

# DIANA-miRPath v4.0: expanding target-based miRNA functional analysis in cell-type and tissue contexts

Spyros Tastsoglou<sup>1,2,\*</sup>, Giorgos Skoufos<sup>1,2</sup>, Marios Miliotis<sup>1,2</sup>, Dimitra Karagkouni<sup>3,4,5</sup>, Ioannis Koutsoukos<sup>1</sup>, Anna Karavangeli<sup>1,2</sup>, Filippos S. Kardaras<sup>1,2</sup> and Artemis G. Hatzigeorgiou<sup>1,2,\*</sup>

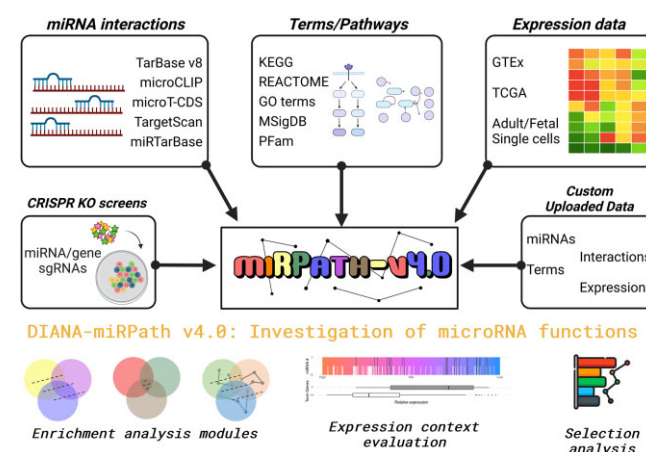
<sup>1</sup>DIANA-Lab, Department of Computer Science and Biomedical Informatics, Univ. of Thessaly, Lamia 35131, Greece, <sup>2</sup>Hellenic Pasteur Institute, Athens 11521, Greece, <sup>3</sup>Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA, USA, <sup>4</sup>Harvard Medical School, Boston, MA, USA and <sup>5</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA

Received February 28, 2023; Revised May 07, 2023; Editorial Decision May 08, 2023; Accepted May 30, 2023

## ABSTRACT

DIANA-miRPath is an online miRNA analysis platform harnessing predicted or experimentally supported miRNA interactions towards the exploration of combined miRNA effects. In its latest version (v4.0, <http://www.microrna.gr/miRPathv4>), DIANA-miRPath breaks new ground by introducing the capacity to tailor its target-based miRNA functional analysis engine to specific biological and/or experimental contexts. Via a redesigned modular interface with rich interaction, annotation and parameterization options, users can now perform enrichment analysis on Gene Ontology (GO) terms, KEGG and REACTOME pathways, sets from Molecular Signatures Database (MSigDB) and PFAM. Included miRNA interaction sets are derived from state-of-the-art resources of experimentally supported (DIANA-TarBase v8.0, miR-TarBase and microCLIP cell-type-specific interactions) or from *in silico* miRNA–target interactions (updated DIANA-microT-CDS and TargetScan predictions). Bulk and single-cell expression datasets from The Cancer Genome Atlas (TCGA), the Genotype-Tissue Expression project (GTEx) and adult/fetal single-cell atlases are integrated and can be used to assess the expression of enriched term components across a wide range of states. A discrete module enabling enrichment analyses using CRISPR knock-out screen datasets enables the detection of selected miRNAs with potentially crucial roles within conditions under study. Notably, the option to upload custom interaction, term, expression and screen sets further expands the versatility of miRPath webserver.

## GRAPHICAL ABSTRACT



## INTRODUCTION

microRNAs (miRNAs) (1) are central post-transcriptional modulators of gene expression and protein production that exhibit abundant yet distinct expression profiles across tissues and cell-types. Robust high throughput methods directly capture functional miRNA Recognition Elements (MREs) for one or more miRNAs in the 3' untranslated region (UTR) and/or the coding (CDS) sequence of most transcripts, providing valuable insights into miRNA regulation of most biological processes. The relative abundance of miRNAs and their targets creates unique RNA landscapes in different cell types, tissues and conditions, in which distinct miRNA regulatory roles are exerted.

