 'View comments/respond' you will see the editor's letter to you together with the referees' comments. You should cut and paste your responses into the text areas at the bottom of the screen. Full instructions are available from  
[http://www.oxfordjournals.org/jnls/list/nar/instauth/msprep\_submission.html#revision](https://outlook.mpimp-golm.mpg.de/owa/redir.aspx?C=905289e2b14d405e856ffc72fd95f2ff&URL=http%3a%2f%2fwww.oxfordjournals.org%2fjnls%2flist%2fnar%2finstauth%2fmsprep_submission.html%23revision" \t "_blank)  
If your revised manuscript has been successfully submitted, you will see a confirmation screen showing your manuscript number; this will be the same as that of your initial submission with the prefix 'R1' (or R2, R3 as appropriate). You will also receive an email confirming the submission. If you do NOT receive this email your revised manuscript will not have been submitted successfully.  
  
\*Manuscript file format  
Files must be in .doc, .rtf or LaTeX format.  
Revised manuscripts are NOT accepted in Portable Document Format (\*.pdf), as this cannot be used for publication.  
  
\*Figures  
600 dpi is required for line drawings (black and white) and 300 dpi for colour and greyscale. Colour figures must be supplied in CMYK not RGB colours.  
  
\*Supplementary data  
This must be uploaded as separate files.  
  
\*References  
Check references very carefully; their accuracy is your responsibility. Please note that full titles of all cited articles are required. Check that you are using a recent version of 'Endnote' or other reference manage package which has the correct NAR format. Manuscripts ‘submitted’ or ‘in preparation’, unpublished results, personal communications, statistical packages, computer programs or web site addresses must go in the text only.  
  
\*Conflicts of interest  
NAR policy requires that authors of all manuscripts reveal any conflicts of interest as detailed in the Instructions to Authors: [http://www.oxfordjournals.org/nar/for\_authors/ed\_policy.html](https://outlook.mpimp-golm.mpg.de/owa/redir.aspx?C=905289e2b14d405e856ffc72fd95f2ff&URL=http%3a%2f%2fwww.oxfordjournals.org%2fnar%2ffor_authors%2fed_policy.html" \t "_blank).  
As corresponding author you are responsible for bringing this to the attention of all co-authors of this manuscript. During submission of your manuscript you may have already completed a declaration on behalf of all authors if no author has a Conflict of Interest to declare. However if this was not possible each author must complete and return an individual copy of the form.  
The form is available at the following link:[http://www.oxfordjournals.org/our\_journals/nar/for\_authors/conflictofinterestform.pdf](https://outlook.mpimp-golm.mpg.de/owa/redir.aspx?C=905289e2b14d405e856ffc72fd95f2ff&URL=http%3a%2f%2fwww.oxfordjournals.org%2four_journals%2fnar%2ffor_authors%2fconflictofinterestform.pdf" \t "_blank)  
When you submit the revised version of the manuscript, please fax the form(s) to the Senior Editorial Office (+44 2380 597 748). If your manuscript is published, this information will be communicated in a statement in the published paper.