Databases and resources for human small non-coding RNAs

Eneritz Agirre[1] and Eduardo Eyras[1,2*]

[1]Department of Computational Genomics, Universitat Pompeu Fabra, Dr. Aiguader 88, E08003 Barcelona, Spain
[2]Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, E08010 Barcelona, Spain
*Correspondence to: Tel: +34 93 316 0502; Fax: +34 93 316 0550; E-mail: eduardo.eyras@upf.edu

Abstract

Recent advances in high-throughput sequencing have facilitated the genome-wide studies of small non-coding RNAs (sRNAs). Numerous studies have highlighted the role of various classes of sRNAs at different levels of gene regulation and disease. The fast growth of sequence data and the diversity of sRNA species have prompted the need to organise them in annotation databases. There are currently several databases that collect sRNA data. Various tools are provided for access, with special emphasis on the well-characterised family of micro-RNAs. The striking heterogeneity of the new classes of sRNAs and the lack of sufficient functional annotation, however, make integration of these datasets a difficult task. This review describes the currently available databases for human sRNAs that are accessible via the internet, and some of the large datasets for human sRNAs from high-throughput sequencing experiments that are so far only available as supplementary data in publications. Some of the main issues related to the integration and annotation of sRNA datasets are also discussed.

sRNA databases

In 2001, three groups published independent reports on the discovery of a new class of small non-coding RNAs (sRNAs), which were named micro-RNAs (miRNAs).[1−3] These comprise a large family of small, ∼22 nucleotide-long, non-coding RNAs that have emerged as key players in post-transcriptional gene regulation.[4] Subsequent years have witnessed the discovery of many new types of sRNAs. In humans, apart from the hundreds of miRNAs detected so far, there are also many endogenous small interfering RNAs (endo-siRNAs)[5] and piwi-interacting RNAs (piRNAs)[6,7]. These and other short non-coding RNA molecules collectively are called 'sRNAs'. They are generally short (∼18−30 nucleotides [nt]); do not code for proteins; exert their function as RNA molecules generally combined with protein factors; and represent a substantial portion of the RNA output of cells. Moreover, sRNAs encompass a diverse, widespread and basal regulatory system: they are known to regulate genes and genomes at different levels, including chromatin structure, transcription, RNA stability and translation.[8−10] Furthermore, they can act as activators or inhibitors and their disruption has been linked to disease.[11] The explosion of information on sRNAs makes necessary its organisation—in terms of their biogenesis, expression properties and functional characteristics—into public databases.

Traditionally, GenBank,[12] the European Molecular Biology Laboratory (EMBL)[13] and the DNA Data Bank of Japan (DDBJ)[14] have been the depository of RNA sequences, while the Gene Expression Omnibus (GEO)[15] database at the National Center for Biotechnology Information (NCBI) compiles high-throughput data for miRNAs and other

sRNAs from publications. Besides these generic resources, there are specialised databases for sRNAs. The most complete ones are those related to miRNAs, since their functional role in RNA metabolism is also the best characterised.[5] The miRBase database[16] (Table 1) is considered the central repository for microRNA sequence information. It contains all published miRNA sequences linked to primary literature and other secondary databases. In miRBase, the user can browse published miRNA

Table 1. sRNA databases

Database	sRNA type	Data available	Reference
miRBase	miRNAs	Sequence, genomic location and predicted targets of published miRNAs with links to references	http://www.mirbase.org/
MirZ	miRNAs	Sequencing-based miRNA expression profiles and predicted targets	http://www.mirz.unibas.ch
IsomiR database	miRNAs and isomiRs	Reads and isomiRs assigned to miRNAs from human 293T cells, with miRNA annotation from miRBase	http://galas.systemsbiology.net/cgi-bin/isomir/find.pl
siRNAdb	siRNAs	Experimentally verified and predicted siRNAs. Sequence information and links to the literature	http://sirna.sbc.su.se/
piRNABank	piRNAs	Sequence information, clusters and homology searches for piRNAs	http://pirnabank.ibab.ac.in/
snoRNA-LBME-db	snoRNAs, scaRNAs	Sequence, expression information and predicted targets with base-pairing information	http://www-snorna.biotoul.fr/
Rfam	snRNAs, snoRNAs, miRNAs, other structural RNAs	Sequence families of structural RNAs. Families represented by a multiple sequence alignment and a probabilistic model	http://rfam.sanger.ac.uk/
NONCODE	miRNAs, piRNAs, snoRNAs, scaRNAs	Sequence information with links to GenBank and functional information	http://noncode.org/
RNAdb	miRNAs, snoRNAs, piRNAs, other ncRNAs	Sequence information with links to literature and other databases	http://jsm-research.imb.uq.edu.au/rnadb/
deepBase	miRNAs, piRNAs, endo-siRNAs, nasRNAs, pasRNAs, easRNAs, rasRNAs	Sequences and clusters for sRNAs from different tissues and for computationally predicted sRNAs	http://deepbase.sysu.edu.cn/
fRNAdb	ncRNAs from various sources (see text)	Annotation of known and predicted non-coding RNAs of various lengths with visualisation in genomic context	http://www.ncrna.org