*De novo* miRNA target prediction methods (2,3) and collections of miRNA interactions (4) supported by robust high-throughput experimental methods (5–7) paved the way for the development of applications conducting target-based analysis of miRNA functions (8–10). Previ-

\*To whom correspondence should be addressed. Tel: +1 30 2421074758; Fax: +1 30 24210 74997; Email: arhatzig@uth.gr  
 Correspondence may also be addressed to Spyros Tastsoglou. Tel: +130 2421074758; Fax: +1 30 2421074997; Email: tastsoglou@uth.gr

ous versions of DIANA-miRPath (11) provided a statistical framework for evaluation of combined miRNA effects on Gene Ontology (GO) terms and KEGG pathways, based on predicted and experimentally supported targets.

DIANA-miRPath v4.0 is the first relevant application to integrate robust experimental (12,13) and predicted (2,3) interactions with phenotypic and gene expression contexts, towards tailored miRNA functional analyses. In the current version, analyses have been dissected into separate modules that each answer different biological questions (Figure 1). Upon providing one or multiple miRNAs and selecting resources of miRNA interactions and terms/pathways/ontologies (hereafter, 'terms'), enrichment analysis can be performed using the miRNA targets union, intersections, or more complex designs (Figure 1A). miRPath v4.0 extends the functionality of previous versions towards context-specific interrogations by applying nonparametric analysis of targeted and non-targeted term components in cell types, tissues and diseases. Nonparametric testing of log-scaled gene expression fold changes has been successfully utilized in the past to assess the functional efficacy of miRNA binding events (2,7). In the current scenario, it is repurposed to enable the identification of contexts and states in which the distribution of the targets is found repressed/de-regulated, transforming miRPath analyses from context-agnostic to context-aware endeavours (Figure 1B). Distribution testing is performed against internal libraries of expression sets, including tissues (healthy or cancerous) and cell types (adult or fetal), outputting results per term, sorted by level of significance. The option to also upload and use custom datasets expands the applicability of context-specific testing even more. In the absence of specific miRNAs of interest, explorations focused instead on terms can be initiated via Term-centric analysis, uncovering potential common miRNA regulators across one or multiple terms (Figure 1C). A plethora of interaction, term and context options are available by default in DIANA-miRPath v4.0 (Figure 1D).

Potential biases in miRNA functional analysis, relevant to the multiplicity of miRNA targeting, as well as the 3'-UTR length and MRE differences among transcripts within terms, have been discussed in the past (11,14). In DIANA-miRPath v3.0, permutation-based testing, which is computationally intensive and yields non-granular p-values, was provided as an alternative testing method in order to close this gap. In the current version, it is replaced by Wallenius non-central hypergeometric testing (15), parameterized to account for bias in targeting events among tested terms.

DIANA-miRPath v4.0 provides a novel module that caters miRNA-specific functional analysis of positive/negative selection results obtained using CRISPR knock-out screen libraries in stressors of interest. Via the relevant module, users can provide tab-separated results from MAGECK analysis (RRA or MLE) of libraries targeting both genes and miRNAs, such as GeCKO v2.0 human (16,17), and obtain significantly enriched terms, miRNAs and targets that are positively/negatively selected in conditions under study.

MIENTURNET (18), miRPathDB 2.0 (19) and miR2GO (20) are three webservers that are also devoted to miRNA functional analysis. In MIENTURNET users

