CancerMIRNome: a web server for interactive analysis and visualization of cancer miRNome data

Ruidong Li[1,2,*], Han Qu[1], Shibo Wang[1], Xuesong Wang[1,2], Yanru Cui[3], Lei Yu[1,2], John M. Chater[1], Rui Zhou[1], Qiong Jia[1,2], Ryan Traband[1], Dongbo Yuan[4], Jianguo Zhu[4,*], Wei-De Zhong[5,*], Zhenyu Jia[1,2,*]

[1] Department of Botany and Plant Sciences, University of California, Riverside, CA, USA
[2] Graduate Program in Genetics, Genomics, and Bioinformatics, University of California, Riverside, CA, USA
[3] College of Agronomy, Hebei Agricultural University, Baoding, China
[4] Department of Urology, Guizhou Provincial People's Hospital, Guizhou, China
[5] Department of Urology, Guangdong Key Laboratory of Clinical Molecular Medicine and Diagnostics, Guangzhou First People's Hospital, School of Medicine, South China University of Technology, Guangzhou, China

*To whom correspondence should be addressed: Ruidong Li at rli012@ucr.edu. Correspondence may also be addressed to Zhenyu Jia at arthur.jia@ucr.edu, Wei-De Zhong at zhongwd2009@live.cn, or Jianguo Zhu at doctorzhujianguo@163.com.

ABSTRACT

MicroRNAs (miRNAs), which play critical roles in gene regulatory networks, have emerged as promising biomarkers for a variety of human diseases, including cancer. In particular, circulating miRNAs that are secreted into circulation exist in remarkably stable forms, and have enormous potential to be leveraged as non-invasive diagnostic biomarkers for early cancer detection. The vast amount of miRNA expression data from tens of thousands of samples in various types of cancers generated by The Cancer Genome Atlas (TCGA) and circulating miRNA data produced by many large-scale circulating miRNA profiling studies provide extraordinary opportunities for the discovery and validation of miRNA signatures in cancer. Novel and user-friendly tools are desperately needed to facilitate the data mining of such valuable cancer miRNome datasets. To fill this void, we developed CancerMIRNome, a web server for interactive analysis and visualization of cancer miRNome data based on TCGA and public circulating miRNome datasets. A series of cutting-edge bioinformatics tools and functions have been packaged in CancerMIRNome, allowing for a pan-cancer analysis of a miRNA of interest across multiple cancer types and a comprehensive analysis of cancer miRNome at the dataset level. The CancerMIRNome web server is freely available at http://bioinfo.jialab-ucr.org/CancerMIRNome.

INTRODUCTION

MicroRNAs (miRNAs) are a class of small endogenous non-coding RNAs of ~22nt in length that negatively regulate the expression of their target protein-coding genes (1). It has been reported that miRNAs are involved in many biological processes, such as cell proliferation, differentiation, and apoptosis (2–5). Mounting evidences have demonstrated that the expression of miRNAs is dysregulated in various types of human cancers (6–10), which makes them potential biomarkers for cancer diagnosis and prognosis. Depending on the specific cellular context, miRNAs may act as oncogenes or tumor suppressors in various cancers (11, 12). Circulating miRNAs represent the miRNAs that are secreted into extracellular body fluids, where they are incorporated in extracellular vesicles (EVs), such as shed microvesicles (sMVs) and exosomes, or in apoptotic bodies, or form complexes with RNA binding proteins, such as Argonates (AGOs). The protected circulating miRNAs remain in remarkably stable forms, rendering potential biomarkers for non-invasive early detection of cancer (13–18).

The vast amount of miRNA expression data from tens of thousands of samples in various types of cancers generated by TCGA provides extraordinary opportunities for the discovery and validation of potential cancer diagnostic and prognostic miRNA biomarkers and signatures. Many large-scale circulating miRNA profiling studies in cancer have also produced tremendous amount of miRNA expression data, which are valuable resources for non-invasive early cancer detection and tissue-of-origin localization studies. An online tool OncomiR has been developed to explore dysregulated miRNAs associated with tumor development and progression based on TCGA data resource (19). While the analytical functions provided by OncomiR are useful, many functions are lacking for the comprehensive analysis of cancer miRNome data. In addition, OncomiR is only designed for TCGA data analysis and it doesn't support data visualization and exportation, which constrains its wide application. Sophisticated and user-friendly web tools have been developed for an interactive analysis and visualisation of gene expression profiles in TCGA and GTEx (20, 21), platforms of the same kind that are specific to the comprehensive analysis of cancer miRNome data are needed to facilitate exploiting miRNome data from human tissues and body fluids of cancer patients. To fill this void, we developed CancerMIRNome, a web server for an interactive analysis and visualization of cancer miRNome data based on 10,998 tumor and normal samples from 33 TCGA projects and 21,993 samples of 32 cancer types from 40 circulating miRNA profiling studies.

