SCIENTIFIC REPORTS

OPEN

miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis

Received: 28 January 2016
Accepted: 21 April 2016
Published: 11 May 2016

Eduardo Andrés-León*,†, Rocío Núñez-Torres*,‡ & Ana M. Rojas

Large-scale RNAseq has substantially changed the transcriptomics field, as it enables an unprecedented amount of high resolution data to be acquired. However, the analysis of these data still poses a challenge to the research community. Many tools have been developed to overcome this problem, and to facilitate the study of miRNA expression profiles and those of their target genes. While a few of these enable both kinds of analysis to be performed, they also present certain limitations in terms of their requirements and/or the restrictions on data uploading. To avoid these restraints, we have developed a suite that offers the identification of miRNA, mRNA and circRNAs that can be applied to any sequenced organism. Additionally, it enables differential expression, miRNA-mRNA target prediction and/or functional analysis. The miARma-Seq pipeline is presented as a stand-alone tool that is both easy to install and flexible in terms of its use, and that brings together well-established software in a single bundle. Our suite can analyze a large number of samples due to its multithread design. By testing miARma-Seq in validated datasets, we demonstrate here the benefits that can be gained from this tool by making it readily accessible to the research community.

The development of high-throughput next-generation sequencing (NGS) technologies has revolutionized the transcriptomics field, paving the way for large-scale RNA sequencing (RNA-Seq)[1]. RNA-Seq can not only be used to study genome-wide transcription but also, it offers the ability to discover new genes and transcripts[2] or to identify additional elements, such as new non-coding RNAs, small interfering RNAs (siRNAs), small nucleolar RNAs (snoRNAs) and micro-RNAs (miRNA). Recently, a new class of RNAs has been described, called circRNAs[3], that are characterized by their ability to form circular RNA through a covalent linkage at the ends of a single RNA molecule. These circRNAs seem to participate in the regulation of gene expression, acting as regulators of miRNAs by specific binding to them. The appearance of these new regulatory molecules has led to the development of new tools for the identification of circRNAs, also through RNA-Seq experiments[4].

There are two important aspects of RNA-Seq experiments, the vast amount of data generated in this kind of study, and the ability to extract and interpret biologically relevant information. These issues are particularly relevant since transcriptomics data analysis can easily become an important experimental bottleneck, especially given the additional constraints that both RNA-Seq and miRNA-Seq analyses impose. Indeed, the combination of different statistical and bioinformatics tools with many customizable parameters often makes such analysis difficult for non-experienced researchers. In addition, the use of different tools may involve time-consuming installations, usually requiring human intervention to proceed to the next step. To alleviate this problem, several tools have been generated for gene expression analysis, like ExpressionPlot[5], GENE-counter[6], RobiNA[7], TCW[8], Grape RNA-Seq[9] or MAP-RSeq[10]. In addition, another set of tools focuses on the analysis of miRNA expression profiles, such as DSAP[11], miRanalyzer[12], miRExpress[13], miRNAkey[14], iMir[15], CAP-miRSeq[16], mirTools 2.0[17] or sRNAtoolbox[18]. Moreover, a few tools have been implemented to perform both RNA-Seq and miRNA-Seq analysis, such as wapRNA[19], eRNA[20], BioVLAB-MMIA-NGS[21] or Omics Pipe[22]. Other available methods integrating

Instituto de Biomedicina de Sevilla (IBIS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Computational Biology and Bioinformatics Group, Seville, Spain. †Present address: Bioinformatics Unit, Instituto de Parasitología y Biomedicina "López Neyra", Consejo Superior de Investigaciones Científicas (IPBLN-CSIC), PTS Granada, Spain. ‡Present address: Unidad de Enfermedades Infecciosas y Microbiología Clínica. Hospital Universitario de Valme. Instituto de Biomedicina de Sevilla (IBIS). Seville. Spain. *These authors contributed equally to this work. Correspondence and requests for materials should be addressed to E.A.-L. (email: eduardo.andres@csic.es) or A.M.R. (email: arojas-ibis@us.es)

