Nucleic Acids Research Advance Access published May 16, 2013

Nucleic Acids Research, 2013, 1–5
doi:10.1093/nar/gkt393

DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows

Maria D. Paraskevopoulou[1,2,*], Georgios Georgakilas[1,2], Nikos Kostoulas[3], Ioannis S. Vlachos[1,4], Thanasis Vergoulis[3], Martin Reczko[1], Christos Filippidis[5], Theodore Dalamagas[3] and A.G. Hatzigeorgiou[1,2,*]

[1]DIANA-Lab, Biomedical Sciences Research Center 'Alexander Fleming', 16672 Vari, Greece, [2]Department of Computer and Communication Engineering, University of Thessaly, 382 21 Volos, Greece, [3]IMIS Institute, 'Athena' Research Center, 11524 Athens, Greece, [4]Laboratory for Experimental Surgery and Surgical Research 'N.S. Christeas', Medical School of Athens, University of Athens, 11527 Athens, Greece and [5]National Center for Scientific Research DEMOKRITOS, Institute of Nuclear and Particle Physics, 15310 Aghia Paraskevi, Greece

Received March 4, 2013; Revised April 15, 2013; Accepted April 18, 2013

ABSTRACT

MicroRNAs (miRNAs) are small endogenous RNA molecules that regulate gene expression through mRNA degradation and/or translation repression, affecting many biological processes. DIANA-microT web server (http://www.microrna.gr/webServer) is dedicated to miRNA target prediction/functional analysis, and it is being widely used from the scientific community, since its initial launch in 2009. DIANA-microT v5.0, the new version of the microT server, has been significantly enhanced with an improved target prediction algorithm, DIANA-microT-CDS. It has been updated to incorporate miRBase version 18 and Ensembl version 69. The *in silico*-predicted miRNA–gene interactions in *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster* and *Caenorhabditis elegans* exceed 11 million in total. The web server was completely redesigned, to host a series of sophisticated workflows, which can be used directly from the on-line web interface, enabling users without the necessary bioinformatics infrastructure to perform advanced multi-step functional miRNA analyses. For instance, one available pipeline performs miRNA target prediction using different thresholds and meta-analysis statistics, followed by pathway enrichment analysis. DIANA-microT web server v5.0 also supports a complete integration with the Taverna Workflow Management System (WMS), using the in-house developed DIANA-Taverna Plug-in. This plug-in provides ready-to-use modules for miRNA target prediction and functional analysis, which can be used to form advanced high-throughput analysis pipelines.

INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNAs (~21–23 nt in length) that post-transcriptionally regulate gene expression by blocking translation or inducing degradation of the targeted mRNA (1). Since their first identification in *Caenorhabditis elegans* in 1993 (2), the number of annotated miRNAs and miRNA-related publications increase in a super linear rate, clearly depicting their central position in the RNA revolution (3).

In silico miRNA target identification is a crucial step in most miRNA experiments, as the miRNA interactome has not yet been adequately mapped, even for the most studied model organisms. Early miRNA-related research efforts have highlighted the necessity of computational analyses in order to assist the experimental identification of miRNA targets. This has resulted to the development of numerous miRNA target prediction algorithms (4), which are now considered indispensable for the design of relevant experiments. These algorithms identify *in silico* miRNA targets as candidates for further experimentation or for computational processing, such as target enrichment analyses. Predictions of the available computational algorithms can be acquired from relevant interaction databases or web servers (4,5).

*To whom correspondence should be addressed. Tel: +30 210 9656310 (ext. 190); Fax: +30 210 9653934; Email: hatzigeorgiou@fleming.gr
Correspondence may also be addressed to Maria D. Paraskevopoulou. Tel: +30 210 9656310 (ext. 190); Fax: +30 210 9653934; Email: paraskevopoulou@fleming.gr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

© The Author(s) 2013. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on May 21, 2013

The DIANA-microT web server v4.0 (6) is focused on providing *in silico* predictions of miRNA–mRNA interactions. It is characterized by a user-friendly interface and provides extensive information for predicted miRNA–target gene interactions such as a global score for each interaction, as well as detailed information for all predicted target sites. Each target site can be visualized, and the user can examine its local prediction score, target site conservation and miRNA–mRNA binding structure. The server also provides connectivity to online biological databases and offers links to nomenclature, sequence and protein databases.

