

sRNAtoolbox: an integrated collection of small RNA research tools

Antonio Rueda¹, Guillermo Barturen², Ricardo Lebrón^{3,4}, Cristina Gómez-Martín^{3,4}, Ángel Alganza^{3,4}, José L. Oliver^{3,4} and Michael Hackenberg^{3,4,*}

¹Genomics and Bioinformatics Platform of Andalusia (GBPA), Edificio INSUR, Calle Albert Einstein, 41092-Sevilla, Spain, ²Centro de Genómica e Investigaciones Oncológicas, Pfizer-Universidad de Granada-Junta de Andalucía, Granada 18016, Spain, ³Dpto. de Genética, Facultad de Ciencias, Universidad de Granada, Campus de Fuentenueva s/n, 18071-Granada, Spain and ⁴Lab. de Bioinformática, Centro de Investigación Biomédica, PTS, Avda. del Conocimiento s/n, 18100-Granada, Spain

Received March 03, 2015; Revised May 02, 2015; Accepted May 15, 2015

ABSTRACT

Small RNA research is a rapidly growing field. Apart from microRNAs, which are important regulators of gene expression, other types of functional small RNA molecules have been reported in animals and plants. MicroRNAs are important in host-microbe interactions and parasite microRNAs might modulate the innate immunity of the host. Furthermore, small RNAs can be detected in bodily fluids making them attractive non-invasive biomarker candidates. Given the general broad interest in small RNAs, and in particular microRNAs, a large number of bioinformatics aided analysis types are needed by the scientific community. To facilitate integrated sRNA research, we developed sRNAtoolbox, a set of independent but interconnected tools for expression profiling from high-throughput sequencing data, consensus differential expression, target gene prediction, visual exploration in a genome context as a function of read length, gene list analysis and blast search of unmapped reads. All tools can be used independently or for the exploration and downstream analysis of sRNAbench results. Workflows like the prediction of consensus target genes of parasite microRNAs in the host followed by the detection of enriched pathways can be easily established. The web-interface interconnecting all these tools is available at <http://bioinfo5.ugr.es/srnatoolbox>

INTRODUCTION

Over the last years, high-throughput sequencing (HTS) technologies changed notably our understanding of the biogenesis and function of small RNA molecules. Apart from microRNAs, which are known for over a decade to play

very important roles in gene regulation (1,2), several other classes of small RNAs have been described. In plants, small interfering RNAs (siRNA), trans-acting RNAs and heterochromatic siRNAs are involved in the regulation of gene expression and the chromatin state by depositing repressive marks (DNA methylation and histone marks) (3). Furthermore, they seem to play important roles in abiotic stress response in plants (4,5). In animals, reproducibly processed small RNA fragments of yRNA, tRNA and snoRNA origin have been identified (6) and especially tRNA fragments might be functional by regulating translation (7).

Over the last years, several other research fields emerged in which microRNAs and other small non-coding RNAs play vital roles. MicroRNAs seem to be crucial in the evasion of the immune response and the regulation of the switch to lytic cycle of virus infections (8,9) and in general might have important functions in host-microbe interactions (10). Furthermore, microRNAs have strong biomarker potential as specific miRNA expression profiles have been associated with different types of cancer, tumour development and allowed to identify the tissue of origin of poorly differentiated tumours (11). Small RNAs are also present in membrane-bound vesicles called exosomes which are secreted by virtually all cell types (12,13). As a consequence, microRNAs can be detected in most studied bodily fluids and are therefore perfect candidates for non-invasive biomarkers (14). Finally, microRNAs have been also detected in exosomes secreted by parasites (15) being probably involved in the modulation of the innate immunity of the host (16).

Given the wide-ranging interest in small RNAs, a broad panoply of bioinformatics tools is needed to assist researchers from different fields in small RNA research. Often the first step consists in the expression profiling of small RNAs by means of HTS approaches. Frequently used programs for this task are: small RNA toolkit (17), miRD-deep (18), miRanalyzer (19,20), SeqBuster (21), DARIO (22),

*To whom correspondence should be addressed. Tel: +34 958 249695; Fax: +34 958 244073; Email: hackenberg@ugr.es

UEA sRNA workbench (23), miRDeep2 (24) and Short-Stack (25). Those tools do generally also include other important aspects like the prediction of novel microRNAs or the isomiR quantification. However, they do generally not offer functional downstream analysis, visualization of the expression pattern in a genome context or the capacity to analyse multi-species assays. Some exceptions are mirTools 2.0 (26), which includes target prediction and functional analysis linked to the expression profiling, UEA sRNA workbench including visualization and sRNAbench (27) which allows the expression profiling in experiments that involve genetic material from more than one species.

