W128–W134 *Nucleic Acids Research, 2016, Vol. 44, Web Server issue* *Published online 20 May 2016*
doi: 10.1093/nar/gkw455

DIANA-mirExTra v2.0: Uncovering microRNAs and transcription factors with crucial roles in NGS expression data

Ioannis S. Vlachos[1,2,3,*], Thanasis Vergoulis[4], Maria D. Paraskevopoulou[1,2], Filopoimin Lykokanellos[4], Georgios Georgakilas[1,2], Penny Georgiou[4], Serafeim Chatzopoulos[4], Dimitra Karagkouni[1,2], Foteini Christodoulou[5], Theodore Dalamagas[4] and Artemis G. Hatzigeorgiou[1,2,*]

[1]DIANA-Lab, Department of Electrical & Computer Engineering, University of Thessaly, 38221 Volos, Greece, [2]Hellenic Pasteur Institute, 127 Vasilissis Sofias Avenue, 11521 Athens, Greece, [3]Laboratory for Experimental Surgery and Surgical Research 'N.S. Christeas', Medical School of Athens, University of Athens, Athens 11527, Greece, [4]'Athena' Research and Innovation Center, 11524 Athens, Greece and [5]Miroculus, San Francisco, CA 94107, USA

Received March 06, 2016; Revised April 29, 2016; Accepted May 11, 2016

ABSTRACT

Differential expression analysis (DEA) is one of the main instruments utilized for revealing molecular mechanisms in pathological and physiological conditions. DIANA-mirExTra v2.0 (http://www.microrna.gr/mirextrav2) performs a combined DEA of mRNAs and microRNAs (miRNAs) to uncover miRNAs and transcription factors (TFs) playing important regulatory roles between two investigated states. The web server uses as input miRNA/RNA-Seq read count data sets that can be uploaded for analysis. Users can combine their data with 350 small-RNA-Seq and 65 RNA-Seq in-house analyzed libraries which are provided by DIANA-mirExTra v2.0.

The web server utilizes miRNA:mRNA, TF:mRNA and TF:miRNA interactions derived from extensive experimental data sets. More than 450 000 miRNA interactions and 2 000 000 TF binding sites from specific or high-throughput techniques have been incorporated, while accurate miRNA TSS annotation is obtained from microTSS experimental/*in silico* framework. These comprehensive data sets enable users to perform analyses based solely on experimentally supported information and to uncover central regulators within sequencing data: miRNAs controlling mRNAs and TFs regulating mRNA or miRNA expression. The server also supports predicted miRNA:gene interactions from DIANA-microT-CDS for 4 species (human, mouse, nematode and fruit fly). DIANA-mirExTra v2.0 has an intuitive user interface and is freely available to all users without any login requirement.

INTRODUCTION

Gene expression is meticulously regulated by an extensive network of interacting molecules and mechanisms, including epigenetic modifications, transcription factors (TFs) and microRNAs (miRNAs) (1,2). TFs can positively or negatively control gene transcription in *cis*, by binding on DNA regulatory regions. miRNAs are short non-coding RNA species acting as potent regulators of gene expression through mRNA destabilization/degradation and/or translation suppression. miRNAs and TFs play important roles in physiological, as well as pathological conditions, including development, differentiation, metabolic disorders and neoplastic diseases (3). Since a single miRNA or TF can regulate the expression of numerous genes, perturbations in crucial molecules often result in significant phenotypic changes, including cell reprogramming or cancer (1,4). miRNAs and TFs exhibiting central regulatory actions between states or in specific conditions are intensely sought after as targets for further investigation or therapeutic interventions.

The expression, as well as the functional impact of miRNAs and TFs is often explored with next generation sequencing (NGS) techniques. These high-throughput methodologies have revolutionized biological research thanks to the variety of available library preparation techniques and their ability to permit hypothesis-free, data-

*To whom correspondence should be addressed. Tel: +30 24210 74758; Fax: +30 24210 74997; Email: ivlachos@lessr.eu
Correspondence may also be addressed to Artemis G Hatzigeorgiou. Tel: +30 24210 74758; Fax: +30 24210 74997; Email: arhatzig@uth.gr

driven investigations on a genome or transcriptome-wide scale (5). Central molecules for each condition are often hidden within vast lists of candidate hits; while the identification of key regulators requires bioinformatics expertise, infrastructure and extensive experimental data sets.

A few tools have already explored the combined analysis of miRNA/mRNA expression data to identify regulators. Implementations such as Sylamer (6) and DIANA-mirExTra (7) utilized expression data in order to increase the accuracy of miRNA target prediction and to uncover miRNAs that actively regulate mRNA expression between two states. Another category of applications including MMIA (8), MAGIA2 (9), cGRNB (10), miRConnX (11), miRTarVis (12) and TFmiR (13) also integrate TF:mRNA and TF:miRNA interactions in their analysis pipelines. These tools aim to identify miRNAs or TFs with central roles and/or to recreate their underlying regulatory networks. Tools falling in the latter category most often cater network visualization modules and relevant statistics.