easRNAs, exon-associated small RNAs; endo-siRNAs, endogenous small interfering RNAs; miRNAs, micro-RNAs; isomiRNAs, nasRNAs, non-coding RNA associated small RNAs; NCRNA, non-coding RNA; pasRNA, promotor-associated small RNAs; piRNAs, piwi-interacting RNAs; rasRNAs, repeat associated small RNAs; scaRNAs, Cajal body-specific RNAs; siRNAs, small interfering RNAs; snRNAs, small nuclear RNAs; snoRNAs, small nucleolar RNAs; SRNA, small RNA.

sequences from several species and can perform searches by name, accession number etc. Another database for miRNAs is MirZ[17] (Table 1), which provides analysis tools for mining various datasets from sequencing projects[18] (Table 2) and miRNA expression profiles. MirZ integrates two previously developed resources, the smiRNAdb miRNA expression atlas[18] and the E1MMo miRNA target prediction algorithm.[32]

Although it is generally assumed that a single precursor miRNA molecule leads to a single functional miRNA, there is evidence that precursors can be processed with heterogeneous ends, giving rise to isomiRs.[33] A recently published database collects isomiR sequences from high-throughput sequencing of miRNAs from human 293T cells[34] (Table 1). This database allows one to retrieve all reads and isomiRs assigned to a specific miRNA, with results linked to miRBase and Ensembl.[35]

The endo-siRNAs, which were first observed in plants, are also a very abundant class of sRNAs.[10] They share some properties regarding biogenesis and function with miRNAs.[36,37] More recently, other sources of endo-siRNAs have been identified, such as convergent mRNA transcripts and sense-antisense pairs.[10] To date, there is only one database specific for siRNAs, siRNAdb,[38] which contains the collections of endogenous and exogenous siRNA molecules from the literature that have been experimentally verified. Moreover, siRNAdb also includes predicted siRNAs based on a combination of computational prediction methods.[39–41] Additionally, the database includes information about targets and experimental sources. A set of target predictions is also available for the non-experimentally verified siRNAs.

In 2006, a new abundant sRNA species of approximately 30 nt was observed in extracts of total RNA

Table 2. Human sRNA datasets from deep sequencing

sRNA	Biology	Datasets
TSSa RNAs	20–90nt sRNAs, localised within −250 to +50 of TSSs. Similar to PASRs. Dataset included in deepBase	GSE13483[19]
PASRs	Promoter-associated small RNAs. 20–200 nt long, with 5′ ends coinciding with the TSSs	GSE14362[20]
tiRNAs	18 nt length sRNAs, localised downstream of TSS	http://fantom.gsc.riken.jp/4/download/ Supplemental_Materials/Taft_et_al_2009[21]
spliRNAs	Nuclear sRNAs, enriched at splice sites	GSE20664[22]
TASRNAs	Termini-associated sRNAs	GSE7576[23]
aTASRNAs	Termini-associated sRNAs from the antisense strand	SRA012676[24]
miRNAs	sRNA sequences from libraries from different human organ systems and cell types. Included in MirZ	GSE7233[18]
miRNAs	miRNAs from HeLa cells. Dataset included in deepBase	GSE10829[25]
miRNAs	miRNAs expressed in human leucocytes	GSE19833[26]
sRNAs	sRNAs associated with AGO1 and AGO2, derived from snoRNAs and with miRNA-like functions. Dataset included in deepBase	GSE13370[27]
sRNAs	Derived from tRNAs, in competition with miRNAs	[28–30]
sRNAs	Derived from snoRNAs, with miRNA-like functions	[31]

