1 **piRPheno: a manually curated database to prioritize and** 2 **analyze human disease related piRNAs**

3 WENLIANG ZHANG ^{1,2,6}, SONG WU ³, HAIYUE ZHANG ³, WEN GUAN ⁴, BINGHUI
4 ZENG ⁵, YANJIE WEI ², GODFREY CHI-FUNG CHAN ^{1,6}, and WEIZHONG LI ^{3,6}

5 ¹ Department of Pediatrics, The University of Hong Kong-Shenzhen Hospital,
6 Shenzhen, 518058, China

7 ² Center for High Performance Computing, Joint Engineering Research Center for
8 Health Big Data Intelligent Analysis Technology, Shenzhen Institutes of Advanced
9 Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, 518055,
10 People's Republic of China.

11 ³ Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510080, China

12 ⁴ Guangdong Key Laboratory of Animal Conservation and Resource
13 Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou,
14 510260.

15 ⁵ Guanghua School of Stomatology, Hospital of Stomatology, Guangdong Provincial
16 Key Laboratory of Stomatology, Sun Yat-sen University, Guangzhou 510055, China.

17 ⁶ Correspondence authors: zhangwl25@mail3.sysu.edu.cn; gcfchan@hku.hk;
18 liweizhong@mail.sysu.edu.cn.

19 **Running title:** piRPheno to prioritize disease related piRNAs.

21   **ABSTRACT**

22   Many studies have uncovered that piRNAs (PIWI-interacting RNA) are associated
 23   with a broad range of diseases and might be a novel type of biomarkers and targets
 24   for precision medicine. However, public resource of high-quality curated human
 25   disease-associated piRNAs remains unavailable. Therefore, we developed the
 26   piRPheno (<http://www.biomedical-web.com/pirpheno>) database to provide an
 27   up-to-date, interactive and extensible data reference for human piRNA-disease
 28   associations. piRPheno includes 9057 experimentally supported associations
 29   between 474 piRNAs and 204 diseases through a manual curation of publications. To
 30   prioritize the piRNA-disease associations, each association in piRPheno is assigned
 31   with a confidence score and clinical correlations based on the experimentally
 32   supported evidences. piRPheno is freely available with user-friendly interface and
 33   novel applications to enable easy exploration and analysis of the human disease
 34   related piRNAs.

35

36   **Keywords:** piRNA, disease phenotype, database, cancer, nervous system disease,
 37   reproductive system disease

38

39 INTRODUCTION

40 PIWI-interacting RNAs (piRNAs) are an animal-specific class of small silencing RNAs
 41 with 21-35 nucleotides in length (Ozata et al., 2019), distinct from microRNAs
 42 (miRNAs) and small interfering RNAs (siRNAs). piRNAs bear 2'-O-methyl-modified 3'
 43 termini and guide PIWI-clade Argonautes (PIWI proteins), while miRNAs and siRNAs
 44 bear AGO-clade proteins involving in gene silencing pathway (Ozata et al., 2019).
 45 piRNAs can guide PIWI proteins to cleave target RNAs, methylate DNA
 46 (Kuramochi-Miyagawa et al., 2008), and promote heterochromatin assembly.
 47 Moreover, due to the architecture of piRNA signaling pathways, piRNAs are able to
 48 regulate expression of conserved host genes, and also provide adaptive immunity
 49 and sequence-based immunity (Ernst et al., 2017; Ozata et al., 2019).