Despite the progress in the field, there are still inherent factors limiting the usefulness of the available implementations. Most pipelines rely on the combined usage of *in silico* miRNA target prediction algorithms, while others utilize predictions intersected with CLIP-Seq peaks for search space reduction. Target prediction algorithms have been proven invaluable tools for the characterization of the miRNA interactome. However, even the most sophisticated implementations still produce a high number of false positive results and decrease the signal to noise ratio of their hosting pipelines (3). miRConnX, MAGIA2 and TFmiR extended the utilized *in silico* predictions with experimentally validated interactions from relevant databases such as TarBase (14) and miRecords (15), in order to circumvent this limitation. At that time, experimental databases contained a very sparse instance of the miRNA interactome, comprising only a few thousand interactions, thus reducing their positive impact.

The lack of accurate means for miRNA promoter identification is also an important hindrance to the discovery of TF:miRNA interactions (16). Pri-miRNAs are readily cleaved upon transcription by Drosha enzyme within the cell nucleus, hindering their detection with conventional techniques (16). Due to this effect, miRNA promoters have just recently started to be accurately characterized. Most available implementations utilize the 5′ of miRNA hairpin precursors (pre-miRNAs) as candidate transcription start sites (TSSs) or incorporate TSS predictions from early relevant algorithms, which were based on broad transcription signals, such as histone marks (17). Both approaches have been shown to significantly differ from experimentally identified TSS positions, often by thousands of kilobases (16).

In this manuscript, we present DIANA-mirExTra v2.0, an online web server which enables state-of-the-art functional investigation of miRNAs and TFs from an intuitive user-friendly interface. It bypasses present limitations in the field, and offers novel functionalities or unique features, including:

(i) An extensive suite for differential expression analysis of RNA-Seq and small-RNA-Seq read count data, hosting multiple algorithms and statistics. The suite also supports advanced methodologies for quality checking of NGS expression data and replicate homogeneity tests, including principal component analysis (PCA) and sample clustering.
(ii) mirExTra can utilize miRNA:mRNA, TF:mRNA and TF:miRNA interactions derived from extensive experimental data sets. It currently hosts more than 450 000 experimentally supported miRNA:gene interactions from DIANA-TarBase v7.0 (www.microrna.gr/tarbase) (18) and more than 2 million TF binding sites (TFBSs) derived from miRGen v3.0 (www.microrna.gr/mirgen) (17) and OregAnno v3.0 (www.oreganno.org) (19). It also offers the option to utilize predicted miRNA:gene interactions from DIANA-microT-CDS (www.microrna.gr/microT-CDS) (20).
(iii) Accurate miRNA TSS annotation using microTSS (16) experimental/*in silico* framework.
(iv) Identification of central regulators: miRNAs controlling mRNAs and TFs controlling (activating, repressing or regulating) mRNA or miRNA expression.
(v) A comprehensive database of ready-to-use NGS expression experiments comprising more than 400 libraries from 4 species (*Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caenorhabditis elegans*). The results are derived by analyzing in-house more than 10 billion reads from RNA-Seq and small RNA-Seq experiments (Supplementary Tables S1 and S2).
(vi) Advanced visualizations, including heat maps and interactive network graphs.

The new mirExTra web server has been redesigned from the ground up and caters an application-like interface that permits sophisticated analyses and graphs without the necessity of bioinformatics expertise. The capabilities of the web server, as well as comprehensive statistics and technical information are presented in the following sections.

MATERIALS AND METHODS

mirExTra database

Expression database: RNA-Seq and small-RNA-Seq analysis. More than 350 small-RNA-Seq and 65 RNA-Seq libraries have been analyzed in-house and incorporated into the mirExTra expression database. The supported experiments cover more than 70 different tissues, 100 cell types and 90 conditions in 4 species (*Homo sapiens*, *Mus musculus*, *Drosophila melanogaster*, *Caernorhabditis elegans*). The data sets were derived from publically available repositories, including Gene Expression Omnibus (GEO) (21), UCSC (22), ENCODE (23,24) and modENCODE (25). Detailed information regarding the analyzed libraries, sample tissue/cell types and their derived sources are presented in Supplementary Tables S1 and S2. The tables also provide the sample accession codes, which can be used to download the raw sequencing files from the relevant repositories. All sequencing libraries were quality-checked using FastQC (26), while contaminants were detected using an in-house developed algorithm and removed with Trimgalore (27) and Trimmomatic (28). Expression at mRNA transcript level was estimated using RSEM (29). Small-RNA-Seq reads were aligned against mature and precursor

miRNA sequences derived from miRBase database (30), as well as against the relevant genomes using Bowtie v1 (31), allowing up to 1 mismatch. Genome assemblies were downloaded from Ensembl database (32).

miRNA and TF interactions. DIANA-mirExTra v2.0 supports analyses using predicted or experimentally validated miRNA:gene interactions. *In silico*-predicted miRNA targets are assessed with DIANA-microT-CDS (20) target prediction algorithm, which enables the identification of miRNA binding sites in 3′UTR and CDS regions. Currently, the database contains interactions for four supported species. All predicted interactions can be browsed and/or downloaded at the microT-CDS web page: www.microrna.gr/microT-CDS. Experimentally supported miRNA:mRNA interactions were obtained by integrating DIANA-TarBase v7.0 to mirExTra v2.0 web server. More than 450 000 experimentally supported interactions have been included for human and mouse, enabling the complete substitution of *in silico* interactions with experimental data. The data set can be examined in depth using the online database interface: www.microrna.gr/tarbase. All interactions can be downloaded for local use from the TarBase website.