AGO, Argonaute; miRNA, micro-RNAs; SRNA, small RNA; snoRNAs, small nucleolar RNAs; tRNA, transfer RNA; TSS, transcription start site.

from mouse testes.[42] These somewhat larger types of sRNA, called piRNAs, were found to exert their function most clearly in the germline[43] and possibly in cancer cell lines.[44,45] In contrast to miRNAs and endo-siRNAs, they derive from single-stranded precursors.[10] The only database exclusively dedicated to piRNAs is piRNAbank,[46] which contains piRNA sequences collected from Genbank[47] and published data, indexed through unique identifiers linked to NCBI and with additional information on gene name and genomic position.

Small nucleolar RNAs (snoRNAs) are longer types of non-coding RNA (ncRNA; 60–300 nt in length), but are also regarded as sRNAs. These are a highly evolutionarily conserved class of RNAs which function mostly in the nucleolus and participate in the chemical modification of other RNAs, mainly ribosomal RNAs (rRNAs).[48] The snoRNA–LBME-db database[49] was created to collect the available information on human snoRNAs. Besides sequence information, it also includes predicted RNA targets and potential base-pairing interactions with these targets. The data have been collected from the literature and can be accessed by the name of the snoRNA or can be downloaded in their entirety. This database includes the Cajal body-specific RNAs (scaRNAs). This is a class of nuclear sRNAs similar to snoRNAs which accumulate specifically in Cajal bodies and guide modifications of small nuclear RNAs (snRNAs). The latter are also considered to be sRNAs and are part of the spliceosome.[50] All of these nuclear sRNAs and other structural RNAs are present in the Rfam database (Table 1), which is a general collection of structured RNA families[51] represented by their sequences and a structural model. Rfam also includes models for miRNAs.

Recent advances in deep-sequencing technologies have yielded a large number of short RNA sequences,[16,37,52,53] leading to an exhaustive detection of both known and novel sRNAs. In order to facilitate access to deep-sequencing data, the Short Read Archive (SRA) of NCBI[54] and the European Nucleotide Archive[55] (ENA) (http://www.ebi.ac.uk/ena/) have been created. They provide centralised access to published sequencing data, including

sRNAs, and, more notably, all the data are accessible through powerful search tools. Many of these sRNAs, however, still await classification. They vary significantly in origin and structure and their function is often unknown, which makes it difficult to develop databases for their storage and analysis.

A growing number of public databases are now available for sRNAs obtained from deep-sequencing experiments that classify them in terms of the experimental origin, function (if known) and genomic localisation. These databases try to integrate novel heterogeneous data and sometimes include curated data. The NONCODE database[56] (Table 1) provides organised information for snoRNAs, piRNAs, miRNAs, scaRNAs and other ncRNA classes present in GenBank, classified according to cellular process. Similarly, RNAdb[57] (Table 1) provides access to data from snoRNA-LBME-db, miRBase, the FANTOM project[58] and the H-invitational project.[59] All the datasets are divided according to the main sRNA classes. Additionally, RNAdb contains predictions based on comparative methods. All RNAdb data are available for downloading.

The deepBase[60] database (Table 1) contains the most heterogeneous collection of sRNAs from deep-sequencing data from different libraries (Table 2).[19,25,27] The deep-sequencing data have been mapped to the human genome assembly and annotated according to various sRNA classes, which are defined by the location of the mapped sRNAs: non-coding RNA-associated small RNAs (nasRNAs); sRNAs that overlap promoter regions (pasRNAs), which overlap with the transcription start site (TSS) of genes; exon-associated small RNAs (easRNAs),[60] which overlap with exons from RefSeq[61] genes; and repeat-associated small RNAs (rasRNAs),[60] which overlap with repeat elements from the University of California, Santa Cruz (UCSC) genome browser.[62] Additionally, deepBase includes: all known miRNAs from miRBase and snoRNAs from snoRNA-LBME; novel predicted miRNAs and snoRNAs; and RNA clusters built from the sRNAs that are proximal in the genomic sequence.