Many important analysis types are currently not available, are linked to a fixed pipeline or are scattered over several independent tools. Given the increasing interest in studying microbe-host interactions, one frequently needed analysis type would be to estimate the impact of parasite secreted microRNAs on the host genes, i.e. which genes and which pathways might be regulated by those exogenous microRNAs. Another example could be the visual exploration of small RNA differential expression in a genome context as a function of read length, given that at least in plants 24 nt long reads have clearly different functions than 21/22 nt long reads.

In order to overcome some of the mentioned lack of analysis types, we developed sRNAtoolbox having in mind two main aims. First, we implemented some analysis types that are currently missing, like the cross-species microRNA target prediction coupled to functional *in silico* analysis, the consensus detection of differentially expressed sRNAs or the visualization of differential expression in a genome context as a function of read length. Second, we implemented workflows in order to make some analysis types easier, opening therefore the analysis to users with less bioinformatics backgrounds. The centre piece of sRNAtoolbox is sRNAbench that allows the expression profiling and prediction of novel microRNAs in a multi-species experimental assay. Most other tools, interconnected through the same web-interface, can be used for downstream and exploratory analyses on sRNAbench output, but also in an independent manner on the appropriate input. This makes sRNAtoolbox a very flexible and easy to use set of small RNA research tools.

WORKFLOW AND SCOPE

sRNAtoolbox consists currently of seven independent but interrelated tools and one workflow addressing several important analysis types. The web-interface interconnecting all these tools is implemented in Django and javascript while the underlying programs are written in Java. The required data are stored either in mongoDB or in MySQL depending on the nature and structure of the data. The results will be stored for 15 days on the server.

Briefly, sRNAtoolbox contains the following tools:

- **sRNAbench**: expression profiling of small RNAs, prediction of novel microRNAs, analysis of isomiRs, genome mapping and read length statistics. The samples are analysed individually.

- **sRNAde**: detection of differentially expressed small RNAs based on three commonly used programs: DESeq (28), edgeR (29) and NOISeq (30). To our knowledge it is the first tool that determines a consensus of differentially expressed sRNA.
- **sRNAblast**: aimed to determine the origin of unmapped or unassigned reads by means of a blast search against several remote databases. The results can either point towards contamination sources or biological meaningful information like the presence of unexpected viral or bacterial RNA molecules.
- **miRNAconsTarget**: consensus target prediction on user provided input data based on Miranda (1), PITA (31) and TargetSpy (32) for animal microRNAs and PsRobot (33) and TAPIR (34) for plant microRNAs.
- **sRNAjBrowser**: visualization of sRNA expression data in a genome context using jBrowse (35).
- **sRNAjBrowserDE**: visualization of differential expression as a function of read length in a genome context using jBrowse.
- **sRNAjfuncTerms**: enrichment of functional annotations in animal and plant target genes not only for a set of microRNAs but also for all its combinations (microRNA modules).
- **sRNAjfuncTargets (workflow)**: cross-species microRNA target prediction and enrichment analysis of functional annotations.

An important point in sRNAtoolbox is that although sRNAbench is the center-piece, most tools can be used also in an independent way. Figure 1 shows the relation between the different tools and mentions briefly the input types that can be used.

DESCRIPTION OF THE INDIVIDUAL TOOLS

Given that sRNAbench was published before (27), in this paragraph we will describe primarily the other tools that comprise the toolbox and the way in which these tools are interconnected, connected to sRNAbench or in which way they can be used independently.

sRNAde: consensus detection of differentially expressed sRNA

The sRNAde module can be used in two different ways. It can be applied either to a user-generated expression matrix or to a number of individual sRNAbench runs indicating the relation between the samples on the input, i.e. defining the experimental groups between which the differentially expressed sRNA should be detected. Independently of the input type, two main results are generated. First, heatmaps are calculated using Hierarchical Clustering (implementation: hclust for R, parameters method = 'complete') which allows the user to visualize the clustering of the samples and to detect outlier samples. Second, the differential expression is assessed using three frequently used programs: edgeR, DESeq and NOISeq. Additionally, the program also generates a consensus differential expression file which gives the user the possibility to increase the stringency using only those microRNAs/sRNA that have been detected as differentially expressed by more than one method.