TFBSs are derived from in-house analyzed DNAse-Seq data sets and entries from the community curated ORegAnno v3.0 database (www.oreganno.org) (19). The analysis of DNAse-Seq data sets was performed using HOMER (33) and Wellington (34), for the detection of enriched regions and hotspot processing, respectively. TF binding coordinates, motifs and binding site annotations were derived from miRGen v3.0 (17). In case of overlapping TFBSs, the binding site with the strongest signal was selected. The finalized TFBS data set comprises more than 2M binding sites for human and mouse.

mRNA annotation was obtained from Ensembl 80 database (GRCh38, GRCm38, BDGP6 and WBcel235). Human and mouse transcripts were filtered to keep only high quality principal coding isoforms. The most 5′ TSS of the remaining isoforms was selected as the gene TSS. miRNA TSSs were identified using microTSS experimental/*in silico* framework (16). TSSs from 9 cell lines and 6 tissues of *Homo sapiens* and *Mus musculus* were derived from miRGen v3.0. In case of alternative tissue-specific miRNA promoters, the most 5′ TSS was selected as the final pri-miRNA TSS. All TF:miRNA interactions and miRNA TSS genomic coordinates can be examined and/or downloaded from the miRGen v3.0 database website (http://www.microrna.gr/mirgen/). The ORegAnno v3.0 data set is also freely available for local download at the relevant website (www.oreganno.org).

Algorithms and methodologies

Implementation. DIANA-mirExTra web interface (Figures 1 and 2) was developed in PHP and JavaScript using the Yii framework. PostgreSQL facilitated the implementation of the main database, comprising information regarding miRNAs, mRNAs and TFs, along with their associations. The schema also supports the storage and retrieval of small-RNA-Seq and RNA-Seq expression data in the miRExtra expression database. All SQL queries were optimized using indices to achieve optimal performance. All analysis scripts were implemented using R/Bioconductor (35,36).

Differential expression analysis module. This module supports differential expression analyses between two conditions using different algorithms and sophisticated statistics. Users can upload and analyze their own experiments or directly utilize the NGS expression data sets present in the mirExTra database to perform exploratory investigations or to complement their own experimental data. DIANA-mirExTra supports miRNA and mRNA differential expression analyses using three reference algorithms: DESeq (37), edgeR (38) and limma (39). Data are automatically normalized using RLE, TMM and TMM + voom algorithms for DESeq, edgeR and limma, respectively. Multiple comparison plots are automatically produced to enhance user understanding of the data: FDR versus log fold change volcano plots, fold change versus log expression MA plots and annotated heat maps (Supplementary Figures S1–S3). The server utilizes also advanced techniques for sample quality and group homogeneity checks. It calculates and plots the first two principal components following PCA for dimensionality reduction (Supplementary Figure S4), and a sample clustering heat map (Supplementary Figure S5) is created based on the expression levels. All methodologies and graphs are equally supported for miRNAs and mRNAs, and results are calculated in real time. The server automatically provides result links, in case the user settings initiate an analysis that will require more than 1 min to complete. The links remain active for 48 h.

Finding miRNAs and TFs with crucial roles. DIANA-mirExTra v2.0 enables the identification of miRNAs and TFs with central regulatory roles in their expression data. This module can directly analyze results of a previously performed differential expression analysis or import user data sets. The module has three distinct functions, depending on the profile of the investigated regulators: (i) miRNAs controlling mRNAs, (ii) TFs regulating mRNAs or (iii) TFs regulating miRNAs. The user interface automatically adjusts to the user's selections and provides tools and options specific to the analysis. Depending on the module, the web server combines differentially expressed regulators (i.e. miRNAs or TFs) with the relevant differentially expressed targets, and performs overrepresentation analyses based on the hypergeometric distribution. miRNA targets are investigated within lists of mRNAs with opposite differential expression results (up-regulated miRNAs versus down-regulated mRNAs, and *vice versa*), while TF targets are selected from different result subsets, depending on the sought TF subtype (activators, repressors or TFs with mixed activities). The definition of 'differential expression' is quite liberal in this module and users can select *P*-value, FDR or fold-change levels as the relevant thresholds, in order to support stringent or sensitive analyses. The module returns extensive statistics, as well as the overrepresentation *P*-value for each regulator; enabling the discovery of molecules with statistically significant functional impact. The module also automatically creates interactive networks