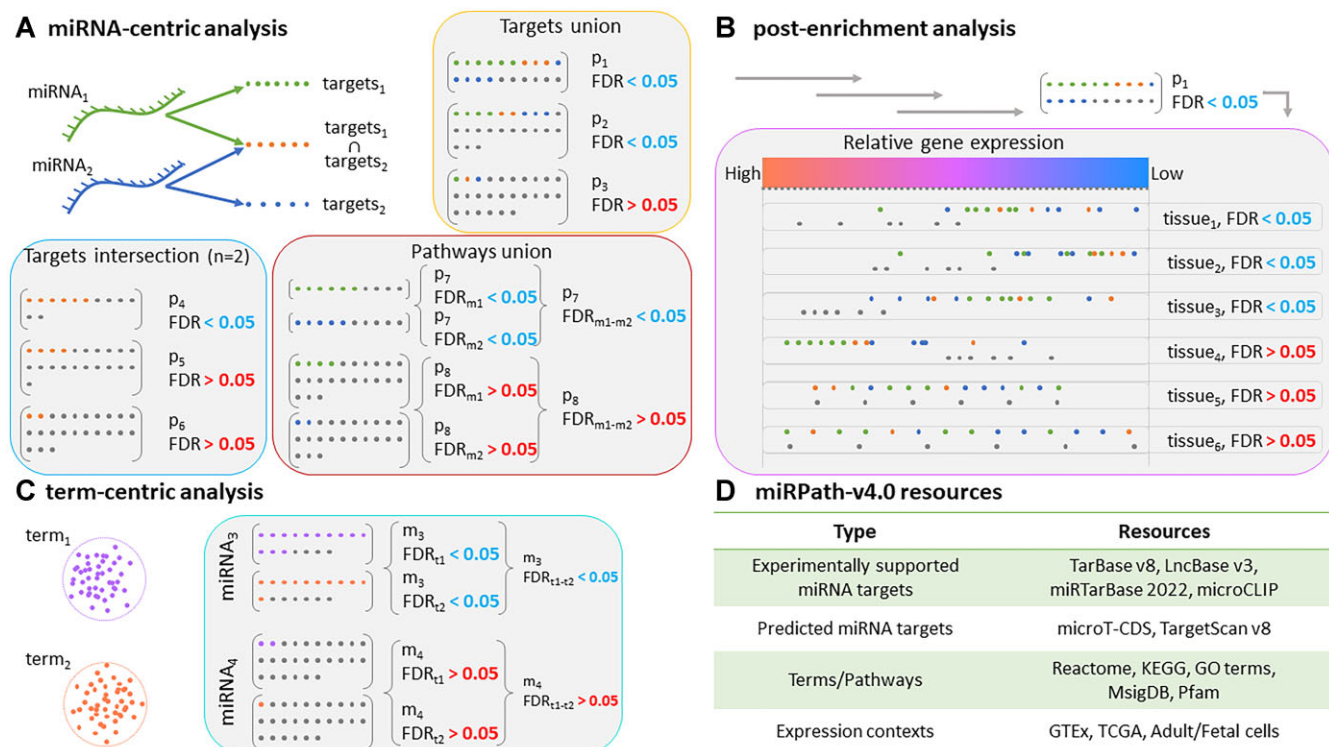
can also produce miRNA-mRNA interaction networks however a term-based analysis is not supported. miRPathDB 2.0 provides a set of downstream information for the queried miRNA (e.g. stem-loop, mature and seed sequence) for both miRBase and miRCarta annotations but supports only human and mouse species and single miRNA queries. On the contrary, the miRNA-centric module of miRPath-v4.0 performs functional enrichment analysis allowing as input both single and multiple miRNAs, to uncover the potential combinatorial effects of miRNAs in terms. miR2GO provides an interface for miRNA functional analysis that also evaluates the functional effects of mutations in miRNA seeds. Even though this feature is not catered as a built-in miRPath capacity, users may still conduct such investigations by making use of the Upload Dashboard module.

miRPath-v4.0 extends the functionality of its previous version (DIANA-miRPath-v3.0) and offers numerous improvements over all the aforementioned implementations. miRPath-v4.0 provides the largest set of resources for both experimentally supported and predicted miRNA-gene interactions, biological terms and pathways and gene expression profiles. Importantly, it is the only method to our knowledge to incorporate gene expression profiles with single-cell resolution, as demonstrated in Supplementary Note S1, Supplementary Figures S1–S4, and Supplementary Tables S1–S3. Two newly introduced key advantages of miRPath-v4.0 are the dedicated workflow for miRNA-centric functional analysis of CRISPR-KO Screen data, for which a specific use-case scenario (21) is provided in Supplementary Note S2 and Supplementary Tables S4–S5, and the Upload Dashboard, where users can unlock the full capacity of miRPath with custom annotations (i.e. miRNAs, genes, terms) for the species/phenotypes of their choice. The case that specific terms are of interest, yet no miRNAs are provided to initiate a miRNA-centric analysis is shown in Supplementary Note S3 and Supplementary Tables S6–S7, which provide a Term-centric miRPath application. The latest version of miRPath is the only relevant implementation to provide support for both miRBase and MirGeneDB miRNA annotation schemes. It hosts the largest number of supported species, offers an option to account for bias in targeting events and provides a completely redesigned, intuitive interface to enhance user experience and enable its ingenious use.