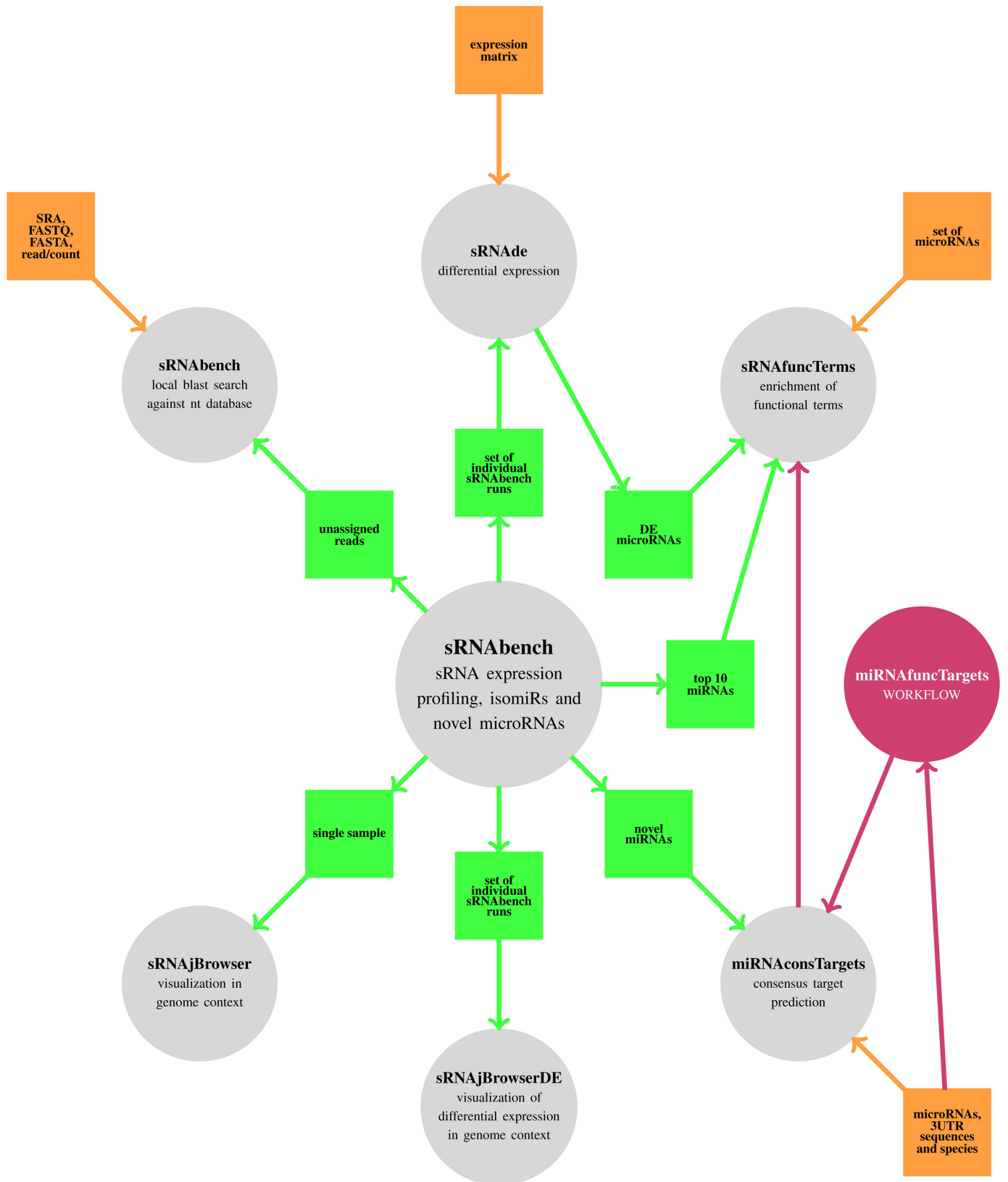


Figure 1. Relation between the different tools. The way the tools interact with sRNAbench (green) and how they can be used independently (orange) is depicted.

