Published online 28 November 2016 *Nucleic Acids Research, 2017, Vol. 45, Database issue* ***D925–D931***
doi: 10.1093/nar/gkw1084

YM500v3: a database for small RNA sequencing in human cancer research

I-Fang Chung[1,†], Shing-Jyh Chang[2,†], Chen-Yang Chen[1], Shu-Hsuan Liu[3,4], Chia-Yang Li[5,6], Chia-Hao Chan[2], Chuan-Chi Shih[2] and Wei-Chung Cheng[3,4,*]

[1]Institute of Biomedical Informatics, National Yang-Ming University, Taipei 11221, Taiwan, [2]Department of Obstetrics and Gynecology, Hsinchu MacKay Memorial Hospital, Hsinchu City 30071, Taiwan, [3]Graduate Institute of Biomedical Sciences, China Medical University, Taichung, 40402, Taiwan, [4]Research Center for Tumour Medical Science, China Medical University, Taichung, 40402, Taiwan, [5]Department of Genome Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung 80708, Taiwan and [6]Center for Infectious Disease and Cancer Research, Kaohsiung Medical University, Kaohsiung 80708, Taiwan

Received September 15, 2016; Revised October 24, 2016; Editorial Decision October 24, 2016; Accepted October 26, 2016

ABSTRACT

We previously presented the YM500 database, which contains >8000 small RNA sequencing (smRNA-seq) data sets and integrated analysis results for various cancer miRNome studies. In the updated YM500v3 database (http://ngs.ym.edu.tw/ym500/) presented herein, we not only focus on miRNAs but also on other functional small non-coding RNAs (sncRNAs), such as PIWI-interacting RNAs (piRNAs), tRNA-derived fragments (tRFs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs). There is growing knowledge of the role of sncRNAs in gene regulation and tumorigenesis. We have also incorporated >10 000 cancer-related RNA-seq and >3000 more smRNA-seq data sets into the YM500v3 database. Furthermore, there are two main new sections, 'Survival' and 'Cancer', in this updated version. The 'Survival' section provides the survival analysis results in all cancer types or in a user-defined group of samples for a specific sncRNA. The 'Cancer' section provides the results of differential expression analyses, miRNA–gene interactions and cancer miRNA-related pathways. In the 'Expression' section, sncRNA expression profiles across cancer and sample types are newly provided. Cancer-related sncRNAs hold potential for both biotech applications and basic research.

INTRODUCTION

Since next generation sequencing (NGS) has become the norm for large-scale genomics research (e.g. The Cancer Genome Atlas, TCGA), small RNA sequencing (smRNA-seq) has shed light on the variations in the expression of small non-coding RNAs (sncRNAs) among different developmental stages and disease states (1). Although the use of smRNA-seq was popularized in genomics studies, most such research has primarily focused on miRNAs, which represent only a subset of all small RNA species. However, the functionality of other sncRNAs, such as PIWI-interacting RNAs (piRNAs), tRNA-derived fragments (tRFs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs), remain an important topic. Increasing evidence has shown that these non-miRNA sncRNAs also play significant roles in regulating cellular processes, such that their dysfunction would consequently contribute to cancer progression (2). Hence, the investigation of dysregulation of other classes of sncRNAs in the context of cancer, as well as of their therapeutic and diagnostic values, is of great importance. For example, a growing number of studies have reported that aberrant piRNA expression is a signature marker across distinct tumor types (3) and that snoRNAs act as oncogenes in tumorigenesis (4–6). The integration of large-scale smRNA-seq data helps researchers study the roles of these functional sncRNAs in cancer progression, but questions remain concerning the optimal methodologies for analysis, translation and utilization of such massive amounts of data (7).

The role of miRNA in cancer progression has been well-investigated in the past decade (8–11). miRNAs can affect gene expression not only by suppressing protein translation but also by reducing the mRNA expression of a target gene, resulting in a correlation between the expression levels of miRNAs and their target genes (12–14). Consequently, the expression relationships between miRNAs and genes are often used to predict miRNA–gene interactions (15–17). Therefore, integrating miRNA and mRNA expression data across different cancer types is another approach

*To whom correspondence should be addressed. Tel: +886 4 2205 2121 (Ext. 7820); Fax: +886 4 2233 7425; Email: cwc0702@gmail.com
†These authors contributed equally to this work as the first authors.

