

pirScan: a webserver to predict piRNA targeting sites and to avoid transgene silencing in *C. elegans*

Wei-Sheng Wu¹, Wei-Che Huang¹, Jordan S. Brown², Donglei Zhang^{2,3}, Xiaoyan Song⁴, Hao Chen⁵, Shikui Tu^{5,6}, Zhiping Weng⁵ and Heng-Chi Lee^{2,*}

¹Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan, ²Department of Molecular Genetics and Cell Biology, University of Chicago, Chicago, IL 60637, USA, ³Department of Biochemistry and Molecular Biology, Huazhong University of Science and Technology, Wuhan, Hubei, China, ⁴Department of Clinical Laboratory, The Third Affiliated Hospital of Zhengzhou University, Zhengzhou, Henan, China, ⁵Program in Bioinformatics and Integrative Biology, University of Mass. Medical School, Worcester, MA 01605, USA and ⁶Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China

Received January 31, 2018; Revised March 28, 2018; Editorial Decision April 01, 2018; Accepted April 04, 2018

ABSTRACT

pirScan is a web-based tool for identifying *C. elegans* piRNA-targeting sites within a given mRNA or spliced DNA sequence. The purpose of our tool is to allow *C. elegans* researchers to predict piRNA targeting sites and to avoid the persistent germline silencing of transgenes that has rendered many constructs unusable. pirScan fulfills this purpose by first enumerating the predicted piRNA-targeting sites present in an input sequence. This prediction can be exported in a tabular or graphical format. Subsequently, pirScan suggests silent mutations that can be introduced to the input sequence that would allow the modified transgene to avoid piRNA targeting. The user can customize the piRNA targeting stringency and the silent mutations that he/she wants to introduce into the sequence. The modified sequences can be re-submitted to be certain that any previously present piRNA-targeting sites are now absent and no new piRNA-targeting sites are accidentally generated. This revised sequence can finally be downloaded as a text file and/or visualized in a graphical format. pirScan is freely available for academic use at <http://cosbi4.ee.ncku.edu.tw/pirScan/>.

INTRODUCTION

PIWI-interacting RNAs (piRNAs) are a class of small non-coding RNAs that guard animal genomes against mutation by silencing transposons (1). piRNAs guide the Argonaute proteins of the PIWI subclass to target mRNAs for silencing (2–5). The majority of the approximately 15,000 piRNAs encoded by the *Caenorhabditis elegans* genome, however, do not match transposon sequences (6,7). It has been

proposed accordingly that piRNAs target additional mRNAs with diverse functions. The rules governing piRNA targeting have proven enigmatic (8,9). It has been shown, however, that piRNAs appear to tolerate at least three mismatches between their sequences and likely prefer pairing at a seed region (10–14). Our group has further refined the piRNA targeting rules in *C. elegans* by examining how a single piRNA recognizes its target, and by analyzing piRNA reporter assays (15). These analyses have confirmed that piRNAs require near perfect matching within the second to seventh nucleotide seed-region, much like microRNAs (16). Unlike microRNAs, however, piRNAs can only tolerate a few mismatches outside of the seed region. Lastly, the first nucleotide does not contribute to piRNA targeting. Therefore, although the piRNA targeting rules share similarities with the miRNA targeting rules, the numerous tools available for miRNA target prediction are insufficient to successfully predict piRNA targeting sites (17–19).

Significantly, the elucidation of the piRNA targeting rules can help resolve a decades-long impediment to *C. elegans* researchers: transgene silencing in the germline. Transgenes carrying GFP or mCherry tags are frequently silenced in *C. elegans* germlines (20,21). Additionally, this transgene silencing phenomenon is dependent on the PIWI Argonaute PRG-1 (22). By introducing silent mutations in silencing-prone transgenes, such as GFP, mCherry and Cas9, at sites predicted to be targeted by endogenous piRNAs, we have successfully observed strong germline expression of these transgenes in wildtype animal (15). In total, the modified transgenes are expressed in all 14 independently isolated transgenic worm lines, while the unmodified transgenes are silenced in all 15 worm lines. Not only do these results validate that our piRNA targeting rules are correct, but they also provide an avenue to avoid the persistent silencing of transgenes in *C. elegans* germlines (15).