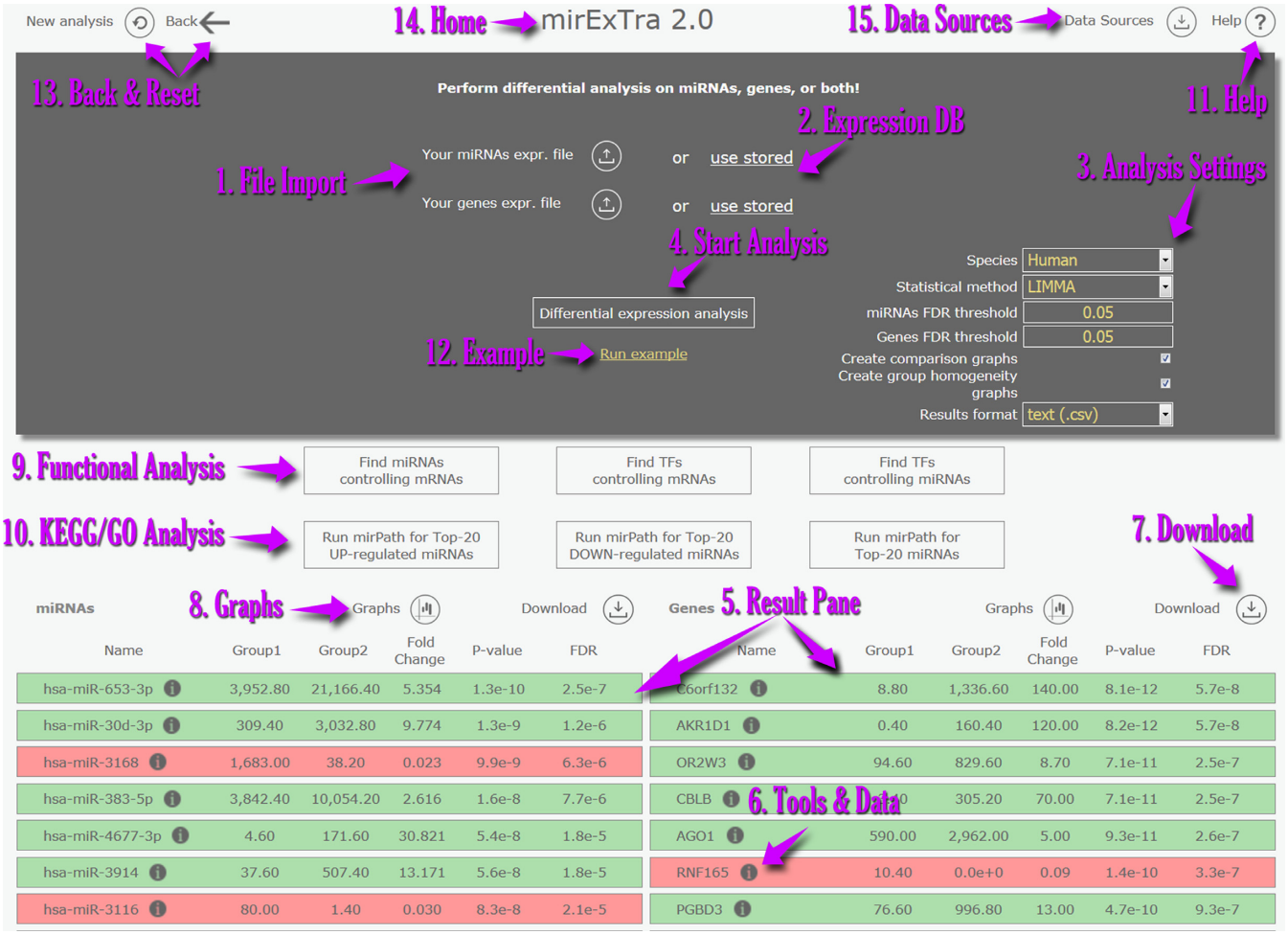

Figure 1. User interface of the Differential Expression Analysis Module. Users can upload miRNA and/or mRNA expression files[1] or select samples from the expression database[2]. miRNAs and mRNAs can be analyzed together or separately. Subsequently, users can select the species annotation, statistical methods, FDR thresholds, file output format and whether they require comparison (volcano, MA, heat map) or homogeneity (PCA, clustering) plots to be created[3]. Following analysis[4], results are presented in a separate pane[5], which is divided into two distinct sections for miRNAs and genes, respectively. It provides information such as transcript name, mean expression in each group, fold change, *P*-values and FDR level. Entries are colored regarding to the comparison outcome and accompanied by an information button[6], which leads to further details, meta-data and connects them with internal (DIANA) or external tools and databases. All results can be downloaded in the user selected format[7]. The requested graphs can be viewed and downloaded by pressing the relevant button[8]. Users can forward the results to mirExTra functional analysis modules[9] or to miRPath v3.0 for pathway and GO analyses[10]. A comprehensive help section[11] and an example run[12] support user interaction with the server. Back, reset[13] and home[14] buttons facilitate the navigation between different mirExTra functionalities. All data sets utilized in mirExTra v2.0 are presented in detail in the 'Data Sources' section[15].

using the Cytoscape.js (40) library presenting the most important regulators and their targets (Figure 3). These interactive networks allow users to easily identify common targets or assess the regulatory function of molecules by their position within the graph.