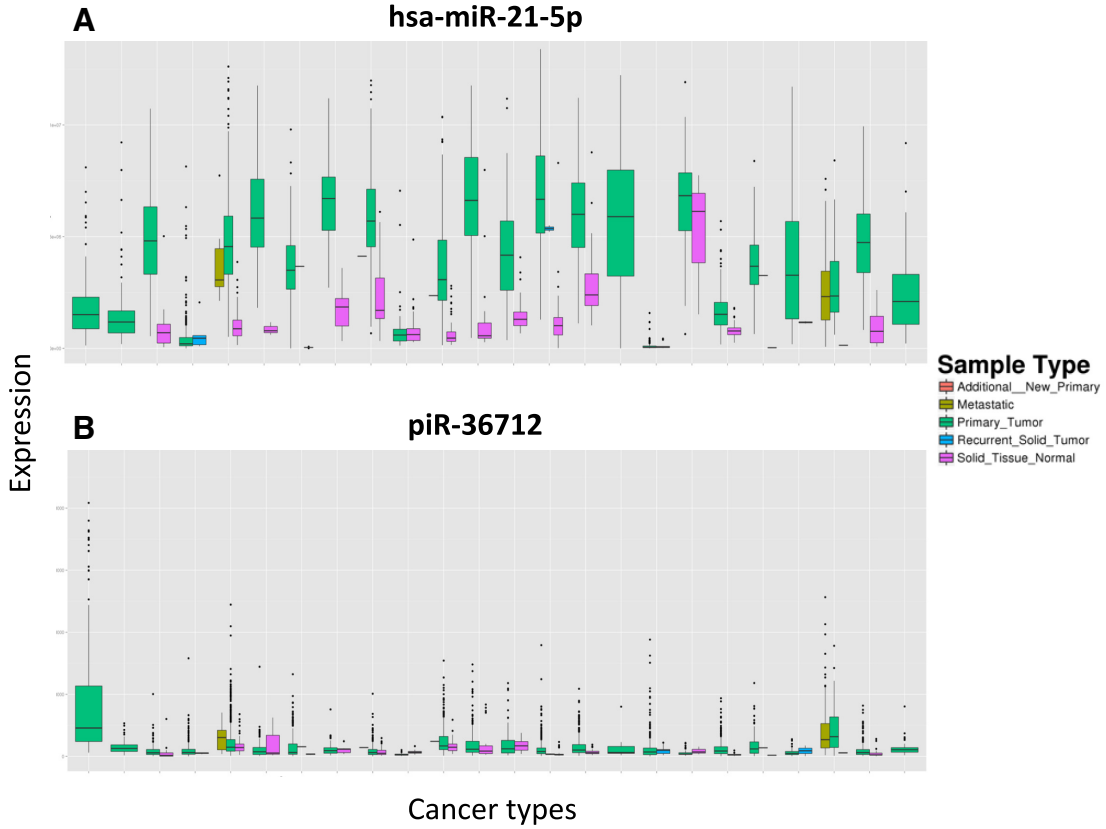

Figure 1. The 'Expression' section. The exemplified expression boxplots of the (**A**) miRNA and the (**B**) piRNA across distinct cancers by sample types.

to providing a global miRNA–gene interactions, including cancer-specific and cancer-wide miRNA–gene regulatory networks. For instance, Meng *et al*. utilized the expressions of miRNAs and mRNA in TGCA to identify miRNA–target interactions (18). Many miRNA markers have been proposed to be predictive of patient prognoses and clinical responses and are being investigated in clinical trials (8). An important step that researchers must take prior to proposing miRNA-based biomarkers for clinical validation is their evaluation in independent patient cohorts, and several web tools, such as SurvMicro (19) and PROGmiR (20), have been developed to help researchers link miRNA expression with cancer outcomes.

Previously, we developed the YM500 database (21,22), a database that contains more than 8000 cancer-related smRNA data sets and includes analysis pipelines for novel miRNA prediction, arm switching discovery, isomiR identification and miRNA quantification from smRNA-seq. The previous version of this database focused only on miRNAs. For the updated version of the database, YM500v3, presented in this study, we also examined other functional sncRNAs in smRNA-seq data sets and incorporated >10 000 cancer-related RNA seq data sets and >3000 more smRNA-seq data sets from TCGA. Moreover, two major new sections, 'Survival' and 'Cancer', are provided in the YM500v3 database. The 'Survival' section provides the survival analysis results for all cancer types or a customer-defined group of samples for a specific sncRNA. The 'Cancer' section provides results regarding the differential expressions of sncRNAs and genes, miRNA–gene regulated networks and cancer miRNA-related pathways.

DATA COLLECTION AND SMALL RNA ANNOTATION

The new smRNA-seq and RNA-seq data sets and clinical data in TCGA were downloaded from CGHub (https://cghub.ucsc.edu/) and pre-processed as described in our previous studies (22–24). In brief, all sequencing data were pre-processed by in-house scripts. The clinical data for each individual was manually curated based on the common data element format, the standard elements of which are used in TCGA. The annotations of miRNA and other sncRNA, such as piRNA, snRNA, snoRNA and tRFs, are based on miRBase database R21 (25) and DASHR database v1.0 (26), respectively. The DASHR database contains 7641 sncRNA gene records and 9703 annotated mature sncRNA product records. Supplementary Table S1 shows the detailed information of sncRNAs in YM500v3.