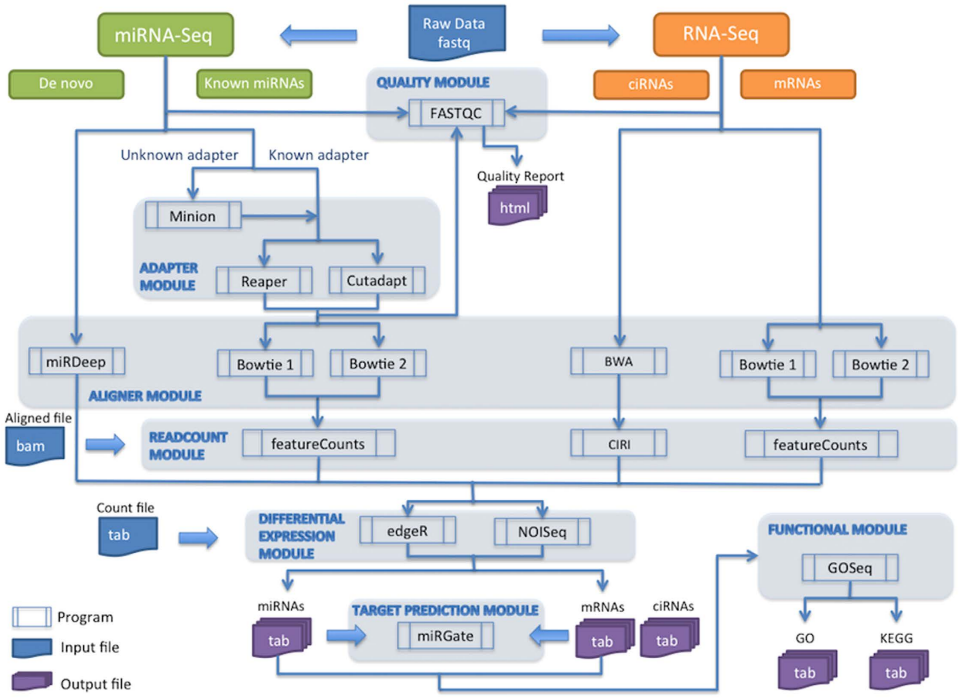

Figure 1. miARma-Seq pipeline workflow. An overview of the modular design of the pipeline. Main modules are indicated by gray background. Output files are indicated by purple background.

several software enabling different type of NGS analyses are GALAXY (https://galaxyproject.org/), QuasR[23], RAP[24], Subread/edgeR[25], while others provide a collection of modules to process files, like the ViennaNGS[26] suite.

Although extremely valuable, the main disadvantage of these tools is that, with some exceptions, they often still rely on manual installation procedures and further human input, steps that have proven difficult to automate. There are also other issues that hamper their wider diffusion and implementation: i) some of the tools have been designed to work on web-based platforms with the consequent restriction on data upload or limited offer of parameter's choice (i.e Galaxy, RAP[24], BioVLAB-MMIA-NGS[21], or DSAP[11]); ii) the analysis pipelines implemented have rigid workflows, so users cannot start the analyses at different steps of the pipeline (i.e. RAP[24], BioVLAB-MMIA-NGS[21]); iii) some of these tools have a large list of pre-requisites for local installation that complicates their use by less experienced researchers (i.e.: Cap-miRSEq[16], Omics Pipe[22], iMir[15], Galaxy, ExpressionPlot[5]); iv) the analysis is usually restricted to a few selected model organisms (i.e. QuasR[23], ExpressionPlot[5], BioVLAB-MMIA-NGS[21]), and iiv) some tools uses in-house code which has not been extensively tested in the NGS community (i.e. Grape RNA-Seq[9] or ExpressionPlot[5]). In addition, to our knowledge, none of these tools has implemented a pipeline for the analysis of circRNAs.

With these limitations in mind, we have developed a comprehensive pipeline analysis suite called "miARma-Seq", which stands for *miRNA-Seq And RNA-Seq Multiprocess Analysis*, that is designed to identify mRNAs, miRNAs and circRNAs, as well as for differential expression, target prediction and functional analysis. Most importantly, it can be applied to any sequenced organism, and it can be initiated at any step of the workflow.