User interface and presentation of results. DIANA-mirExTra v2.0 interface is designed to provide an application-like environment. It automatically adjusts to user selections without having to load (or reload) the relevant web pages. The initial screen leads to the two main mirExTra modules. In the differential expression module, users can select to analyze any stored miRNA/mRNA expression data or upload their own experimental results. The interface of mirExTra expression database facilitates the selection of data sets by providing extensive meta-data, including species, tissue, cell type, conditions, developmental stages and time points. All samples are accompanied by their accession IDs, the relevant repository and a direct link for in-depth investigation of the specific data set or local download of the raw sequencing data. Following the upload or selection of the expression data, users can select the analysis algorithm (DESeq, edgeR or limma), the significance threshold (FDR) and whether they also require advanced comparison (volcano, heatmaps and MA plots) or group homogeneity graphs (PCA and clustering). The module can serve as an autonomous application for differential expression analysis of miRNA and mRNA expression data sets, since all results can be directly exported to popular file formats (tabulated texts or MS Excel spreadsheets). The exported results are enhanced with rich metadata, including normalized read counts for each

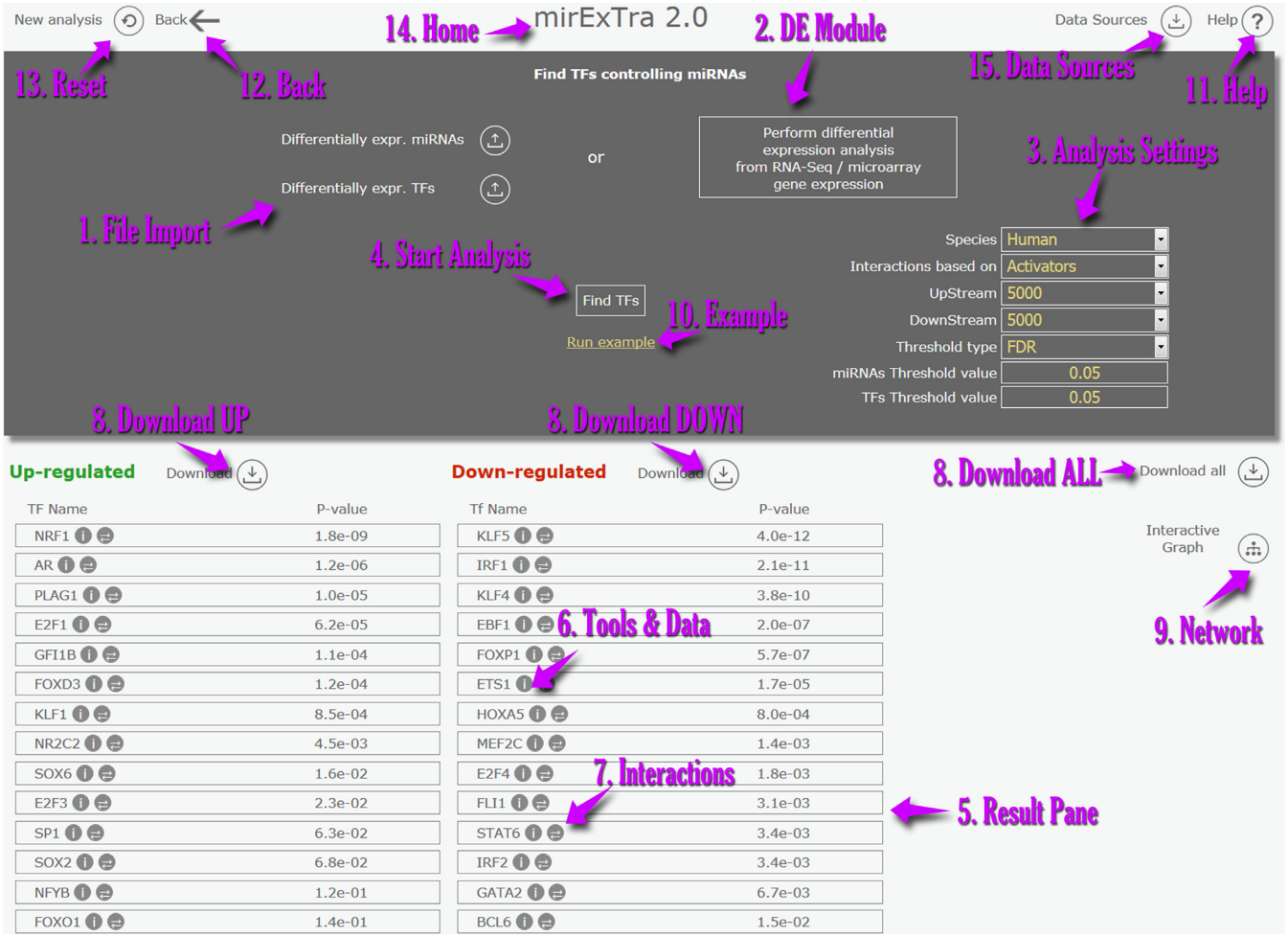

Figure 2. User interface of the Functional Analysis Module. Users can directly upload files for analysis[1] or perform the preparatory differential expression analysis directly from mirExTra[2]. Depending on the investigated regulators (miRNAs, TFs), the module offers different setup options[3], including (for TFs) the analyzed species, TF function, assessed promoter region, as well as the relevant thresholds for the regulator and their targets. Following analysis[4], the results are presented in a relevant pane below[5]. Regulators with increased or decreased expression are placed in separate columns. The information link[6] leads to internal and external tools or data, while the interactions link[7] permits users to examine all identified interactions for each regulator. All results can be downloaded[8] and saved for future reference. The module also automatically creates interactive network graphs[9] for the regulators with the lowest *P*-values. User interaction with the interface is supported by the example[10], help[11] and navigation[12–14] buttons. All utilized data sets are presented in the 'Data Sources' section[15].