DIFFERENTIAL EXPRESSION AND MIRNA-TARGET INTERACTIONS

For differential expression analysis, we utilized an R/Bioconductor package, DESeq (27,28) to identify differentially expressed miRNAs, other non-miRNA sncRNA and genes. The miRNA-target interactions in the YM500v3 database can be grouped into three types, including 'Validated', 'Predicated' and 'Without any evidence'.

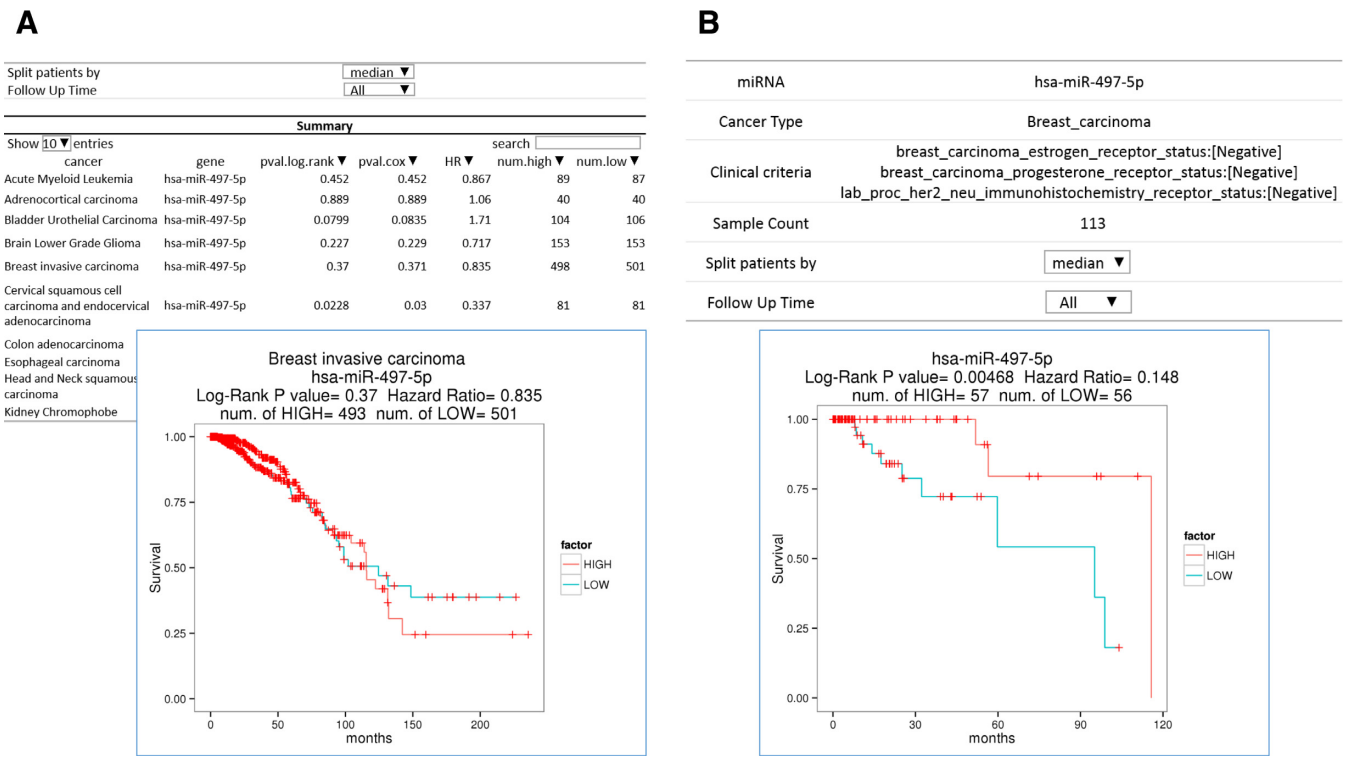

Figure 2. Two features of the 'Section' section. (**A**) 'All cancer types' contains a summary table for all the cancers and a Kaplan–Meier plot for each individual cancer type. (**B**) 'Specific sample group' helps investigators define a subgroup of patients in a cancer type and provide a Kaplan–Meier plot for the subgroup. Both of the two features contains two menu bars to control the stratification method and the follow-up time.

The 'Validated' interactions are based on the information from miRTarBase database Release 6.1 (29), which contains >366 000 interactions. The predicted miRNA targets were identified by 12 miRNA target prediction tools, including DIANA-microT (30), MicroT4 (31), miRBridge (32), miRDB (33), miRMap (34), PITA (35), RNAhybrid (36), TargetScan (37), PICTAR2 (38), RNA22 (39), miRWalk (40) and miRanda (41). Only the targets that were identified by at least six tools were retained to improve the reliability of the prediction results. In the YM500v3 database, for a specific cancer type, only the differentially expressed miRNAs and genes, as identified by DESeq with $q < 0.05$ and fold change > 2, would be further calculated for the Pearson, Spearman and Kendall correlations for each miRNA–gene pair. The maximum absolute correlation coefficient, max(|R|) and the minimum *P*-value of the three correlation tests were also calculated for further filtration.