Here, we present DIANA-microT web server v5.0, a significantly updated version, which hosts the state-of-the-art target prediction algorithm, DIANA-microT-CDS (7). microT-CDS is the only algorithm available online, specifically designed to identify miRNA targets both in 3′ untranslated region (3′UTR) and in coding sequences (CDS). The new server detects miRNA targets in mRNA sequences of *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster* and *C.elegans*. Furthermore, it has been updated to miRBase v18 (8) and Ensembl v69 (9). Specific attention has been paid to the web server interface, which follows the DIANA design framework, to be instantly familiar to users of previous versions, or other DIANA tools (Figure 1). On the other hand, online help, informative tooltips and easy-to-use menus minimize the learning curve of new users. The fifth version of the DIANA-microT web server focuses also on advanced users and laboratories requiring support for sophisticated pipelines. The server provides programmatic access to services of multiple DIANA algorithms and a complete integration with the Taverna Workflow Management System (WMS) (10), using the in-house developed DIANA-Taverna Plug-in. Furthermore, a new section of the web interface hosts ready-made advanced workflows that can perform extensive miRNA-related analyses on results derived from high-throughput techniques, such as microarrays or Next-Generation Sequencing (NGS).

MATERIALS AND METHODS

Integration of DIANA-microT-CDS

Initial research efforts have unveiled that miRNAs regulated gene expression through their binding on the 3′UTR of protein-coding genes (2). However, accumulated experimental evidence has revealed that miRNA-binding sites within coding sequences are also functional in controlling gene expression (11). The new algorithm microT-CDS (7) can identify miRNA targets in 3′UTR, as well as in CDS regions. Further details on the microT-CDS algorithm and the utilized training sets can be found in the relevant publication by Reczko *et al.* (7). DIANA-microT-CDS provides increased accuracy and the highest sensitivity at any level of specificity over other available state-of-the-art implementations, when tested against pulsed stable isotope labeling by amino acids in cell culture (pSILAC) proteomics data sets (12).

The selection of DIANA-microT-CDS as its core algorithm renders the new web server the only available online resource capable of incorporating miRNA targets in 3′UTR as well as in CDS. The new web server enables users to attain high quality predicted miRNA–gene interactions in all relevant *in silico* pipelines.

Web server update and extension

DIANA-microT server has been updated to miRBase v18 and Ensembl v69. The server is compatible with the new miRNA nomenclature (3p/5p) introduced in miRBase v18, as well as with previous miRNA naming conventions. It currently supports 7.3×10^6 *H.sapiens*, 3.5×10^6 *M.musculus*, 4.4×10^5 *D.melanogaster* and 2.5×10^5 *C.elegans* interactions between 3876 miRNAs and 64 750 protein-coding genes. Gene (9) and miRNA (13) expression annotation has been incorporated into the web server, enabling the user to perform advanced result filtering based on tissue expression. Furthermore, users can also restrict predictions between uploaded lists of expressed genes and/or miRNAs. For example, this feature can be used to identify interactions between a list of repressed (or overexpressed) genes and overexpressed (or repressed) miRNAs, in the case of a differential expression analysis pipeline.

Moreover, the web server hosts an updated version of the KEGG database providing a relevant search module based on KEGG pathway descriptions (14). A redesigned optional user space has also been implemented, which provides personalized features and facilitates the interconnectivity between the web server and the available DIANA software and databases (Figure 1).

Web server support of advanced pipelines

As high-throughput data have become the new backbone of biological research, there is an increasing need to support advanced high-throughput analysis pipelines. The new DIANA-microT web server aims to facilitate users, not having access to extensive computational infrastructures and support, to perform ready-to-use sophisticated pipelines.

DIANA-microT web server v5.0 hosts numerous integrated analyses in the form of ready-made advanced pipelines, covering a wide range of inquiries regarding predicted or validated miRNA–gene interactions and their impact on metabolic and signalling pathways. These pipelines can be used to analyse user data derived from small scale and high-throughput experiments directly from the DIANA-microT web server interface, without the necessity to install or implement any kind of software.

For instance, one of the available workflows (Figure 2) can analyse mRNA and miRNA expression data (expression and fold change). Suppressed genes are automatically matched with overexpressed miRNAs (and vice versa). The workflow performs enrichment analysis of experimentally validated targets derived from DIANA-TarBase v6.0 (3) or/and predicted interactions from microT-CDS. This step is considered crucial to identify miRNAs that significantly regulate the differentially expressed genes.