Finally, fRNAdb[63] (Table 1) is a searchable database, with functional and genomic annotation for ncRNAs

of various lengths from snoRNA–LBME–db, miRBase, the FANTOM and the H–invitational projects, NONCODE and RNAdb. At the same URL, the authors provide visualisation of the annotated and predicted ncRNAs with a mirror of the UCSC browser.

sRNA datasets

The sRNA world is expanding: it includes many new sRNA sequences that share many features but also differ in others, which makes it difficult to integrate them into the same resource. Despite the current wealth of public databases, there are many recently discovered sRNA classes that are not present in any specific or general database. Below, we enumerate some of the datasets that we cannot find in the previously mentioned databases, and which are only available directly from the publications.

Deep sequencing has led to the detection of at least three new classes of sRNAs linked to the region proximal to the promoter and TSS of genes: promoter-associated sRNAs, which are hypothesised to result from the transcription of independent capped short transcripts or as cleavage products of longer RNAs and which were specifically called 'PASRs';[20] TSS-associated RNAs (TSSa RNAs),[19] included in deepBase and similar to PASRs (Table 2); and transcription initiation RNAs (tiRNAs)[21] (Table 2), which are predominantly 18 nt in length and originate mostly from the region downstream of the TSS, possibly from the backtracking of RNA polymerase II (RNAPII) at the start of transcription.[64] Interestingly, a new class of nuclear sRNAs, also of about 18 nt in length, recently has been found to be associated with the 5′ splice sites of genes (spliRNAs)[22] (Table 2). PASRs, TSSaRNAs and tiRNAs datasets are available at NCBI GEO, and spliRNAs at the FANTOM web page.[65] A different class of sRNAs, also with positional biases, are the termini-associated short RNAs (TASRs)[23] (Table 2) and their antisense counterparts (aTASRs),[24] obtained via a change of the sequencing protocol. These are found to be located antisense of the 3′-untranslated regions of genes. They are only available from the SRA at NCBI (Table 2).

Interestingly, some authors have proposed that some miRNAs and other sRNAs could stem from the processing of snoRNAs by a still unknown mechanism.[66–68] Moreover, there is also some evidence from deep-sequencing experiments that other sRNAs also could originate from transfer RNAs (tRNAs).[28–30] The sRNAs derived from snoRNAs[31,68] and tRNAs[28–30] (see Table 2) are not available in any of the databases mentioned above.

Additionally, there are datasets of sRNAs expressed in specific cell lines, like novel miRNAs obtained from human leucocytes[26] (Table 2), which have not yet been integrated into any database. These examples of new classes of sRNAs display diverse positional and length biases and their function is mostly uncharacterised. They are therefore difficult to integrate with each other and within the existing sRNA databases.

Discussion

The sRNA data currently available—especially those obtained from deep-sequencing experiments—are very heterogeneous. Ideally, these sRNA datasets should be compiled and integrated together, thereby facilitating functional and computational downstream analyses. One of the main problems with the integration of sRNA data is the lack of functional information. While miRNAs and siRNAs are well characterised in terms of how they interact with their targets,[10,36,69] nothing is known so far about the mode of action of many of the new classes of sRNAs. This is further complicated by the fact that sRNAs may have multiple functions. For instance, exogenous siRNAs may affect gene expression[70] and also splicing[71] by inducing chromatin changes. Likewise, miRNAs can also affect gene expression through a similar pathway affecting chromatin,[72] and other sRNAs originating from sense-antisense transcription can also trigger similar mechanisms.[73]

For many of the new sRNA classes, then, we still lack a function and a target definition. Interestingly, the Argonaute (AGO) family of proteins is known to be essential for the function of

miRNAs, piRNAs and endo-siRNAs.[74] It may be possible, therefore, that some of the new classes of sRNAs exert their function through one or more members of the AGO family as well, and may interact with DNA or RNA in a similar way. Further evidence on the proteins that interact with the new classes of sRNAs will help in defining their function and possible targets.

Another important issue for the classification of sRNAs is the characterisation of their biogenesis. Interestingly, some of the new sRNAs have similar positional biases. The sRNAs localised near the TSS of genes have been linked to transcription activity.[21] Likewise, TSSa RNAs and PASRs have been associated with transcription and processing of other RNAs.[20] These similarities may be indicative of a common biogenesis mechanism. Furthermore, many snRNAs are found to be independent of the processing machinery responsible for the biogenesis of miRNAs.[75] Thus, there must be other factors involved in the generation of these sRNAs. Recent experiments have shown that there is considerable endonucleolytic activity associated with still undetermined proteins.[76] These proteins may be the key factors responsible for the biogenesis of some of the novel classes of sRNAs.