Results

miARma-Seq main features. The most important aspect of the suite is that it is a stand-alone tool that is both easy to install and extremely flexible in terms of its use as compared to other methods (Supplementary Table S1). It brings together well-established software in a single bundle, allowing a complete analysis from raw data (Fig. 1). All the capabilities can be easily and simultaneously enabled at will, regardless the step in the workflow. We have tested miARma-Seq using different published datasets of miRNAs, mRNAs, and circRNAs to illustrate some of the tool capabilities and to show that the tool works as expected (Supplementary Tables S2 and S3, Supplementary Results).

miARma-Seq computational performance. We computed the performance in terms of time used to run each analysis (Table 1). To illustrate miARma-Seq's capabilities in the analysis of miRNA-Seq data, we processed the SRR873382 sample from the GSE47602 experiment with our pipeline in a standard computer, with 8 GB of RAM memory and a 1.7 GHz CPU. The sample selected has over 35 million reads and it was analyzed using the multithread option (4 threads). Total analysis of the sample took 19 minutes and it included: a quality analysis (2 min); pre-processing of the sample that included adapter removal with Reaper[27] (6 min); alignment against a reference genome with Bowtie 1[28] (8 min); and summarizing of the read counts (3 min). Similarly, the SRR873382 sample was also processed with miARma-Seq to identify novel miRNAs (DeNovo analysis). The total analysis of

Type	EXP	Sample	Scope	Reads	T*	Q*	PRE*	ALIGN*	SUMM*
miRNA	GSE47602	SRR873382	D	34686701	19	2	6 (Reaper)	8 (Bowtie 1)	3 (FeatureCounts)
miRNA	GSE47603	SRR873383	I, dN	34686701	48	2	<1 (mirDeep2)	45 (mirDeep2 uses Bowtie 1 for the alignment)	
mRNA	GSE52778	SRR1039508	mD	45871042	172	6	**	161 (TopHat/Bowtie 2)	5 (FeatureCounts)
circRNAs	GSE49321	SRR1051292	cD	6767745	48	3	**	35 (BWA)	10 (CIRI)

Table 1. Performance of miARma-Seq tool. The analyses have been performed in an average computer with 8 GB of RAM memory, 1.7 GHz CPU, and 4 threads. **EXP**: Experiment identifier. **SCOPE**: D stands for "detection" of known miRNAS, ID indicates "identification", dN indicates "de novo" prediction of miRNAS, mD indicates "detection" of mRNAS, cD indicates "detection" of circRNAS. **Reads:** Number of reads per sample. **T**: Total analyses time, **Q**: Quality analyses time (FastQC). **PRE**: Preprocessing time (Software used). **ALIGN**: Alignment time (Software used). **SUMM**: Summarization of read counts time (Software used). *Time in minutes. **The use of pre-processing stage will depend on the sequencing process.

the sample took 48 minutes including: a quality analysis (2 min), pre-processing of the sample (<1 min); alignment against reference genomes; and summarizing the read counts with miRDeep 2[29] (45 min).

The performance of miARma-Seq for RNA-Seq data analysis was also evaluated using a paired-end SRR1039508 sample of nearly 44 million reads from the GSE52778 experiment and using 4 threads on a standard computer as describe above. The total analysis of the sample took 172 minutes and it included: a quality analysis (6 min), alignment against reference genomes with TopHat[30]/Bowtie 2[31] (161 min) and the summarizing of the read counts (5 min).

In order to demonstrate how miARma-Seq can be implemented to analyze circRNAs we processed the SRR1051292 sample from the GSE49321 experiment on a standard computer set up as indicated above. The sample selected was a paired-end sample with over 7 million reads and it was analyzed using the multithread option (4 threads). The total analysis of the sample took 48 minutes and it included: a quality analysis (3 min), an alignment against a reference genome with BWA[32] (35 min) and the summarizing of the read counts with CIRI[4] (10 min).

Evaluation of miARma-Seq. Proper validation studies to produce meaningful statistics cannot be conducted with the available experimental data, which in its current state does not provide complete information of false positives (FP) or false negatives (FN). Therefore, precision cannot be calculated. Nonetheless, we calculated correlations with previous studies as a proxy to control whether or not the tool was performing as expected.

In this regard, we used miARma-Seq to analyze three case scenarios. In miRNA transcriptome analyses, we analyzed the expression analysis of regulated miRNAs from a time course experiment under different hypoxic conditions, and miRNA novel detection as well. We next run mRNA genome-wide analyses, and finally we performed analyses to detect circRNAs from RNA-seq data. All these analyses showed a strong overlapping with previous studies (Supplementary Tables S2 and S3, Supplementary Fig. SF1), confirming that the pipeline is working as it should be expected.

Discussion

High throughput NGS technology is now being widely employed for many purposes in scientific research. Although the vast majority of the existing tools are very valuable and efficient in analyzing such data (Supplementary Table S1), most of them are difficult to be tested by users without basic programming and computing skills. In this regard, the NGS computational-based community has made a tremendous effort to make these tools usable. For instance, software's interoperability is no longer an issue, as most of them enable local installation for different operative systems (Supplementary Table S1).