The prediction score threshold can significantly affect the analysis steps that follow. In the case of predicted interactions, the pipeline can be optimized by automatic

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on May 21, 2013

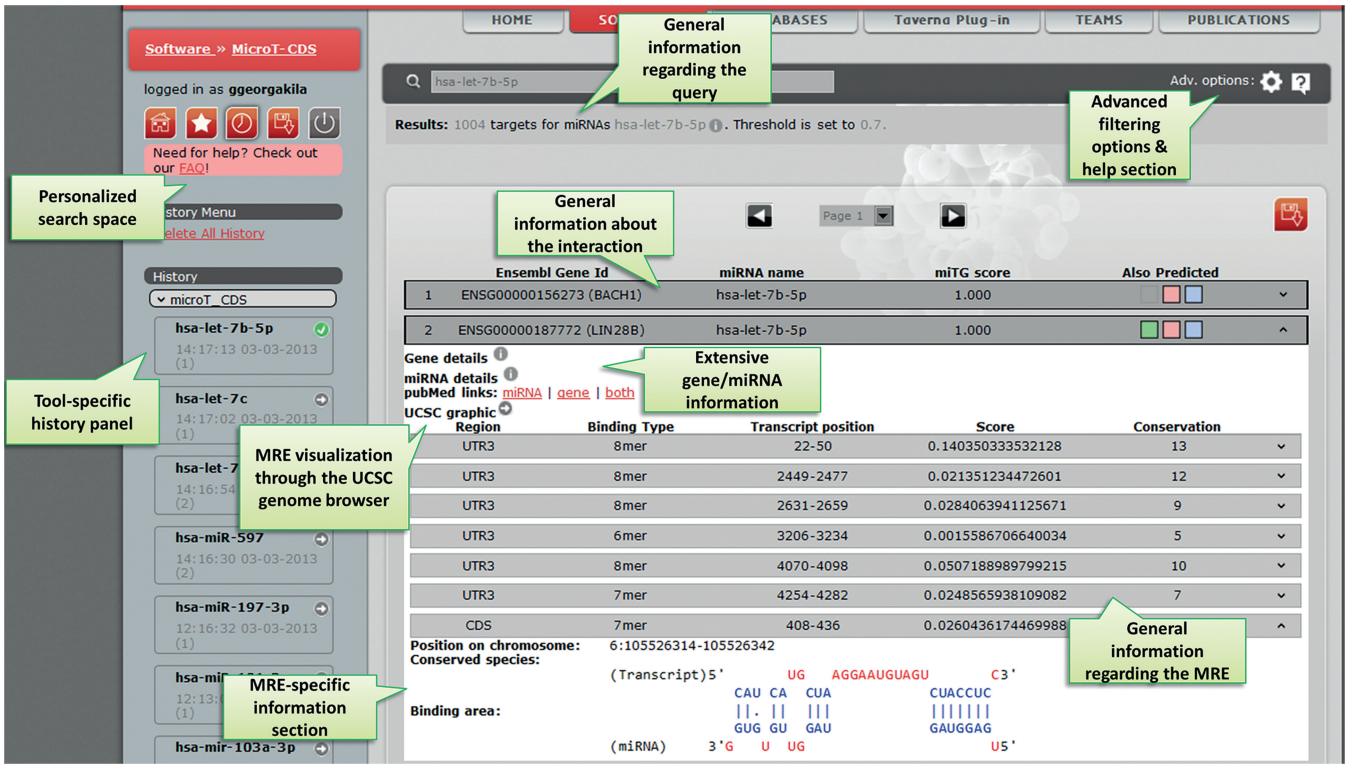

Figure 1. Example of a submitted query in the DIANA-microT web server v5.0. The interface presents information regarding the specified predicted miRNA–mRNA interactions. miRNA and gene-related information, as well as the advanced search options have been expanded. Links to external databases, graphical representation of the binding sites, as well as miRNA-recognition elements (MREs) conservation and prediction scores are displayed in the relevant sections. The left side of the page is devoted to personal user space, reporting latest searches and bookmarks.

repetitions of different prediction thresholds (from sensitive to more stringent), to minimize the effect of the selected settings to the result. By using meta-analysis statistics, the server combines the *P*-values from each repetition into a total *P*-value for each miRNA, signifying its effect on the selected genes for all used thresholds (15,16). In the last step of the pipeline, the identified miRNAs are subjected to a functional analysis, where pathways controlled by the combined action of these miRNAs are detected using DIANA-miRPath v2.1 (16).

Other available pipelines can handle miRNA and gene lists, to perform the enrichment analysis, or even select the type of used interactions (predicted or experimentally validated). In the latter workflow, the algorithm 'personalizes' the target identification module for each miRNA. It initially identifies the number of available interactions in DIANA-TarBase and DIANA-microT-CDS (validated versus predicted) and automatically selects to use validated targets only in the cases of well-annotated miRNAs. Computationally identified interactions are used otherwise.