In summary, high-throughput methods—and especially deep-sequencing technologies—provide a unique opportunity to explore the wealth of RNA species in diverse cellular contexts. Different types of RNA have been characterised using these technologies, giving rise to new sRNA categories. Further experiments will be necessary to determine the function and biogenesis of these new molecules. Considering the large amount of data that are being generated, we are just touching the tip of the iceberg regarding the sRNA world. The coming years will witness new and exciting developments in molecular and computational biology in the area of sRNAs.

References

1. Lagos-Quintana, M., Rauhut, R., Lendeckel, W. and Tuschl, T. (2001), 'Identification of novel genes coding for small expressed RNAs', *Science* Vol. 294, pp. 853–858.

2. Lau, N.C., Lim, L.P., Weinstein, E.G. and Bartel, D.P. (2001), 'An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*', *Science* Vol. 294, pp. 858–862.

3. Lee, R.C. and Ambros, V. (2001), 'An extensive class of small RNAs in *Caenorhabditis elegans*', *Science* Vol. 294, pp. 862–864.

4. Fabian, M.R., Sonenberg, N. and Filipowicz, W. (2010), 'Regulation of mRNA translation and stability by microRNAs', *Annu. Rev. Biochem.* Vol. 79, pp. 351–379.

5. Okamura, K., Balla, S., Martin, R., Liu, N. *et al.* (2008), 'Two distinct mechanisms generate endogenous siRNAs from bidirectional transcription in *Drosophila melanogaster*', *Nat. Struct. Mol. Biol.* Vol. 15, pp. 581–590.

6. Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M. *et al.* (2006), 'A novel class of small RNAs bind to MILI protein in mouse testes', *Nature* Vol. 442, pp. 203–207.

7. Girard, A., Sachidanandam, R., Hannon, G.J. and Carmell, M.A. (2006), 'A germline-specific class of small RNAs binds mammalian Piwi proteins', *Nature* Vol. 442, pp. 199–202.

8. Van Wolfswinkel, J.C. and Ketting, R.F. (2010), 'The role of small non-coding RNAs in genome stability and chromatin organization', *J. Cell Sci.* Vol. 123, pp. 1825–1839.

9. Bühler, M. and Moazed, D. (2007), 'Transcription and RNAi in heterochromatic gene silencing', *Nat. Struct. Mol. Biol.* Vol. 14, pp. 1041–1048.

10. Carthew, R.W. and Sontheimer, E.J. (2009), 'Origins and mechanisms of miRNAs and siRNAs', *Cell* Vol. 136, pp. 642–655.

11. Taft, R.J., Pang, K.C., Mercer, T.R., Dinger, M. *et al.* (2010), 'Non-coding RNAs: Regulators of disease', *J. Pathol.* Vol. 220, pp. 126–139.

12. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. *et al.* (2004), 'GenBank: Update', *Nucleic Acids Res.* Vol. 32, pp. D23–D26.

13. Stoesser, G., Baker, W., van der Broek, A., Camon, E. *et al.* (2002), 'The EMBL Nucleotide Sequence Database', *Nucleic Acids Res.* Vol. 30, pp. 21–26.

14. Kaminuma, E., Kosuge, T., Kodama, Y., Aono, H. *et al.* (2010), 'DDBJ progress report', *Nucleic Acids Res.* Vol. 39 (Database issue), pp. D22–D27.

15. Barret, T., Troup, D.B., Wilhite, S.E., Ledoux, P. *et al.* (2009), 'NCBI GEO: Archive for high-throughput functional genomic data', *Nucleic Acids Res.* Vol. 37, pp. D885–D890.

16. Griffiths-Jones, S. (2010), 'miRbase: MicroRNA sequences and annotation', *Curr. Protoc. Bioinformatics* Chapter 12, Unit 12.9.1–12.9.10, Wiley online at: http://onlinelibrary.wiley.com/doi/10.1002/0471250953. bil209s29/abstract;jsessionid=8A10CDA0953FDEE1C0A384F3FF7AC1A6. d03t02.