If sRNAbench output is used as input, additional features are available. For example, the module generates a summary of sequencing, adapter removal and genome mapping statistics, both in table and graphic format. In this way, poor samples can be easily spotted and probably removed from the analysis.

sRNAblast

Frequently it happens that only a minor fraction of the reads can be mapped to the genome or that a high fraction of the reads cannot be assigned to any annotation. Two basic explanations exist for unmapped reads: bad sequencing quality or contamination. While the first possibility can be assessed by means of tools like fastQC, the search for possible contamination sources is more daunting. sRNAblast uses a local installation of blast+ (36) and allows to detect the reads in several remote databases hosted at NCBI: nr (all GenBank, RefSeq Nucleotides, EMBL, DDBJ and PDB sequences), refseq_rna (RNA entries from NCBI's Reference Sequence project), est (Database of GenBank, EMBL, DDBJ sequences from EST Divisions) and env_nt (Nucleotide sequences from environmental samples) in order to aid the identification of unmapped/unassigned reads. Given that the mapping with blast of thousands of reads is a CPU consuming task, sRNAblast is mainly intended for the annotation of unmapped reads from a previous sRNAbench run. It can, however, also be used as an independent tool. It shares the sRNAbench preprocessing and accepts therefore fastq, sra, fasta and read/count format and is able to perform adapter trimming.

miRNAconsTargets: consensus microRNA target prediction

Currently many of the most frequently used microRNA target prediction programs are available over web-server programs or databases and do therefore not allow a user customized prediction of microRNA targets. miRNAconsTargets allows the user to upload a set of target sequences or choosing a 3' UTR set from our database together with a set of microRNAs. TargetSpy, Miranda and PITA are available for the prediction of animal microRNA targets and PsRobot and TAPIR are used for plants. The target prediction pipelines generate both, individual predictions for each program and a more reliable consensus prediction. Furthermore, the parameters of the different programs can be changed by the user making the consensus target prediction very flexible.

sRNAfuncTerms

One of the first questions after having obtained a number of differentially expressed microRNAs or even a set of highly expressed microRNAs in a single sample is on its putative functions. A generally applied approach consists of determining the target genes and performing an enrichment/depletion analysis of functional annotations on this gene list. This can give important insight into the molecular functions or biological processes of the regulated genes.

sRNAfuncTerms uses Gene Ontology as functional annotations (37) and is based on AnnotationModules (38)

to extract those functional annotations that are enriched among the target genes in a statistically significant way.

sRNAfuncTerms can be used on sRNAbench, sRNAde results and independently by providing microRNA names and the species. For sRNAbench, the 10 most expressed microRNAs are analysed while for sRNAde input the statistically significant differentially expressed microRNAs detected by any of the methods are analysed.

In a first step, the target genes are assigned by means of an underlying database which holds experimentally verified targets from TarBase v6 (39), predictions obtained by means of the psRNAtarget web-server for plants (40) and predictions using the miRNAconsTargets pipeline explained in the last section (both for plants and animals). The frequency of each functional annotation in the list of target genes is then compared to the background of all annotated genes by means of the local implementation of the AnnotationModules tool.

As a novel feature, sRNAfuncTerms does not only determine the enriched annotations for the complete set of microRNAs, but also for microRNA modules up to a certain depth or module size.

Briefly, sRNAfuncTerms determines the enrichment for all individual microRNAs, and all combinations (or modules) of two and three microRNAs. If a functional annotation is detected for more than one microRNA module, the one with the highest enrichment ratio is reported. In this way, the researcher obtains the knowledge on which concrete microRNAs might be involved in a certain pathway or function.

sRNAfuncTargets

sRNAconsTargets and sRNAfuncTerms can be used directly as a pipeline. This enables prediction of the target genes of a set of microRNAs (independently of the origin) in any of the species included in sRNAtoolbox (either using our database or user provided target sequences). The consensus target genes are then used by sRNAfuncTerms to determine the enriched functional annotations. This allows the user to determine which genes and pathways might be regulated, for example by parasite microRNAs within a host cell. Please note that cross-species target prediction can be carried out also by the psRNAtarget server for plants, however without a functional downstream analysis.

sRNAjBrowser and sRNAjBrowserDE

The visualization of expression values in a genome context can give important information. For example in virus research, the mapping profile can indicate the existence of a previously unknown small non-coding RNA. Both tools share the basic workflow, but while sRNAjBrowser visualizes the expression values reads per million (RPM) of a single sample, sRNAjBrowserDE depicts the differential expression values together with the logarithm of the mean expression values of each experimental group.