The new DIANA-microT web server enables users to perform such analyses directly from the on-line user interface, or create even more extensive pipelines programmatically or by using visual tools (Taverna WMS).

Web server integration with Taverna WMS

DIANA-microT web server v5.0 was completely redesigned to provide the necessary building blocks for easily incorporating miRNA functional analysis in complex pipelines. The new DIANA-microT web server aims to facilitate advanced users in creating novel or enhancing existing pipelines with miRNA target identification and functional analysis tools. To this end, DIANA-microT web server v5.0 provides a complete integration with the Taverna WMS, using our in-house developed DIANA-Taverna Plug-in.

DIANA-Taverna Plug-in enables the user to directly access our target prediction server (microT-CDS) from the graphic interface of Taverna and incorporate advanced miRNA analysis functionalities into custom pipelines. Furthermore, the plug-in enables the extension of such pipelines through the use of other DIANA tools and databases, providing access to the most extensive collection of validated miRNA targets (DIANA-TarBase v6.0) and to DIANA-miRPath v2.1, a tool designed for the identification of miRNA targeted pathways. Furthermore, the web server also supports direct programmatic access to all aforementioned utilities in the form of services, to facilitate users having already implemented pipelines using scripting or programming languages.

CONCLUSION

High-throughput techniques have significantly increased the number of users requiring in-depth analysis of

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on May 21, 2013

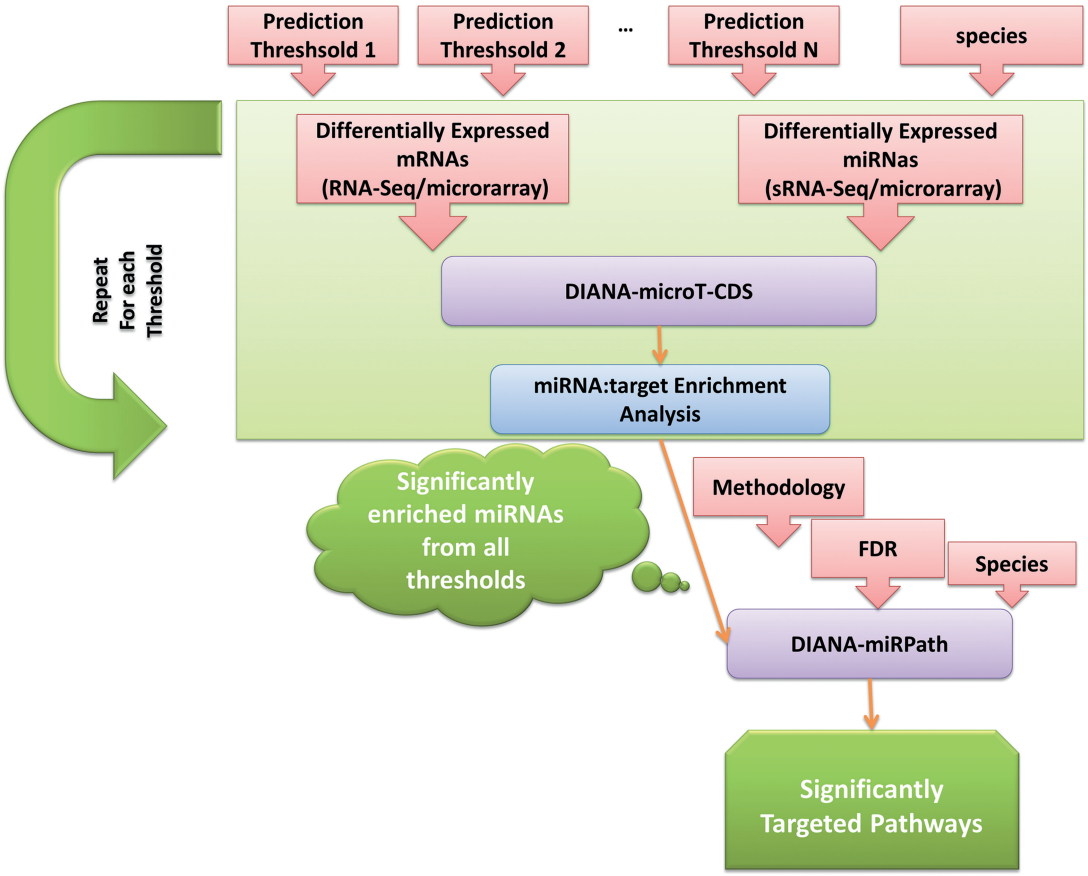

Figure 2. Flowchart depicting an analysis pipeline directly available from the web server interface. Interactions between user-defined miRNA and gene sets are *in silico* identified in 3′UTR and CDS regions using DIANA-microT-CDS. A subsequent miRNA target enrichment analysis identifies miRNAs controlling significantly the sets of differentially expressed genes. The pipeline is automatically repeated for different prediction thresholds (from more sensitive, to more stringent). By using meta-analysis statistics, the server combines the *P*-values from each repetition into a total *P*-value for each miRNA, signifying its effect on the selected genes for all used thresholds. In the last step of the pipeline, miRNA-targeted pathway analysis is implemented with DIANA-miRPath v2.