17. Hausser, J., Berninger, P., Rodak, C., Jantscher, Y. *et al.* (2009), 'MirZ: An integrated microRNA expression atlas and target prediction resource', *Nucleic Acids Res.* Vol. 37, pp. 266–272.

18. Landgraf, P., Rusu, M., Sheridan, R., Sewer, A. *et al.* (2007), 'A mammalian microRNA expression atlas based on small RNA library sequencing', *Cell* Vol. 129, pp. 1401–1414.

19. Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W. *et al.* (2008), 'Divergent transcription from active promoters', *Science* Vol. 322, pp. 1849–1851.

20. Fejes-Toth, K., Sotirova, V., Sachidanandam, R., Assaf, G. *et al.* (2009), 'Affymetrix ENCODE Transcriptome Project: Cold Spring Harbor Laboratory ENCODE Transcriptome Project. Post-transcriptional processing generates a diversity of 5′-modified long and short RNAs', *Nature* Vol. 457, pp. 1028–1032.

21. Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C. *et al.* (2009), 'Tiny RNAs associated with transcription start sites in animals', *Nat. Genet.* Vol. 41, pp. 572–578.

22. Taft, R.J., Simons, C., Nahkuri, S., Oey, H. *et al.* (2010), 'Nuclear-localized tiny RNAs are associated with transcription initiation and splice sites in metazoans', *Nat. Struct. Mol. Biol.* Vol. 17, pp. 1030–1034.

23. Kapranov, P., Cheng, J., Dike, S., Nix, D.A. *et al.* (2007), 'RNA maps reveal new RNA classes and a possible function for pervasive transcription', *Science* Vol. 316, pp. 1484–1488.

24. Kapranov, P., Ozsolak, F., Kim, S.W., Foissac, S. et al. (2010), 'New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism', Nature Vol. 466, pp. 642–646.

25. Friedländer, M.R., Chen, W., Adamidi, C., Maaskola, J. et al. (2008), 'Discovering microRNAs from deep sequencing data using miRDeep', Nat. Biotechnol. Vol. 26, pp. 407–415.

26. Vaz, C., Ahmad, H.M., Sharma, P., Gupta, R. et al. (2010), 'Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood', BMC Genomics Vol. 11, p. 288.

27. Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M. et al. (2008), 'A human snoRNA with microRNA-like functions', Mol. Cell. Vol. 32, pp. 519–528.

28. Lee, Y.S., Shibata, Y., Malhotra, A. and Dutta, A. (2009), 'A novel class of small RNAs: tRNA-derived RNA fragments (tRFs)', Genes Dev. Vol. 23, pp. 2639–2649.

29. Cole, C., Sobala, A., Lu, C., Thatcher, S.R. et al. (2009), 'Filtering of deep sequencing data reveals the existence of abundant Dicer-dependet small RNAs derived from tRNAs', RNA Vol. 15, pp. 2147–2160.

30. Haussecker, D., Huang, Y., Lau, A., Parameswaran, P. et al. (2010), 'Human tRNA-derived small RNAs in the global regulation of RNA silencing', RNA Vol. 16, pp. 673–695.

31. Brameier, M., Herwig, A., Reinhardt, R., Walter, L. et al. (2010), 'Human box C/D snoRNAs with miRNA like functions: Expanding the range of regulatory RNAs', Nucleic Acids Res. Vol. 39, pp. D675–D686.

32. Gaidatzis, D., van Nimwegen, E., Hausser, J. and Zavolan, M. (2007), 'Inference of miRNA targets using evolutionary conservation and pathway analysis', BMC Bioinformatics Vol. 8, p. 69.

33. Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F. et al. (2008), 'Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells', Genome Res. Vol. 18, pp. 610–621.

34. Lee, L.W., Zhang, S., Etheridge, A., Ma, L. et al. (2010), 'Complexity of the microRNA repertoire revealed by next generation sequencing', RNA Vol. 16, pp. 2170–2180.

35. Hubbard, T.J., Aken, B.L., Ayling, S., Ballester, B. et al. (2009), 'Ensembl 2009', Nucleic Acids Res. Vol. 37, pp. D690–D697.

36. Bartel, D.P. (2009), 'MicroRNAs: Target recognition and regulatory functions', Cell Vol. 136, pp. 215–230.

37. Mattick, J.S. and Makunin, I.V. (2005), 'Small regulatory RNAs in mammals', Hum. Mol. Genet. Vol. 1, pp. R121–R132.

38. Chalk, A.M., Wahlestedt, C. and Sonnhammer, E.L. (2004), 'Improved and automated prediction of effective siRNA', Biochem. Biophys. Res. Commun. Vol. 319, pp. 264–274.

