piRBase: integrating piRNA annotation in all aspects

Jiajia Wang[1,2,3,4,†], Yirong Shi[3,5,†], Honghong Zhou[1,2,3,†], Peng Zhang [3,6], Tingrui Song[3], Zhiye Ying[1,2], Haopeng Yu[1,2], Yanyan Li[1,2,3], Yi Zhao [1,2,7], Xiaoxi Zeng[1,2,*], Shunmin He [1,2,3,4,6,*] and Runsheng Chen[1,2,3,6,8,*]

[1]West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, China, [2]Med-X Center for Informatics, Sichuan University, Chengdu 610041, China, [3]Key Laboratory of RNA Biology, Center for Big Data Research in Health, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, [4]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China, [5]University of Chinese Academy of Sciences, Beijing 100049, China, [6]National Genomics Data Center, Chinese Academy of Sciences, Beijing 100101, China, [7]Key Laboratory of Intelligent Information Processing, Advanced Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China and [8]Guangdong Geneway Decoding Bio-Tech Co. Ltd, Foshan 528316, China

ABSTRACT

Piwi-interacting RNAs are a type of small noncoding RNA that have various functions. piRBase is a manually curated resource focused on assisting piRNA functional analysis. piRBase release v3.0 is committed to providing more comprehensive piRNA related information. The latest release covers >181 million unique piRNA sequences, including 440 datasets from 44 species. More disease-related piRNAs and piRNA targets have been collected and displayed. The regulatory relationships between piRNAs and targets have been visualized. In addition to the reuse and expansion of the content in the previous version, the latest version has additional new content, including gold standard piRNA sets, piRNA clusters, piRNA variants, splicing-junction piRNAs, and piRNA expression data. In addition, the entire web interface has been redesigned to provide a better experience for users. piRBase release v3.0 is free to access, browse, search, and download at http://bigdata.ibp.ac.cn/piRBase.

INTRODUCTION

Piwi-interacting RNAs (piRNAs) are a type of small noncoding RNA that bind to Piwi proteins (1,2) to form a complex and perform vital functions in the germline, somatic tissues and diseases. Unlike the other two types of small RNAs, microRNAs (miRNAs) and siRNAs, piRNAs are 24–31 nucleotides (nt) in length, and 2′-O-methylated at

the 3′ terminus (3–5). Recent studies have shown that piRNAs loaded with the ovary-specific Piwi protein, PIWIL3, have no 2′-O-methylation at the 3′ end (6,7). Many piRNAs originate from large genomic regions called piRNA clusters (2,8,9), which can be transcribed into long single-strand precursors and further processed into mature piRNAs (10–14). It was reported that some piRNA precursors contain introns (15).

Numerous studies have shown that piRNAs have many important biological functions (16). piRNAs are necessary for the maintenance of genome stability. The Piwi/piRNA complex can regulate the activity of transposons at genomic level through heterochromatin formation to silence transcription (17–21) and at the posttranscriptional level by repressing expression (9,22–26). Moreover, piRNAs are also involved in sex determination (27,28) and the regulation of mRNA and lncRNA expression (28–38). Accumulating studies have shown that piRNAs are expressed not only in germline cells but also in cancer cells (39–45) and somatic cells (46,47). piRNAs that have abnormal expression in various cancers can be used as potential diagnostic and prognostic biomarkers. In addition, Fu *et al.* found that the SNP rs132306 in piR-021285 is significantly associated with an increased risk of breast cancer (48). In addition to cancers, some piRNAs are aberrantly expressed in other diseases. In Alzheimer's disease (AD), Qiu *et al.* found that >100 piRNAs are differentially expressed in diseased versus normal brain tissue, and most of these piRNAs are associated with risk SNPs (49). In addition, piRNAs can also serve as potential biomarkers for AD (50). piRNAs also play roles that cannot be ignored in other noncancer diseases, such as cardiovascular disease (51–53). Previous studies have shown

---

*To whom correspondence should be addressed. Tel: +86 10 6488 7032; Fax: +86 10 6488 7842; Email: heshunmin@ibp.ac.cn
Correspondence may also be addressed to Runsheng Chen. Tel: +86 10 6488 8543; Fax: +86 10 6487 1293; Email: crs@ibp.ac.cn
Correspondence may also be addressed to Xiaoxi Zeng. Tel: +86 28 85422819; Fax: +86 28 85582944; Email: zengxiaoxi@wchscu.cn
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as joint First Authors

that defects in the Piwi protein, which influence the piRNA pathway, can lead to male sterility (22). Recent studies have revealed that the production of fertile oocytes in female golden hamsters depends on piRNAs (54–56). These studies have extended our understanding of piRNA functions.

piRBase is currently the largest and most comprehensive piRNA database, and has been imported into RNAcentral (57) as an expert database. The number of different piRNAs in the current release has reached over 181 million, including 440 datasets from 44 species. Information about newly added piRNAs is annotated and displayed based on the previous modules. More information about piRNA-like small RNAs (piRNA-like sRNAs) and diseases, including cancers and other types of diseases, has been collected. Moreover, piRBase release v3.0 visualizes the regulatory network among the piRNAs and their targets. In addition to the expansion of the existing content, we also added some new content in piRBase release v3.0. At present, there are too many (up to tens of millions) piRNA sequences from various sequencing methods including chromatography, Piwi protein IP, Piwi protein CLIP-seq, oxidization and sRNA-seq. In order to get smaller and more representative piRNA set, we introduced the concept of the gold standard piRNA set in the piRBase release v3.0 according to the characteristics of piRNAs. We offer the gold standard piRNA set for six species, which will help users study piRNAs more effectively. Additionally, information on piRNA clusters, piRNA variants, splicing-junction piRNAs and piRNA expression profiles of various datasets was also added to the new version. Moreover, the web interface of piRBase release v3.0 has been redesigned.

DATA COLLECTION AND PROCESSING

Data collection and piRBase ID allocation

According to the data collection method in the previous versions of piRBase (58), new sequences defined as piRNAs in the corresponding article were collected from the literature, supplementary files and related GEOs (59). In addition, other relevant information, such as piRNA clusters, piRNA targets, piRNAs/piRNA-like sRNAs and diseases, was extracted from the published literature.