WEB INTERFACE

Expression

This section now contains not only miRNAs but also other functional sncRNA annotated in the DASHR database. Several statistical charts are added to the 'Expression' section to help researchers realize the expression profile of a given sncRNA across distinct cancer types. For example, the expression profiles of the miRNA and piRNA across different cancers by sample types are illustrated by boxplots in Figure 1A and B, respectively. Supplementary Figure S1A and B indicate the log2 ratio (tumors compared to adjacent normal tissues) distribution across cancer types and the expression boxplot by sample types for each cancer type, respectively. Moreover, a given sncRNA may have different IDs in different sources. As such, we also provide a sequence search function in the new database to overcome any inconsistencies in the IDs used by different sources.

Survival

This new section has two features: 'All Cancer Types' and 'Specific Sample Group'. 'All Cancer Type' displays the survival analysis of a specific sncRNA (either miRNAs or other sncRNAs) in all different cancer types (Figure 2A), including a summary table for all the cancers and a Kaplan–Meier plot for each individual cancer type. In addition, we also provide two menu bars to control the stratification method, such as 'mean' and 'median', and the follow-up time and to display the results immediately. The default setting uses the median expression value to divide the patients into two groups in addition to using the entire follow-up time. 'Specific Sample Group' helps researchers define a subgroup of samples in a single cancer type, such as triple negative breast cancer, to perform survival analysis according to dozens of clinical characteristics. Figure 2 shows that the high expression of hsa-miR-497-5p is related to good prognosis in triple negative breast cancer (Figure 2B) but it does not significantly correlate with good prognosis in all breast cancer patients (Figure 2A).