*To whom correspondence should be addressed. Tel: +1 773 702 4684; Email: hengchilee@uchicago.edu

To make our piRNA targeting rules available to the worm community, we have developed pirScan (<http://cosbi4.ee.ncku.edu.tw/pirScan/>), which predicts the endogenous piRNA targeting sites in an input sequence using our established targeting rules. pirScan intuitively displays where piRNA targeting sites are located within an input sequence, as well as the pairing information at each piRNA targeting site. The results of each piRNA target prediction can subsequently be downloaded. pirScan then provides suggestions of how to break piRNA targeting rules and therefore render a given sequence more likely to be expressed. Once the suggested silent mutations are chosen by the user using the graphical interface, the revised sequence is re-scanned to verify that it is now free of piRNA-targeting sites. The final modified sequences can then be downloaded.

TOOL DESCRIPTION

General framework

pirScan is a tool that predicts the presence of piRNA targeting sites within a given sequence in *C. elegans* and suggests silent mutations that can be introduced to avoid silencing of transgenes by piRNAs (Figure 1). The user can input a mature mRNA or spliced DNA sequence and choose to use default or customized piRNA targeting rules to search for targeting sites. The results are provided with a concise graphical and tabular representation displaying all predicted piRNA targeting sites within the input sequence. Each of these views can be downloaded in portable network graphic (.png) and comma separated value (.csv) format, respectively. The user can then choose between suggested silent mutations that will allow the input sequence to escape predicted piRNA targeting, but still retain the same amino acid sequence. If more than one piRNA targeting site were present in the input sequence, then each piRNA will have its own separate table complete with the possible silent mutations to introduce. Upon the user's completion of the proposed mutagenesis, the modified sequence are immediately re-submitted for scanning to verify that the sequence will successfully escape piRNA targeting and ensure that new targeting sites were not inadvertently introduced. Once the user is satisfied with the adjusted input sequence, the modified version can be downloaded as a text file (.txt).

Input

pirScan will accept an input sequence in plain text format (Figure 1, 1. Input). The user must also specify an input name, which will be included in the output following modifications. The input can be a mature mRNA sequence or spliced DNA sequence. By default, pirScan will assume that the first nucleotide present is the start of the open reading frame (see Example 1 in the input window on pirScan). If the first nucleotide present is not the start of the open reading frame, users can also specify the numerical positions of the open reading frame to provide coding information of the input sequence by changing the selection in the input window (see Example 2 in the input window on pirScan). As piRNAs can also target 5' and 3' untranslated regions (UTRs), we suggest the user to include the sequence of 5'

and/or 3' UTRs and simply specify the coding region. The input sequence should not contain introns. 'None' can be selected as the coding sequence region if the input is a non-coding RNA/DNA.

Target stringency

pirScan will identify piRNA sites in *C. elegans* using the default target stringency setting if not specified (Figure 1, 1. Input). The default setting will predict the confident piRNA targeting sites according to our reporter assay which allows zero non-GU mismatches in seed, less than or equal to two GU pairs in seed, less than or equal to two non-GU mismatches in non-seed, less than or equal to three GU pairs in non-seed, and less than or equal to six mismatches and GU pairs in seed and non-seed regions combined (15). To predict more piRNA targeting sites, we have used a slightly relaxed rule (zero non-GU pairs in seed, less than or equal to two GU pairs in seed, less than or equal to three non-GU pairs in non-seed, less than or equal to four GU pairs in non-seed, and less than or equal to six mismatches in seed and non-seed regions combined). Both stringencies have enabled our group to obtain transgenic strains that stably express silencing-prone transgenes in the germline, including GFP::CDK-1, mCherry::ANI-1 and Cas9 (15,23,24). The stringency of the targeting rules can be adjusted by the user prior to the first submission. If an adjustment to the target stringency is made, then those revised targeting rules are automatically used for rescanning of a modified sequence. If the chosen stringency results in over 100 predicted targeting sites, then the search will abort and ask the user to either strengthen the stringency or shorten in input sequence.