## METHODS AND RESULTS

### Statistical methods

**Main analysis engine.** Enrichment analysis is performed by parameterizing function *kegga()* from *limma* package (22) to user-selected resources. Depending on the scenario, the union or intersection(s) of miRNA targets is calculated and subjected to testing. In the case of Pathways union, the targets of each miRNA are tested separately and significantly enriched instances are merged per-term using Fisher's method (23), as implemented in *metap* package. The option to account for targeting differences among tested terms through Wallenius non-central hypergeometric testing is integrated using packages *GOseq* (15) and *Biase-dUrn*.



**Figure 1.** Primary DIANA-miRPath v4.0 analysis modules. (A) In miRNA-centric analysis, miRPath v4.0 evaluates term overrepresentation in targets of the input miRNAs, via one-sided Fisher's exact test. The Targets Union module tests the combined effect of all miRNAs. Identification of  $n$  miRNAs compensating each other's function by targeting the same mRNAs within specific terms ( $n$ , user-defined) can be performed under Targets Intersection module. To highlight potent term regulators, Pathways Union module assesses the individual significance of each input miRNA against each term and *a posteriori* combines separate p-values into a merged per-term significance level. (B) Significantly enriched terms obtained through miRNA-centric analysis can be further investigated, to uncover cell types and healthy/cancerous tissues where the expression of miRNA-targeted genes is shifted considerably relative to non-targeted genes. One-sided MWU tests are the default option, enabling the detection of instances where targets are expressed in lower levels; however the miRPath engine accommodates various exploratory scenarios, also permitting KS tests and two-sided testing. (C) Under term-centric analysis, miRPath v4.0 accepts as input one or multiple terms of interest and utilizes overrepresentation statistics in sets of miRNA targets, to identify miRNAs with significantly more targets in terms of interest than would be expected by chance. In the case of more than one input term, each miRNA-term combination is assessed separately and significance levels are combined per-miRNA, similarly to Pathways union module in miRNA-centric analysis (A). (D) Resources of miRNA interactions, terms and expression contexts that are currently integrated. miRPath v4.0 also supports analyses with user-provided data.

**Context specific capacities.** Post-enrichment analysis compares miRNA-targeted versus non-targeted pathway components across a selected context set (Section 'Data collections' below) using either a Mann Whitney U test (default) or a two-sample Kolmogorov-Smirnov test, as implemented in base R stats package (functions *wilcox.test()* and *ks.test()*), to assess difference in the mean ranks of miRNA-targeted and non-targeted genes, or if they are drawn from the same distribution, respectively. By default, one-tailed tests are used, in order to assess the repressive effect of miRNAs, however two-sided testing can also be selected to accommodate, for example, scenarios where the abundance of miRNAs under question is limited.

**miRNA-centric analysis of phenotype selection screens.** Primary term analysis of the top positively or negatively selected miRNAs (user-defined) is performed taking the union of their targets and employing hypergeometric testing, as described in, 'Main analysis engine'. For significantly enriched pathways, two downstream options are provided: the default option performs competitive, rank-based CAMERA testing (proposed and described by Wu and Smyth (24)) to identify significant left/right trends of targets within