miRNA–mRNA interactions. These techniques provide great numbers of differentially expressed miRNAs and genes, which have to be analysed in sophisticated pipelines. The new DIANA-microT web server aims to further increase the target prediction accuracy and usability of the server interface, while facilitating users aiming to perform complex miRNA-related analyses.

Compared with the previous version, the new web server has received a major upgrade and is currently up to date with key miRNA/gene repositories, such as Ensembl (v69) and miRBase (v18). The Taverna Plug-in and the integration of cutting-edge workflows in the web interface provide the missing link between experimental results and state-of-the-art functional meta-analyses.

FUNDING

The project [09 SYN-13-1055] ‘MIKRORNA’ by the Greek General Secretariat for Research and Technology and the project ‘TOM’ which is implemented under the ‘ARISTEIA’ Action of the ‘OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING’ and is co-funded by the European Social Fund (ESF) and National Resources. Funding for open access charge: Projects TOM and MIKRORNA.

Conflict of interest statement. None declared.

REFERENCES

1. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
2. Lee,R.C., Feinbaum,R.L. and Ambros,V. (1993) The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, **75**, 843–854.
3. 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.
4. Alexiou,P., Maragkakis,M., Papadopoulos,G.L., Reczko,M. and Hatzigeorgiou,A.G. (2009) Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*, **25**, 3049–3055.
5. Witkos,T.M., Koscianska,E. and Krzyzosiak,W.J. (2011) Practical aspects of microRNA target prediction. *Curr. Mol. Med.*, **11**, 93–109.
6. Maragkakis,M., Vergoulis,T., Alexiou,P., Reczko,M., Plomaritou,K., Gousis,M., Kourtis,K., Koziris,N., Dalamagas,T.

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on May 21, 2013

and Hatzigeorgiou,A.G. (2011) DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Res.*, **39**, W145–W148.
7. Reczko,M., Maragkakis,M., Alexiou,P., Grosse,I. and Hatzigeorgiou,A.G. (2012) Functional microRNA targets in protein coding sequences. *Bioinformatics*, **28**, 771–776.
8. Kozomara,A. and Griffiths-Jones,S. (2011) miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.*, **39**, D152–D157.
9. Flicek,P., Amode,M.R., Barrell,D., Beal,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fairley,S., Fitzgerald,S. *et al.* (2012) Ensembl 2012. *Nucleic Acids Res.*, **40**, D84–D90.
10. Hull,D., Wolstencroft,K., Stevens,R., Goble,C., Pocock,M.R., Li,P. and Oinn,T. (2006) Taverna: a tool for building and running workflows of services. *Nucleic Acids Res.*, **34**, W729–W732.
11. Tay,Y., Zhang,J., Thomson,A.M., Lim,B. and Rigoutsos,I. (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, **455**, 1124–1128.
12. Selbach,M., Schwanhausser,B., Thierfelder,N., Fang,Z., Khanin,R. and Rajewsky,N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
13. Landgraf,P., Rusu,M., Sheridan,R., Sewer,A., Iovino,N., Aravin,A., Pfeffer,S., Rice,A., Kamphorst,A.O., Landthaler,M. *et al.* (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, **129**, 1401–1414.
14. Kanehisa,M., Goto,S., Sato,Y., Furumichi,M. and Tanabe,M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.*, **40**, D109–D114.
15. Ramon,C.L. and Folks,J.L. (1971) Asymptotic optimality of Fisher's method of combining independent tests. *J. Am. Stat. Assoc.*, **66**, 802–806.
16. Vlachos,I.S., Kostoulas,N., Vergoulis,T., Georgakilas,G., Reczko,M., Maragkakis,M., Paraskevopoulou,M.D., Prionidis,K., Dalamagas,T. and Hatzigeorgiou,A.G. (2012) DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic Acids Res.*, **40**, W498–W504.

Downloaded from http://nar.oxfordjournals.org/ at Universitatsbibliothek der Technischen Universitaet Muenchen Zweigbibliothe on May 21, 2013