New piRNA sequences of existing species in previous versions were assigned sequential piRBase IDs. For the new species added to the current piRBase, the naming rules of the piRNA sequence are piR + three characters (representing a species) + sequential numbers. Similarly, the IDs of piRNA clusters in piRBase are clus + three characters (representing a species) + sequential numbers. The list of three letter abbreviations and species is shown in Supplementary Table S1.

Genome alignment, splicing-junction piRNAs and annotation

To obtain the piRNA alignment information, Bowtie (60) was used to map the piRNA sequences to the corresponding genome with parameters '-v 1 -a -m 10 –best –strata' as described in the previous versions of piRBase. However, some piRNAs cannot be mapped to the genome based on the previous alignment strategy. Considering the possibility of splicing, the unmapped piRNAs in Bowtie were re-

aligned by STAR (61) with transcript annotation in piRBase release v3.0. The results show that there are many piRNAs that span the intron regions. We called this kind of piRNA splicing-junction piRNAs. Based on the genomic loci, piRNAs are annotated into two categories, gene- or repeat-related piRNAs, as in previous versions of piRBase.

piRNA variant annotation

Based on the alignment results, not all piRNAs have perfect matches in the genome. There are some mismatches between piRNAs and genome sequences that could be induced by genomic variants. To extract variant information, the unique mapped records with one mismatch were selected from the Bowtie results. A mismatch site was designated as a piRNA variant if >10 nonredundant piRNA sequences supported the same mismatch site. Human piRNA variants are annotated by dbSNP (62) and associated with RS IDs if they exist.

piRNA expression profiles and gold standard piRNA sets

To determine the expression of piRNAs in various tissues, piRNA read counts in various datasets were collected, and these data are shown in piRBase (Supplementary Table S2). In this version, the CPM (counts per million) in each dataset was calculated for each piRNA sequence based on the total read count in the dataset and the read count of each sequence.

piRNAs interact with the Piwi protein (1,2) and have $2'$-O-methylation at their $3'$ termini (3–5). Although some piRNAs have no $2'$-O-methylation (6,7), the $2'$-O-methylation enriched sequences are considered a subset of piRNAs. These properties allow enrichment of piRNA sequences through methods such as immunoprecipitation (IP) of Piwi proteins and oxidation treatment (3). piRBase release v3.0 defines gold standard piRNA sets based on support from different kinds of piRNA enrichment datasets. Matched Piwi protein IP or oxidation treatment versus small RNA sequencing data were collected. We used a binomial model to calculate the P values of piRNA enrichment based on piRNA counts in enriched (n_e) and matched control small RNA (n_c) datasets. If piRNA has equal proportions in the two libraries, then n_e/n_c should be equal to N_e/N_c, where N_e and N_c are the total read counts of the two libraries. Let $p = N_e/(N_e + N_c)$ and $n = n_e + n_c$; then, n_e follows a binomial distribution of $X \sim B(n, p)$. The enrichment P value can be calculated as the probability of $X \geq n_e$. To include datasets lacking matched controls, the distributions of significantly enriched piRNAs at different abundance levels were examined for matched datasets. Considering the impact of library depth, piRNAs were sorted by counts in the enriched library, and the proportion of significantly ($P < 0.05$) enriched piRNAs was calculated in the most abundant piRNAs whose count sum exceeded a certain proportion of the total count in the enriched library (Figure 1). Based on the distributions, we chose a threshold of 0.2 to select the top piRNAs as the gold standard piRNAs in the enriched datasets lacking matched controls. For matched libraries, the top piRNAs with the same threshold were also selected regardless of significance, while for piR-
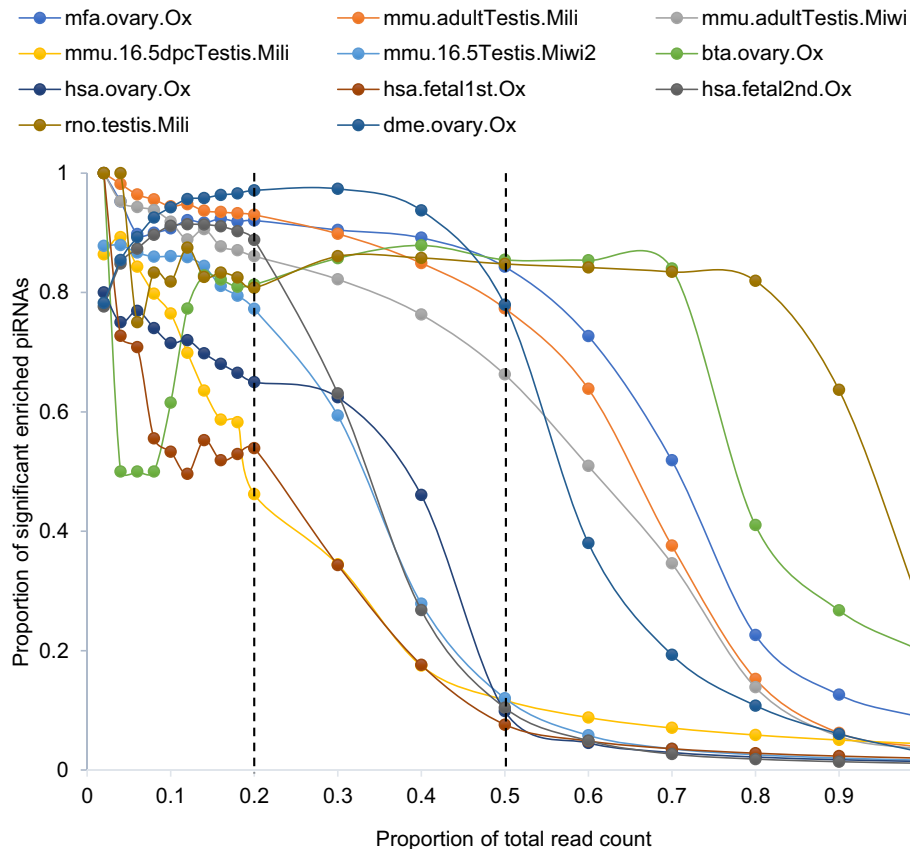
Figure 1. The proportion of significantly enriched piRNAs with different cutoffs of the most abundant piRNAs. The piRNAs in each enriched dataset were sorted by counts, and the cutoffs of the top piRNAs were determined when the sum of the top piRNA counts exceeded a certain proportion of the total read count in the enriched dataset. The matched libraries are shown in Supplementary Table S3. 16.5dpc, 16.5 days post-coitum. fetal1st, 1st trimester embryos. fetal2nd, 2nd trimester embryos. Ox, oxidation treatment. Mili, Mili IP. Miwi, Miwi IP. Miwi2, Miwi2 IP.