Output of targeting site

Upon submission of an input sequence, the user can download a table in .csv format that contains the list of predicted piRNA targeting sites in the input sequence, the targeting score, the location of targeting sites, the numbers of mismatches, and some aspects about those mismatches (Figure 1, 2. Target sites). The targeting score is based on the results of our reporter assay (15) and is calculated using a penalty system where a perfect targeting site is assigned a score of 10. Each non-GU mismatch in the seed region, GU pair in the seed region, non-GU mismatch in the non-seed region, and GU pair in the non-seed region incurs a -7, -1.5, -2 and -1.5 penalty, respectively. We recommend changing any site predicted by the default targeting rules which results in targeting scores of 0 and above, as these are confident sites that are more likely to be critical for silencing. In fact, we have successfully expressed a silencing-prone *gfp::cdk-1* transgene by modifying all four confident sites predicted with default targeting rules and one additional site found in the junction between the *gfp* and *cdk* sequence (Figure 2 and Table 1). This suggests that, in contrast to the more relaxed rules used in our previous study (15), it is possible to selectively mutate only a few sites and still observe transgene expression. Additionally, a visual representation of the predicted piRNA targeting to the input sequence is available for downloading in .png format. The pairing between the

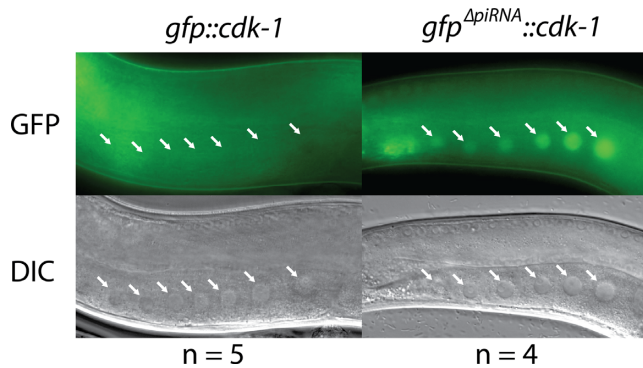


Figure 2. Avoiding only five piRNAs predicted to target GFP results in germline expression of a *gfp::cdk-1* transgene. *gfp::cdk-1* is not expressed prior to introduction of silent mutations, but *gfp^{ΔpiRNA}::cdk-1* is expressed when five predicted piRNA target sites are modified to break the targeting rules. Germline nuclei are indicated by white arrows.

21 nucleotide long piRNA and the corresponding position in the input sequence is displayed for clarity, with the target sites of the input sequence listed according to targeting score. Either of these outputs can allow a user to customize his/her input sequence outside of our pirScan framework if he/she chooses.

Modification of input sequences

At the end of the output page of piRNA target sites, a link is offered for users to introduce silent mutations in their input sequences to avoid piRNA recognition (Figure 1, 3. Introduce silent mutations). The system displays piRNA targeting sites by their targeting score from highest to lowest and lists up to six suggested changes at each site. Each piRNA will have its own separate table complete with the possible silent mutations to introduce. The user can select/deselect any modifications or select multiple modifications at each piRNA site. The system also indicates which rule(s) each mutation will break. Note that sometimes multiple mutations are needed for avoiding piRNA recognition. Importantly, pirScan will not suggest mutations that will result in the incorporation of a rarely used codon, defined as a codon present in <10% of the occurrences of a given amino acid (25). Once the user chooses silent mutations to introduce into the input sequence, the program will automatically rescan any sites with modified sequence(s) and display a table summarizing the changes (Figure 1, 4. Check for new/remaining sites). pirScan will display 'Success' if no piRNA targeting sites are found in the modified sequence, or 'Fail' if additional piRNA targeting sites are found. The modified sequences (either success or failure) can be downloaded in .txt format. In the case of failure, the user can choose to refine the design and the program will first display the piRNA sites that remain to allow further modification. We suggest the user to choose additional sites if a single mutation is not sufficient to avoid piRNA recognition, or to choose different sites if a new piRNA targeting site is found. Additionally, a visual representation which highlights the chosen substitutions can be downloaded in .png format.