In general, the tools are based on the genome mappings that sRNAbench generates by means of the Bowtie aligner (41). In a first step, the mappings are converted into internal bedGraph files. Reads that map to several positions

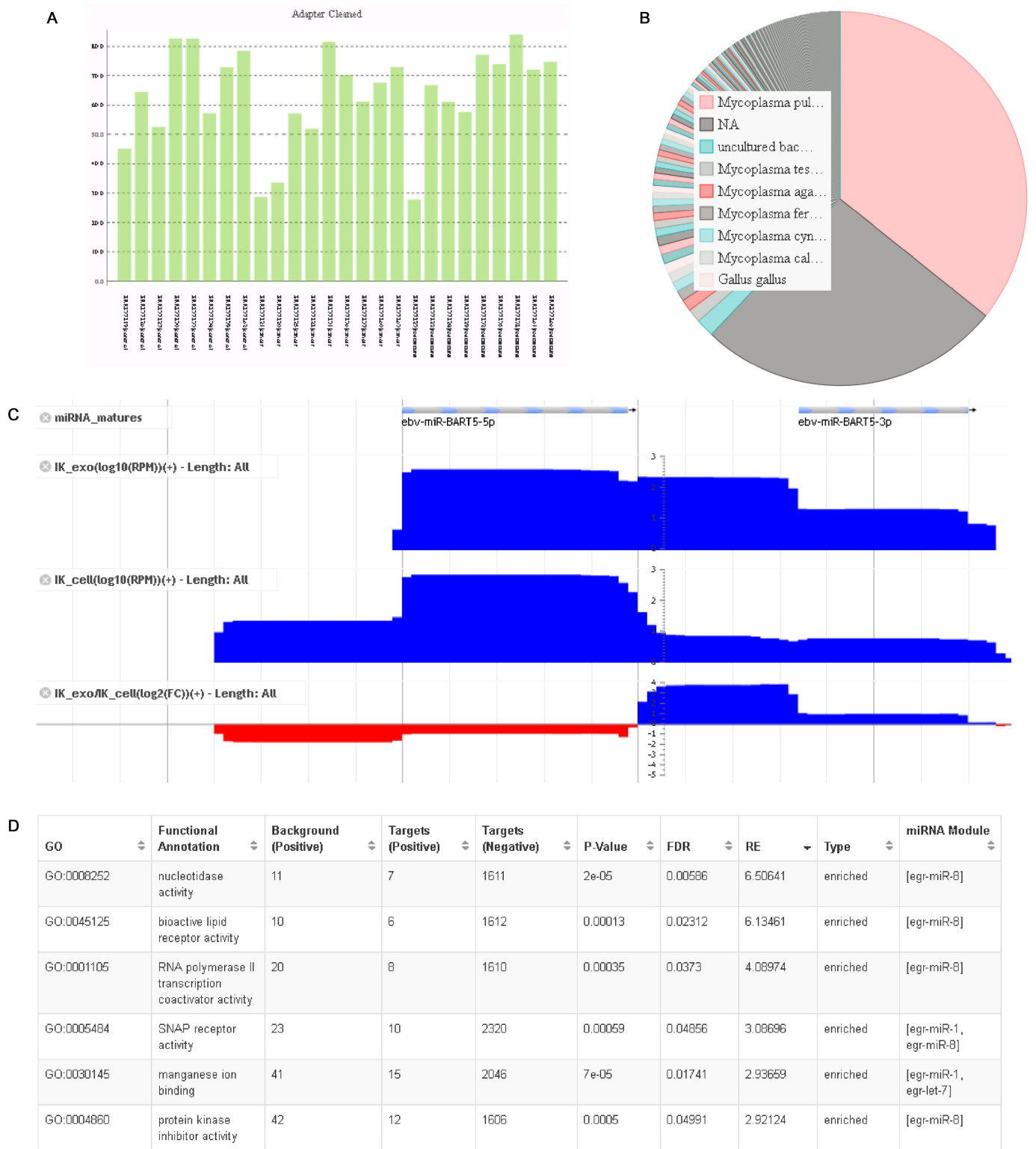


Figure 2. Results for four different working examples. (A) Percentage of adapter trimmed reads in 26 samples of colorectal cancer; (B) graphical representation of the blast search of reads not mapped to the human genome from a SRA data set (SRR1057370); (C) differential expression between small RNAs in cells and exosomes from a LCL cell line (IK). The bottom line shows the \log_2 of the fold-change indicating that the most differentially expressed region is the loop region of viral microRNA mir-BART5; (D) a screen-shot of the most enriched molecular functions among the cow target genes of three *E. granulosus* microRNAs.

in the genome are fully assigned to each locus, i.e. by default, no proportional assignment is applied. The number of mapped reads per position is then converted into RPM using the total number of genome mapped reads for normalization. By means of the bedGraphToBigWig tool provided by UCSC, the bedGraph files are then converted into bigwig format which can be visualized by means of the implemented jBrowse (35).

The sRNAjBrowserDE tool allows furthermore the analysis as a function of read length. It sums first the mappings of all samples of one condition before assigning a RPM value to each position. In a second step, the log₂ of the fold-change is calculated for each position. In order to avoid division by zero, a fixed value of 10 is added to each RPM before applying the log₂ to the case/control ratio (fold-change).

The bigwig files can then be visualized together with several genome annotation tracks including microRNA (both mature and precursor sequences), genes and repetitive elements.

WORKING EXAMPLES

Figure 2 summarizes the results of four working examples in order to demonstrate the usefulness of sRNAtoolbox. Figure 2A depicts the summary of adapter cleaning of 26 samples for colorectal cancer (42) which is used as test data set for sRNAde. The sRNAde tool gives several summary graphics at an experiment level which can help the user to easily detect problematic samples. In this case, three samples can be seen with only around 30% of detected adapters while on average this figure is around 70% in the whole experiment.

