RNAdb 2.0—an expanded database of mammalian non-coding RNAs

Ken C. Pang[1,2], Stuart Stephen[1], Marcel E. Dinger[1], Pär G. Engström[3,4], Boris Lenhard[3] and John S. Mattick[1,*]

[1]ARC Special Research Centre for Functional and Applied Genomics, Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia, [2]T cell Laboratory, Ludwig Institute for Cancer Research, Melbourne Centre for Clinical Sciences, Austin Hospital, Heidelberg, Victoria 3084, Australia, [3]Computational Biology Unit, Bergen Center for Computational Science, University of Bergen, Bergen, Norway and [4]Programme for Genomics and Bioinformatics, Department of Cell and Molecular Biology, Karolinska Institutet, Stockholm, Sweden

ABSTRACT

RNAdb is a comprehensive database of mammalian non-protein-coding RNAs (ncRNAs). There is increasing recognition that ncRNAs play important regulatory roles in multicellular organisms, and there is an expanding rate of discovery of novel ncRNAs as well as an increasing allocation of function. In this update to RNAdb, we provide nucleotide sequences and annotations for tens of thousands of non-housekeeping ncRNAs, including a wide range of mammalian microRNAs, small nucleolar RNAs and larger mRNA-like ncRNAs. Some of these have documented functions and/or expression patterns, but the majority remain of unclear significance, and include PIWI-interacting RNAs, ncRNAs identified from the latest rounds of large-scale cDNA sequencing projects, putative antisense transcripts, as well as ncRNAs predicted on the basis of structural features and alignments. Improvements to the database comprise not only new and updated ncRNA datasets, but also provision of microarray-based expression data and closer interface with more specialized ncRNA resources such as miRBase and snoRNA-LBME-db. To access RNAdb, visit http://research.imb.uq. edu.au/RNAdb.

INTRODUCTION

The mammalian genome encodes thousands of non-protein-coding RNAs (ncRNAs). Ribosomal RNAs (rRNAs), transfer RNAs (tRNAs) and small nuclear RNAs (snRNAs), fulfil mainly housekeeping roles in mRNA translation and splicing. Small nucleolar RNAs (snoRNAs) and the related small Cajal body-specific RNAs (scaRNAs) guide modifications of other RNAs (1–3). MicroRNAs (miRNAs) regulate gene expression by controlling mRNA translation and turnover (4,5), PIWI-interacting RNAs (piRNAs) are thought to be important in spermatogenesis (6,7), while larger ncRNAs have been discovered to be developmentally regulated (8–10) and to function in a range of processes including genomic imprinting, intracellular protein trafficking and brain development (11–16). The abundance of ncRNAs has only become apparent in the past few years and was largely unexpected. Although many recently identified ncRNAs remain of unknown function, and appear to be evolving rapidly (15,17), it is increasingly clear that ncRNAs represent a diverse and important class of functional output from mammalian genomes.

RNAdb is a comprehensive database of mammalian ncRNAs. The focus of the database is on ncRNAs that have restricted expression and whose function is likely to be regulatory. Housekeeping RNAs (rRNAs, tRNAs, snRNAs) are not included and are covered elsewhere (18,19). The aim of the database is to provide a nucleotide sequence-based platform to facilitate both bioinformatic and experimental research in the burgeoning field of RNomics. Already, RNAdb has been used to develop machine-learning algorithms for identifying ncRNAs (20), annotate transcripts from a large-scale transcriptome project (8) and examine ncRNA evolution (17). In addition to containing sequence data, individual ncRNA entries in RNAdb are annotated based upon publicly available information in the literature or secondary databases. In this way, the database can also be browsed or searched by the casual user interested in learning more about particular ncRNAs.

Since the original release of RNAdb two years ago (21), the number of known mammalian ncRNAs has grown

considerably. In recognition of this growth, we have updated the database to include tens of thousands of novel ncRNAs. Some of these have been characterized in isolation, continuing the trend of *ad hoc* discovery by which many earlier ncRNAs were identified. The majority, however, comes from large-scale cloning and sequencing studies or structural alignment-based predictions. As well as incorporating new ncRNA datasets, the current release of RNAdb provides other enhancements, including microarray-based expression data, closer interface with specialized ncRNA resources such as miRBase and snoRNA-LBME-db (3,22), and the availability of data for use as custom tracks on the UCSC Genome Browser (23).

DATABASE ACCESS

RNAdb is available on-line at http://research.imb.uq.edu.au/ RNAdb. Currently, datasets are stored in relational form in a Microsoft SQL2005 database. The web application is multilayered with the primary presentation layer implemented in C# 2 under the ASP.NET 2.0 framework. The application layer is implemented as a mixture of C# and C++ modules with dataset normalization performed through SQL stored procedures.