CancerMIRNome provides a suite of advanced functions to facilitate the interactive analysis and visualization of cancer miRNome data. When querying a miRNA of interest, pan-cancer analysis based on all available TCGA projects including differential expression (DE) analysis, receiver operating characteristic (ROC) analysis, survival analysis, miRNA-target correlation analysis, and functional enrichment analysis, as well as circulating miRNA expression analysis in public circulating miRNome datasets will be automatically performed. Moreover, CancerMIRNome allows users to perform a comprehensive analysis including the identification of highly expressed miRNAs, DE analysis between two user-defined groups, ROC analysis, feature selection using a machine learning algorithm, principal component analysis (PCA), and survival analysis in a TCGA project or in a

circulating miRNome dataset. Advanced visualizations are supported in CancerMIRNome to produce publication-quality vector images in PDF format which can be easily downloaded. All the data and results generated by CancerMIRNome are exportable, allowing for further analysis by the end users (Figure 1).

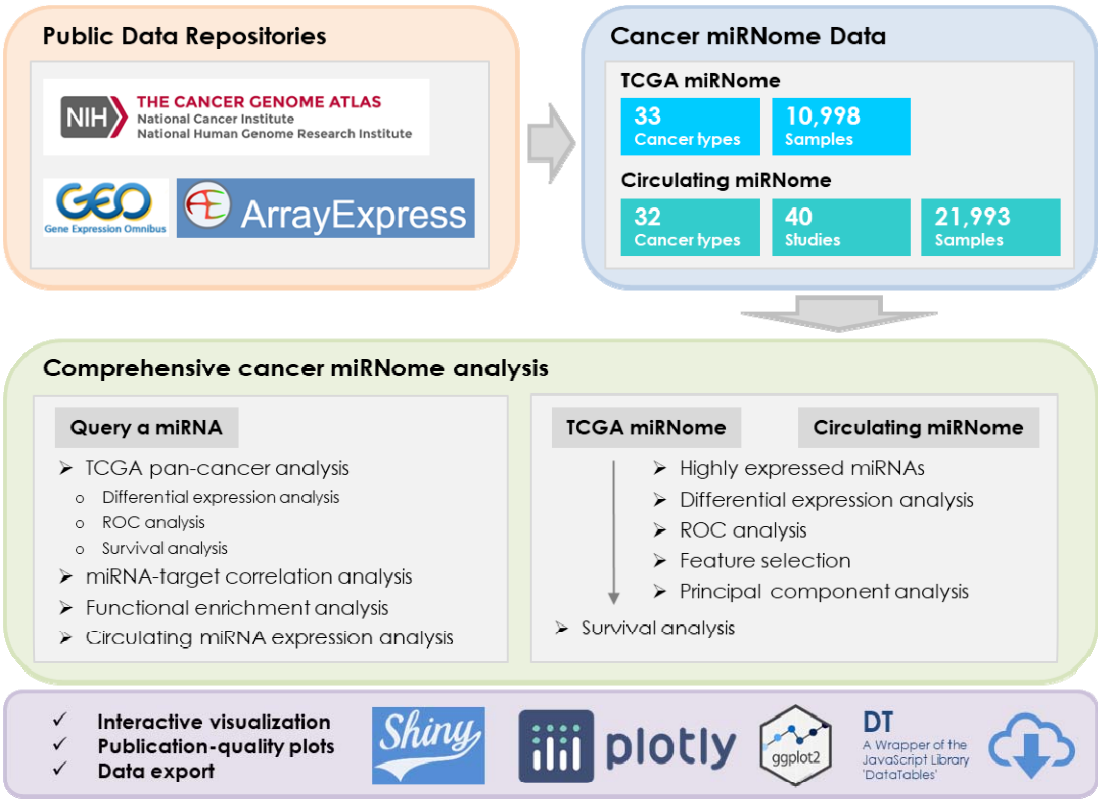

Figure 1. Overview of the CancerMIRNome web server.

MATERIAL AND METHODS

Data collection and processing

TCGA miRNome datasets

The R/Bioconductor package GDCRNATools (22) was used to download and process the miRNA expression data and clinical data of 33 cancer types in TCGA program. Isoform expression quantification data of miRNA-Seq were downloaded from National Cancer Institute (NCI) Genomic Data Commons (GDC) using the *gdcRNADownload* function. The expression data from the same project were merged to a single expression matrix using the *gdcRNAMerge* function in the R package GDCRNATools, followed by a normalization with the Trimmed Mean of M-values (TMM) normalization method implemented in the R package edgeR (23). Clinical information including age, tumor stages, and overall survival, etc. were retrieved from the XML file of each sample with the *gdcClinicalMerge* function in GDCRNATools.