To show the usefulness of sRNAblast we reanalysed a publicly available data set of *Salmonella enteritidis* (strain SE2472) infected HT-29 cells (43). After processing the data (SRR1057370) with sRNAbench, we analysed the unmapped reads with sRNAblast. Figure 2B shows that most unmapped reads were assigned to *Mycoplasma pulmonis* and not to *Salmonella enteritidis*. Figure 2C shows an example of differential expression between an Epstein-Barr virus (EBV) infected Lymphoblastoid Cell Line (LCL cell line) and its secreted exosomes (35). The most differentially expressed region in EBV is shown (bottom of Figure 2C) and interestingly coincides with the loop region of ebv-mir-BART5 which is over 10-fold more frequent in exosomes. The existence of a miRNA-offset RNA (moRNA) in cells can be also observed which is nearly absent in exosomes. As neither loop nor moRNA sequences are generally used as annotations, such findings would be normally missed which shows that the visual exploration of differential expression data can give important information and might trigger or orientate additional analysis. Finally, we analysed the putative functional role of three *Echinococcus granulosus* microRNAs (egr-miR-1, egr-let-7 and egr-miR-8) in the cow host by means of the miRNAfuncTargets workflow. Figure 2D shows the most enriched molecular functions among the cow target genes. It can be seen that many are regulated only by one microRNA but among the most enriched terms are also several that might be regulated by a combination of microRNAs (for example SNAP receptor activity by [egr-miR-1, egr-miR-8]).

CONCLUSIONS AND FUTURE PLANS

sRNAtoolbox is a flexible and easy to use set of tools for small RNA research. Most tools can be used independently or for downstream analysis of sRNAbench output data. The underlying database contains currently only the most important model species from animals and plants. We are planning to continuously increase the number of species, encouraging all users to request the inclusion of required species. An integrated suite of standalone tools is also in preparation at this moment, which will give users with high workloads the possibility to perform the analysis locally.