A widely used strategy to optimize usability is the development of web-based versions of particular tools, which in one hand comes with a cost in terms of several restrictions. For instance, most of web-based systems (i.e., Galaxy, RAP[24]), may present issues intrinsically related to web-traffic (i.e. bandwith limits, data privacy policies affecting users, and/or queue saturation, among others). Even when these systems enable their local installation, the process can be very demanding in terms of required expertise to handle the installation (i.e., Galaxy). They also can be limited in the number of analyses or data uploads (i.e., RAP[24]), or they may be built upon in-house code, which has not been extensively tested in the NGS community (i.e., DSAP[11]).

However, the main issue that hampers the difusion of these methods, relies on difficulties related to overcome software dependencies when installing, which usually requires advanced computing skills. All together, these restrictions make the full potential of any tool too difficult to be properly exploited by non-experienced users (QuasR[23], Subread/edgeR[25], Cap-MirSeq[16], Omics pipe[22], etc).

Here, we present miARma-Seq, a comprehensive pipeline analysis for RNA-Seq and miRNA-Seq data suitable to identify mRNAs, miRNAs and circRNAs in any organism with a sequenced genome, and to analyze their Differential Expression. This tool aims to resolve some of the main problems that researchers might encounter when analyzing high throughput sequencing data: i) integrated capabilities that allow data from both miRNA and mRNA expression experiments to be analyzed; ii) easy installation, without having to check otherwise tedious requirements that make the analysis more difficult for non-experienced researchers; iii) no restrictions regarding the reference organism as it is not restricted to the analysis of only a few model organisms but rather, it can be used with any reference organism as long as its genome has been sequenced; iv) flexibility in terms of fitting any experimental design, allowing the user to start the analysis at different steps of the pipeline; v) Speed of execution,

allowing the pipeline to be executed in a standard computer or in a cluster environment (profiting from its parallelizing properties); vi) reliability, as the tool performs the analysis with the most of the standard tools available; vii) Wide coverage, which not only includes miRNA-Seq and RNA-Seq data analysis to identify miRNAs, mRNAs or circRNAs but also, the possibility to carry out differential expression analysis, target prediction and functional analysis, and to predict novel miRNAs. We have tested miARma-Seq using different published datasets and the results obtained correlated well with the original sources.

A particularly interesting characteristic of miARma-Seq, is the possibility to detect and identify circRNAs from RNA-Seq data, and to analyze their differential expression. The recent description of these molecules means only a few tools have been developed to detect circRNAs and to our knowledge, no pipeline has included this kind of analysis.

One of the main goals of our pipeline is to offer a fast and a flexible tool for miRNA and RNA-Seq data analysis. In this sense, we demonstrated that using a standard computer that can be found in any research laboratory, the median time required to detect known miRNAs, novel miRNAs, mRNAs and circRNAs in a sample is 19, 48, 172 and 48 minutes, respectively. Obviously, these are illustrative processing times, since the time required will depend on aspects such as the depth of the sample, the tool selected for analysis, the steps included in the analysis and the computer employed.

In contrast to some web-based tools, miARma-Seq is designed for local execution on small computers or at big computing infrastructures that take advantage of multithread analysis, which is especially recommended to rapidly analyze huge amounts of data.

This kind of analysis is usually faster than that offered by web-based tools, since one of the main problems in these tools is that the processing time depends on the number of jobs pending in the queue and on the server's capacity.

In order to make it easier to use our pipeline, we have developed a stand-alone tool with extremely easy to install software (downloading only one unique file), with minimum software requirements. Therefore, we consider that miARma-Seq offers the speed of a stand-alone tool with the ease of a web-based tool.

In conclusion, miARma-Seq is a powerful and very flexible tool for transcriptomics (miRNA, mRNA and circRNAs), which offers additional capabilities including the identification of differentially expressed entities, miRNA-mRNA target prediction, or functional analysis (Supplementary Figs S1 and S2). The miARma-Seq pipeline can perform a fast and reliable analysis of a large number of samples due to its multithread design and the established quality of the software included (Supplementary Table S2). The tool can be used by a wide range of users, from those having little experience in programming and computing in both installation and usage to advanced users more accustomed to the command line handling. It runs in three different operative systems, and it has been tested for reliance on software updates.

Notably, it successfully provided well-correlated data when compared with validated and published results obtained from different transcriptomics analyses (Supplementary Tables S2 and S3), demonstrating the utility of this tool for the research community.