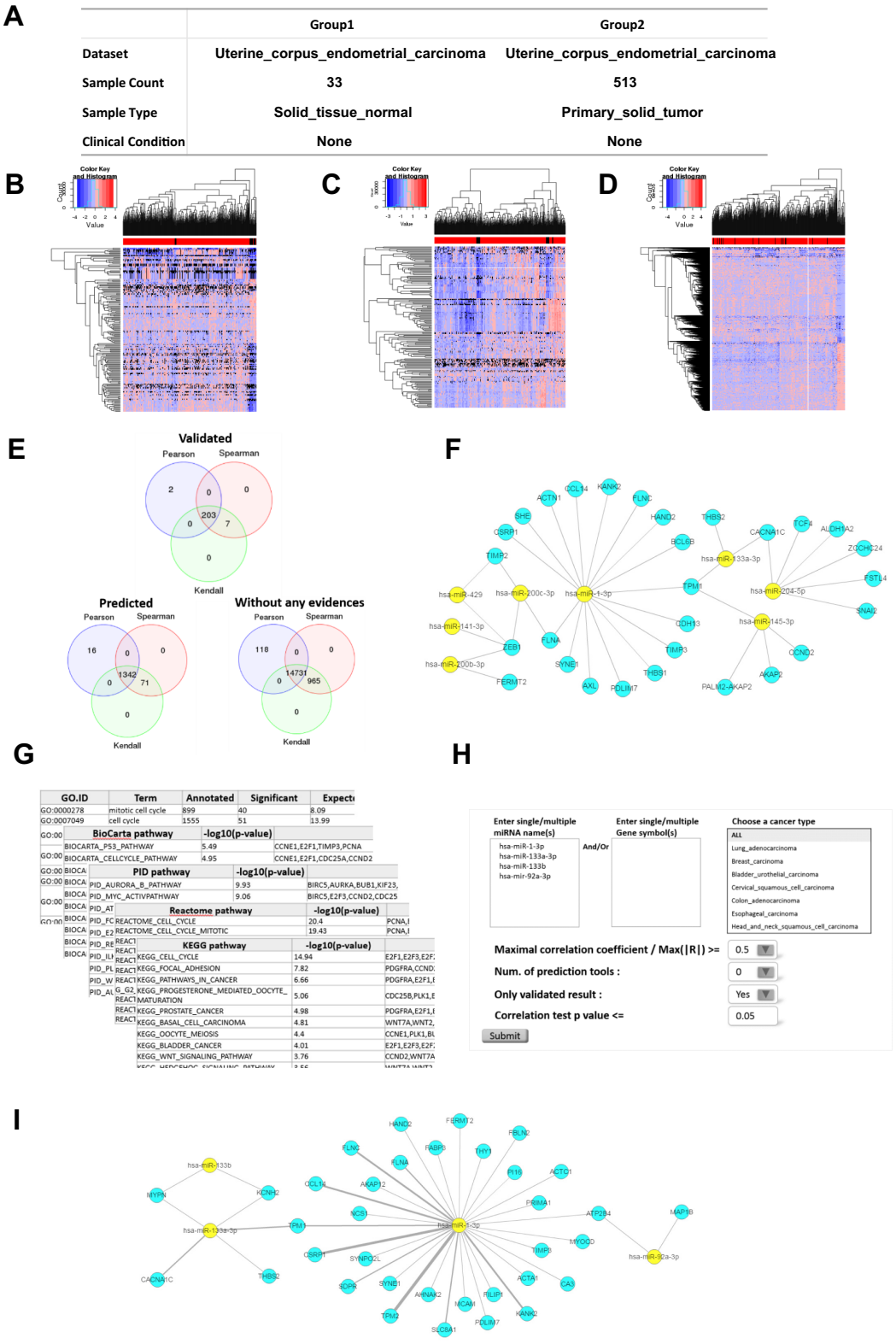

Figure 3. The 'Cancer' section. This section stores the calculated results by (**A**) cancer types that contains the results of differential expression analysis, including (**B**) miRNAs, (**C**) non-miR sncRNAs, (**D**) mRNAs. The correlations of each miRNA–gene pair were calculated and divided into three groups, namely, (**E**) 'Validated', 'Predicated' and 'Without any evidence', as well as displayed by an (**F**) interactive network visualization. (**G**) The cancer miRNA-related pathways were identified by the miRNA-interacted genes through functional enrichment analysis. The another feature, 'Specific miRNA-gene pairs', help researchers examine the interactions between miRNAs and genes by (**H**) user-defined criteria and then the (**I**) miRNA–gene pairs are displayed immediately. The width of the line in (I) indicates the number of records.

Cancer

The 'Cancer' section stores the calculated results of differential expression analyses, miRNA–gene interactions and cancer miRNA-related pathways for a specific cancer type that contains the smRNA-seq and RNA-seq data of normal and tumor tissues for the same individuals. Figure 3 shows the results of uterine corpus endometrial carcinoma in TCGA for 33 adjacent normal and 513 primary tumor tissues (Figure 3A). There are 175 miRNAs (Figure 3B), 170 other sncRNAs (Figure 3C) and 3148 genes (Figure 3D) differentially expressed between normal and tumor tissues. The correlations of each miRNA–gene pair between the differentially expressed miRNAs and genes were calculated and divided into three groups, namely, 'Validated', 'Predicated' and 'Without any evidence' (Figure 3E). In order to illustrate the many-to-many relationships between miRNA–gene interactions (Figure 3F), the Cytoscape Web (48) tool is embedded for interactive network visualization. The genes that interacted with miRNAs were further functionally analyzed to address the cancer miRNA-related pathways (Figure 3G). Detailed information regarding the functional enrichment analysis method was presented in our previous studies (24). Two menu bars are also provided to control the criteria, the max(|R|) and the number of prediction tools used in order to display the corresponding results.

We also provide another feature, 'Specific miRNA-gene pairs', in the 'Cancer' section in order to help researchers examine the interactions between miRNAs and genes by user-defined criteria. Researchers can enter multiple miRNAs and/or genes, and can also define the interactions according to max(|R|), minimum *P*-value, the number of prediction tools and the validated information for the miRNA-gene pairs (Figure 3H). After a query is submitted, the miRNA–gene pairs identified according to the user-defined criteria are then displayed immediately (Figure 3I). For the interactions supported by multiple cancer types, the width of the line indicates the number of records.

DISCUSSION

The library construction in smRNA-seq selects RNAs by their lengths rather than their types. The libraries obtained for smRNA-seq contain a variety of species of sncRNAs, indicating that miRNAs represent only a subset of the species obtained by size selection. Although miRNAs are only one of the many sncRNA species in smRNA-seq data sets, miRNAs remain the most popular class to study, largely because their biogenesis is relatively well understood and because the regulatory mechanism in post-transcription is known (42). However, more and more evidence shows that other non-miRNA sncRNAs also play important roles in gene regulation and certain diseases, such as cancers (5,7,43–45). For instance, there is an increasing amount of knowledge regarding the role of snoRNAs in cancer progression, and the information obtained thus far suggests that snoRNAs hold considerable potential for use as novel biomarkers and therapeutic targets in cancer treatment (4–6). It has also been reported that tRFs exhibit features of functional regulatory molecules (46–48), and they have a relatively well described role in disease and infection (49–51). Unfortunately, many researchers ignore the numerous non-miRNA sncRNA species present in the smRNA-seq data. A common barrier is often the lack of genomic annotations for these non-miRNA species. In this updated version of the YM500 database, however, we not only focus on miRNAs in smRNA-seq but also on other non-miRNA sncRNA according to the well-annotated sncRNA database, DASHR. Several functions in the updated database, including 'Expression', 'Survival' and 'Cancer', can assist researchers in investigating sncRNAs.