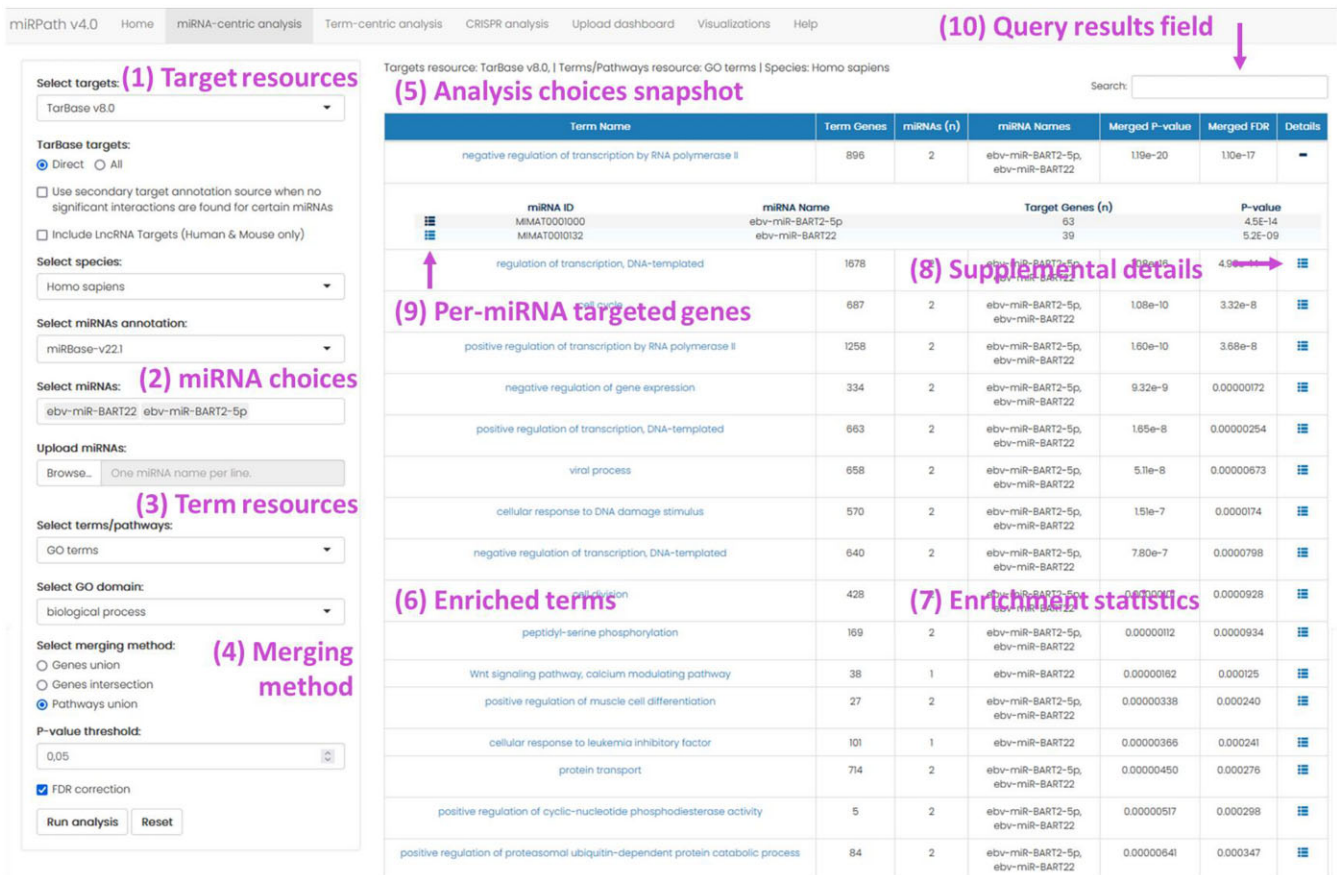
the opposite ranking (e.g. for positively selected miRNAs the negative ranking of genes is assessed); the second option also requires setting the number of top opposite-ranked genes and performs one-sided Fisher's exact test to evaluate the enrichment on the rank's extreme in miRNA targets.

### Data collections

Gene identifiers were retrieved via Biomart from Ensembl v102 (25). miRNA annotations were retrieved from miRBase v22.1 (26) and also matched to instances from MirGeneDB v2.1 (27).

Experimentally verified interactions were retrieved from reference resources DIANA-TarBase v8 (12), DIANA-LncBase v3 (28), and miRTarBase 2022 (13). Context-aware interaction sets were obtained by analysing publicly available CLIP-Seq datasets in 35 human and 6 mouse tissues with microCLIP (7). Predicted interactions were obtained directly from TargetScan 8 and by performing *de novo* DIANA-microT-CDS executions (3) on Ensembl v102 and miRBase v22.1 annotation. microT interactions include for the first time the capacity to query host targets for virally encoded miRNAs regarding 20 viruses infect-





**Figure 2.** Depiction of miRNA-centric mode of DIANA-miRPath v4.0. On the left panel, users may choose among (1) different miRNA interaction resources and (2) miRNA annotations. Enrichment can be performed in sets of terms (3) after selecting the most appropriate results methodology (4) and significance thresholds. Upon conducting an analysis, (5) a brief annotation of selected options is provided. The results table is organized per-term (6), offering metrics regarding the numbers of miRNAs, total genes and targeted genes in each entry, along with the corresponding significance values (7). For terms of interest, supplementary details can be unfolded using the respective buttons (8), highlighting the interactions that yielded each specific result (9), which can also be retrieved locally. The entire results table can also be retrieved locally by Download buttons in the bottom of the page, while terms/genes/miRNAs of interest can be provided into the filter-down button (10), in order to briefly query the results table for existence of entries harbouring these terms.