The database can be accessed or queried in various ways. Users can casually browse the collection. Specific searches can be performed using keywords (with or without Boolean operators) and/or by applying filters across nominated fields. BLAST searches permit users to locate regions of similarity between sequences of interest and those stored in the database.

To facilitate links with other on-line resources, users can now directly go to a detailed view of an entry by using the following URL and substituting the RNAdb unique identifier of interest for <RNAdbID>: http://jsm-research.imb.uq.edu. au/rnadb/default.aspx?ncrna=<RNAdbID>. For example, if a user wishes to look at the detailed view for MIR1004, one would use http://jsm-research.imb.uq.edu.au/rnadb/default. aspx?ncrna=MIR1004.

The entire database is available for download in either FASTA or XML format via the website. Specific datasets are also provided as custom tracks for loading directly into the UCSC Genome Browser (23). This feature allows users to easily take a defined subset of the database and readily apply it to the UCSC Genome Browser's extensive set of comparative and analytical tools.

DATABASE CONTENT

ncRNAs in RNAdb are divided into several distinct datasets (Table 1). This decision was made not only to reflect the different ways in which ncRNAs have been identified, but also in recognition that users may want to separately query and download each set. A description of each dataset is provided below.

miRNAs

Over 1800 mammalian miRNAs are found within RNAdb. These sequences were obtained from the latest release of

Table 1. Summary of RNAdb 2.0 datasets

Dataset	Organism	Number of sequences
miRNAs	Human	462
	Mouse	358
	Other	1035
snoRNAs and scaRNAs	Human	375
	Mouse	175
	Other	24
piRNAs	Human	32 046
	Mouse	30 024
	Rat	26 568
Other ncRNAs from the literature	Human	531
	Mouse	242
	Other	167
FANTOM3 ncRNAs	Mouse	34 030
H-Invitational ncRNAs	Human	1794
ncRNAs predicted from structural alignments		
RNAz	Human	35 984
Non-coding RNA Search	Human	1273
	Mouse	1252
EvoFold	Human	47 509
Predicted antisense ncRNAs	Human	1068 (in 919 TUs)[a]
	Mouse	1615 (in 1395 Tus)[a]

[a]Sequences were clustered into transcriptional units (TUs) by joining any two or more sequences that had overlapping exons on the same genomic strand.

miRBase (release 8.2, July 2006) (22). miRBase is the central repository for miRNA data on the web and is regularly maintained. We have elected to directly link the RNAdb miRNA entries to miRBase, so as to keep abreast with the most recent annotations and updates.

snoRNAs and scaRNAs

RNAdb contains more than 500 mammalian snoRNAs and scaRNAs. The snoRNAs fall into two general classes, C/D box and H/ACA snoRNAs, which classically guide ribose methylation and pseudo-uridylation of rRNAs, respectively. Interestingly, some snoRNAs appear to regulate other RNAs, including HBII-52 which regulates the alternative splicing of the serotonin receptor 2C and is implicated in the pathogenesis of Prader-Willi syndrome (24). Human snoRNAs and scaRNAs in RNAdb were derived from snoRNA-LBME-db (release 3, August 2006) (3), and annotations for these sequences are maintained by linking out to this informative and specialized resource.

piRNAs

The PIWI family of proteins is known to be important for germ cell development. PIWI proteins were recently discovered to bind thousands of small RNAs, termed piRNAs (6,7). piRNAs have been identified in testis, 26–31 nt in length, and are distinct from miRNAs. Over 88 000 piRNA candidates have been cloned and sequenced from mouse, human and rat, and are included for the first time in the current release of RNAdb.

Other ncRNAs from the literature

This dataset contains more than 900 unique ncRNA sequences which have been identified and manually curated based upon extensive literature review. A majority of ncRNAs listed

here are much longer than those listed above. Altogether, 36 mammalian organisms are represented but most ncRNAs are either from mouse or human. Although some of these transcripts have documented biological roles, most are transcripts of unknown function. As well as sequence data, additional information–including GenBank accessions, references, chromosomal location, transcript length, splicing status, conservation notes, function, disease associations, antisense relationships, imprinting status and tissue expression patterns-is provided wherever possible in separate searchable fields.