piRNA database and pirScan webservice construction

pirScan identifies the piRNA targets by searching for homology between the input sequences and the *C. elegans* piRNA sequence data, which includes 15364 type I piRNAs that can be downloaded from WormMine (<http://intermine.wormbase.org/tools/wormmine/>) and 2485 type II piRNAs, which was collected from PRG-1 IP data (26) with at least 1 read per million reads in (27). For *Caenorhabditis briggsae* piRNAs, a 14453 piRNA-containing database is composed of piRNAs that are cloned at least 1 read from one of three small RNA libraries separately on three strains, i.e., adult hermaphrodites, adult males, and embryos, and contain an upstream motif score of at least 5, based on the data in (28). The search for piRNA target sites is performed using our Python script. The web interface of pirScan was constructed using Django, a Python web framework that encourages rapid web development. All figures in pirScan were generated using D3.js, a JavaScript library which provides powerful visualization components. All tables in pirScan were generated using DataTables, a table enhancing plugin for the jQuery Javascript library which adds sorting, paging and filtering abilities to plain HTML tables with minimal effort. pirScan is available at <http://cosbi4.ee.ncku.edu.tw/pirScan/> (main site), <http://cosbi2.ee.ncku.edu.tw/pirScan/> (backup site 1) or <http://cosbi5.ee.ncku.edu.tw/pirScan/> (backup site 2).

POTENTIAL APPLICATIONS

Our goal in developing pirScan is to make our discoveries concerning the rules of piRNA targeting available to the scientific community and to provide researchers with a method of avoiding the germline silencing of transgenes in *C. elegans*. pirScan can be employed for similar purposes, as well as for more disparate analyses. Examples of tangential goals are outlined below.

Express transgenes containing various foreign sequences in *C. elegans* germline

We have described our success in expressing GFP, mCherry and Cas9 in *C. elegans* germline (15). pirScan will be useful to recode other foreign sequences, such as RNA/DNA binding proteins, enzymes or additional fluorescent markers. A recent study has reported an alternative approach by adding periodic A_n/T_n clusters (PATCs) to the promoter and/or introns of the transgene (29). Although the function of PATCs in gene expression is not understood, this strategy has also been employed by our group to stimulate gene expression in the germline (15,29). It remains to be determined whether transgene expression can be further promoted by combining both strategies.

Analyze previously used constructs for piRNA targeting

Interpreting the lack of expression from a previously published or previously used transgenic construct can be difficult. For example, when miRNA promoters are used to drive GFP expression, GFP is detectable in most tissue types with the exception of the germline (30). This result

Table 1. List of mutations in *gfp::cdk-1* to avoid targeting by 5 piRNAs in *gfp^{ΔpiRNA}::cdk-1*

piRNA	Targeted region in spliced GFP	Mutation(s)	Targeting score before/after
21ur-9758	241–261	G245A, T255C	6/–3
21ur-9756	556–576	T570A, C573A	3.5/–9
21ur-7942	390–410	A408C	1.5/–5.5
21ur-4841	338–358	T354A, T357C	0/–14
21ur-1953 ^a	703–723	G722C	2.5/–4.5

^a21ur-1953 is present in the linker region between *gfp* and *cdk-1*.

suggests that either miRNAs are not expressed in the *C. elegans* germline, or that the transgenes used in this study are prone to germline silencing. pirScan can be applied to these constructs to suggest the best interpretation of this negative result or a similar case.