NAs falling into the 0.2–0.5 count proportion range, significantly ($P < 0.05$) enriched piRNAs were also selected into the gold standard piRNA set.

DATABASE CONTENT

piRBase release v3.0 not only expanded on previous content and modules but also added some new content that could assist piRNA function studies more comprehensively. Given the large number of piRNAs, gold standard piRNA sets and piRNA clusters are provided in piRBase release v3.0. In addition to general information such as piRNA loci, piRNA targets and some basic annotations, more information on piRNAs is introduced into the current piRBase. Variants in piRNAs increase the diversity of piRNA sequences, and they may result in abnormal biological functions. Relevant information on splicing-junction piRNAs is shown on the detailed information page and is visualized in JBrowse (63). The collected piRNA-related data and the results processed by further analyses are released in v3.0.

New piRNA records and statistics

piRBase has the largest number of piRNA sequences and the most species compared to other similar databases. More than 181 million unique piRNA sequences from 44 species

are included in piRBase release v3.0, and the number of datasets reaches 440. Figure 2 shows the statistics for the data in the current piRBase. In Figure 2A, the number of piRNA sequences for each species is presented. Figure 2B and C shows the number of datasets and the length distribution of piRNA sequences from different species in piRBase release v3.0, respectively.

piRBase release v3.0 contains the most piRNA sequences compared with other piRNA-related databases. piRBase includes all piRNA sequences in piRNAQuest (64) and piRNABank (65) except for platypus, which has no piRNA sequences provided in the related literature. In addition, there are 16 datasets in piRNAdb (https://www.pirnadb.org/), 4 of which are not in the 440 datasets of piRBase because these 4 datasets do not meet our inclusion criteria.

Gold standard piRNA set

To get more representative piRNA set, we tried to define the gold standard piRNA set based on two features of piRNAs: binding to Piwi protein and 2′-*O*-methylation at the 3′ end. We collected matched high-throughput sequencing data from six species, including human, mouse, rat, *Drosophila melanogaster*, cow, and crab-eating macaque, and marked the enriched sequences in Piwi IP- or oxidation-treated libraries as the gold standard piRNA set.
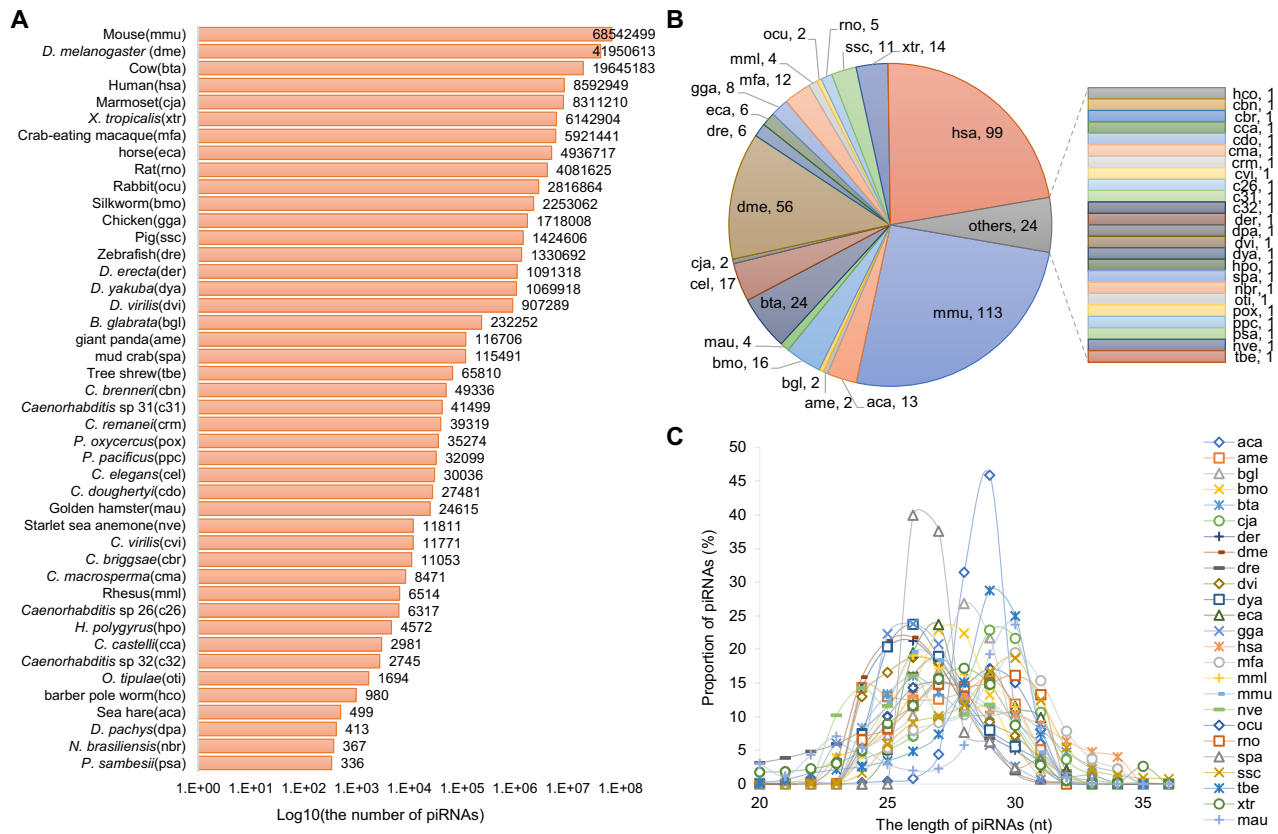
Figure 2. Statistics of piRNA data in piRBase release v3.0. (**A**) The number of piRNA sequences from each species. (**B**) Dataset composition of piRBase release v3.0. Three characters represent species (refer to Figure 2A), and the numbers represent the number of datasets. (**C**) The length distribution of piRNAs from each species.