Circulating miRNome datasets

An extensive search for circulating miRNA expression data in cancer was performed in public databases, including NCBI Gene Expression Omnibus (GEO) and ArrayExpress. A total of 40 public circulating miRNA expression datasets with over 1000 miRNAs and more than 10 samples in each dataset were identified by searching the keywords ‘circulating’, ‘whole blood’, ‘serum’, ‘plasma’, ‘extracellular vesicle’, and ‘exosome’, in combination with ‘miRNA’ or ‘microRNA’, and with ‘cancer’, ‘tumor’, and ‘carcinoma’. Both expression data and metadata were downloaded by the *getGEO* function in the R package GEOquery (24). Metadata of the samples were processed using custom scripts with manual inspections.

Functionalities

Query a miRNA of interest

Users can query a miRNA of interest by typing the miRNA accession number, miRNA ID of miRBase release 22.1 (25), or previous miRNA IDs in the ‘Search a miRNA’ field and selecting this miRNA from the dropdown list. In addition to the general information including IDs and sequence of the queried miRNA, links to five miRNA-target databases including ENCORI (26), miRDB (27), miTarBase (28), TargetScan (29), and Diana-TarBase (30) are also provided.

A suite of advanced analyses can be interactively performed for a selected miRNA of interest, including (1) pan-cancer differential expression (DE) analysis, receiver operating characteristic (ROC) analysis, and Kaplan Meier (KM) survival analysis in TCGA; (2) DE analysis, ROC analysis, and KM survival analysis in a selected TCGA project; (3) miRNA-target correlation analysis; (4) functional enrichment analysis of miRNA targets; and (5) circulating miRNA expression analysis (Figure 2).

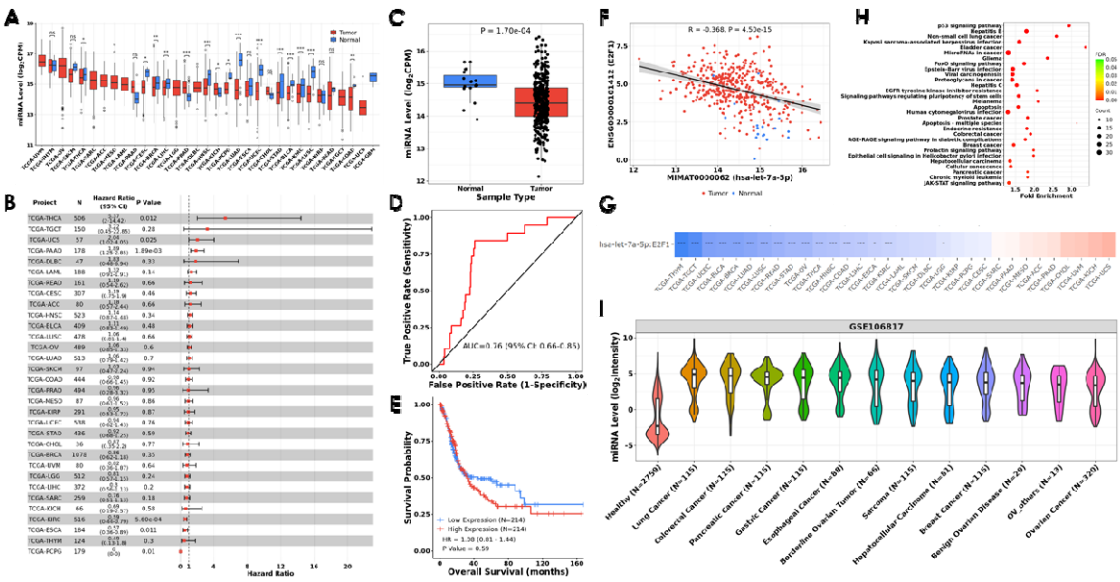

Figure 2. Examples of CancerMIRNome outputs from the query of a miRNA of interest. (A) Pan-cancer differential expression analysis across all TCGA projects. (B) A forest plot visualizing pan-cancer survival analysis across all TCGA projects. (C) Boxplot of the miRNA expression in tumor and normal samples from the selected TCGA project. (D) An ROC curve illustrating the diagnostic ability of the miRNA in the selected TCGA project. (E) Kaplan Meier analysis of overall survival between tumor samples with high and low expression of the miRNA of interest defined by its median expression value in the selected TCGA project. (F) Correlation analysis of the miRNA with one of its targets in a TCGA project. (G) An interactive heatmap visualizing miRNA-target correlations across all TCGA projects. (H) A bubble plot visualizing the functional enrichment of target genes for the miRNA. (I) A violin plot visualizing the circulating miRNA expression in a selected cancer circulating miRNome dataset.