The concept behind precision medicine is intuitive: individual patients are better modeled by a subgroup of patients, rather than a larger, more general population of patients (52,53). In seeking to adhere to this concept, the 'Survival' and 'Meta-analysis' sections provide functions to help investigators define specific sample groups according to dozens of clinical characteristics. In the 'Survival' section, this concept has been exemplified by hsa-miR-497-5p that has been reported as a 'protective' miRNA in triple negative breast cancer (54). Our analysis shows that the high expression of hsa-miR-497-5p is significantly related to good prognosis in triple negative breast cancer but its expression does not correlate with prognosis in all breast cancer patients (Figure 2). Furthermore, the 'Meta-analysis' section contains the same types of results in the 'Cancer' section, including differential expression analyses, miRNA–gene interaction and miRNA-related pathway results, but the results in the 'Meta-analysis' section are based on the two customer-defined groups. The 'Cancer' section only stores the calculated results based on the two groups, the adjacent normal and primary tumor tissues in the same cancer type. For example, if a miRNA–gene interaction only exists in some specific sample groups, it cannot be found in the 'Cancer' section but might be identified in the 'Meta-analysis' section. Moreover, the 'Specific miRNA–gene interactions' function in the 'Cancer' section helps researchers investigate specific interactions according to a list of criteria that they themselves have defined, with the width of lines in the interactive network indicating the confidence.

It is currently a golden era in the field of genomics. Due to the rapidly decreasing costs of sequencing, the obstacles to performing genomic-scale NGS do not lie in the area of data generation, but rather are obstacles affecting data analysis and storage (42). Although researchers in the field of genomics are certainly aware of the mining of novel sncRNAs, many investigators currently choose not to fully analyze the sncRNAs in their smRNA-seq. Nonetheless, there are still many complexities to be discovered in the sncRNA transcriptome. To achieve this goal, we will continue to update the smRNA-seq data sets and sncRNA annotations to provide a comprehensive overview of up-to-date sncRNAs in cancer research.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors are grateful to the National Center for High performance Computing for computer time and facilities

and thank the TCGA research network for the availability of data.

FUNDING

Ministry of Science and Technology (MOST) [MOST 105-2320-B-039-006 and 105-2628-E-010-002-MY2]; China Medical University [CMU105-N-06]; Hsinchu MacKay Memorial Hospital, Taiwan [MMH-HB-10515 and 1558213]. Funding for open access charge: Ministry of Science and Technology in Taiwan [MOST 105-2320-B-039-006 and 105-2628-E-010-002-MY2], and Hsinchu MacKay Memorial Hospital, Taiwan [MMH-HB-10515 and 1558213].

Conflict of interest statement. None declared.