FUNDING

Ministerio de Economía y Competitividad from the Spanish Government [AGL2013-49090-C2-2-R to J.L.O.].

Conflict of interest statement. None declared.

REFERENCES

- John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
- He, L. and Hannon, G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.*, **5**, 522–531.
- Axtell, M.J. (2013) Classification and comparison of small RNAs from plants. *Annu. Rev. Plant Biol.*, **64**, 137–159.
- Ferdous, J., Hussain, S.S. and Shi, B.-J. (2015) Role of microRNAs in plant drought tolerance. *Plant Biotechnol. J.*, **13**, 293–305.
- Hackenberg, M., Huang, P.-J., Huang, C.-Y., Shi, B.-J., Gustafson, P. and Langridge, P. (2013) A comprehensive expression profile of microRNAs and other classes of non-coding small RNAs in barley under phosphorous-deficient and -sufficient conditions. *DNA Res.*, **20**, 109–125.
- Tuck, A.C. and Tollervey, D. (2011) RNA in pieces. *Trends Genet.*, **27**, 422–432.
- Ivanov, P., Emara, M.M., Villen, J., Gygi, S.P. and Anderson, P. (2011) Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol. Cell.*, **43**, 613–623.
- Gottwein, E., Corcoran, D.L., Mukherjee, N., Skalsky, R.L., Hafner, M., Nusbaum, J.D., Shamulilatpam, P., Love, C.L., Dave, S.S., Tuschl, T. et al. (2011) Viral microRNA targetome of KSHV-infected primary effusion lymphoma cell lines. *Cell Host Microbe*, **10**, 515–526.
- Skalsky, R.L. and Cullen, B.R. (2010) Viruses, microRNAs, and host interactions. *Annu. Rev. Microbiol.*, **64**, 123–141.
- Harris, J.F., Micheva-Viteva, S., Li, N. and Hong-Geller, E. (2013) Small RNA-mediated regulation of host-pathogen interactions. *Virulence*, **4**, 785–795.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. et al. (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
- Nolte-’t Hoen, E.N.M., Buermans, H.P.J., Waasdorp, M., Stoorvogel, W., Wauben, M.H.M. and ’t Hoen, P.A.C. (2012) Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. *Nucleic Acids Res.*, **40**, 9272–9285.
- Koppers-Lalic, D., Hackenberg, M., Bijnsdorp, I.V., van Eijndhoven, M.A.J., Sadek, P., Sie, D., Zini, N., Middeldorp, J.M., Ylstra, B., de Menezes, R.X. et al. (2014) Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. *Cell Rep.*, **8**, 1649–1658.
- Kosaka, N., Iguchi, H. and Ochiya, T. (2010) Circulating microRNA in body fluid: a new potential biomarker for cancer diagnosis and prognosis. *Cancer Sci.*, **101**, 2087–2092.
- Bernal, D., Trellis, M., Montaner, S., Cantalapiedra, F., Galiano, A., Hackenberg, M. and Marcilla, A. (2014) Surface analysis of *Dicrocoelium dendriticum*. The molecular characterization of exosomes reveals the presence of miRNAs. *J. Proteomics*, **105**, 232–241.