1. Pan-cancer analysis

Pan-cancer DE analysis and ROC analysis of a miRNA between tumor and normal samples can be performed in 33 cancer types from TCGA. Wilcoxon rank sum test is used for DE analysis. The expression levels and statistical significances of the miRNA in all the TCGA projects can be visualized in a box plot. ROC analysis is performed to measure the diagnostic ability of the miRNA in classifying tumor and normal samples. A forest plot with the number of tumor and normal samples, area under the curve (AUC), and 95% confidence interval (CI) of the AUC for each TCGA project is used to visualize the result. Prognostic ability of a miRNA can be evaluated by performing KM survival analysis of overall survival (OS) between tumor samples with high and low expression of the miRNA of interest defined by its median expression value. A forest plot displaying the number of tumor samples, hazard ratio (HR), 95% CI of the HR, and p value for each cancer type in TCGA is used to visualize the result of pan-cancer survival analysis.

2. miRNA analysis in a TCGA project

CancerMIRNome provides functions to focus the DE analysis, ROC analysis, and KM survival analysis for the miRNA of interest in a selected TCGA project. For example, if a user is only interested in the differential expression, diagnostic ability, and prognostic ability of the miRNA in prostate cancer, the user can simply select ‘TCGA-PRAD’ in the ‘TCGA Project’ field. A box plot with miRNA expression and p value of wilcoxon rank-sum test between tumor and normal samples, an ROC curve, and a KM survival curve for the selected project will be displayed.

3. miRNA-target correlation analysis

Pearson correlation between a miRNA and its targets in tumor and normal tissues of TCGA projects can be queried in CancerMIRnome. The miRNA-target interactions are based on miRTarBase 2020 (28), an experimentally validated miRNA-target interactions database. The expression correlations between a miRNA and all of its targets in a selected TCGA project are listed in an interactive data table. Users can select an interested interaction between miRNA and mRNA target in the data table to visualize a scatter plot showing their expression pattern and correlation metrics. An interactive

heatmap is also available to visualize and compare such miRNA-target correlations across all TCGA projects.

4. Functional enrichment analysis of miRNA targets

Functional enrichment analysis of the target genes for a miRNA can be performed using clusterProfiler (31) in CancerMIRNome. CancerMIRNome supports functional enrichment analysis with many pathway/ontology knowledgebases including Kyoto Encyclopedia of Genes and Genomes (KEGG) (32), Gene Ontology (GO) (33), Reactome (34), Disease Ontology (DO) (35), Network of Cancer Gene (NCG) (36), DisGeNET (37), and Molecular Signatures Database (MSigDB) (38, 39). A data table is produced to summarize the significantly enriched pathways/ontologies in descending order based on their significance levels, as well as the number and proportion of enriched genes and the gene symbols in each pathway/ontology term. The top enriched pathways/ontologies are visualized using both bar plot and bubble plot.

5. Circulating miRNA expression

Expression of the interested miRNA in whole blood, serum, plasma, extracellular vesicles, or exosomes in both healthy and different cancer types can be conveniently explored in CancerMIRnome on the basis of 40 circulating miRNome datasets. Users can select one or more datasets for an analysis, through which violin plots are displayed for visualization and comparison of circulating miRNA expression between samples or datasets.

Comprehensive TCGA miRNome analysis

CancerMIRNome is equipped with well-designed functions which can perform comprehensive dataset-level analysis of cancer miRNome for each of the 33 TCGA projects. When a TCGA dataset (or project) is selected, the summary of important clinical features of patients in this dataset, including sample type, tumor stages, ages, overall survival, and etc., will be displayed. The dataset-level analysis of cancer miRNome includes: (1) identification of highly expressed miRNAs; (2) DE analysis between two user-defined subgroups; (3) ROC analysis between tumor and normal samples; (4) selection of diagnostic miRNA markers; (5) principal component analysis; and (6) identification of prognostic miRNA biomarkers and construction of prognostic models (Figure 3).