ing human (EBV, KSHV, HSV1-2, HCMV, SFV, HBV, HIV1, HHV6B, MCPV, JCV, BKV, SV40 and TTV), mouse (MGHV, MCMV) and chicken (MDV1-2, HVT, ILTV). For TarBase, LncBase, miRTarBase and TargetScan, gene IDs were converted to Ensembl v102 using Ensembl ID History Converter tool or Biomart (miRTarBase) and miRNA names were manually converted to miRBase v22.1. Supplemental confidence information, including annotation on direct/indirect proof of interaction by TarBase, high/low interaction confidence annotation by miRTarBase, and interaction scores by target prediction tools, was retained for collections.

Annotation of GO terms (29,30) against Gene IDs was retrieved through Ensembl. Pathview package (31) was used to retrieve KEGG (32)-to-Entrez Gene ID relationships, which were annotated to Ensembl Gene IDs (Biomart). Reactome (33) pathway annotations were retrieved and Ensembl Gene IDs were retained and converted to Ensembl v102. Expertly curated Molecular Signatures Database (MSigDB) gene sets (34,35) regarding Hallmark (H), Oncogenic, (C7), Immunologic (C8) and Cell type (C9) signatures

were retrieved using the msigdb package (36) and annotated to Ensembl Gene IDs (Biomart). Protein family annotations were retrieved from PFam (37).

Adult and fetal single-cell RNA-Seq data was collected from the Han *et al.* human cell atlas (38). Log<sub>2</sub>(Fold Change) (log<sub>2</sub>FC) gene expression values were obtained for each cell type by estimating the percent of individual cells exhibiting at least one read within each annotated cell type and contrasting it to the median of percentages of the remaining cell types. Transcripts-Per-Million (TPM-) level gene expression data were obtained for each cancer project in TCGA (39) via FireBrowse (<https://firebrowse.org>). Adjacent healthy and blood-derived normal samples were filtered out. Healthy tissue expression data (TPM-level) were retrieved from the GTEx Consortium (40). TPM-level data were used to denote a per-gene per-tissue log<sub>2</sub>FC, similarly to the single-cell data, depicting the percentage of samples presenting at least 10 TPM in a tissue relative to the median of the percentages of the remaining tissues. This metric was adopted as a balanced estimator to depict consistent and cell-/tissue-specific non-zero expression. For all three re-

sources (single-cell atlas, TCGA, GTEx), gene names were converted to Ensembl v102 Gene IDs using Ensembl ID History Converter tool and R.

## Implementation

The user interface of miRPath has been upgraded to provide further flexibility without sacrificing user-friendliness. Figure 2 briefly showcases the input menu and results from the tab enabling miRNA-centric functional analysis. All interaction, term and context resources included in DIANA-miRPath v4.0 were converted to Mongo collections and stored locally in a MongoDB schema (v4.4.2). DIANA-miRPath v4 was set up as an R Shiny application (<https://shiny.rstudio.com/>), and deployed using ShinyProxy (<https://www.shinyproxy.io>). The data access layer is implemented using *mongolite* package. The presentation layer relies primarily on packages *shiny*, *shinyalert*, *shinyjs*, *markdown* (41), *tippy*, while visualizations are created making use of packages *ggplot2* (42), *limma*, *Pathview* (31) and *pheatmap* (v1.0.12, Kolde, 2019). R Markdown was used to create a Help Section, while all functionalities requiring user input are accompanied with example data to further aid user experience.

## DISCUSSION

DIANA-miRPath v4.0 constitutes an upgrade both in terms of interface, which now holds a modular application-like character, and of content. Available visualization options were also expanded. Importantly, a User Upload Dashboard is introduced to further augment tailored use and flexibility. By it, the full arsenal of miRPath modules is unlocked for application with custom, user-provided annotations e.g. non-model organisms, modified/edited miRNAs, or even other non-coding RNAs. DIANA-miRPath v4.0 constitutes a valuable resource, delivering numerous computational methods to initiate or consolidate miRNA-related investigations.

## DATA AVAILABILITY

DIANA-miRPath is freely available, without registration or login, at <http://www.microrna.gr/miRPathv4>.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank former and current members of DIANA-Lab for useful discussions and being involved in testing of the online platform.