39. Chalk, A.M., Warfinge, R.E., Georgii-Hemming, P. and Sonnhammer, E.L. (2005), 'SiRNAdb: A database of siRNA sequences', Nucleic Acids Res. Vol. 33, pp. D131–D134.

40. Reynolds, A., Leake, D., Boese, Q., Scaringe, S. et al. (2004), 'Rational siRNA design for RNA interference', Nat. Biotechnol. Vol. 22, pp. 326–330.

41. Ui-Tei, K., Naito, Y., Takahashi, F., Haraguchi, T. et al. (2004), 'Guidelines for the selection of highly effective siRNA sequence for mammalian and chick RNA interference', Nucleic Acids Res. Vol. 32, pp. 936–948.

42. Grivna, S.T., Beyret, E., Wang, Z. and Lin, H. (2006), 'A novel class of small RNAs in mouse spermatogenic cells', Genes Dev. Vol. 20, pp. 1709–1714.

43. Thomson, T. and Lin, H. (2009), 'The biogenesis and function of PIWI proteins and piRNAs: Progress and prospect', Annu. Rev. Cell Dev. Biol. Vol. 25, pp. 355–376.

44. Lee, H.J., Schütte, D., Wulf, G., Füzesi, L. et al. (2006), 'Stem-cell protein PIWIL2 is widely expressed in tumors and inhibits apoptosis through activation of Stat3/Bcl-XL pathway', Hum. Mol. Genet. Vol. 15, pp. 201–211.

45. Lu, Y., Li, C., Zhang, K., Sun, H. et al. (2010), 'Identification of piRNAs in Hela cells by massive parallel sequencing', BMB Reports Vol. 43, pp. 635–641.

46. Lakshmi, S.S. and Agrawal, S. (2008), 'piRNABank: A web resource on classified and clustered Piwi-interacting RNAs', Nucleic Acids Res. Vol. 36, pp. D173–D177.

47. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. et al. (2009), 'GenBank. 2009', Nucleic Acids Res. Vol. 37, pp. D26–D31.

48. Matera, A.G., Terns, R.M. and Terns, M.P. (2007), 'Non-coding RNAs: Lessons from the small nuclear and small nucleolar RNAs', Nat. Rev. Mol. Cell Biol. Vol. 8, pp. 209–220.

49. Lestrade, L. and Weber, M.J. (2006), 'SnoRNA-LBME-db, a comprenhensive database of human H/ACA and C/D box snoRNAs', Nucleic Acids Res. Vol. 34, pp. D158–D162.

50. Darzacq, X., Jády, B.E., Verheggen, C., Kiss, A.M. et al. (2002), 'Cajal body-specific small nuclear RNAs: A novel class of 2′-O-methylation and pseudoridylation guide RNAs', EMBO J. Vol. 21, pp. 2746–2756.

51. Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P. et al. (2009), 'Rfam: Updates to the RNA families database', Nucleic Acids Res. Vol. 37, pp. D136–D140.

52. Malone, C. and Hannon, G. (2009), 'Small RNAs as guardians of the genome', Cell Vol. 136, pp. 656–668.

53. Ghildiyal, M. and Zamore, P.D. (2009), 'Small silencing RNAs: An expanding universe', Nat. Rev. Genet. Vol. 10, pp. 94–108.

54. Leinonen, R., Sugawara, H. and Shumway, M., on behalf of the International Nucleotide Sequence Database Collaboration (2010), 'The sequence read archive', Nucleic Acids Res. Vol. 39 (Database issue), pp. D19–D21.

55. Leinonen, R., Akhtar, R., Birney, E., Bower, L. et al. (2010), 'The European Nucleotide Archive', Nucleic Acids Res. Vol. 39 (Database issue), pp. D28–D31.

56. He, S., Liu, C., Skogerbø, G., Zhao, H. et al. (2008), 'NONCODE v2.0: Decoding the non-coding', Nucleic Acids Res. Vol. 36, pp. D170–D172.