New additions to this dataset include multiple long mRNA-like ncRNAs whose functions have recently become apparent. For instance, NRON, an ncRNA repressor of the nuclear factor of activated T cells (NFAT), regulates nuclear trafficking of NFAT (13). Taurine upregulated gene 1 (TUG1) is required for photoreceptor development in the eye (25). Saf, which lies antisense to the death receptor Fas, alters the expression of alternative Fas isoforms and increases resistance to Fas-induced apoptosis (26). Evf-2, an ncRNA derived from an ultraconserved element, cooperates with the homeodomain protein Dlx-2 to augment the transcriptional activity of a nearby enhancer (27). Serving as a salient reminder that many ncRNAs are poorly conserved (17) is HAR1F, which is expressed specifically in Cajal-Retzius neurons in the developing human neocortex and has evolved rapidly in the human lineage (15).

FANTOM3 ncRNAs

Using full-length cDNA cloning and sequencing strategies, the Functional Annotation of Mouse (FANTOM) project has identified thousands of novel transcripts from the mouse genome (8). In the most recent round of annotation, 34 030 cDNAs were manually annotated as putative ncRNAs (28), a subset of which were subsequently shown to be derived fragments of very long ncRNAs (29). Since both cloning and manual human annotation is subject to variation and error, the true number of ncRNAs remains unclear. To this end, we provide the results of various computational prediction strategies for use as additional filters in identifying ncRNAs (Supplementary Data 1). In addition to sequence data, details such as the Riken clone identifier, GenBank accession, genomic location, transcript length, likely imprinting status and library of origin are provided. RNAdb also incorporates expression information from publicly available microarray datasets such as GNF SymAtlas (30) (Supplementary Data 2). Although limited to only a small proportion of FANTOM3 ncRNAs, this information allows the identification of transcripts that are dynamically expressed across various tissues and cell types, and is expected to provide a useful starting point for their further characterization.

As indicated earlier, the vast majority of ncRNAs identified from large-scale cDNA sequencing projects are of unknown significance. A recent screen of several hundred, well-conserved FANTOM ncRNAs identified not only NRON, but also seven other functional ncRNA genes essential for cell viability or involved in Hedgehog signalling (13,31). Given that this strategy employed only a limited number of cell-based assays and that only a tiny proportion of ncRNAs were examined, it would appear likely that many more

functional ncRNAs from the FANTOM collection remain to be uncovered in the future.

H-Invitational ncRNAs

This dataset contains more than 1700 putative ncRNAs from the latest round of the Human Full-length cDNA Annotation Invitational (H-Invitational) project (release 3.4, August 2006) (32). Non-protein-coding transcripts are defined in this dataset by the absence of any open reading frame and by not belonging to the pseudogene classification. In addition to the sequence data, details such as the GenBank accession no., genomic location, transcript length, library of origin and expression data (based upon publicly available microarray data where present; see Supplementary Data 2) are also listed.

ncRNAs predicted from structural alignments

Recently, a number of studies have identified thousands of putative ncRNAs based upon predicted structural features and alignments using novel comparative genomics tools. The datasets resulting from three independent approaches, RNAz (33), Non-coding RNA Search (34) and EvoFold (35), are included here. RNAz combines a comparative approach (scoring conservation of secondary structure) with the observation that ncRNAs are thermodynamically more stable than expected by chance. Using sequences conserved in at least human, mouse, rat and dog, over 35 000 structured elements were identified in the human genome (36). Non-coding RNA Search uses syntenic regions between human and mouse that are unalignable and then utilizes the FOLDALIGN algorithm to identify regions with conserved secondary structure. Finally, EvoFold utilizes a comparative genomics method based on phylogenetic stochastic context-free grammars to identify functional RNAs. Using an eight-way genome-wide alignment of human, chimpanzee, mouse, rat, dog, chicken, zebrafish and Fugu, over 47 000 candidate RNA structures were identified in the human genome.

Comprehensive antisense ncRNA dataset

Natural antisense transcription is now recognized as being a common occurrence in the mammalian transcriptome and a means by which gene expression can be regulated (37–39). Data from tiling array experiments and sequencing of short tags representing 5′ and 3′ ends of transcripts suggest that more than 60% of all human and mouse loci may be transcribed on both strands and give rise to complementary transcripts (9,37). In its original release, RNAdb contained a dataset of putative antisense ncRNAs identified from cDNA and EST databases for human and mouse using a computational pipeline (21). Coinciding with the current release, we have recently re-developed the pipeline that searches for antisense RNAs and experimentally validated a subset of its predictions (40). We will continue to use the improved pipeline in regular updates of the antisense ncRNA dataset.