16. Buck, A.H., Coakley, G., Simbari, F., McSorley, H.J., Quintana, J.F., Le Bihan, T., Kumar, S., Abreu-Goodger, C., Lear, M., Harcus, Y. *et al.* (2014) Exosomes secreted by nematode parasites transfer small RNAs to mammalian cells and modulate innate immunity. *Nat. Commun.*, **5**, 5488.
17. Moxon, S., Schwach, F., Dalmay, T., Maclean, D., Studholme, D.J. and Moulton, V. (2008) A toolkit for analysing large-scale plant small RNA datasets. *Bioinformatics*, **24**, 2252–2253.
18. Friedländer, M.R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S. and Rajewsky, N. (2008) Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.*, **26**, 407–415.
19. Hackenberg, M., Sturm, M., Langenberger, D., Falcón-Pérez, J.M. and Aransay, A.M. (2009) miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res.*, **37**, W68–W76.
20. Hackenberg, M., Rodríguez-Ezpeleta, N. and Aransay, A.M. (2011) miRanalyzer: an update on the detection and analysis of microRNAs in high-throughput sequencing experiments. *Nucleic Acids Res.*, **39**, W132–W138.
21. Pantano, L., Estivill, X. and Martí, E. (2010) SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells. *Nucleic Acids Res.*, **38**, e34.
22. Fasold, M., Langenberger, D., Binder, H., Stadler, P.F. and Hoffmann, S. (2011) DARIO: a ncRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res.*, **39**, W112–W117.
23. Stocks, M.B., Moxon, S., Mapleson, D., Woolfenden, H.C., Mohorianu, I., Folkes, L., Schwach, F., Dalmay, T. and Moulton, V. (2012) The UEA sRNA workbench: a suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. *Bioinformatics*, **28**, 2059–2061.
24. Friedländer, M.R., Mackowiak, S.D., Li, N., Chen, W. and Rajewsky, N. (2012) miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.*, **40**, 37–52.
25. Axtell, M.J. (2013) ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA*, **19**, 740–751.
26. Wu, J., Liu, Q., Wang, X., Zheng, J., Wang, T., You, M., Sheng Sun, Z. and Shi, Q. (2013) mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing. *RNA Biol.*, **10**, 1087–1092.
27. Barturen, G., Rueda, A., Hamberg, M., Alganza, A., Lebron, R., Kotsyfakis, M., Shi, B.-J., Koppers-Lalic, D. and Hackenberg, M. (2014) sRNAbench: profiling of small RNAs and its sequence variants in single or multi-species high-throughput experiments. *Methods Next Gener. Seq.*, **1**, 21–31.
28. Anders, S. and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
29. Robinson, M.D., McCarthy, D.J. and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
30. Tarazona, S., Garcia-Alcalde, F., Dopazo, J., Ferrer, A. and Conesa, A. (2011) Differential expression in RNA-seq: a matter of depth. *Genome research*, **21**, 2213–2223.
31. Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U. and Segal, E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
32. Sturm, M., Hackenberg, M., Langenberger, D. and Frishman, D. (2010) TargetSpy: a supervised machine learning approach for microRNA target prediction. *BMC Bioinformatics*, **11**, 292.
33. Wu, H.-J., Ma, Y.-K., Chen, T., Wang, M. and Wang, X.-J. (2012) PsRobot: a web-based plant small RNA meta-analysis toolbox. *Nucleic Acids Res.*, **40**, W22–W28.
34. Bonnet, E., He, Y., Billiau, K. and Van de Peer, Y. (2010) TAPIR, a web server for the prediction of plant microRNA targets, including target mimics. *Bioinformatics*, **26**, 1566–1568.
35. Skinner, M.E. and Holmes, I.H. (2010) Setting up the JBrowse genome browser. *Curr. Protoc. Bioinformatics*, **32**, doi/10.1002/0471250953.bi0913s32.
36. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
37. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
38. Hackenberg, M. and Matthiesen, R. (2008) Annotation-Modules: a tool for finding significant combinations of multisource annotations for gene lists. *Bioinformatics*, **24**, 1386–1393.
39. Vergoulis, T., Vlachos, I.S., Alexiou, P., Georgakilas, G., Maragkakis, M., Reczko, M., Gerangelos, S., Koziris, N., Dalamagas, T. and Hatzigeorgiou, A.G. (2012) TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.*, **40**, D222–D229.
40. Dai, X. and Zhao, P.X. (2011) psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.*, **39**, W155–W159.
41. Langmead, B., Trapnell, C., Pop, M. and Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
42. Röhr, C., Kerick, M., Fischer, A., Kühn, A., Kashofer, K., Timmermann, B., Daskalaki, A., Meinel, T., Drichel, D., Börno, S.T. *et al.* (2013) High-throughput miRNA and mRNA sequencing of paired colorectal normal, tumor and metastasis tissues and bioinformatic modeling of miRNA-1 therapeutic applications. *PLoS One*, **8**, e67461.
43. Gu, H., Zhang, T., Li, D., Liu, F., Zhang, C.-Y. and Zen, K. (2014) A Salmonella-encoded microRNA-like RNA facilitates bacterial invasion and intracellular replication via suppressing host cell inducible nitric oxide synthase (MPF3P809). *J. Immunol.*, **192**, 132.