Eliminate particular endogenous piRNAs for allowing transgene expression

pirScan is designed so that the input sequence can ultimately be modified to escape silencing of the indicated endogenous piRNAs. In some cases, the user may be unable to introduce even silent mutations in his/her transgenic sequence. pirScan can be used to identify which piRNAs are responsible for silencing a particular construct, and the user can apply CRISPR/Cas9 or a similar method to delete or modify the indicated piRNA sequence in the *C. elegans* genome.

Identify piRNA targeting sites on endogenous genes of interest

pirScan accepts sequences of any origin as input. Therefore, the piRNA targeting site prediction is not limited to exogenously constructed genes. If a user hypothesizes that the lack of germline expression or germline-specific regulation of an endogenous gene could be a result of piRNA targeting, then he/she can simply submit the endogenous sequence and look for piRNA targeting sites.

CONCLUSIONS

The discovery of small noncoding RNAs has changed how we understand the factors contributing to gene expression. In a similar vein to the numerous miRNA target prediction tools that are available to predict miRNA binding, we have developed a tool to predict piRNA targeting sites using the piRNA targeting rules. This piRNA target prediction tool is the first of its kind for *C. elegans* and we believe it will be highly valuable in assisting the *C. elegans* community to circumvent the decades old impediment of germline-specific transgene silencing. Our piRNA targeting rules have only been experimentally validated in *C. elegans*; however, we will expand pirScan once the piRNA rules are identified in other animal systems. In addition, we can update our tool to incorporate any additional factors that influence targeting. For example, the influence of piRNA abundance on silencing efficiency is not yet understood, but our tool can be easily updated in the future to consider expression level if this is established as a deterministic factor. As the piRNA targeting rules are likely to be operant in *C. briggsae* as well,

we have also provided the option of using the piRNA data of *C. briggsae* by simply choosing *C. briggsae* in the input page (28). As PIWI proteins are highly conserved among animals, pirScan offers an exciting platform of piRNA target prediction that can be expanded to various organisms in the future.

FUNDING

Ministry of Science of Technology of Taiwan [MOST-105-2221-E-006-203-MY2 and MOST-106-2628-E-006-006-MY2 to W.-S.W.]; NIH predoctoral training grant [T32 GM07197 to J.B.]; NIH P01 grant [HD078253 to Z.W.]; NIH R00 grant [GM108866 to H.-C.L.]. Funding for open access charge: NIGMS [GM108866].

Conflict of interest statement. None declared.

REFERENCES

- Weick,E.-M., Miska,E.A. and Team,A.S. (2014) piRNAs: from biogenesis to function. *Development*, **141**, 3458–3471.
- 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.
- 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.
- Grivna,S.T. (2006) A novel class of small RNAs in mouse spermatogenic cells. *Genes Dev.*, **20**, 1709–1714.
- Lau,N.C. (2006) Characterization of the piRNA complex from rat testes. *Science*, **313**, 363–367.
- Batista,P.J., Ruby,J.G., Claycomb,J.M., Chiang,R., Fahlgren,N., Kasschau,K.D., Chaves,D.A., Gu,W., Vasale,J.J., Duan,S. *et al.* (2008) PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. *Mol. Cell*, **31**, 67–78.
- Ruby,J.G., Jan,C., Player,C., Axtell,M.J., Lee,W., Nusbaum,C., Ge,H. and Bartel,D.P. (2006) Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell*, **127**, 1193–1207.
- 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.
- Toombs,J.A., Sytnikova,Y.A., Chirn,G., Ang,I., Lau,N.C. and Blower,M.D. (2017) Xenopus Piwi proteins interact with a broad proportion of the oocyte transcriptome. *RNA*, **23**, 504–520.
- 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.
- 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.
- Lee,H.-C., Gu,W., Shirayama,M., Youngman,E., Conte,D. and Mello,C.C. (2012) *C. elegans* piRNAs mediate the genome-wide surveillance of germline transcripts. *Cell*, **150**, 78–87.