CONCLUSIONS

The total number of mammalian ncRNA sequences contained in RNAdb has increased ~10-fold since the database's inception 2 years ago. Such growth in content reflects the high

level of interest and activity in the field over this period. Nevertheless, most of the newly added sequences represent putative ncRNAs and their biological roles, if any, remain unclear. The fact that most of the mammalian genome is transcribed suggests that there is either a great deal of transcriptional noise or that these RNAs are fulfilling some unexpected functions in mammalian biology (11,41). Although the search for new ncRNAs is far from exhausted (42), one of today's principal challenges in the field of RNomics is to explore the functional significance of this abundant non-coding transcription. If this challenge is to be successfully met in coming years, then experimental proof of function will be paramount. Such proof often comes slowly and incrementally, and having bioinformatic resources such as RNAdb will be essential to guide and facilitate future discovery.

FUTURE DIRECTIONS

As new ncRNAs are discovered, we will continue to update RNAdb. Submissions of new mammalian ncRNAs are invited, and should be sent to RNAdb@imb.uq.edu.au. We also plan to regularly synchronize our datasets with miRBase and snoRNA-LBME-db, as these resources are updated. Currently, publicly available microarray-based expression data for ncRNAs remain limited, but is likely to be significantly expanded in the future (K. Pang and M. Dinger, unpublished data). Once new expression data are released, they will subsequently be incorporated into RNAdb.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Bachellerie,J.P., Cavaille,J. and Huttenhofer,A. (2002) The expanding snoRNA world. *Biochimie*, **84**, 775–790.
2. Meier,U.T. (2005) The many facets of H/ACA ribonucleoproteins. *Chromosoma*, **114**, 1–14.
3. Lestrade,L. and Weber,M.J. (2006) snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.*, **34**, D158–D162.
4. Mattick,J.S. and Makunin,I.V. (2005) Small regulatory RNAs in mammals. *Hum. Mol. Genet.*, **14** (Suppl. 1), R121–R132.
5. Wienholds,E. and Plasterk,R.H. (2005) MicroRNA function in animal development. *FEBS Lett.*, **579**, 5911–5922.
6. 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.
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*, **442**, 199-202.
8. Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Maeda,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
9. Cheng,J., Kapranov,P., Drenkow,J., Dike,S., Brubaker,S., Patel,S., Long,J., Stern,D., Tammana,H., Helt,G. *et al.* (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–1154.
10. Ravasi,T., Suzuki,H., Pang,K.C., Katayama,S., Furuno,M., Okunishi,R., Fukuda,S., Ru,K., Frith,M.C., Gongora,M.M. *et al.* (2006) Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.*, **16**, 11–19.
11. Mattick,J.S. and Makunin,I.V. (2006) Non-coding RNA. *Hum. Mol. Genet.*, **15**, R17–R29.
12. Prasanth,K.V., Prasanth,S.G., Xuan,Z., Hearn,S., Freier,S.M., Bennett,C.F., Zhang,M.Q. and Spector,D.L. (2005) Regulating gene expression through RNA nuclear retention. *Cell*, **123**, 249–263.
13. Willingham,A.T., Orth,A.P., Batalov,S., Peters,E.C., Wen,B.G., Aza-Blanc,P., Hogenesch,J.B. and Schultz,P.G. (2005) A strategy for probing the function of noncoding RNAs finds a repressor of NFAT. *Science*, **309**, 1570–1573.
14. Ginger,M.R., Shore,A.N., Contreras,A., Rijnkels,M., Miller,J., Gonzalez-Rimbau,M.F. and Rosen,J.M. (2006) A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc. Natl Acad. Sci. USA*, **103**, 5781–5786.
15. Pollard,K.S., Salama,S.R., Lambert,N., Lambot,M.A., Coppens,S., Pedersen,J.S., Katzman,S., King,B., Onodera,C., Siepel,A. *et al.* (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, **443**, 167–172.
16. Mehler,M.F. and Mattick,J.S. (2006) Non-coding RNA in the nervous system. *J. Physiol.*, **575**, 333–341.
17. Pang,K.C., Frith,M.C. and Mattick,J.S. (2006) Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends Genet.*, **22**, 1–5.
18. Griffiths-Jones,S., Moxon,S., Marshall,M., Khanna,A., Eddy,S.R. and Bateman,A. (2005) Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.*, **33**, D121–D124.
19. Wuyts,J., Perriere,G. and Van De Peer,Y. (2004) The European ribosomal RNA database. *Nucleic Acids Res.*, **32**, D101–D103.
20. Liu,J., Gough,J. and Rost,B. (2006) Distinguishing protein-coding from non-coding RNAs through support vector machines. *PLoS Genet.*, **2**, e29.
21. Pang,K.C., Stephen,S., Engstrom,P.G., Tajul-Arifin,K., Chen,W., Wahlestedt,C., Lenhard,B., Hayashizaki,Y. and Mattick,J.S. (2005) RNAdb—a comprehensive mammalian noncoding RNA database. *Nucleic Acids Res.*, **33**, D125–D130.
22. Griffiths-Jones,S., Grocock,R.J., van Dongen,S., Bateman,A. and Enright,A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
23. Hinrichs,A.S., Karolchik,D., Baertsch,R., Barber,G.P., Bejerano,G., Clawson,H., Diekhans,M., Furey,T.S., Harte,R.A., Hsu,F. *et al.* (2006) The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.*, **34**, D590–D598.
24. Kishore,S. and Stamm,S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230–232.
25. Young,T.L., Matsuda,T. and Cepko,C.L. (2005) The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr. Biol.*, **15**, 501–512.
26. Yan,M.D., Hong,C.C., Lai,G.M., Cheng,A.L., Lin,Y.W. and Chuang,S.E. (2005) Identification and characterization of a novel gene Saf transcribed from the opposite strand of Fas. *Hum. Mol. Genet.*, **14**, 1465–1474.