50 With the advances of high throughput sequencing technologies and
 51 bioinformatics methods, many piRNAs have been identified and thus the piRNAs data
 52 are accumulating rapidly into computational resources, such as piRBase (Wang et al.,
 53 2019; Zhang et al., 2014), piRNABank (Sai and Agrawal, 2008), piRNA cluster
 54 (Rosenkranz, 2016), piRNAQuest (Sarkar et al., 2014), COMPSRA(Li et al., 2020),
 55 piRTarBase (Wu et al., 2019), PingPongPro (Uhrig and Klein, 2019), pirScan (Wu et
 56 al., 2018), and IsopiRBank (Zhang et al., 2018). Most of these resources focus on
 57 systematically integrating various piRNA associated data to support piRNA functional
 58 analysis, biological annotations, and expression profiling. Recently, as piRNAs are
 59 implicated in transposon and host gene regulation, many studies have uncovered that

60 piRNA dysfunctions are associated with a broad range of human diseases, such as
61 various cancers (Lee et al., 2016; Mei et al., 2015; Moyano and Stefani, 2015),
62 nervous system disorders (Millan, 2017; Qiu et al., 2017; Roy et al., 2017), and
63 reproductive system disease (Hong et al., 2016).

64 piRNAs could be a novel type of potential biomarkers and targets for human
65 disease diagnosis, therapy, and prognosis (Millan, 2017; Moyano and Stefani, 2015;
66 Romano et al., 2017). Public resource of manually curated human disease-associated
67 piRNAs remains unavailable. Therefore, in the end of 2018, we started to develop the
68 piRPheno database, which manually curated piRNA-disease phenotype association
69 data from publications. Currently, piRPheno provides 9057 experimentally supported
70 associations between 474 piRNAs and 204 human diseases. To prioritize the
71 piRNA-disease associations, each association in piRPheno is assigned with a
72 confidence score and clinical correlation base on the experimentally supported
73 evidences. In order to enable users exploration and application of the piRNA-disease
74 association data easily, piRPheno (<http://www.biomedical-web.com/pirpheno/>)
75 provides user-friendly interface and novel visualizations to prioritize and analyze
76 disease related piRNAs online.

77 **RESULTS**

78 **Data contents**

79 piRPheno provides 9057 experimentally supported associations among 474 piRNAs,
80 204 disease phenotypes, 26 targets, 16 pathways and 2 treatments. 605 of the

81 associations were manually curated from more than 200 publications and 8470 of
 82 them were derived by using the disease parent-child relationships in the Experimental
 83 Factor Ontology (EFO) resource (Malone et al., 2010). The methods of data curation
 84 and new association derivation are detailed in the “Materials and methods” section
 85 (Fig. 1). For piRNAs annotation, 97.5 % (462/474) of piRNAs are consistently and
 86 comprehensively annotated by piRBase (Fig. 1). For diseases annotation, all of
 87 diseases in piRPheno are annotated by the EFO resource (Fig. 1). The piRPheno
 88 database covers 6 disease subtypes associated with piRNAs dysregulation, including
 89 neoplasm, nervous system disease, reproductive system disease, respiratory system
 90 disease, skeletal system disease, and immune system disease. Other than piRNA
 91 dysregulation in expression dysregulation, piRPheno also offers 7 single nucleotide
 92 polymorphisms (SNPs) on piRNAs are associated with the risk of cancers.

FIGURE 1. The data curation and annotation framework of piRPheno.

95   **Search and Browse**

96   The piRPheno database provides user friendly, open access web interfaces and
 97   applications to enable users to search, browse, analyze, and prioritize the
 98   piRNA-disease association data, as well as to download and submit new associations
 99   for further integration (Fig. 1).

100   To promptly prioritize the piRNA-disease phenotype associations, the search and
 101   browse applications were developed in piRPheno. The applications allow users to
 102   quickly prioritize piRNA-disease associations through retrieving piRNA and disease
 103   phenotype. The search application facilitates smart assistance with keyword tips of
 104   expected piRNA and disease phenotype. The prioritizing association data is shown in
 105   a brief table, showing key information of association identifiers (IDs), piRNAs, disease
 106   phenotypes with ontology identifiers in EFO, confidence scores, and correlations (Fig.
 107   2A). In addition, the prioritizing data allows sorting by confidence scores and filtering
 108   by specific piRNA and disease phenotype (Fig. 2A). Moreover, the prioritizing data of
 109   a disease search can be optionally visualized in word-cloud diagrams (Fig. 2B), while
 110   the prioritizing data of a piRNA search can be optionally visualized in disease-tree and
 111   disease-network diagrams (Supplemental_Fig_S1.pptx). Larger sizes and more
 112   central locations of the symbols in the word-cloud diagrams indicate higher
 113   confidence scores between the piRNAs and disease phenotypes (Fig. 2B).
 114   Furthermore, the association IDs in the table, the piRNAs in the word-cloud diagrams,
 115   and the circle nodes in the disease-tree diagrams link to further information of the