Splicing-junction piRNAs

It was reported that some piRNA precursors contain introns (15). In piRBase release v3.0, piRNAs not mapped to the genome in the first alignment step are further mapped using known splicing junctions. Some unmapped piRNAs could be located based on splicing junctions (Figure 3A). The relevant information is presented in the 'Location' panel of the piRNA detailed information page. The genes and repeats overlapping with splicing-junction piRNAs are also shown in the same panel. In addition, users can browse information near a piRNA in JBrowse (63) by clicking the link on the location, such as gene annotation, adjacent piRNAs and so on.

piRNA clusters

A substantial fraction of piRNAs originating from genomic loci are termed piRNA clusters. We manually collected information on piRNA clusters from the related literature and assigned piRBase cluster IDs. In piRBase release v3.0, information on piRNA clusters of four species, including information on genomic location, strand, the number of piRNAs among the cluster region, genome version and species, is incorporated. The general information of a specific cluster and the piRNAs in the cluster are presented on the detailed information page. All the related piRNAs in this cluster can be browsed, and the detailed information page of each piRNA can be accessed through the piRNA name.

piRNAclusterDB (66) includes data for many small RNA sequences that are not specific to piRNA studies, and piRNA clusters are predicted using proTRAC (67), while the piRNA clusters in piRBase were collected from high-quality literature. piRBase integrated only the information of piRNA clusters from different literature sources.

piRNA variants

Single nucleotide polymorphisms (SNPs) are a common type of genomic variant, and many SNPs are associated with phenotypes and diseases (68). The updated piRBase version shows the variants in piRNAs based on the genome position, the type of base substitution, and the number of different piRNAs containing a specific variant. In the detailed information page of piRNA variants, the general information and the related piRNAs containing the variant are displayed. For human, piRNA variants are annotated by RS IDs with links to other SNP function annotation databases, such as dbSNP (62), HaploReg (69), eQTL (70), PancanQTL (71) and piRNA-eQTL (72).
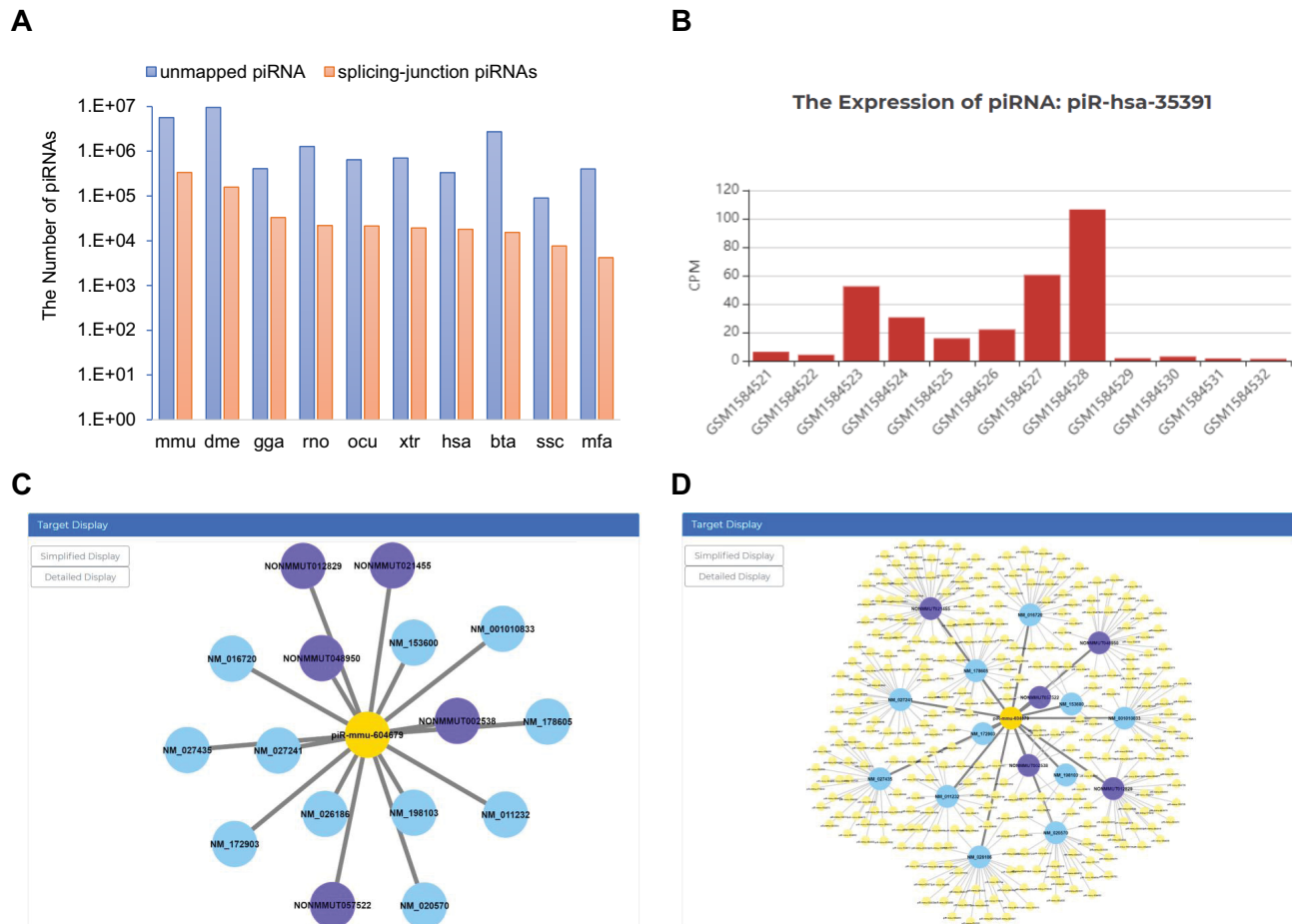
Figure 3. New content and visualizations in piRBase release v3.0. (**A**) The number of splicing-junction piRNAs in the unmapped piRNAs during the Bowtie alignment step. (**B**) An example of an expression profile for one piRNA in human. (**C, D**) piRNA regulatory relationships in simplified mode (C) and detailed mode (D).