Methods

Code availability. A stand-alone version has been developed to minimize software requirements and to simplify its installation on any computer, while allowing non-expert users to perform complex analyses. The tool has been tested in the latest Fedora 23, Centos 7.2, Ubuntu (from 10.4 to 14.04) and Debian jessie. It has been also tested in Apple computers from 10.9 to 10.11.

The tool's performance is not affected by updates of the included software and the current version contains the most updated software. Nonetheless, previous versions of miARma-Seq are also available with outdated versions of the included software, so the user can compare results (i.e. to check reproducibility among different versions of the software). In the case of CIRI[4], we offer the choice of methods via configuration file.

The stable code in miARma-Seq is freely available at https://sourceforge.net/projects/miarma/.

The development code in miARma-Seq is freely available at https://bitbucket.org/cbbio/miarma/src.

A complete guide for the installation, analysis of miRNA, mRNA and circRNA data is provided, along with different examples of its use, at http://miarmaseq.cbbio.es/.

Description and implementation. To facilitate the installation process, we provide three different possibilities of installing the software for Mac and Linux (Source code, four Docker images, and a virtual Image), and two possibilities for Windows 7, 8, 8.1 and 10 (four Docker images and a virtual Image). Detailed information on how to install is available here (http://miarmaseq.cbbio.es/installation).

The miARma-Seq tool was implemented as a combination of Perl and R scripts for Unix environments. To further simplify its use, the miARma-Seq pipeline has preconfigured and customized parameters that can be executed with a simple command line program and a configuration file. Nevertheless, miARma-Seq can be used by more experienced bioinformaticians, since Perl modules are freely available and can be configured in the custom pipelines.

The miARma-Seq tool has a highly flexible modular structure designed to perform the different stages of analysis (Fig. 1). This structure offers the possibility to start the analysis at any point in the workflow by simply providing a configuration file. This characteristic is most useful for the analysis of pre-processed data from public databases, which usually implies aligned bam or tab separated read-count files. In addition, the configuration file allows the user to define which software to use at each step. Alternative software can be used as a single election or simultaneously. Nevertheless, the resources included in miARma-Seq have been selected by comparing and evaluating software widely used in RNA-Seq and miRNA-Seq data analysis[33–36], or based on the advantages offered for the analysis (i.e.: adapter prediction or replicate simulation).

Although the analysis can start at any point in the pipeline, the typical input files in sequenced transcriptome experiments are fastq files, a common format used in miRNA, mRNA and circRNAs studies. These files are processed by miARma-Seq with specific software for each data type, while the main analysis stages are shared.

Quality assessment and pre-processing. Initialy, all the fastq files are submitted for quality evaluation using FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), which provides a detailed report of the quality of the reads. In order to customize the pre-processing of the reads, miARma-Seq offers different tools for customization (for details see Supplementary methods). Cutadapt[37] and Reaper[27] are included for trimming purposes. The Minion software[27] enables to perform an adapter sequence prediction. In addition, miARma-Seq also includes an in-house tool to remove a specific number of nucleotides from the 3′ or 5′ end that usually contain low quality information.

As a result of the read trimming, miARma-Seq generates plots with the length distribution of the reads in each sample (Supplementary Fig. S2). After pre-processing, miARma-Seq offers the possibility of performing additional quality control analyses with FastQC to assess the quality of the processed reads.

Alignment. This stage differs according to the kind of the data to be analyzed. For miRNA expression studies, read alignment can be performed with Bowtie[28], Bowtie2[31] or both. In addition, miARma-Seq includes mirDeep2[29] to predict novel mirnas (for details see Supplementary methods). TopHat[30] is available, which can also execute the miRNA alignment software. In both cases, miARma-Seq offers the possibility to construct genome indices from fasta files offering the possibility to analyze any organism with a sequenced genome. For circRNAs detection the BWA aligner[32] is available (for details see Supplementary methods).

Entity quantification. FeatureCounts[38] was implemented in miARma-Seq to summarize reads. In addition, novel miRNAs can also be quantified using mirDeep2[29]. The quantification of circRNA reads is achieved with CIRI[4] (for details see Supplementary methods). A tab separated read-count file is generated as an output at this stage.