116 association, piRNA, disease phenotype, and the supporting evidences in publications
 117 ([Supplemental_Fig_S2.pptx](#)). External links to other reference resources, such as
 118 piRBase, EFO, and the NCBI PubMed, and Gene database are also provided
 119 ([Supplemental_Fig_S2.pptx](#)).

120

121 **FIGURE 2.** The search and word-cloud application interfaces in piRPheno. (A) Searching
 122 results shows breast carcinoma related information in a tabular profile and provides a tabular
 123 profile to prioritize piRNAs dysregulation. The “Positive” means that the piRNAs are positively
 124 associated with the disease phenotypes, while the “Negative” means that the piRNAs are
 125 negatively associated with the disease phenotypes. The “Contradictory” means the
 126 piRNA-disease associations with conflicting evidences from different publications. (B) A
 127 word-cloud diagram displays the prioritized breast carcinoma related piRNAs. Larger sizes
 128 and more central locations of the piRNAs indicate a higher confidence score between the
 129 piRNAs and breast carcinoma.

130 piRNAs prioritization on disease related piRNAs datasets

131 A typical case-control microarray assay and piRNA sequencing can find hundreds of
 132 significant piRNA dysregulations, but identifying their clinical significance remains
 133 challenging. For example, Chu et al investigated and shown that the expression levels
 134 of 106 piRNAs were significantly up-regulated and 91 were significantly

135 down-regulated in bladder cancer tissues compared with their corresponding adjacent
136 tissues (Chu et al., 2015). However, how to promptly identify and prioritize the
137 experimentally validated bladder cancer-related piRNAs from these large-scale
138 piRNAs is not a trivial task. To copy with this challenge, a piRNAs prioritization
139 application was developed in piRPheno to analyze and prioritize experimental
140 validated disease phenotype related piRNAs from a set of piRNAs (Fig. 3). We upload
141 197 piRNAs with bladder cancer phenotype in the piRNAs prioritization application.
142 The application completed the analysis in a few seconds and shown that the piRNA
143 most significantly associated with bladder cancer is piR-hsa-24274. The result table
144 also allows data sorting based on confidence scores and data filtering by specific
145 piRNA (Fig. 3), and it provides links to further webpages for detailed information
146 (Supplemental_Fig_S2.pptx).

147
148 **FIGURE 3.** piRNAs prioritization application to promptly prioritize disease related piRNAs from
149 large scale dataset.

150 **Network visualization to explore the relationships between piRNAs and disease**

151 **phenotypes**

152 A network visualization application was developed in piRPheno to explore the
153 relationships between piRNAs and disease phenotypes. The application allows user
154 to input a set of piRNAs and disease phenotypes, and to generate interaction
155 networks to display the association data. For example, we entered “breast cancer,
156 lung cancer” in the input box and generated an interaction network to explore the
157 relationships between the two cancers (Fig. 4A). The network clearly indicates that
158 piR-hsa-1077 (Fig. 4A, blue box) is positively associated with the risk of lung cancer,
159 but it is negatively associated with breast cancer and with conflicting evidence from
160 different publications. Similarly, we entered “piR-hsa-8674, piR-hsa-1963” in the input
161 box and generated an interaction network to explore the relationships between the
162 two piRNAs (Fig. 4B). Interestingly, the network clearly indicates that the two piRNAs
163 are both associated with cancer, but piR-hsa-8674 is positively associated with lung
164 cancer and piR-hsa-1963 is negatively associated with colorectal cancer (Fig. 4B).