piRNA expression profiles

Numerous studies have shown that piRNAs are not specific to the germline. They also exist in many somatic tissues (73,74). In piRBase release v3.0, the CPM expressions of piRNAs from five species (Supplementary Table S2) was calculated, and it can be visualized for specific piRNAs on the piRNA details page (Figure 3B). In addition, users can select samples of interest to calculate the *P* value of the differential expression of piRNAs. This data provides the basis for the study of piRNA functions in different tissues.

piRNA targets

In the previous version, mRNA targets of piRNAs in mouse, silkworm, *Caenorhabditis elegans* and predicted lncRNA targets in mouse testis (58,75) were included. In this update, more piRNA targets were extracted from the published literature and added to piRBase. The target information of five species was included. As described in the previous version, piRNA target records can be browsed by selecting a species and accessing links to the detailed information page of piRNAs by piRBase ID. Moreover, on the detailed information page, there is a network that visualizes

the regulatory relationships between a specific piRNA and its targets, which can be displayed in both simplified and detailed mode (Figure 3C, D).

piRNA/piRNA-like sRNA and disease

An increasing number of piRNA/piRNA-like sRNAs are closely associated with diseases (45). In piRBase release v3.0, we further manually collected not only cancer-, but also other types of disease-related piRNA/piRNA-like sRNAs from the related literature. 17 types of diseases (13 types of cancers, cardiovascular disease, stroke, Parkinson, and Alzheimer) were incorporated, and related information is exhibited on the corresponding pages.

Optimization of online tools

piRBase release v3.0 provides homepage search function. Users can search piRBase collected information by entering the piRBase ID. The 'Name Convert Tool' under the 'Tools' menu on the piRBase website allows one to convert piRNA names from not only the NCBI database (76) but also piRNABank (65), piRNAclusterDB (66), and other databases into piRBase names.

User interface and visualization

In piRBase release v3.0, the entire web interface has been redesigned to provide a better experience for users. All modules and functionality in previous versions have been kept and optimized. The new modules for piRNA variants and piRNA clusters have been added to the current version. In the detailed information page for each piRNA, there are panels for visualizing the expression of the piRNA in different datasets and the regulatory relationships between piRNAs and targets.

In the current piRBase, we used the JBrowse (63) genome browser to visualize piRNA positions and annotations in different genomic regions. JBrowse is fast and easily embedded into websites. All species whose genome and annotation files are available can be visualized by selecting the appropriate species in JBrowse. In addition, information such as piRNA variants, piRNA target sites and piRNA-related epigenetic data can be linked to JBrowse in corresponding pages.

CONCLUSION

piRBase is a manually curated resource for piRNAs, that focuses on piRNA function analyses. piRBase release v3.0 covers more species and unique piRNA sequences than the previous version. According to the characteristics of piRNAs, piRBase release v3.0 provides the gold standard piRNA sets. To further expand the research on piRNA functions, potential information on splicing-junction piRNAs and piRNA variants is included in piRBase release v3.0. As potential biomarkers in diseases, disease-related piRNA/piRNA-like sRNAs information is collected not only for cancers but also for other diseases, such as cardiovascular disease, stroke, Pankinson and Alzheimer's disease. In addition to the further collection of piRNA targets, piRBase release v3.0 visualizes the regulation of piRNA targets, which makes the regulatory network of piRNAs more intuitive. In addition, the current version also provides the expression profiles of piRNAs in various tissues at different developmental stages and supports research on piRNAs in different tissues. In addition, the web interface of piRBase release v3.0 has been redesigned to provide a better user experience. We believe that piRBase release v3.0 will make valuable contributions to the field of piRNA.