sample and group, fold changes, *P*-values, FDR levels, database identifiers (i.e. miRBase and Ensembl accessions), genomic locations and gene descriptions.

The same modular approach is also followed when users select: *'Find miRNAs and TFs with crucial roles'*. The interface can directly lead to the relevant analysis page from within the Differential Expression Analysis module or it can be used as an autonomous suite for functional investigation. The interface automatically adjusts to the use-case scenario and prompts the user to import relevant data sets if the module is used autonomously. The desired type of threshold for the analysis (FDR, *P*-value, fold-change) as well as the source of interactions can be subsequently selected. In case of miRNA functional investigation they can choose between experimentally supported or predicted interactions (TarBase or microT-CDS), while in TF role assessment, options specific to TF function (activator, repressor, mixed activity) and promoter search region (upstream and downstream bases from the TSS) are provided.

All results are enhanced with active links to internal and external tools and sources of metadata. For instance, mirExTra users can select individual miRNAs to query against the DIANA website for predicted or experimentally validated targets in mRNAs (microT-CDS and TarBase v7.0, respectively) or lncRNAs (lncBase v2.0 (41)). Importantly, miRNAs with lowest *P*-values and strongest functional support can be directly exported to DIANA-miRPath v3.0 (42) for pathway and gene ontology (GO) (43) analyses. All results from this module can be also downloaded for future reference or downstream analysis. Similar to all DIANA-mirExTra v2.0 modules, it can be used as an autonomous suite and downloaded results are accompanied by rich meta-data. Interactive network graph visualizations are created automatically and can be viewed upon user selection.

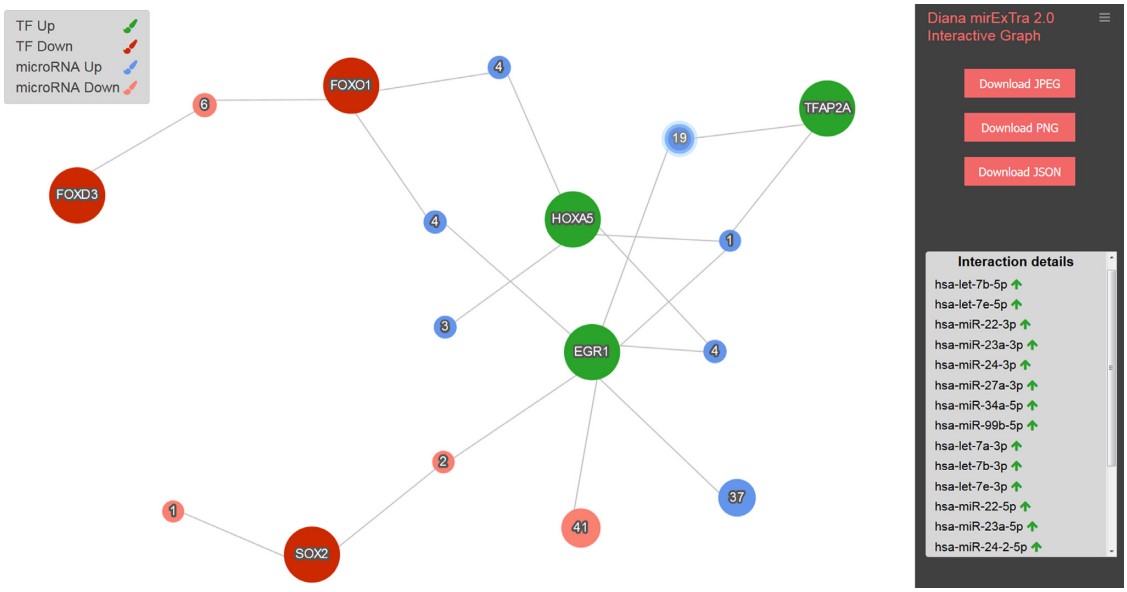

Figure 3. Interactive network graph output from mirExTra v2.0 functional analysis module. The network depicts up- and down-regulated TFs (green and red, respectively) along with their miRNA targets (blue and pink). The module was set to identify TFs functioning as activators and therefore targets and regulators tend to be correspondingly differentially expressed. Each regulator node (green/red nodes) hosts the regulator name, while target nodes (blue/pink nodes) depict the number of grouped targets. By selecting a target node all grouped targets are presented in the relevant pane (bottom right).