165 **Data access and submission**

166 piRPheno provides web service APIs for programmatically access of the association
167 data. The resulted data through the APIs are available in the universal JSON formats.
168 Documentation for the use of APIs is available on the “web service” webpage. In
169 addition, all association data in piRPheno is freely available to downloaded and used.
170 Moreover, piRPheno encourages users to submit their new piRNA-disease

171 association data for future data integration. The submitted records will be checked by
172 our professional curators and approved by our submission review committee for the
173 future release. Furthermore, a detailed tutorial is available on the 'Help' webpage.

174

175 **FIGURE 4.** Network visualization to explore the relationships between piRNAs and disease
176 phenotypes. (A) An interaction network to explore the relationships among breast cancer, lung
177 cancer and their related piRNAs. (B) An interaction network to explore the relationships among
178 piR-hsa-8674, piR-hsa-1963 and their related diseases. The red lines indicate that the piRNAs
179 are positively associated with the disease phenotypes, while the green lines indicate that the
180 piRNAs are negatively associated with the disease phenotypes. The dash lines indicate the
181 piRNA-disease associations with conflicting evidences from different publications.

182 DISCUSSION

183 As many studies uncovered that piRNA dysfunctions are associated with a broad
184 range of human diseases, piRNAs is becoming a novel type of potential biomarkers
185 and targets for human disease diagnosis, therapy, and prognosis. Recently, many
186 piRNAs have been identified and several computational resources (Li et al., 2020;
187 Uhrig and Klein, 2019; Wang et al., 2019; Wu et al., 2019; Wu et al., 2018) have been
188 developed to systematically integrate various piRNA associated data to support
189 piRNA functional analysis. Compared with these resources, our piRPheno database
190 not only aims to provide comprehensive and up-to-date data of piRNAs-disease
191 phenotypes association, but also provides novel web applications to analyze and
192 prioritize disease related piRNAs.

193 The latest update of piRBase (Wang et al., 2019) has collated piRNA-cancer
194 associations from cancer related publications. Compared with piRBase, our piRPheno
195 database not only manually curates piRNA-disease associations from publications,
196 but also derives new associations from the manually curated associations by using
197 the EFO parent-child relationship data. The number of associations in piRPheno is
198 approximately 40-fold of those in piRBase (9057vs. 227). In addition, compared with
199 the latest piRBase, each association in piRPheno is assigned with a confidence score
200 and a clinical correlation base on the experimentally supporting evidences to prioritize
201 and interpret the RNA dysregulation. Importantly, piRPheno provides several novel
202 applications and visualizations to enable easy identification of piRNA dysregulation

203 associated with disease phenotypes for disease diagnosis and therapeutic
204 development, including piRNAs prioritization, disease-tree, word-cloud visualization,
205 and network visualizations.

206 The piRPheno is updated every 6 months to include new association data and
207 applications. We plan to enrich new association data by analyzing multi-omic data in
208 TCGA (Weinstein et al., 2013) and ICGC (Hudson et al., 2010), and integrate novel
209 bioinformatic tools for further analyzing the piRNA-disease associations in piRPheno.

210 **MATERIALS AND METHODS**

211 **Data collection and annotation**

212 As previously described (Li et al., 2014; Ning et al., 2016; Zhao et al., 2018), to obtain
213 the all available publications describing the associations between piRNAs and human
214 diseases, we made a query in the National Center for Biotechnology Information
215 (NCBI) PubMed database with the keywords of “(((Piwi-interacting RNA[Title/Abstract]
216 OR Piwi interacting RNA[Title/Abstract] OR piRNA[Title/Abstract]))) NOT
217 review[Publication Type] AND (Humans[Mesh])”. The query resulted in more than
218 200 publications (before November 2018). We downloaded all of these publications
219 and extracted experimentally supported piRNA-disease association data by manually
220 curation from these publications. Researchers were assigned to double-check all of
221 the collected piRNA-disease associations. In this step, we extracted the piRNA
222 symbol, disease name, experimental evidence, samples, NCBI PubMed ID (PMID),
223 dysfunction status, direct targets, pathway, and treatment (Fig. 1). The clinical

significances of piRNA dysfunctions and the experimentally supported evidence levels are also assigned for the piRNA-disease associations (Fig. 1). The clinical significances of piRNA dysregulations are consistently assigned to four status including decreasing risk, increasing risk, decreasing risk with good prognosis and increasing risk with poor prognosis. The assignment of experimentally supported evidence levels of each publication for the piRNA-disease association are shown in detail in Table 1.