DATA AVAILABILITY

piRBase release v3.0 is free to access, browse, search and download at http://bigdata.ibp.ac.cn/piRBase. The source code of analyses has been provided as Supplementary file 1.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Girard,A., Sachidanandam,R., Hannon,G.J. and Carmell,M.A. (2006) A germline-specific class of small RNAs binds mammalian Piwi proteins. *Nature*, **442**, 199–202.
2. Aravin,A., Gaidatzis,D., Pfeffer,S., Lagos-Quintana,M., Landgraf,P., Iovino,N., Morris,P., Brownstein,M.J., Kuramochi-Miyagawa,S., Nakano,T. *et al.* (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature*, **442**, 203–207.
3. Ohara,T., Sakaguchi,Y., Suzuki,T., Ueda,H., Miyauchi,K. and Suzuki,T. (2007) The 3′ termini of mouse Piwi-interacting RNAs are 2′-O-methylated. *Nat. Struct. Mol. Biol.*, **14**, 349–350.
4. Kirino,Y. and Mourelatos,Z. (2007) Mouse Piwi-interacting RNAs are 2′-O-methylated at their 3′ termini. *Nat. Struct. Mol. Biol.*, **14**, 347–348.
5. Lau,N.C., Seto,A.G., Kim,J., Kuramochi-Miyagawa,S., Nakano,T., Bartel,D.P. and Kingston,R.E. (2006) Characterization of the piRNA complex from rat testes. *Science*, **313**, 363–367.
6. Yang,Q., Li,R., Lyu,Q., Hou,L., Liu,Z., Sun,Q., Liu,M., Shi,H., Xu,B., Yin,M. *et al.* (2019) Single-cell CAS-seq reveals a class of short PIWI-interacting RNAs in human oocytes. *Nat. Commun.*, **10**, 3389.
7. Ishino,K., Hasuwa,H., Yoshimura,J., Iwasaki,Y.W., Nishihara,H., Seki,N.M., Hirano,T., Tsuchiya,M., Ishizaki,H., Masuda,H. *et al.* (2021) Hamster PIWI proteins bind to piRNAs with stage-specific size variations during oocyte maturation. *Nucleic Acids Res.*, **49**, 2700–2720.
8. Aravin,A.A., Sachidanandam,R., Bourc'his,D., Schaefer,C., Pezic,D., Toth,K.F., Bestor,T. and Hannon,G.J. (2008) A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell*, **31**, 785–799.
9. Aravin,A.A., Sachidanandam,R., Girard,A., Fejes-Toth,K. and Hannon,G.J. (2007) Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science*, **316**, 744–747.
10. Ipsaro,J.J., Haase,A.D., Knott,S.R., Joshua-Tor,L. and Hannon,G.J. (2012) The structural biochemistry of Zucchini implicates it as a nuclease in piRNA biogenesis. *Nature*, **491**, 279–283.
11. Nishimasu,H., Ishizu,H., Saito,K., Fukuhara,S., Kamatani,M.K., Bonnefond,L., Matsumoto,N., Nishizawa,T., Nakanaga,K., Aoki,J. *et al.* (2012) Structure and function of Zucchini endoribonuclease in piRNA biogenesis. *Nature*, **491**, 284–287.
12. Mohn,F., Handler,D. and Brennecke,J. (2015) Noncoding RNA. piRNA-guided slicing specifies transcripts for Zucchini-dependent, phased piRNA biogenesis. *Science*, **348**, 812–817.
13. Han,B.W., Wang,W., Li,C., Weng,Z. and Zamore,P.D. (2015) Noncoding RNA. piRNA-guided transposon cleavage initiates Zucchini-dependent, phased piRNA production. *Science*, **348**, 817–821.
14. Gainetdinov,I., Colpan,C., Arif,A., Cecchini,K. and Zamore,P.D. (2018) A single mechanism of biogenesis, initiated and directed by PIWI proteins, explains piRNA production in most animals. *Mol. Cell*, **71**, 775–790.

15. Li,X.Z., Roy,C.K., Dong,X., Bolcun-Filas,E., Wang,J., Han,B.W., Xu,J., Moore,M.J., Schimenti,J.C., Weng,Z. *et al.* (2013) An ancient transcription factor initiates the burst of piRNA production during early meiosis in mouse testes. *Mol. Cell*, **50**, 67–81.

16. Ozata,D.M., Gainetdinov,I., Zoch,A., O'Carroll,D. and Zamore,P.D. (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.*, **20**, 89–108.

17. Vagin,V.V., Sigova,A., Li,C., Seitz,H., Gvozdev,V. and Zamore,P.D. (2006) A distinct small RNA pathway silences selfish genetic elements in the germline. *Science*, **313**, 320–324.

18. Sienski,G., Donertas,D. and Brennecke,J. (2012) Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. *Cell*, **151**, 964–980.

19. Pezic,D., Manakov,S.A., Sachidanandam,R. and Aravin,A.A. (2014) piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes Dev.*, **28**, 1410–1428.

20. Teixeira,F.K., Okuniewska,M., Malone,C.D., Coux,R.X., Rio,D.C. and Lehmann,R. (2017) piRNA-mediated regulation of transposon alternative splicing in the soma and germ line. *Nature*, **552**, 268–272.

21. Yu,Y., Gu,J., Jin,Y., Luo,Y., Preall,J.B., Ma,J., Czech,B. and Hannon,G.J. (2015) Panoramix enforces piRNA-dependent cotranscriptional silencing. *Science*, **350**, 339–342.

22. Czech,B., Munafo,M., Ciabrelli,F., Eastwood,E.L., Fabry,M.H., Kneuss,E. and Hannon,G.J. (2018) piRNA-guided genome defense: from biogenesis to silencing. *Annu. Rev. Genet.*, **52**, 131–157.

23. Houwing,S., Kamminga,L.M., Berezikov,E., Cronembold,D., Girard,A., van den Elst,H., Filippov,D.V., Blaser,H., Raz,E., Moens,C.B. *et al.* (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell*, **129**, 69–82.

24. Ku,H.Y., Gangaraju,V.K., Qi,H., Liu,N. and Lin,H. (2016) Tudor-SN interacts with Piwi antagonistically in regulating spermatogenesis but synergistically in silencing transposons in Drosophila. *PLos Genet.*, **12**, e1005813.

25. De Fazio,S., Bartonicek,N., Di Giacomo,M., Abreu-Goodger,C., Sankar,A., Funaya,C., Antony,C., Moreira,P.N., Enright,A.J. and O'Carroll,D. (2011) The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature*, **480**, 259–263.

26. Reuter,M., Berninger,P., Chuma,S., Shah,H., Hosokawa,M., Funaya,C., Antony,C., Sachidanandam,R. and Pillai,R.S. (2011) Miwi catalysis is required for piRNA amplification-independent LINE1 transposon silencing. *Nature*, **480**, 264–267.

27. Kiuchi,T., Koga,H., Kawamoto,M., Shoji,K., Sakai,H., Arai,Y., Ishihara,G., Kawaoka,S., Sugano,S., Shimada,T. *et al.* (2014) A single female-specific piRNA is the primary determiner of sex in the silkworm. *Nature*, **509**, 633–636.

28. Tang,W., Seth,M., Tu,S., Shen,E.Z., Li,Q., Shirayama,M., Weng,Z. and Mello,C.C. (2018) A sex chromosome piRNA promotes robust dosage compensation and sex determination in *C. elegans*. *Dev. Cell*, **44**, 762–770.

29. Watanabe,T., Cheng,E.C., Zhong,M. and Lin,H. (2015) Retrotransposons and pseudogenes regulate mRNAs and lncRNAs via the piRNA pathway in the germline. *Genome Res.*, **25**, 368–380.