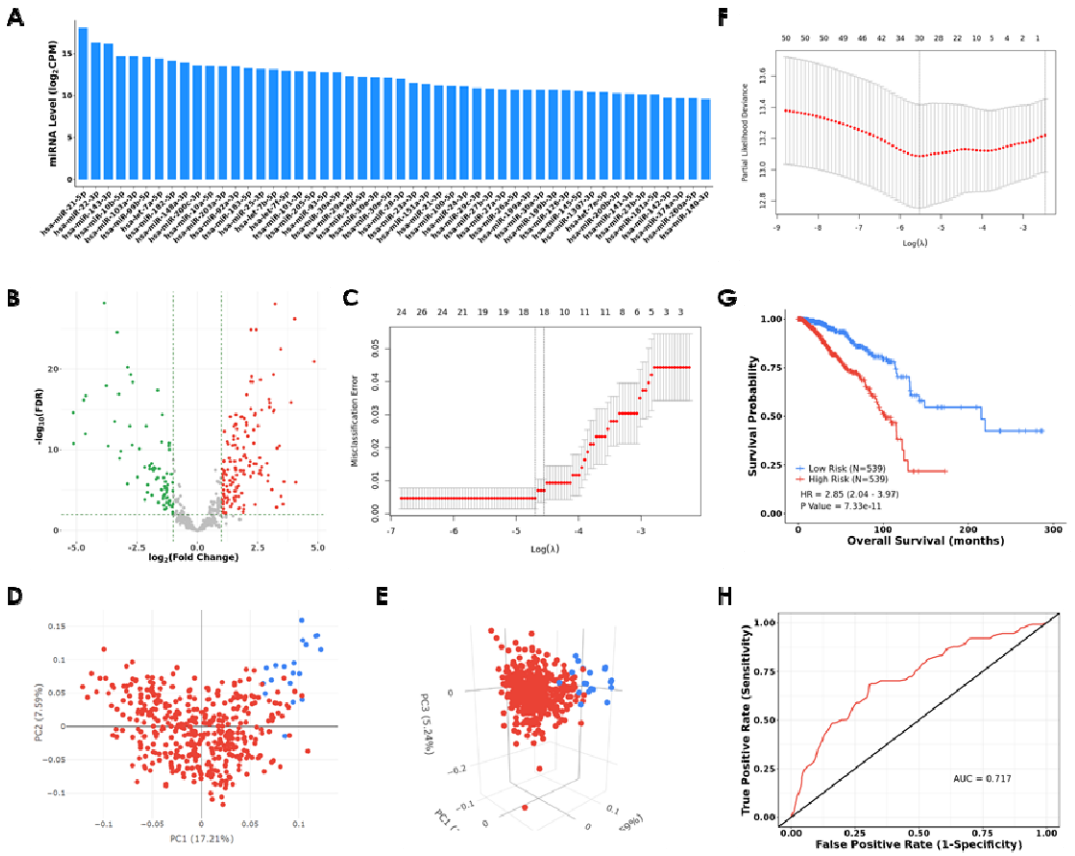

Figure 3. Examples of CancerMIRNome outputs from the comprehensive miRNome data analysis of an interested TCGA project. (A) Barplot of highly expressed miRNAs. (B) A volcano plot visualizing the differentially expressed miRNAs between two user-defined groups. (C) Selection of diagnostic miRNA biomarkers using LASSO. (D) 2D interactive visualization of principal component analysis result using the first two principal components. (E) 3D interactive visualization of principal component analysis result using the first three principal components. (F) Selection of prognostic miRNA biomarkers using a regularized Cox regression model with LASSO penalty to develop a prognostic model. (G) Kaplan Meier survival analysis evaluating the prognostic ability of the miRNA-based prognostic model. (H) Time-dependent ROC analysis evaluating the prognostic ability of the miRNA-based prognostic model.

1. Highly expressed miRNAs

miRNAs with counts per million (CPM) greater than 1 in more than 50% of the samples in a TCGA project of interest are reported as highly expressed miRNAs. The miRNAs are ranked by the median expression values and the top 50 of the highly expressed miRNAs are visualized with a bar plot.

2. DE analysis

The DE analysis of highly expressed miRNAs at the dataset-level allows users to identify miRNAs that are differentially expressed between two user-defined subgroups in a TCGA project. Metadata,

including sample type, tumor stages, gender, and etc., may be used to group the samples. For examples, the DE analysis can be performed not only between tumor and normal samples, but also between patients at early and late tumor stages. Both limma (40) and wilcoxon rank-sum test are used for the identification of differentially expressed miRNAs.

3. ROC analysis

The ROC analysis is carried out to screen the highly expressed miRNAs in a selected TCGA dataset for the diagnostic biomarkers that can distinguish tumor samples from normal samples. All the miRNAs are ranked in a data table based on their AUC values.