To make the piRNA symbols and disease names consistent with other public databases, the piRBase database offers identifiers, piRNA sequences, alias and links for piRNAs (Fig. 1). Finally, we used a standardized classification scheme, the EFO resource (Malone et al., 2010) to annotate each disease. The annotations of diseases include official disease name, definition, EFO identifier, diseases parent-child relationships, and alternative names (Fig. 1).

Deriving new piRNA-disease associations

Referred to the Open Target Platform (Koscielny et al., 2017), we also used the EFO parent-child relationship data to derived new piRNA-disease associations, which may not have direct supporting publications, from known piRNA-disease associations with supporting publications (Fig. 1). For example, the non-small cell lung carcinoma and lung adenocarcinoma are both a lung carcinoma. The direct evidence of piRNAs associated to non-small cell lung carcinoma and lung adenocarcinoma are propagated to the higher level of lung carcinomas to allow users to find common

245 piRNAs across groups of related diseases. Other piRNA-disease associations can
246 also be derived based on EFO inferred-by-property classification: disease location
247 (e.g. brain, lung and colon) and disease phenotypes (e.g. azoospermia in male
248 infertility). These two approaches enable driving and propagating new piRNA-disease
249 associations.

250 **Confidence score**

251 To prioritize and interpret the piRNA dysregulations associated with different diseases
252 in piRPheno, a confidence score for each association is assigned in piRPheno based
253 on two evidential metrics. These evidential metrics include the evidential value in
254 publication (E_p) and the number of publications. The assignment of confidence score
255 consists of three steps:

256 **Step 1:** In principle, validation experiments of mechanism and functional
257 analyses provide more reliable evidence than throughput expression analyses. Based
258 on the validation experiments in publications, we defined and assigned the
259 experimentally supported evidence levels into six levels, as detailed in Table 1. The
260 evidential value in publication (E_p) for supporting piRNA-disease association is
261 empirically defined and calculated, as indicated in Table 1. D_i ($D_i \in \{-1, 1\}$) in Table 1
262 represents the changing direction of a piRNA associated with a disease. If a piRNA is
263 increased (or obtain function) in a disease, D_i equates to 1; if a piRNA is decreased
264 (or loss function) in a disease, D_i equates to -1.

265 **Step 2:** A large number of publications can enhance the evidential values (Score)

for supporting the same piRNA-disease association. To dampen the effect of large number of publications, a harmonic sum function (Hagen, 2008; Koscielny et al., 2017) was used to account Score and abs_Score. The Score and abs_Score are respectively calculating as following equation:

$$\text{Score} = E_{p1} + E_{p2} / 2 + E_{p3} / 3 + \dots + E_{pn} / n \quad (1)$$

$$\text{abs\_Score} = |E_{p1}| + |E_{p2} / 2| + |E_{p3} / 3| + \dots + |E_{pn} / n| \quad (2)$$

As indicated in equation (1) and (2), “*n*” is the total number of supporting publications, and E_{p1} , E_{p2} , E_{p3} , ..., E_{pn} are the sorted evidential values of different supporting publications in descending order. The Score of an association less than zero indicates that the piRNA dysfunction is negatively associated with the development of disease, and thus the clinical correlation of the association was assigned as “Negative”. On the contrary, the Score of an association greater than zero indicates that the piRNA dysfunction is positively associated with the development of disease, and thus the clinical correlation of the association was assigned with “Positive”. In addition, if the absolute of Score ($|\text{Score}|$) of an association is less than the abs_Score of the association, the clinical correlation of the association was assigned with “Contradictory”. The assignment of “Contradictory” means that the association have conflicting evidence supported.