Exploratory analysis of the data. This analysis generates an exhaustive PDF report that allows every aspect of the processed data to be inspected in detail. This report includes a boxplot and a density plot of the normalized and non-normalized data, which facilitates the inspection of the read distribution, as well as a multidimensional scaling (MDS) plot, a principal component analysis (PCA) plot, a heatmap (Supplementary Fig. S2), and clustering plots.

Differential expression analysis. Two alternative software tools that can be combined: edgeR[39] and Noiseq[40] (for details see Supplementary methods). As a result of the analysis, easy to explore tabulated files are generated with the DE elements (Supplementary Fig. S2) and the main statistic values, as well as exploratory plots for each experimental condition evaluated, such as volcano or expression plots (Supplementary Fig. S2).

Target prediction. Through miRGate, miARma-Seq offers a miRNA-mRNA target prediction module for DE miRNAs or genes[41]. miRGate is a database containing novel predicted miRNA-mRNA pairs that are calculated using well-established algorithms, including miRanda[42], Pita[43], RNAhybrid[44] or MicroTar[45], among others. In addition, miRGate includes experimental validated miRNA-mRNA pairs that provide miARma-Seq a highly reliable tool for gene target prediction. Notably, miARma-Seq not only provides the potential targets of the DE miRNAs or mRNAs detected but if both sets of data are provided, negative correlations between miRNAs and mRNAs can be performed to extract DE mRNAs targeted by DE miRNAs. Thus, a detailed report is generated by miARma-Seq that includes the DE miRNAs and their corresponding mRNA targets, as well as statistical information on the number of tools in which there is an agreement for each prediction and the experimental validated targets.

Gene list analysis. A gene list analysis is implemented in miARma-Seq using the Goseq tool[46]. Goseq allows gene ontologies (GO) and metabolic pathways (KEGG) to be identified in genome-wide expression analyses. This enables the user to understand the biological processes affected in the experiment, either a metabolic pathway or a cellular complex, and it reduces complexity by highlighting the biological processes. This analysis generates a tabulated file (excel compatible) with information regarding the GO terms identified and their corresponding categories, ordered for relevance (p-value) of the up- and down-regulated entities.

Final summary report. Transcriptome expression analysis involves long, and multistep process that generates a wide variety of important information. To help understand all this relevant data, miARma-Seq generates an easy to read summary of the entire analysis requested. This summary is an excel compatible file that includes a section for each step of the analysis, presenting the main statistics: number of reads processed, total identified entities, number of aligned reads, number of DE entities, etc. An example of this summary report is presented in the Supplementary Fig. S2.

Datasets used to compute performance of miARma-Seq. *miRNA transcriptome analysis.* In order to evaluate the ability of miARma-Seq to detect, identify and assess the Differentitally expressed (DE) miRNAs in time-course experiments, a miRNA expression dataset[47] was analysed (GEO experiment GSE47602). Briefly, this experiment measures miRNAs regulated under different hypoxic conditions in the MCF7 cell line. It contains 2 replicates in normal conditions and 2 replicates taken after 16, 32 and 48 hours in hypoxic conditions. In addition to known miRNAs, miARma-Seq allows novel miRNAs to be identified and analyzed. The miRNA-Seq data from the hypoxic cell indicated above was used to detect novel miRNAs.

mRNA genome-wide expression analysis. The expression data available from the GSE52778 GEO experiment was used to test the capacity of our pipeline for RNA-Seq analysis[48]. This experiment contains mRNA samples obtained from four primary human airway smooth muscle cell lines untreated or treated with dexamethasone, albuterol, dexamethasone+albuterol. To facilitate the understanding of the pipeline, only samples treated with dexamethasone and control samples (3 replicates for each condition) were analyzed.

Detection of circRNAs from RNA-Seq data. The detection and identification of circRNAs from RNA-Seq data is a recent incorporation into the RNA-Seq analysis setting. To verify our circRNA analysis implementation, we used the data available from the GEO GSE49321 experiment[33]. Briefly, different cell types were used to identify the circRNAs expressed specifically in seven samples from HEK293T cells.