4. Feature selection

The least absolute shrinkage and selection operator (LASSO) (41, 42), a machine-learning method, can be used to analyse the entire set of miRNAs in a selected TCGA project for the identification of prognostic miRNAs, and use the miRNA signature to develop a classification model for differentiating tumor and normal samples.

5. Principal component analysis

Principal component analysis can be utilized to analyse the highly expressed miRNAs in a selected TCGA project such that all patient samples, including tumor and/or normal samples, may be visualized in a 2D and 3D interactive plot using the first two and three principal components, respectively.

6. Survival analysis

CancerMIRNome supports both Cox Proportional-Hazards (CoxPH) regression analysis and Kaplan-Meier (KM) survival analysis at a dataset level to identify prognostic miRNA biomarkers in a TCGA project. The miRNAs with p values less than 0.05 in the univariate CoxPH analysis will be jointly analysed using a regularized Cox regression model with LASSO penalty to develop a prognostic model (42). The prognostic model, which is a linear combination of the finally selected miRNA variables with the LASSO-derived regression coefficients, will be used to calculate a risk score for each patient. All the patients will be divided into either high-risk group or low-risk group based on the median risk value in the cohort. The KM survival analysis and time-dependent ROC analysis can be performed to evaluate the prognostic ability of the miRNA-based prognostic model.

Comprehensive cancer circulating miRNome analysis

A set of similar functions are available for the comprehensive analysis of circulating miRNome at a dataset level to identify diagnostic miRNA biomarkers for non-invasive early cancer detection. The summary of a selected dataset includes the distribution of cancer types, the distribution of subgroups of the patients, and an embedded webpage (from either GEO or ArrayExpress) housing the public dataset. Users can perform various analyses for circulating miRNome in CancerMIRNome, including (1) identification of highly expressed miRNAs; (2) DE analysis; (3) ROC analysis; (4) feature selection; and (5) principal component analysis.

1. Highly expressed miRNAs

Since almost all the circulating miRNome datasets were based on the microarray assays, all the miRNAs in a dataset are ranked by the median expression values and the top 500 miRNAs are considered as highly expressed miRNAs in this dataset. The top 50 highly expressed miRNAs in a selected dataset are visualized in a bar plot.

2. DE analysis

The limma and the wilcoxon rank sum test can be used to identify DE miRNA biomarkers between two user-defined subgroups in the dataset. Similar to DE analysis in TCGA projects, only the highly expressed circulating miRNAs are included in the DE analysis.

3. ROC analysis

The ROC analysis of the highly expressed circulating miRNAs between two user-defined subgroups of samples in a selected dataset can be performed to identify diagnostic biomarkers for non-invasive early cancer detection or cancer type classification. The circulating miRNA biomarkers are ranked in a data table by their AUC values.

4. Feature selection

Similar to the feature selection function for miRNome analysis in TCGA projects, LASSO can be also used in a selected dataset to identify the circulating miRNA biomarkers for non-invasive early cancer detection or cancer type classification.

5. Principal component analysis

Principal component analysis can be utilized to analyse the highly expressed circulating miRNAs in a selected dataset such that all the subjects, including healthy individuals and patients with various types of cancers, may be visualized in a 2D and 3D interactive plots using the first two and three principal components, respectively.

DISCUSSION

In this project, we present to the cancer research community a user-friendly web tool, cancerMIRNome, for an interactive analysis and visualization of cancer miRNome by leveraging 10,998 tumor and normal samples from 33 TCGA projects and 21,993 samples of 32 cancer types from 40 public circulating miRNA profiling studies. A suite of well-designed functions is provided in cancerMIRNome to facilitate the interactive analysis and visualization at both miRNA level and miRNome level (or dataset level). For example, a comprehensive characterization of an interested miRNA, including pan-cancer differential expression analysis, ROC analysis, survival analysis, miRNA-target correlation analysis, functional enrichment analysis, and circulating miRNA expression analysis, may be simply carried out by querying this miRNA on the webpage of cancerMIRNome. This is tremendously helpful when users are interested in the characterization of a miRNA across all cancer types in TCGA; otherwise, the users will have to download, process and analyze all the data

including expression data and clinical data from the TCGA projects, to reach the same results, which requires advanced bioinformatics programming skills and takes time and effort. Users can also choose to perform comprehensive analyses, including identification of highly expressed miRNAs, DE analysis between two user-defined subgroups, ROC analysis, feature selection using a machine learning algorithm, principal component analysis, and survival analysis, in a TCGA project or in a public circulating miRNome dataset to identify potential diagnostic and prognostic biomarkers in cancer. Advanced visualizations are supported in CanerMIRNome and the publication-quality vector images can be easily created and downloaded. Moreover, all the data and results are exportable, allowing for further local analyses by the end users. While CancerMIRNome is diligently serving the cancer research community, we are open to any feedback from users and will constantly maintain and improve this web tool. New datasets, analytical methods, and visualization functions will be included in CancerMIRNome as soon as they are available. We expect that CancerMIRNome would become a valuable online resource for a comprehensive analysis of cancer miRNome data not only for experimental biologists, but also for bioinformatics scientists in the field.