## FUNDING

ELIXIR-GR: The Greek Research Infrastructure for Data Management and Analysis in Life Sciences [MIS-5002780] which is implemented under the Action ‘Reinforcement of the Research and Innovation Infrastructure’, funded by the

Operational Programme ‘Competitiveness, Entrepreneurship and Innovation’ [NSRF 2014–2020]; Greece and the European Union (European Regional Development Fund). The study was also supported by the Hellenic Foundation for Research and Innovation (HFRI) under the 1<sup>st</sup> Call for HFRI Research Projects to Support Faculty Members & Researchers and Procure High-Value Research Equipment grant (Project Number 2563). Funding for open access charge: ELIXIR-GR Infrastructure [MIS-5002780].  
*Conflict of interest statement.* None declared.

## REFERENCES

- Bartel,D.P. (2018) Metazoan micromas. *Cell*, **173**, 20–51.
- Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife*, **4**, e05005.
- Tastsoglou,S., Alexiou,A., Karagkouni,D., Skoufos,G., Zacharopoulou,E. and Hatzigeorgiou,A.G. (2023) DIANA-microT 2023: including predicted targets of virally encoded miRNAs. *Nucleic Acids Res.*, gkad283.
- Vlachos,I.S., Paraskevopoulou,M.D., Karagkouni,D., Georgakilas,G., Vergoulis,T., Kanellos,I., Anastasopoulos,I.-L., Maniou,S., Karathanou,K. and Kalfakakou,D. (2015) DIANA-TarBase v7. 0: indexing more than half a million experimentally supported miRNA: mRNA interactions. *Nucleic Acids Res.*, **43**, D153–D159.
- Hafner,M., Landthaler,M., Burger,L., Khorshid,M., Hausser,J., Berninger,P., Rothballer,A., Ascano Jr,M., Jungkamp,A.-C. and Munschauer,M. (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
- Chi,S.W., Zang,J.B., Mele,A. and Darnell,R.B. (2009) Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps. *Nature*, **460**, 479–486.
- Paraskevopoulou,M.D., Karagkouni,D., Vlachos,I.S., Tastsoglou,S. and Hatzigeorgiou,A.G. (2018) microCLIP super learning framework uncovers functional transcriptome-wide miRNA interactions. *Nat. Commun.*, **9**, 1–16.
- Nam,S., Kim,B., Shin,S. and Lee,S. (2007) miRator: an integrated system for functional annotation of microRNAs. *Nucleic Acids Res.*, **36**, D159–D164.
- Wang,X. (2008) miRDB: a microRNA target prediction and functional annotation database with a wiki interface. *RNA*, **14**, 1012–1017.
- Kern,F., Fehlmann,T., Solomon,J., Schwed,L., Grammes,N., Backes,C., Van Keuren-Jensen,K., Craig,D.W., Meese,E. and Keller,A. (2020) miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems. *Nucleic Acids Res.*, **48**, W521–W528.
- 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.
- Karagkouni,D., Paraskevopoulou,M.D., Chatzopoulos,S., Vlachos,I.S., Tastsoglou,S., Kanellos,I., Papadimitriou,D., Kavakiotis,I., Maniou,S. and Skoufos,G. (2018) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res.*, **46**, D239–D245.
- Huang,H.-Y., Lin,Y.-C.-D., Cui,S., Huang,Y., Tang,Y., Xu,J., Bao,J., Li,Y., Wen,J. and Zuo,H. (2022) miRTarBase update 2022: an informative resource for experimentally validated miRNA–target interactions. *Nucleic Acids Res.*, **50**, D222–D230.
- Bleazard,T., Lamb,J.A. and Griffiths-Jones,S. (2015) Bias in microRNA functional enrichment analysis. *Bioinformatics*, **31**, 1592–1598.
- Young,M.D., Wakefield,M.J., Smyth,G.K. and Oshlack,A. (2010) Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.*, **11**, 1–12.
- Li,W., Köster,J., Xu,H., Chen,C.-H., Xiao,T., Liu,J.S., Brown,M. and Liu,X.S. (2015) Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol.*, **16**, 1–13.

17. Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M. and Liu, X.S. (2014) MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.*, **15**, 1–12.
18. Licursi, V., Conte, F., Fisco, G. and Paci, P. (2019) MIENTURNET: an interactive web tool for microRNA–target enrichment and network-based analysis. *BMC Bioinf.*, **20**, 1–10.
19. Kehl, T., Kern, F., Backes, C., Fehlmann, T., Stöckel, D., Meese, E., Lenhof, H.-P. and Keller, A. (2020) miRPathDB 2.0: a novel release of the miRNA Pathway Dictionary Database. *Nucleic Acids Res.*, **48**, D142–D147.
20. Bhattacharya, A. and Cui, Y. (2015) miR2GO: comparative functional analysis for microRNAs. *Bioinformatics*, **31**, 2403–2405.
21. Wang, Y., Wu, J., Chen, H., Yang, Y., Xiao, C., Yi, X., Shi, C., Zhong, K., He, H. and Li, Y. (2021) Genome-wide CRISPR–Cas9 screen identified KLF11 as a druggable suppressor for sarcoma cancer stem cells. *Sci. Adv.*, **7**, eabe3445.
22. 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–e47.
23. Fisher, R.A. (1992) *Statistical Methods for Research Workers*. Springer.
24. Wu, D. and Smyth, G.K. (2012) Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.*, **40**, e133–e133.
25. Howe, K.L., Achuthan, P., Allen, J., Allen, J., Alvarez-Jarreta, J., Amode, M.R., Armean, I.M., Azov, A.G., Bennett, R. and Bhai, J. (2021) Ensembl 2021. *Nucleic Acids Res.*, **49**, D884–D891.
26. Kozomara, A., Birgaoanu, M. and Griffiths-Jones, S. (2019) miRBase: from microRNA sequences to function. *Nucleic Acids Res.*, **47**, D155–D162.
27. Fromm, B., Høye, E., Domanska, D., Zhong, X., Aparicio-Puerta, E., Ovchinnikov, V., Umu, S.U., Chabot, P.J., Kang, W. and Aslanzadeh, M. (2022) MirGeneDB 2.1: toward a complete sampling of all major animal phyla. *Nucleic Acids Res.*, **50**, D204–D210.
28. Karagkouni, D., Paraskevopoulou, M.D., Tastsoglou, S., Skoufos, G., Karavangeli, A., Pierros, V., Zacharopoulou, E. and Hatzigeorgiou, A.G. (2020) DIANA-LncBase v3: indexing experimentally supported miRNA targets on non-coding transcripts. *Nucleic Acids Res.*, **48**, D101–D110.
29. Gene Ontology Consortium (2021) The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.*, **49**, D325–D334.
30. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S. and Eppig, J.T. (2000) Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
31. Luo, W. and Brouwer, C. (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*, **29**, 1830–1831.
32. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
33. Jassal, B., Matthews, L., Viteri, G., Gong, C., Lorente, P., Fabregat, A., Sidiropoulos, K., Cook, J., Gillespie, M. and Haw, R. (2020) The reactome pathway knowledgebase. *Nucleic Acids Res.*, **48**, D498–D503.
34. Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P. and Tamayo, P. (2015) The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.*, **1**, 417–425.
35. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P. and Mesirov, J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.
36. Dolgalev, I. (2020) msigdb: mSigDB gene sets for multiple organisms in a tidy data format. R Package Version 7.
37. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L., Tosatto, S.C., Paladin, L., Raj, S. and Richardson, L.J. (2021) Pfam: the protein families database in 2021. *Nucleic Acids Res.*, **49**, D412–D419.
38. Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W. et al. (2020) Construction of a human cell landscape at single-cell level. *Nature*, **581**, 303–309.
39. Tomczak, K., Czerwińska, P. and Wiznerowicz, M. (2015) The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology*, **19**, A68.
40. Consortium, G., Ardlie, K.G., Deluca, D.S., Segrè, A.V., Sullivan, T.J., Young, T.R., Gelfand, E.T., Trowbridge, C.A., Maller, J.B. and Tukiainen, T. (2015) The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.
41. Xie, Y., Allaire, J.J. and Grolmund, G. (2018) *R Markdown: The Definitive Guide*. Chapman and Hall/CRC.
42. Wickham, H., Chang, W. and Wickham, M.H. (2016) Package ‘ggplot2’. *Create Elegant Data Visualisations Using the Grammar of Graphics*. Vol. 2, pp. 1–189.