References

1. Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities. *Nature reviews Genetics* **12,** 87–98, doi: 10.1038/nrg2934 (2011).
2. Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5,** 621–628, doi: 10.1038/nmeth.1226 (2008).
3. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495,** 333–338, doi: 10.1038/nature11928 (2013).
4. Gao, Y., Wang, J. & Zhao, F. CIRI: an efficient and unbiased algorithm for de novo circular RNA identification. *Genome biology* **16,** 4, doi: 10.1186/s13059-014-0571-3 (2015).
5. Friedman, B. A. & Maniatis, T. ExpressionPlot: a web-based framework for analysis of RNA-Seq and microarray gene expression data. *Genome biology* **12,** R69, doi: 10.1186/gb-2011-12-7-r69 (2011).
6. Cumbie, J. S. *et al.* GENE-counter: a computational pipeline for the analysis of RNA-Seq data for gene expression differences. *Plos one* **6,** e25279, doi: 10.1371/journal.pone.0025279 (2011).
7. Lohse, M. *et al.* RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. *Nucleic acids research* **40,** W622–627, doi: 10.1093/nar/gks540 (2012).
8. Soderlund, C., Nelson, W., Willer, M. & Gang, D. R. TCW: transcriptome computational workbench. *Plos one* **8,** e69401, doi: 10.1371/journal.pone.0069401 (2013).
9. Knowles, D. G., Roder, M., Merkel, A. & Guigo, R. Grape RNA-Seq analysis pipeline environment. *Bioinformatics* **29,** 614–621, doi: 10.1093/bioinformatics/btt016 (2013).
10. Kalari, K. R. *et al.* MAP-RSeq: Mayo Analysis Pipeline for RNA sequencing. *BMC bioinformatics* **15,** 224, doi: 10.1186/1471-2105-15-224 (2014).
11. Huang, P. J. *et al.* DSAP: deep-sequencing small RNA analysis pipeline. *Nucleic acids research* **38,** W385–391, doi: 10.1093/nar/gkq392 (2010).
12. Hackenberg, M., Rodriguez-Ezpeleta, N. & Aransay, A. M. miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic acids research* **39,** W132–138, doi: 10.1093/nar/gkr247 (2011).
13. Wang, W. C. *et al.* miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC bioinformatics* **10,** 328, doi: 10.1186/1471-2105-10-328 (2009).
14. Ronen, R. *et al.* miRNAkey: a software for microRNA deep sequencing analysis. *Bioinformatics* **26,** 2615–2616, doi: 10.1093/bioinformatics/btq493 (2010).
15. Giurato, G. *et al.* iMir: an integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-Seq. *BMC bioinformatics* **14,** 362, doi: 10.1186/1471-2105-14-362 (2013).
16. Sun, Z. *et al.* CAP-miRSeq: a comprehensive analysis pipeline for microRNA sequencing data. *BMC genomics* **15,** 423, doi: 10.1186/1471-2164-15-423 (2014).
17. Wu, J. *et al.* mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing. *RNA biology* **10,** 1087–1092, doi: 10.4161/rna.25193 (2013).
18. Rueda, A. *et al.* sRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic acids research* **43,** W467–473, doi: 10.1093/nar/gkv555 (2015).
19. Zhao, W. *et al.* wapRNA: a web-based application for the processing of RNA sequences. *Bioinformatics* **27,** 3076–3077, doi: 10.1093/bioinformatics/btr504 (2011).
20. Yuan, T. *et al.* eRNA: a graphic user interface-based tool optimized for large data analysis from high-throughput RNA sequencing. *BMC genomics* **15,** 176, doi: 10.1186/1471-2164-15-176 (2014).
21. Chae, H., Rhee, S., Nephew, K. P. & Kim, S. BioVLAB-MMIA-NGS: microRNA-mRNA integrated analysis using high-throughput sequencing data. *Bioinformatics* **31,** 265–267, doi: 10.1093/bioinformatics/btu614 (2015).
22. Fisch, K. M. *et al.* Omics Pipe: a community-based framework for reproducible multi-omics data analysis. *Bioinformatics* **31,** 1724–1728, doi: 10.1093/bioinformatics/btv061 (2015).
23. Gaidatzis, D., Lerch, A., Hahne, F. & Stadler, M. B. QuasR: quantification and annotation of short reads in R. *Bioinformatics* **31,** 1130–1132, doi: 10.1093/bioinformatics/btu781 (2015).
24. D'Antonio, M. *et al.* RAP: RNA-Seq Analysis Pipeline, a new cloud-based NGS web application. *BMC genomics* **16,** S3, doi: 10.1186/1471-2164-16-S6-S3 (2015).
25. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research* **41,** e108, doi: 10.1093/nar/gkt214 (2013).
26. Wolfinger, M. T., Fallmann, J., Eggenhofer, F. & Amman, F. ViennaNGS: A toolbox for building efficient next- generation sequencing analysis pipelines. *F1000Res* **4,** 50, doi: 10.12688/f1000research.6157.2 (2015).
27. Davis, M. P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. & Enright, A. J. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods* **63,** 41–49, doi: 10.1016/j.ymeth.2013.06.027 (2013).
28. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology* **10,** R25, doi: 10.1186/gb-2009-10-3-r25 (2009).
29. Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N. miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research* **40,** 37–52, doi: 10.1093/nar/gkr688 (2012).
30. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14,** R36, doi: 10.1186/gb-2013-14-4-r36 (2013).
31. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* **9,** 357–359, doi: 10.1038/nmeth.1923 (2012).
32. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26,** 589–595, doi: 10.1093/bioinformatics/btp698 (2010).
33. Fan, X. *et al.* Single-cell RNA-seq transcriptome analysis of linear and circular RNAs in mouse preimplantation embryos. *Genome biology* **16,** 148, doi: 10.1186/s13059-015-0706-1 (2015).