CONCLUSION

NGS techniques have revolutionized biomedical research due to their unprecedented scope and yield. Most often, significant findings and results are 'lost' within genome and transcriptome-wide result lists, while further functional analyses require bioinformatics expertise, time and computational resources. Until recently, even the best custom analysis pipelines were obstructed by the lack of adequate experimentally validated miRNA:mRNA interactions and miRNA promoter locations, since *in silico* predictions still return a high number of false positive results.

DIANA-mirExTra v2.0 is an online web server enabling users perform A-to-Z functional analyses, starting from NGS expression data to the identification of important regulators with crucial roles in the investigated libraries. Users can analyze their own experiments or utilize the extensive mirExTra NGS expression library, in order to assess the role of miRNAs and TFs in various states, diseases and conditions. DIANA-mirExTra v2.0 permits complete substitution of *in silico* predictions with experimentally supported interactions and TSS positions for human and mouse. Importantly, the new web server performs sophisticated methodologies and advanced visualizations from a user-friendly interface. The multifaceted modular structure of this web application permits numerous different use-case scenarios and enables researchers to utilize DIANA-mirExTra v2.0 as a one stop shop for differential expression, functional or investigative analyses.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Konstantinos Zagganas for his technical guidance during the development of mirExTra Web interface, as well as Joanna-Elzbieta Handzlik and Marielena Georgaki for their valuable assistance in the creation of the mirExTra expression database.
A significant part of the computations were performed on 'ARIS' National HPC Infrastructure of the Greek Research and Technology Network.

FUNDING

This work has been supported by the project "NGS-infect" from the Greek General Secretary of Research and Technology and by a Fondation Santé grant to Artemis Hatzigeorgiou. Funding for open access charges: Fondation Santé grant (Artemis Hatzigeorgiou).
Conflict of interest statement. None declared.

REFERENCES

1. Hobert,O. (2008) Gene regulation by transcription factors and microRNAs. *Science*, **319**, 1785–1786.
2. Chen,K. and Rajewsky,N. (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat. Rev. Genet.*, **8**, 93–103.
3. Vlachos,I.S. and Hatzigeorgiou,A.G. (2013) Online resources for miRNA analysis. *Clin. Biochem.*, **46**, 879–900.
4. Luningschror,P., Hauser,S., Kaltschmidt,B. and Kaltschmidt,C. (2013) MicroRNAs in pluripotency, reprogramming and cell fate induction. *Biochim. Biophys. Acta*, **1833**, 1894–1903.
5. Shendure,J. and Aiden,E.L. (2012) The expanding scope of DNA sequencing. *Nat. Biotechnol.*, **30**, 1084–1094.
6. van Dongen,S., Abreu-Goodger,C. and Enright,A.J. (2008) Detecting microRNA binding and siRNA off-target effects from expression data. *Nat. Methods*, **5**, 1023–1025.
7. Alexiou,P., Maragkakis,M., Papadopoulos,G.L., Simmosis,V.A., Zhang,L. and Hatzigeorgiou,A.G. (2010) The DIANA-mirExTra web