AVAILABILITY

The CancerMIRNome web server is available at http://bioinfo.jialab-ucr.org/CancerMIRNome. R shiny source code for the web server is also publicly available at https://github.com/rli012/CancerMIRNome.

FUNDING

This work was supported by UC Riverside Faculty Start-up Fund and UC Cancer Research Coordinating Committee Competition Award to Z.J.

CONFLICT OF INTEREST

The authors declare that they have no competing interests.

REFERENCES

1. Bartel,D.P. (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *cell*, **116**, 281–297.

2. Lu,Y., Thomson,J.M., Wong,H.Y.F., Hammond,S.M. and Hogan,B.L. (2007) Transgenic over-expression of the microRNA miR-17-92 cluster promotes proliferation and inhibits differentiation of lung epithelial progenitor cells. *Developmental biology*, **310**, 442–453.

3. Wang,Y., Baskerville,S., Shenoy,A., Babiarz,J.E., Baehner,L. and Blelloch,R. (2008) Embryonic stem cell–specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nature genetics*, **40**, 1478–1483.

4. Chang,T.-C., Wentzel,E.A., Kent,O.A., Ramachandran,K., Mullendore,M., Lee,K.H., Feldmann,G., Yamakuchi,M., Ferlito,M. and Lowenstein,C.J. (2007) Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Molecular cell*, **26**, 745–752.

5. Vidigal,J.A. and Ventura,A. (2015) The biological functions of miRNAs: lessons from in vivo studies. *Trends in cell biology*, **25**, 137–147.

6. Lu,J., Getz,G., Miska,E.A., Alvarez-Saavedra,E., Lamb,J., Peck,D., Sweet-Cordero,A., Ebert,B.L., Mak,R.H. and Ferrando,A.A. (2005) MicroRNA expression profiles classify human cancers. *nature*, **435**, 834–838.

7. Thomson,J.M., Newman,M., Parker,J.S., Morin-Kensicki,E.M., Wright,T. and Hammond,S.M. (2006) Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes & development*, **20**, 2202–2207.

8. He,L., He,X., Lim,L.P., De Stanchina,E., Xuan,Z., Liang,Y., Xue,W., Zender,L., Magnus,J. and Ridzon,D. (2007) A microRNA component of the p53 tumour suppressor network. *Nature*, **447**, 1130–1134.

9. O'Donnell,K.A., Wentzel,E.A., Zeller,K.I., Dang,C.V. and Mendell,J.T. (2005) c-Myc-regulated microRNAs modulate E2F1 expression. *Nature*, **435**, 839–843.

10. Peng,Y. and Croce,C.M. (2016) The role of MicroRNAs in human cancer. *Signal Transduction and Targeted Therapy*, **1**, 1–9.

11. Shenouda,S.K. and Alahari,S.K. (2009) MicroRNA function in cancer: oncogene or a tumor suppressor? *Cancer and metastasis reviews*, **28**, 369.

12. Svoronos,A.A., Engelman,D.M. and Slack,F.J. (2016) OncomiR or tumor suppressor? The duplicity of microRNAs in cancer. *Cancer research*, **76**, 3666–3670.

13. Schwarzenbach,H., Nishida,N., Calin,G.A. and Pantel,K. (2014) Clinical relevance of circulating cell-free microRNAs in cancer. *Nature reviews Clinical oncology*, **11**, 145.

14. Sohel,M.H. (2016) Extracellular/circulating microRNAs: release mechanisms, functions and challenges. *Achievements in the Life Sciences*, **10**, 175–186.

15. Mitchell,P.S., Parkin,R.K., Kroh,E.M., Fritz,B.R., Wyman,S.K., Pogosova-Agadjanyan,E.L., Peterson,A., Noteboom,J., O'Briant,K.C. and Allen,A. (2008) Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences*, **105**, 10513–10518.

16. Hayes,J., Peruzzi,P.P. and Lawler,S. (2014) MicroRNAs in cancer: biomarkers, functions and therapy. *Trends in molecular medicine*, **20**, 460–469.

17. Xu,R., Rai,A., Chen,M., Suwakulsiri,W., Greening,D.W. and Simpson,R.J. (2018) Extracellular vesicles in cancer—implications for future improvements in cancer care. *Nature reviews Clinical oncology*, **15**, 617–638.