REFERENCES

1. Cech,T.R. and Steitz,J.A. (2014) The noncoding RNA revolution-trashing old rules to forge new ones. *Cell*, **157**, 77–94.
2. Galasso,M., Sana,M.E. and Volinia,S. (2010) Non-coding RNAs: a key to future personalized molecular therapy? *Genome Med.*, **2**, 12.
3. Ng,K.W., Anderson,C., Marshall,E.A., Minatel,B.C., Enfield,K.S., Saprunoff,H.L., Lam,W.L. and Martinez,V.D. (2016) Piwi-interacting RNAs in cancer: emerging functions and clinical utility. *Mol. Cancer*, **15**, 5.
4. Chen,W.D. and Zhu,X.F. (2013) Small nucleolar RNAs (snoRNAs) as potential non-invasive biomarkers for early cancer detection. *Clin. J. Cancer*, **32**, 99–101.
5. Thorenoor,N. and Slaby,O. (2015) Small nucleolar RNAs functioning and potential roles in cancer. *Tumour Biol.*, **36**, 41–53.
6. Mannoor,K., Liao,J. and Jiang,F. (2012) Small nucleolar RNAs in cancer. *Biochim. Biophys. Acta*, **1826**, 121–128.
7. Guo,Y., Bosompem,A., Mohan,S., Erdogan,B., Ye,F., Vickers,K.C., Sheng,Q., Zhao,S., Li,C.I., Su,P.F. *et al.* (2015) Transfer RNA detection by small RNA deep sequencing and disease association with myelodysplastic syndromes. *BMC Genomics*, **16**, 727.
8. Hayes,J., Peruzzi,P.P. and Lawler,S. (2014) MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol. Med.*, **20**, 460–469.
9. Bandyopadhyay,S., Mitra,R., Maulik,U. and Zhang,M.Q. (2010) Development of the human cancer microRNA network. *Silence*, **1**, 6.
10. Lu,J., Getz,G., Miska,E.A., Alvarez-Saavedra,E., Lamb,J., Peck,D., Sweet-Cordero,A., Ebert,B.L., Mak,R.H., Ferrando,A.A. *et al.* (2005) MicroRNA expression profiles classify human cancers. *Nature*, **435**, 834–838.
11. Chang,L.C. and Yu,Y.L. (2016) Dietary components as epigenetic-regulating agents against cancer. *Biomedicine (Taipei)*, **6**, 9-16.
12. Bagga,S., Bracht,J., Hunter,S., Massirer,K., Holtz,J., Eachus,R. and Pasquinelli,A.E. (2005) Regulation by let-7 and lin-4 miRNAs results in target mRNA degradation. *Cell*, **122**, 553–563.
13. Lim,L.P., Lau,N.C., Garrett-Engele,P., Grimson,A., Schelter,J.M., Castle,J., Bartel,D.P., Linsley,P.S. and Johnson,J.M. (2005) Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, **433**, 769–773.
14. Guo,H., Ingolia,N.T., Weissman,J.S. and Bartel,D.P. (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, **466**, 835–840.
15. Huang,J.C., Babak,T., Corson,T.W., Chua,G., Khan,S., Gallie,B.L., Hughes,T.R., Blencowe,B.J., Frey,B.J. and Morris,Q.D. (2007) Using expression profiling data to identify human microRNA targets. *Nat. Methods*, **4**, 1045–1049.
16. Gennarino,V.A., Sardiello,M., Avellino,R., Meola,N., Maselli,V., Anand,S., Cutillo,L., Ballabio,A. and Banfi,S. (2009) MicroRNA target prediction by expression analysis of host genes. *Genome Res.*, **19**, 481–490.
17. Jacobsen,A., Silber,J., Harinath,G., Huse,J.T., Schultz,N. and Sander,C. (2013) Analysis of microRNA-target interactions across diverse cancer types. *Nat. Struct. Mol. Biol.*, **20**, 1325–1332.
18. Meng,X., Wang,J., Yuan,C., Li,X., Zhou,Y., Hofestadt,R. and Chen,M. (2015) CancerNet: a database for decoding multilevel molecular interactions across diverse cancer types. *Oncogenesis*, **4**, e177.
19. Aguirre-Gamboa,R. and Trevino,V. (2014) SurvMicro: assessment of miRNA-based prognostic signatures for cancer clinical outcomes by multivariate survival analysis. *Bioinformatics*, **30**, 1630–1632.
20. Goswami,C.P. and Nakshatri,H. (2012) PROGmiR: a tool for identifying prognostic miRNA biomarkers in multiple cancers using publicly available data. *J. Clin. Bioinformatics*, **2**, 23.
21. Cheng,W.C., Chung,I.F., Huang,T.S., Chang,S.T., Sun,H.J., Tsai,C.F., Liang,M.L., Wong,T.T. and Wang,H.W. (2013) YM500: a small RNA sequencing (smRNA-seq) database for microRNA research. *Nucleic Acids Res.*, **41**, D285–D294.
22. Cheng,W.C., Chung,I.F., Tsai,C.F., Huang,T.S., Chen,C.Y., Wang,S.C., Chang,T.Y., Sun,H.J., Chao,J.Y., Cheng,C.C. *et al.* (2015) YM500v2: a small RNA sequencing (smRNA-seq) database for human cancer miRNome research. *Nucleic Acids Res.*, **43**, D862–D867.
23. Cheng,W.C., Chung,I.F., Chen,C.Y., Sun,H.J., Fen,J.J., Tang,W.C., Chang,T.Y., Wong,T.T. and Wang,H.W. (2014) DriverDB: an exome sequencing database for cancer driver gene identification. *Nucleic Acids Res.*, **42**, D1048–D1054.
24. Chung,I.F., Chen,C.Y., Su,S.C., Li,C.Y., Wu,K.J., Wang,H.W. and Cheng,W.C. (2016) DriverDBv2: a database for human cancer driver gene research. *Nucleic Acids Res.*, **44**, D975–D979.
25. Kozomara,A. and Griffiths-Jones,S. (2013) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
26. Leung,Y.Y., Kuksa,P.P., Amlie-Wolf,A., Valladares,O., Ungar,L.H., Kannan,S., Gregory,B.D. and Wang,L.S. (2016) DASHR: database of small human noncoding RNAs. *Nucleic Acids Res.*, **44**, D216–D222.
27. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, R106.
28. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
29. Chou,C.H., Chang,N.W., Shrestha,S., Hsu,S.D., Lin,Y.L., Lee,W.H., Yang,C.D., Hong,H.C., Wei,T.Y., Tu,S.J. *et al.* (2016) miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.*, **44**, D239–D247.
30. 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.
31. Maragkakis,M., Vergoulis,T., Alexiou,P., Reczko,M., Plomaritou,K., Gousis,M., Kourtis,K., Koziris,N., Dalamagas,T. and Hatzigeorgiou,A.G. (2011) DIANA-microT Web server upgrade supports Fly and Worm miRNA target prediction and bibliographic miRNA to disease association. *Nucleic Acids Res.*,**39**, W145–W148.
32. Tsang,J.S., Ebert,M.S. and van Oudenaarden,A. (2010) Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures. *Mol. Cell*, **38**, 140–153.
33. Wong,N. and Wang,X. (2014) miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.*, **43**, D146–D152.
34. Vejnar,C.E. and Zdobnov,E.M. (2012) miRmap: Comprehensive prediction of microRNA target repression strength. *Nucleic Acids Res.*, **40**, 11673–11683.
35. Kertesz,M., Iovino,N., Unnerstall,U., Gaul,U. and Segal,E. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.*, **39**, 1278–1284.
36. Kruger,J. and Rehmsmeier,M. (2006) RNAhybrid: microRNA target prediction easy, fast and flexible. *Nucleic Acids Res.*, **34**, W451–W454.
37. Shin,C., Nam,J.W., Farh,K.K., Chiang,H.R., Shkumatava,A. and Bartel,D.P. (2010) Expanding the microRNA targeting code: functional sites with centered pairing. *Mol. Cell*, **38**, 789–802.
38. Blin,K., Dieterich,C., Wurmus,R., Rajewsky,N., Landthaler,M. and Akalin,A. (2015) DoRiNA 2.0–upgrading the doRiNA database of RNA interactions in post-transcriptional regulation. *Nucleic Acids Res.*, **43**, D160–D167.
39. Miranda,K.C., Huynh,T., Tay,Y., Ang,Y.S., Tam,W.L., Thomson,A.M., Lim,B. and Rigoutsos,I. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, **126**, 1203–1217.