57. Pang, K.C., Stephen, S., Dinger, M.E., Engström, P.G. et al. (2007), 'RNAdb 2.0: An expanded database of mammalian non-coding RNAs', Nucleic Acids Res. Vol. 35, pp. D178–D182.

58. Ajioka, I., Maeda, T. and Nakajima, K. (2006), 'Large-scale correlation of DNA accession numbers to the cDNAs in the FANTOM full-length mouse cDNA clone set', Keio J. Med. Vol. 55, pp. 107–110.

59. Yamasaki, C., Murakami, K., Takeda, J., Sato, Y. et al. (2010), 'H-InvDB in 2009: Extended database and data mining resources for human genes and transcripts', Nucleic Acids Res. Vol. 38, pp. D626–D632.

60. Yang, J.-H., Shao, P., Zhou, H., Chen, Y.-Q. et al. (2009), 'deepBase: A database for deeply annotating and mining deep sequencing data', Nucleic Acids Res. Vol. 38, pp. D123–D130.

61. Pruitt, K.D., Tatusova, T., Klimke, W. and Maglott, D.R. (2009), 'NCBI Reference Sequences: Current status, policy and new initiatives', Nucleic Acids Res. Vol. 37, pp. D32–D36.

62. Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S. et al. (2010), 'The UCSC Genome Browser database: Update 2010', Nucleic Acids Res. Vol. 38, pp. D613–619.

63. Mituyama, T., Yamada, K., Hattori, E., Okida, H. et al. (2009), 'The functional RNA Database 3.0: Databases to support mining and annotation of functional RNAs', Nucleic Acids Res. Vol. 37, pp. D89–D92.

64. Taft, R.J., Kaplan, C.D., Simons, C. and Mattick, J.S. (2009), 'Evolution, biogenesis and function of promoter-associated RNAs', Cell Cycle Vol. 8, pp. 2332–2338.

65. Kawaji, H., Severin, J., Lizio, M., Waterhouse, A. et al. (2009), 'The FANTOM web resource: From mammalian transcriptional landscape to its dynamic regulation', Genome Biol. Vol. 10, pp. R40.

66. Kishore, S., Khanna, A., Zhang, Z., Hui, J. et al. (2010), 'The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing', Hum. Mol. Genet. Vol. 19, pp. 1153–1164.

67. Khanna, A. and Stamm, S. (2010), 'Regulation of alternative splicing by short non-coding nuclear RNAs', RNA Biol. Vol. 7, p. 4.

68. Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y. et al. (2009), 'Small RNAs derived from snoRNAs', RNA Vol. 15, pp. 1233–1240.

69. Shin, C., Nam, J.-W., Farh, K.K.-H., Chiang, H.R. et al. (2010), 'Expanding the microRNA targeting code: Functional sites with centered pairing', Mol. Cell. Vol. 38, pp. 789–802.

70. Morris, K.V., Chan, S.W., Jacobsen, S.E. and Looney, D.J. (2004), 'Small interfering RNA-induced transcriptional gene silencing in human cells', Science Vol. 305, pp. 1289–1292.

71. Alló, M., Buggiano, V., Fededa, J.P., Petrillo, E. *et al.* (2009), 'Control of alternative splicing through siRNA-mediated transcriptional gene silencing', *Nat. Struct. Mol. Biol.* Vol. 16, pp. 717–724.

72. Kim, D.H., Saetrom, P, Snove, O., Jr and Rossi, J.J. (2008), 'MicroRNA-directed transcriptional gene silencing in mammalian cells', *Proc. Natl. Acad. Sci. USA* Vol. 105, pp. 16230–16235.

73. Yu, W., Gius, D., Onyango, P., Muldoon-Jacobs, K. *et al.* (2008), 'Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA', *Nature* Vol. 451, pp. 202–206.

74. Kim, V.N., Han, J. and Siomi, M.C. (2009), 'Biogenesis of small RNAs in animals', *Nature* Vol. 12, pp. 126–139.

75. Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P. *et al.* (2008), 'Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, Dicer-dependent small RNAs', *Genes Dev.* Vol. 22, pp. 2773–2785.

76. Karginov, F.V., Cheloufi, S., Chong, M.M., Stark, A. *et al.* (2010), 'Diverse endonucleolytic cleavage sites in the mammalian transcriptome depend upon microRNAs, Drosha, and additional nucleases', *Mol. Cell* Vol. 38, pp. 781–788.