Step 3: The Score above was normalized to limit the range of confidence score from 0 to 1.0.

$$\text{Confidence score} = 1 - \frac{1}{e^{|Score|}} \quad (3)$$

In equation [3], 'e' represents the natural constant e.

Web implementation

The piRPheno website was built with the technologies of Spring MVC and jQuery AJAX framework. Data in piRPheno were organized into a local MySQL database. The programs for data processing were written in Java. The web interface was built by using JavaScript, HTML5, and CSS3. The D3.js widget (<http://d3js.org/d3.v3.min.js>) and The vis.js widget (<http://www.visjs.org>) were implemented to display disease-tree visualization and networks on the webpages, respectively. The web service is deployed to an Apache Tomcat web server. .

DATA DEPOSITION

The data in piRPheno is available at <http://www.biomedical-web.com/pirpheno>.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article at <http://www.biomedical-web.com/pirpheno/suppl.jsp>.

ACKNOWLEDGMENTS

This work has been supported by Sanming Project of Medicine (Shenzhen) [SZSM201911016], the National Key R&D Program of China [2016YFC0901604 & 2018YFC0910401], the National Natural Science Foundation of China [31771478],

305 the Fundamental Research Funds for the Central Universities, Sun Yat-sen University
306 (No.19ykpy86), and the China Postdoctoral Science Foundation (No.2020M673023).

307 **DISCLOSURE STATEMENT**

308 No potential conflict of interest was reported by the authors.

309 **Table 1.** The assignment of experimentally supported evidence levels and the
310 calculation of evidential values

Experimental evidences in publications	Ep
Only different expression analysis has been screened to support the piRNA-disease associations by using high throughput technologies such as RNA-seq, microarray, SNP array etc.	0.05 * D _i
The screened significance different expression piRNA has been confirmed by qPCR or RT-PCR etc.	0.1 * D _i
Having two situations: (1) Knockdown or overexpression of the piRNA has been conducted to verify the impacts of the piRNA in cellular physiology (cell viability, cell proliferation and cell apoptosis) in the disease cell line(s) or tissue(s). (2) Knockdown or overexpression of the piRNA has been conducted to verify the impacts of the piRNA in molecular signaling pathway in the disease cell line(s) or tissue(s).	0.2 * D _i
Knockdown or overexpression of the piRNA has been conducted to verify the impacts of the piRNA in cellular physiology and in molecular signaling pathway using the disease cell line(s) or tissue(s).	0.4 * D _i
Having two situations: (1) Knockdown or overexpression of the piRNA affects the cellular physiology and molecular signaling pathway in the disease cell line(s) or tissue(s). And the dependent target of the piRNA is discovered involving in the development of the disease by further complex experiments. (2) Not having level 3 or 4 evidences, the physiological function of the	0.6 * D _i

---

piRNA is documented by organism models in *vivo* such as xenograft or transgenic model, etc.

Other than level 3 or 4 evidences, the physiological function of the piRNA $0.8 * D_i$ is documented by organism models in *vivo* such as xenograft or transgenic model etc.

---

311 The E_p defined as the evidential value of each publication for supporting the piRNA
 312 dysregulations associated with diseases based on the assignation of evidence levels
 313 in publications. Where D_i ($D_i \in \{-1,1\}$) represents the changing direction of a piRNA
 314 associated with a disease. If a piRNA expression is increased (or obtain function) in a
 315 disease, $D_i = 1$; if a piRNA expression is decreased (or loss function) in a disease, $D_i =$
 316 -1 .