server: from gene expression data to microRNA function. *PLoS One*, **5**, e9171.
8. Li,Z., Wang,L., Xu,J. and Yang,Z. (2015) MiRNA expression profile and miRNA-mRNA integrated analysis (MMIA) during podocyte differentiation. *Mol. Genet. Genomics*, **290**, 863–875.
9. Bisognin,A., Sales,G., Coppe,A., Bortoluzzi,S. and Romualdi,C. (2012) MAGIA2: from miRNA and genes expression data integrative analysis to microRNA–transcription factor mixed regulatory circuits (2012 update). *Nucleic Acids Res.*, **40**, W13–W21.
10. Xu,H., Yu,H., Tu,K., Shi,Q., Wei,C., Li,Y.Y. and Li,Y.X. (2013) cGRNB: a web server for building combinatorial gene regulatory networks through integrated engineering of seed-matching sequence information and gene expression datasets. *BMC Syst. Biol.*, **7**(Suppl. 2), S7.
11. Huang,G.T., Athanassiou,C. and Benos,P.V. (2011) mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic Acids Res.*, **39**, W416–W423.
12. Jung,D., Kim,B., Freishtat,R.J., Giri,M., Hoffman,E. and Seo,J. (2015) miRTarVis: an interactive visual analysis tool for microRNA-mRNA expression profile data. *BMC Proc.*, **9**, S2.
13. Hamed,M., Spaniol,C., Nazarieh,M. and Helms,V. (2015) TFmiR: a web server for constructing and analyzing disease-specific transcription factor and miRNA co-regulatory networks. *Nucleic Acids Res.*, **43**, W283–W288.
14. Papadopoulos,G.L., Reczko,M., Simossis,V.A., Sethupathy,P. and Hatzigeorgiou,A.G. (2009) The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res.*, **37**, D155–D158.
15. Xiao,F., Zuo,Z., Cai,G., Kang,S., Gao,X. and Li,T. (2009) miRecords: an integrated resource for microRNA-target interactions. *Nucleic Acids Res.*, **37**, D105–D110.
16. Georgakilas,G., Vlachos,I.S., Paraskevopoulou,M.D., Yang,P., Zhang,Y., Economides,A.N. and Hatzigeorgiou,A.G. (2014) microTSS: accurate microRNA transcription start site identification reveals a significant number of divergent pri-miRNAs. *Nat. Commun.*, **5**, 5700.
17. Georgakilas,G., Vlachos,I.S., Zagganas,K., Vergoulis,T., Paraskevopoulou,M.D., Kanellos,I., Tsanakas,P., Dellis,D., Fevgas,A., Dalamagas,T. *et al.* (2016) DIANA-miRGen v3.0: accurate characterization of microRNA promoters and their regulators. *Nucleic Acids Res.*, **44**, D190–D195.
18. Vlachos,I.S., Paraskevopoulou,M.D., Karagkouni,D., Georgakilas,G., Vergoulis,T., Kanellos,I., Anastasopoulos,I.L., Maniou,S., Karathanou,K., Kalfakakou,D. *et al.* (2015) DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.*, **43**, D153–D159.
19. Lesurf,R., Cotto,K.C., Wang,G., Griffith,M., Kasaian,K., Jones,S.J., Montgomery,S.B., Griffith,O.L. and The Open Regulatory Annotation, Consortium. (2016) ORegAnno 3.0: a community-driven resource for curated regulatory annotation. *Nucleic Acids Res.*, **44**, D126–D132.
20. Paraskevopoulou,M.D., Georgakilas,G., Kostoulas,N., Vlachos,I.S., Vergoulis,T., Reczko,M., Filippidis,C., Dalamagas,T. and Hatzigeorgiou,A.G. (2013) DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.*, **41**, W169–W173.
21. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.
22. Rosenbloom,K.R., Armstrong,J., Barber,G.P., Casper,J., Clawson,H., Diekhans,M., Dreszer,T.R., Fujita,P.A., Guruvadoo,L., Haeussler,M. *et al.* (2015) The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.*, **43**, D670–D681.
23. Yue,F., Cheng,Y., Breschi,A., Vierstra,J., Wu,W., Ryba,T., Sandstrom,R., Ma,Z., Davis,C., Pope,B.D. *et al.* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, **515**, 355–364.
24. Consortium,E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
25. Celniker,S.E., Dillon,L.A., Gerstein,M.B., Gunsalus,K.C., Henikoff,S., Karpen,G.H., Kellis,M., Lai,E.C., Lieb,J.D., MacAlpine,D.M. *et al.* (2009) Unlocking the secrets of the genome. *Nature*, **459**, 927–930.
26. Andrews,S. (2015) FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
27. Krueger,F. (2015) Trim Galore!: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
28. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
29. Li,B. and Dewey,C. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
30. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
31. 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.
32. Yates,A., Akanni,W., Amode,M.R., Barrell,D., Billis,K., Carvalho-Silva,D., Cummins,C., Clapham,P., Fitzgerald,S., Gil,L. *et al.* (2016) Ensembl 2016. *Nucleic Acids Res.*, **44**, D710–D716.
33. Heinz,S., Benner,C., Spann,N., Bertolino,E., Lin,Y.C., Laslo,P., Cheng,J.X., Murre,C., Singh,H. and Glass,C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
34. Piper,J., Elze,M.C., Cauchy,P., Cockerill,P.N., Bonifer,C. and Ott,S. (2013) Wellington: a novel method for the accurate identification of digital genomic footprints from DNase-seq data. *Nucleic Acids Res.*, **41**, e201.
35. Core Team,R. (2015) R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*. Vienna, https://www.R-project.org.
36. Gentleman,R.C., Carey,V.J., Bates,D.M., Bolstad,B., Dettling,M., Dudoit,S., Ellis,B., Gautier,L., Ge,Y., Gentry,J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.*, **5**, R80.
37. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
38. 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.
39. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
40. Franz,M., Lopes,C.T., Huck,G., Dong,Y., Sumer,O. and Bader,G.D. (2016) Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, **32**, 309–311.
41. Paraskevopoulou,M.D., Vlachos,I.S., Karagkouni,D., Georgakilas,G., Kanellos,I., Vergoulis,T., Zagganas,K., Tsanakas,P., Floros,E., Dalamagas,T. *et al.* (2016) DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.*, **44**, D231–D238.
42. Vlachos,I.S., Zagganas,K., Paraskevopoulou,M.D., Georgakilas,G., Karagkouni,D., Vergoulis,T., Dalamagas,T. and Hatzigeorgiou,A.G. (2015) DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.*, **43**, W460–W466.
43. 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.