34. Williamson, V. *et al.* Detecting miRNAs in deep-sequencing data: a software performance comparison and evaluation. *Briefings in bioinformatics* **14,** 36–45, doi: 10.1093/bib/bbs010 (2013).
35. Nookaew, I. *et al.* A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in Saccharomyces cerevisiae. *Nucleic acids research* **40,** 10084–10097, doi: 10.1093/nar/gks804 (2012).
36. Fonseca, N. A., Marioni, J. & Brazma, A. RNA-Seq gene profiling–a systematic empirical comparison. *Plos one* **9,** e107026, doi: 10.1371/journal.pone.0107026 (2014).
37. Creighton, C. J., Nagaraja, A. K., Hanash, S. M., Matzuk, M. M. & Gunaratne, P. H. A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. *Rna* **14,** 2290–2296, doi: 10.1261/rna.1188208 (2008).
38. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930, doi: 10.1093/bioinformatics/btt656 (2014).
39. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140, doi: 10.1093/bioinformatics/btp616 (2010).
40. Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. & Conesa, A. Differential expression in RNA-seq: a matter of depth. *Genome research* **21,** 2213–2223, doi: 10.1101/gr.124321.111 (2011).
41. Andres-Leon, E., Gonzalez Pena, D., Gomez-Lopez, G. & Pisano, D. G. miRGate: a curated database of human, mouse and rat miRNA-mRNA targets. *Database : the journal of biological databases and curation* **2015,** doi: 10.1093/database/bav035 (2015).
42. Betel, D., Koppal, A., Agius, P., Sander, C. & Leslie, C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome biology* **11,** R90, doi: 10.1186/gb-2010-11-8-r90 (2010).
43. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. & Segal, E. The role of site accessibility in microRNA target recognition. *Nature genetics* **39,** 1278–1284, doi: 10.1038/ng2135 (2007).
44. Kruger, J. & Rehmsmeier, M. RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic acids research* **34,** W451–454, doi: 10.1093/nar/gkl243 (2006).
45. Thadani, R. & Tammi, M. T. MicroTar: predicting microRNA targets from RNA duplexes. *BMC bioinformatics* **7** Suppl 5, S20, doi: 10.1186/1471-2105-7-S5-S20 (2006).
46. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome biology* **11,** R14, doi: 10.1186/gb-2010-11-2-r14 (2010).
47. Camps, C. *et al.* Integrated analysis of microRNA and mRNA expression and association with HIF binding reveals the complexity of microRNA expression regulation under hypoxia. *Molecular cancer* **13,** 28, doi: 10.1186/1476-4598-13-28 (2014).
48. Himes, B. E. *et al.* RNA-Seq transcriptome profiling identifies CRISPLD2 as a glucocorticoid responsive gene that modulates cytokine function in airway smooth muscle cells. *Plos one* **9,** e99625, doi: 10.1371/journal.pone.0099625 (2014).

Acknowledgements

E.A.-L. was funded by the European Union grant FP7-REGPOT-2012-2013-1. We acknowledge Ildefonso Cases at IBIS for invaluable input.

Author Contributions

Software design: A.M.R. and E.A.-L. Code and benchmarking analyses: E.A.-L. and R.N.-T. Tool documentation: R.N.-T. Software webpage: E.A.-L. Writers and interpretation of results: E.A.-L., R.N.-T. and A.M.R.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Andrés-León, E. *et al.* miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis. *Sci. Rep.* **6**, 25749; doi: 10.1038/srep25749 (2016).