13. 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.
14. Goh, W.S.S., Falcatori, I., Tam, O.H., Burgess, R., Meikar, O., Kotaja, N., Hammell, M. and Hannon, G.J. (2015) piRNA-directed cleavage of meiotic transcripts regulates spermatogenesis. *Genes Dev.*, **29**, 1032–1044.
15. 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.
16. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
17. Krek, A., Grün, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.
18. 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.
19. Coronello, C. and Benos, P. V. (2013) ComiR: Combinatorial microRNA target prediction tool. *Nucleic Acids Res.*, **41**, W159–W164.
20. Kelly, W.G. and Fire, A. (1998) Chromatin silencing and the maintenance of a functional germline in *Caenorhabditis elegans*. *Development*, **125**, 2451–2456.
21. Merritt, C., Gallo, C.M., Rasoloson, D. and Seydoux, G. (2010) Transgenic solutions for the germline. *WormBook*, ed. The *C. elegans* Research Community. *WormBook*, doi:10.1895/wormbook.1.148.1.
22. Shirayama, M., Seth, M., Lee, H.-C.C., Gu, W., Ishidate, T., Conte, D. and Mello, C.C. (2012) PiRNAs initiate an epigenetic memory of nonself RNA in the *C. elegans* germline. *Cell*, **150**, 65–77.
23. Tse, Y.C., Piekny, A. and Glotzer, M. (2011) Anillin promotes astral microtubule-directed cortical myosin polarization. *Mol. Biol. Cell*, **22**, 3165–3175.
24. Waaijers, S., Portegijs, V., Kerver, J., Lemmens, B.B.L.G., Tijsterman, M., van den Heuvel, S. and Boxem, M. (2013) CRISPR/Cas9-targeted mutagenesis in *Caenorhabditis elegans*. *Genetics*, **195**, 1187–1191.
25. Sharp, P.M. and Bradnam, K.R. (1997) 2nd edn. Riddle, D.L., Blumenthal, T., Meyer, B.J. and Priess, J.R. (eds). Cold Spring Harbor Laboratory Press, NY.
26. Gu, W., Lee, H.-C., Chaves, D., Youngman, E.M., Pazour, G.J., Conte, D. and Mello, C.C. (2012) CapSeq and CIP-TAP identify Pol II start sites and reveal capped small RNAs as *C. elegans* piRNA Precursors. *Cell*, **151**, 1488–1500.
27. Tang, W., Tu, S., Lee, H.-C.C., Weng, Z. and Mello, C.C. (2016) The RNase PARN-1 trims piRNA 3' ends to promote transcriptome surveillance in *C. elegans*. *Cell*, **164**, 974–984.
28. Shi, Z., Montgomery, T.A., Qi, Y. and Ruvkun, G. (2013) High-throughput sequencing reveals extraordinary fluidity of miRNA, piRNA, and siRNA pathways in nematodes. *Genome Res.*, **23**, 497–508.
29. Frøkjær-Jensen, C., Jain, N., Hansen, L., Davis, M.W., Li, Y., Zhao, D., Rebora, K., Millet, J.R.M., Liu, X., Kim, S.K. *et al.* (2016) An abundant class of non-coding DNA can prevent stochastic gene silencing in the *C. elegans* germline. *Cell*, **166**, 343–357.
30. Martinez, N.J., Ow, M.C., Reece-Hoyes, J.S., Barrasa, M.I., Ambros, V.R. and Walhout, A.J.M. (2008) Genome-scale spatiotemporal analysis of *Caenorhabditis elegans* microRNA promoter activity. *Genome Res.*, **18**, 2005–2015.