30. Zhang,D., Tu,S., Stubna,M., Wu,W.S., Huang,W.C., Weng,Z. and Lee,H.C. (2018) The piRNA targeting rules and the resistance to piRNA silencing in endogenous genes. *Science*, **359**, 587–592.

31. Zhang,P., Kang,J.Y., Gou,L.T., Wang,J., Xue,Y., Skogerboe,G., Dai,P., Huang,D.W., Chen,R., Fu,X.D. *et al.* (2015) MIWI and piRNA-mediated cleavage of messenger RNAs in mouse testes. *Cell Res.*, **25**, 193–207.

32. Gou,L.T., Dai,P., Yang,J.H., Xue,Y., Hu,Y.P., Zhou,Y., Kang,J.Y., Wang,X., Li,H., Hua,M.M. *et al.* (2014) Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res.*, **24**, 680–700.

33. Rouget,C., Papin,C., Boureux,A., Meunier,A.C., Franco,B., Robine,N., Lai,E.C., Pelisson,A. and Simonelig,M. (2010) Maternal mRNA deadenylation and decay by the piRNA pathway in the early Drosophila embryo. *Nature*, **467**, 1128–1132.

34. Vourekas,A., Alexiou,P., Vrettos,N., Maragkakis,M. and Mourelatos,Z. (2016) Sequence-dependent but not sequence-specific piRNA adhesion traps mRNAs to the germ plasm. *Nature*, **531**, 390–394.

35. Lee,H.C., Gu,W., Shirayama,M., Youngman,E., Conte,D. Jr. and Mello,C.C. (2012) C. elegans piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell*, **150**, 78–87.

36. Bagijn,M.P., Goldstein,L.D., Sapetschnig,A., Weick,E.M., Bouasker,S., Lehrbach,N.J., Simard,M.J. and Miska,E.A. (2012) Function, targets, and evolution of Caenorhabditis elegans piRNAs. *Science*, **337**, 574–578.

37. Shen,E.Z., Chen,H., Ozturk,A.R., Tu,S., Shirayama,M., Tang,W., Ding,Y.H., Dai,S.Y., Weng,Z. and Mello,C.C. (2018) Identification of piRNA binding sites reveals the argonaute regulatory landscape of the *C. elegans* germline. *Cell*, **172**, 937–951.

38. Wang,C. and Lin,H. (2021) Roles of piRNAs in transposon and pseudogene regulation of germline mRNAs and lncRNAs. *Genome Biol.*, **22**, 27.

39. Cheng,J., Guo,J.M., Xiao,B.X., Miao,Y., Jiang,Z., Zhou,H. and Li,Q.N. (2011) piRNA, the new non-coding RNA, is aberrantly expressed in human cancer cells. *Clin. Chim. Acta*, **412**, 1621–1625.

40. Chu,H., Hui,G., Yuan,L., Shi,D., Wang,Y., Du,M., Zhong,D., Ma,L., Tong,N., Qin,C. *et al.* (2015) Identification of novel piRNAs in bladder cancer. *Cancer Lett.*, **356**, 561–567.

41. Zhang,H., Ren,Y., Xu,H., Pang,D., Duan,C. and Liu,C. (2013) The expression of stem cell protein Piwil2 and piR-932 in breast cancer. *Surg. Oncol.*, **22**, 217–223.

42. Cheng,J., Deng,H., Xiao,B., Zhou,H., Zhou,F., Shen,Z. and Guo,J. (2012) piR-823, a novel non-coding small RNA, demonstrates in vitro and in vivo tumor suppressive activity in human gastric cancer cells. *Cancer Lett.*, **315**, 12–17.

43. Muller,S., Raulefs,S., Bruns,P., Afonso-Grunz,F., Plotner,A., Thermann,R., Jager,C., Schlitter,A.M., Kong,B., Regel,I. *et al.* (2015) Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. *Mol. Cancer*, **14**, 94.

44. Mai,D., Ding,P., Tan,L., Zhang,J., Pan,Z., Bai,R., Li,C., Li,M., Zhou,Y., Tan,W. *et al.* (2018) PIWI-interacting RNA-54265 is oncogenic and a potential therapeutic target in colorectal adenocarcinoma. *Theranostics*, **8**, 5213–5230.

45. Liu,Y., Dou,M., Song,X., Dong,Y., Liu,S., Liu,H., Tao,J., Li,W., Yin,X. and Xu,W. (2019) The emerging role of the piRNA/piwi complex in cancer. *Mol. Cancer*, **18**, 123.

46. Yan,Z., Hu,H.Y., Jiang,X., Maierhofer,V., Neb,E., He,L., Hu,Y., Hu,H., Li,N., Chen,W. *et al.* (2011) Widespread expression of piRNA-like molecules in somatic tissues. *Nucleic Acids Res.*, **39**, 6596–6607.

47. Lu,Y., Li,C., Zhang,K., Sun,H., Tao,D., Liu,Y., Zhang,S. and Ma,Y. (2010) Identification of piRNAs in Hela cells by massive parallel sequencing. *BMB Rep*, **43**, 635–641.

48. Fu,A., Jacobs,D.I., Hoffman,A.E., Zheng,T. and Zhu,Y. (2015) PIWI-interacting RNA 021285 is involved in breast tumorigenesis possibly by remodeling the cancer epigenome. *Carcinogenesis*, **36**, 1094–1102.

49. Qiu,W., Guo,X., Lin,X., Yang,Q., Zhang,W., Zhang,Y., Zuo,L., Zhu,Y., Li,C.R., Ma,C. *et al.* (2017) Transcriptome-wide piRNA profiling in human brains of Alzheimer's disease. *Neurobiol. Aging*, **57**, 170–177.

50. Jain,G., Stuendl,A., Rao,P., Berulava,T., Pena Centeno,T., Kaurani,L., Burkhardt,S., Delalle,I., Kornhuber,J., Hull,M. *et al.* (2019) A combined miRNA-piRNA signature to detect Alzheimer's disease. *Translational psychiatry*, **9**, 250.