40. Dweep,H. and Gretz,N. (2015) miRWalk2.0: a comprehensive atlas of microRNA-target interactions. *Nat. Methods*, **12**, 697–698.
41. Enright,A.J., John,B., Gaul,U., Tuschl,T., Sander,C. and Marks,D.S. (2003) MicroRNA targets in Drosophila. *Genome Biol.*, **5**, R1.
42. Vickers,K.C., Roteta,L.A., Hucheson-Dilks,H., Han,L. and Guo,Y. (2015) Mining diverse small RNA species in the deep transcriptome. *Trends Biochem. Sci.*, **40**, 4–7.
43. Keam,S.P. and Hutvagner,G. (2015) tRNA-Derived Fragments (tRFs): Emerging new roles for an ancient RNA in the regulation of gene expression. *Life (Basel)*, **5**, 1638–1651.
44. Su,J., Liao,J., Gao,L., Shen,J., Guarnera,M.A., Zhan,M., Fang,H., Stass,S.A. and Jiang,F. (2016) Analysis of small nucleolar RNAs in sputum for lung cancer diagnosis. *Oncotarget*, **7**, 5131–5142.
45. Su,H., Xu,T., Ganapathy,S., Shadfan,M., Long,M., Huang,T.H., Thompson,I. and Yuan,Z.M. (2014) Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene*, **33**, 1348–1358.
46. Haussecker,D., Huang,Y., Lau,A., Parameswaran,P., Fire,A.Z. and Kay,M.A. (2010) Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA*, **16**, 673–695.
47. Sobala,A. and Hutvagner,G. (2013) Small RNAs derived from the 5′ end of tRNA can inhibit protein translation in human cells. *RNA Biol.*, **10**, 553–563.
48. Couvillion,M.T., Bounova,G., Purdom,E., Speed,T.P. and Collins,K. (2012) A Tetrahymena Piwi bound to mature tRNA 3′ fragments activates the exonuclease Xrn2 for RNA processing in the nucleus. *Mol. Cell*, **48**, 509–520.
49. Deng,J., Ptashkin,R.N., Chen,Y., Cheng,Z., Liu,G., Phan,T., Deng,X., Zhou,J., Lee,I., Lee,Y.S. *et al.* (2015) Respiratory syncytial virus utilizes a tRNA Fragment to suppress antiviral responses through a novel targeting mechanism. *Mol. Ther.*, **23**, 1622–1629.
50. Blanco,S., Dietmann,S., Flores,J.V., Hussain,S., Kutter,C., Humphreys,P., Lukk,M., Lombard,P., Treps,L., Popis,M. *et al.* (2014) Aberrant methylation of tRNAs links cellular stress to neuro-developmental disorders. *EMBO J.*, **33**, 2020–2039.
51. Selitsky,S.R., Baran-Gale,J., Honda,M., Yamane,D., Masaki,T., Fannin,E.E., Guerra,B., Shirasaki,T., Shimakami,T., Kaneko,S. *et al.* (2015) Small tRNA-derived RNAs are increased and more abundant than microRNAs in chronic hepatitis B and C. *Sci. Rep.*, **5**, 7675.
52. Alemi,F., Erdman,H., Griva,I. and Evans,C.H. (2009) Improved statistical methods are needed to advance personalized medicine. *Open Transl. Med. J.*, **1**, 16–20.
53. Gaile,D.P. and Miecznikowski,J.C. (2013) From small studies to precision medicine: prioritizing candidate biomarkers. *Genome Med.*, **5**, 104.
54. Cascione,L., Gasparini,P., Lovat,F., Carasi,S., Pulvirenti,A., Ferro,A., Alder,H., He,G., Vecchione,A., Croce,C.M. *et al.* (2013) Integrated microRNA and mRNA signatures associated with survival in triple negative breast cancer. *PLoS One*, **8**, e55910.