18. Grasedieck,S., Sorrentino,A., Langer,C., Buske,C., Döhner,H., Mertens,D. and Kuchenbauer,F. (2013) Circulating microRNAs in hematological diseases: principles, challenges, and perspectives. *Blood, The Journal of the American Society of Hematology*, **121**, 4977–4984.

19. Wong,N.W., Chen,Y., Chen,S. and Wang,X. (2018) OncomiR: an online resource for exploring pan-cancer microRNA dysregulation. *Bioinformatics*, **34**, 713–715.

20. Tang,Z., Li,C., Kang,B., Gao,G., Li,C. and Zhang,Z. (2017) GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic acids research*, **45**, W98–W102.

21. Tang,Z., Kang,B., Li,C., Chen,T. and Zhang,Z. (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic acids research*, **47**, W556–W560.

22. Li,R., Qu,H., Wang,S., Wei,J., Zhang,L., Ma,R., Lu,J., Zhu,J., Zhong,W.-D. and Jia,Z. (2018) GDCRNATools: an R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC. *Bioinformatics*, **34**, 2515–2517.

23. 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.

24. Davis,S. and Meltzer,P.S. (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, **23**, 1846–1847.

25. Kozomara,A., Birgaoanu,M. and Griffiths-Jones,S. (2019) miRBase: from microRNA sequences to function. *Nucleic acids research*, **47**, D155–D162.

26. Li,J.-H., Liu,S., Zhou,H., Qu,L.-H. and Yang,J.-H. (2014) starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research*, **42**, D92–D97.

27. Chen,Y. and Wang,X. (2020) miRDB: an online database for prediction of functional microRNA targets. *Nucleic acids research*, **48**, D127–D131.

28. Huang,H.-Y., Lin,Y.-C.-D., Li,J., Huang,K.-Y., Shrestha,S., Hong,H.-C., Tang,Y., Chen,Y.-G., Jin,C.-N. and Yu,Y. (2020) miRTarBase 2020: updates to the experimentally validated microRNA–target interaction database. *Nucleic acids research*, **48**, D148–D154.

29. Agarwal,V., Bell,G.W., Nam,J.-W. and Bartel,D.P. (2015) Predicting effective microRNA target sites in mammalian mRNAs. *elife*, **4**, e05005.

30. 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 research*, **46**, D239–D245.

31. Yu,G., Wang,L.-G., Han,Y. and He,Q.-Y. (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology*, **16**, 284–287.

32. Kanehisa,M. and Goto,S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, **28**, 27–30.

33. Consortium,G.O. (2019) The gene ontology resource: 20 years and still GOing strong. *Nucleic acids research*, **47**, D330–D338.

34. Fabregat,A., Jupe,S., Matthews,L., Sidiropoulos,K., Gillespie,M., Garapati,P., Haw,R., Jassal,B., Korninger,F. and May,B. (2018) The reactome pathway knowledgebase. *Nucleic acids research*, **46**, D649–D655.

35. Schriml,L.M., Mitraka,E., Munro,J., Tauber,B., Schor,M., Nickle,L., Felix,V., Jeng,L., Bearer,C. and Lichenstein,R. (2019) Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic acids research*, **47**, D955–D962.

36. Repana,D., Nulsen,J., Dressler,L., Bortolomeazzi,M., Venkata,S.K., Tourna,A., Yakovleva,A., Palmieri,T. and Ciccarelli,F.D. (2019) The Network of Cancer Genes (NCG): a comprehensive catalogue of known and candidate cancer genes from cancer sequencing screens. *Genome biology*, **20**, 1.

37. Piñero,J., Ramírez-Anguita,J.M., Saüch-Pitarch,J., Ronzano,F., Centeno,E., Sanz,F. and Furlong,L.I. (2020) The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic acids research*, **48**, D845–D855.

38. Liberzon,A., Subramanian,A., Pinchback,R., Thorvaldsdóttir,H., Tamayo,P. and Mesirov,J.P. (2011) Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, **27**, 1739–1740.

39. Liberzon,A., Birger,C., Thorvaldsdóttir,H., Ghandi,M., Mesirov,J.P. and Tamayo,P. (2015) The molecular signatures database hallmark gene set collection. *Cell systems*, **1**, 417–425.

40. Smyth,G.K. (2005) Limma: linear models for microarray data. In *Bioinformatics and computational biology solutions using R and Bioconductor*. Springer, pp. 397–420.

41. Tibshirani,R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**, 267–288.

42. Friedman,J., Hastie,T. and Tibshirani,R. (2010) Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, **33**, 1.