51. Vella,S., Gallo,A., Lo Nigro,A., Galvagno,D., Raffa,G.M., Pilato,M. and Conaldi,P.G. (2016) PIWI-interacting RNA (piRNA) signatures in human cardiac progenitor cells. *Int. J. Biochem. Cell Biol.*, **76**, 1–11.

52. Rajan,K.S., Velmurugan,G., Pandi,G. and Ramasamy,S. (2014) miRNA and piRNA mediated Akt pathway in heart: antisense expands to survive. *Int. J. Biochem. Cell Biol.*, **55**, 153–156.

53. Li,M., Yang,Y., Wang,Z., Zong,T., Fu,X., Aung,L.H.H., Wang,K., Wang,J.X. and Yu,T. (2021) Piwi-interacting RNAs (piRNAs) as potential biomarkers and therapeutic targets for cardiovascular diseases. *Angiogenesis*, **24**, 19–34.

54. Zhang,H., Zhang,F., Chen,Q., Li,M., Lv,X., Xiao,Y., Zhang,Z., Hou,L., Lai,Y., Zhang,Y. *et al.* (2021) The piRNA pathway is essential for generating functional oocytes in golden hamsters. *Nat. Cell Biol.*, **23**, 1013–1022.

55. Hasuwa,H., Iwasaki,Y.W., Au Yeung,W.K., Ishino,K., Masuda,H., Sasaki,H. and Siomi,H. (2021) Production of functional oocytes requires maternally expressed PIWI genes and piRNAs in golden hamsters. *Nat. Cell Biol.*, **23**, 1002–1012.

56. Loubalova,Z., Fulka,H., Horvat,F., Pasulka,J., Malik,R., Hirose,M., Ogura,A. and Svoboda,P. (2021) Formation of spermatogonia and fertile oocytes in golden hamsters requires piRNAs. *Nat. Cell Biol.*, **23**, 992–1001.

57. The RNAcentral Consortium (2019) RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res.*, **47**, D1250–D1251.

58. Wang,J., Zhang,P., Lu,Y., Li,Y., Zheng,Y., Kan,Y., Chen,R. and He,S. (2019) piRBase: a comprehensive database of piRNA sequences. *Nucleic Acids Res.*, **47**, D175–D180.

59. Barrett,T., Wilhite,S.E., Ledoux,P., Evangelista,C., Kim,I.F., Tomashevsky,M., Marshall,K.A., Phillippy,K.H., Sherman,P.M., Holko,M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets–update. *Nucleic Acids Res.*, **41**, D991–D995.

60. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

61. Dobin,A., Davis,C.A., Schlesinger,F., Drenkow,J., Zaleski,C., Jha,S., Batut,P., Chaisson,M. and Gingeras,T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.

62. Sherry,S.T., Ward,M.H., Kholodov,M., Baker,J., Phan,L., Smigielski,E.M. and Sirotkin,K. (2001) dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.*, **29**, 308–311.

63. Buels,R., Yao,E., Diesh,C.M., Hayes,R.D., Munoz-Torres,M., Helt,G., Goodstein,D.M., Elsik,C.G., Lewis,S.E., Stein,L. *et al.* (2016) JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.*, **17**, 66.

64. Sarkar,A., Maji,R.K., Saha,S. and Ghosh,Z. (2014) piRNAQuest: searching the piRNAome for silencers. *BMC Genomics*, **15**, 555.

65. Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, **36**, D173–D177.

66. Rosenkranz,D., Zischler,H. and Gebert,D. (2021) piRNAclusterDB 2.0: update and expansion of the piRNA cluster database. *Nucleic Acids Res.*, https://doi.org/10.1093/nar/gkab622.

67. Rosenkranz,D. and Zischler,H. (2012) proTRAC–a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*, **13**, 5.

68. Visscher,P.M., Wray,N.R., Zhang,Q., Sklar,P., McCarthy,M.I., Brown,M.A. and Yang,J. (2017) 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.*, **101**, 5–22.

69. Ward,L.D. and Kellis,M. (2016) HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.*, **44**, D877–D881.

70. Keen,J.C. and Moore,H.M. (2015) The Genotype-Tissue Expression (GTEx) project: linking clinical data with molecular analysis to advance personalized medicine. *J Pers Med*, **5**, 22–29.

71. Gong,J., Mei,S., Liu,C., Xiang,Y., Ye,Y., Zhang,Z., Feng,J., Liu,R., Diao,L., Guo,A.Y. *et al.* (2018) PancanQTL: systematic identification of cis-eQTLs and trans-eQTLs in 33 cancer types. *Nucleic Acids Res.*, **46**, D971–D976.

72. Xin,J., Du,M., Jiang,X., Wu,Y., Ben,S., Zheng,R., Chu,H., Li,S., Zhang,Z. and Wang,M. (2021) Systematic evaluation of the effects of genetic variants on PIWI-interacting RNA expression across 33 cancer types. *Nucleic Acids Res.*, **49**, 90–97.

73. Malone,C.D., Brennecke,J., Dus,M., Stark,A., McCombie,W.R., Sachidanandam,R. and Hannon,G.J. (2009) Specialized piRNA pathways act in germline and somatic tissues of the Drosophila ovary. *Cell*, **137**, 522–535.

74. Li,C., Vagin,V.V., Lee,S., Xu,J., Ma,S., Xi,H., Seitz,H., Horwich,M.D., Syrzycka,M., Honda,B.M. *et al.* (2009) Collapse of germline piRNAs in the absence of Argonaute3 reveals somatic piRNAs in flies. *Cell*, **137**, 509–521.

75. Zhang,P., Si,X., Skogerbo,G., Wang,J., Cui,D., Li,Y., Sun,X., Liu,L., Sun,B., Chen,R. *et al.* (2014) piRBase: a web resource assisting piRNA functional study. *Database (Oxford)*, **2014**, bau110.

76. Benson,D.A., Cavanaugh,M., Clark,K., Karsch-Mizrachi,I., Ostell,J., Pruitt,K.D. and Sayers,E.W. (2018) GenBank. *Nucleic Acids Res.*, **46**, D41–D47.