317

318

319

320 References

- 321 Chu H, Hui G, Yuan L, Shi D, Wang Y, Du M, Zhong D, Ma L, Tong N, Qin C, et al. 2015. Identification
322 of novel piRNAs in bladder cancer. *Cancer Lett* **356**: 561-7.doi:10.1016/j.canlet.2014.10.004
- 323 Ernst C, Odom DT, Kutter C. 2017. The emergence of piRNAs against transposon invasion to preserve
324 mammalian genome integrity. *Nat Commun* **8**: 1411.doi:10.1038/s41467-017-01049-7
- 325 Hagen NT. 2008. Harmonic allocation of authorship credit: source-level correction of bibliometric bias
326 assures accurate publication and citation analysis. *Plos One* **3**: e4021.doi:10.1371/journal.pone.0004021
- 327 Hong Y, Wang C, Fu Z, Liang H, Zhang S, Lu M, Sun W, Ye C, Zhang CY, Zen K, et al. 2016. Systematic
328 characterization of seminal plasma piRNAs as molecular biomarkers for male infertility. *Sci Rep* **6**:
329 24229.doi:10.1038/srep24229
- 330 Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, Bernabe RR, Bhan MK, Calvo F, Eerola I,
331 Gerhard DS, et al. 2010. International network of cancer genome projects. *Nature* **464**:
332 993-8.doi:10.1038/nature08987
- 333 Koscielny G, An P, Carvalho-Silva D, Cham JA, Fumis L, Gasparyan R, Hasan S, Karamanis N, Maguire
334 M, Papa E, et al. 2017. Open Targets: a platform for therapeutic target identification and validation.
335 *Nucleic Acids Res* **45**: D985-D994.doi:10.1093/nar/gkw1055
- 336 Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, Asada N, Kojima K,
337 Yamaguchi Y, Ijiri TW, et al. 2008. DNA methylation of retrotransposon genes is regulated by Piwi
338 family members MILI and MIWI2 in murine fetal testes. *Genes Dev* **22**:
339 908-17.doi:10.1101/gad.1640708
- 340 Lee YJ, Moon SU, Park MG, Jung WY, Park YK, Song SK, Ryu JG, Lee YS, Heo HJ, Gu HN, et al. 2016.
341 Multiplex bioimaging of piRNA molecular pathway-regulated theragnostic effects in a single breast
342 cancer cell using a piRNA molecular beacon. *Biomaterials* **101**:
343 143-55.doi:10.1016/j.biomaterials.2016.05.052
- 344 Li J, Kho AT, Chase RP, Pantano L, Farnam L, Amr SS, Tantisira KG. 2020. COMPSRA: a
345 COMprehensive Platform for Small RNA-Seq data Analysis. *Sci Rep* **10**:
346 4552.doi:10.1038/s41598-020-61495-0
- 347 Li Y, Qiu C, Tu J, Geng B, Yang J, Jiang T, Cui Q. 2014. HMDD v2.0: a database for experimentally
348 supported human microRNA and disease associations. *Nucleic Acids Res* **42**:
349 D1070-4.doi:10.1093/nar/gkt1023
- 350 Malone J, Holloway E, Adamusiak T, Kapushesky M, Zheng J, Kolesnikov N, Zhukova A, Brazma A,

351 Parkinson H. 2010. Modeling sample variables with an Experimental Factor Ontology. *Bioinformatics* **26**:
352 1112-8.doi:10.1093/bioinformatics/btq099

353 Mei Y, Wang Y, Kumari P, Shetty AC, Clark D, Gable T, MacKerell AD, Ma MZ, Weber DJ, Yang AJ, et
354 al. 2015. A piRNA-like small RNA interacts with and modulates p-ERM proteins in human somatic cells.
355 *Nat Commun* **6**: 7316.doi:10.1038/ncomms8316

356 Millan MJ. 2017. Linking deregulation of non-coding RNA to the core pathophysiology of Alzheimer's
357 disease: An integrative review. *Prog Neurobiol* **156**: 1-68.doi:10.1016/j.pneurobio.2017.03.004

358 Moyano M, Stefani G. 2015. piRNA involvement in genome stability and human cancer. *J Hematol*
359 *Oncol* **8**: 38.doi:10.1186/s13045-015-0133-5

360 Ning S, Zhang J, Wang P, Zhi H, Wang J, Liu Y, Gao Y, Guo M, Yue M, Wang L, et al. 2016. Lnc2Cancer:
361 a manually curated database of experimentally supported lncRNAs associated with various human
362 cancers. *Nucleic Acids Res* **44**: D980-5.doi:10.1093/nar/gkv1094

363 Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, Zamore PD. 2019. PIWI-interacting RNAs: small RNAs
364 with big functions. *Nat Rev Genet* **20**: 89-108.doi:10.1038/s41576-018-0073-3

365 Qiu W, Guo X, Lin X, Yang Q, Zhang W, Zhang Y, Zuo L, Zhu Y, Li CR, Ma C, et al. 2017.
366 Transcriptome-wide piRNA profiling in human brains of Alzheimer's disease. *Neurobiol Aging* **57**:
367 170-177.doi:10.1016/j.neurobiolaging.2017.05.020

368 Romano G, Veneziano D, Acunzo M, Croce CM. 2017. Small non-coding RNA and cancer.
369 *Carcinogenesis* **38**: 485-491.doi:10.1093/carcin/bgx026

370 Rosenkranz D. 2016. piRNA cluster database: a web resource for piRNA producing loci. *Nucleic Acids*
371 *Res* **44**: D223-30.doi:10.1093/nar/gkv1265

372 Roy J, Sarkar A, Parida S, Ghosh Z, Mallick B. 2017. Small RNA sequencing revealed dysregulated
373 piRNAs in Alzheimer's disease and their probable role in pathogenesis. *Mol Biosyst* **13**:
374 565-576.doi:10.1039/c6mb00699j

375 Sai LS, Agrawal S. 2008. piRNABank: a web resource on classified and clustered Piwi-interacting
376 RNAs. *Nucleic Acids Res* **36**: D173-7.doi:10.1093/nar/gkm696

377 Sarkar A, Maji RK, Saha S, Ghosh Z. 2014. piRNAQuest: searching the piRNAome for silencers. *Bmc*
378 *Genomics* **15**: 555.doi:10.1186/1471-2164-15-555

379 Uhrig S, Klein H. 2019. PingPongPro: a tool for the detection of piRNA-mediated transposon-silencing
380 in small RNA-Seq data. *Bioinformatics* **35**: 335-336.doi:10.1093/bioinformatics/bty578

381 Wang J, Zhang P, Lu Y, Li Y, Zheng Y, Kan Y, Chen R, He S. 2019. piRBase: a comprehensive database

382 of piRNA sequences. *Nucleic Acids Res* **47**: D175-D180.doi:10.1093/nar/gky1043

383 Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C,
384 Stuart JM. 2013. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**:
385 1113-20.doi:10.1038/ng.2764

386 Wu WS, Brown JS, Chen TT, Chu YH, Huang WC, Tu S, Lee HC. 2019. piRTarBase: a database of
387 piRNA targeting sites and their roles in gene regulation. *Nucleic Acids Res* **47**:
388 D181-D187.doi:10.1093/nar/gky956

389 Wu WS, Huang WC, Brown JS, Zhang D, Song X, Chen H, Tu S, Weng Z, Lee HC. 2018. pirScan: a
390 webserver to predict piRNA targeting sites and to avoid transgene silencing in *C. elegans*. *Nucleic Acids*
391 *Res* **46**: W43-W48.doi:10.1093/nar/gky277

392 Zhang H, Ali A, Gao J, Ban R, Jiang X, Zhang Y, Shi Q. 2018. IsopiRBank: a research resource for
393 tracking piRNA isoforms. *Database (Oxford)* **2018**.doi:10.1093/database/bay059

394 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
395 web resource assisting piRNA functional study. *Database (Oxford)* **2014**:
396 bau110.doi:10.1093/database/bau110

397 Zhao Z, Wang K, Wu F, Wang W, Zhang K, Hu H, Liu Y, Jiang T. 2018. circRNA disease: a manually
398 curated database of experimentally supported circRNA-disease associations. *Cell Death Dis* **9**:
399 475.doi:10.1038/s41419-018-0503-3

400

401