27. Feng,J., Bi,C., Clark,B.S., Mady,R., Shah,P. and Kohtz,J.D. (2006) The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.*, **20**, 1470–1484.

28. Maeda,N., Kasukawa,T., Oyama,R., Gough,J., Frith,M., Engstrom,P.G., Lenhard,B., Aturaliya,R.N., Batalov,S., Beisel,K.W. *et al.* (2006) Transcript annotation in FANTOM3: mouse gene catalog based on physical cDNAs. *PLoS Genet.*, **2**, e62.

29. Furuno,M., Pang,K.C., Ninomiya,N., Fukuda,S., Frith,M.C., Bult,C., Kai,C., Kawai,J., Carninci,P., Hayashizaki,Y. *et al.* (2006) Clusters of internally primed transcripts reveal novel long noncoding RNAs. *PLoS Genet.*, **2**, e37.

30. Su,A.I., Wiltshire,T., Batalov,S., Lapp,H., Ching,K.A., Block,D., Zhang,J., Soden,R., Hayakawa,M., Kreiman,G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6067.

31. Mattick,J.S. (2005) The functional genomics of noncoding RNA. *Science*, **309**, 1527–1528.

32. Imanishi,T., Itoh,T., Suzuki,Y., O'Donovan,C., Fukuchi,S., Koyanagi,K.O., Barrero,R.A., Tamura,T., Yamaguchi-Kabata,Y., Tanino,M. *et al.* (2004) Integrative annotation of 21,037 human genes validated by full-length cDNA clones. *PLoS Biol.*, **2**, E162.

33. Washietl,S., Hofacker,I.L. and Stadler,P.F. (2005) Fast and reliable prediction of noncoding RNAs. *Proc. Natl Acad. Sci. USA*, **102**, 2454–2459.

34. Torarinsson,E., Sawera,M., Havgaard,J.H., Fredholm,M. and Gorodkin,J. (2006) Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res.*, **16**, 885–889.

35. Pedersen,J.S., Bejerano,G., Siepel,A., Rosenbloom,K., Lindblad-Toh,K., Lander,E.S., Kent,J., Miller,W. and Haussler,D. (2006) Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput. Biol.*, **2**, e33.

36. Washietl,S., Hofacker,I.L., Lukasser,M., Huttenhofer,A. and Stadler,P.F. (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.*, **23**, 1383–1390.

37. Katayama,S., Tomaru,Y., Kasukawa,T., Waki,K., Nakanishi,M., Nakamura,M., Nishida,H., Yap,C.C., Suzuki,M., Kawai,J. *et al.* (2005) Antisense transcription in the mammalian transcriptome. *Science*, **309**, 1564–1566.

38. Faghihi,M.A. and Wahlestedt,C. (2006) RNA interference is not involved in natural antisense mediated regulation of gene expression in mammals. *Genome Biol.*, **7**, R38.

39. Wahlestedt,C. (2006) Natural antisense and noncoding RNA transcripts as potential drug targets. *Drug Discov. Today*, **11**, 503–508.

40. Engstrom,P.G., Suzuki,H., Ninomiya,N., Akalin,A., Sessa,L., Lavorgna,G., Brozzi,A., Luzi,L., Tan,S.L., Yang,L. *et al.* (2006) Complex loci in human and mouse genomes. *PLoS Genet.*, **2**, e47.

41. Mattick,J.S. (2003) Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays*, **25**, 930–939.

42. Berezikov,E., van Tetering,G., Verheul,M., van de Belt,J., van Laake,L., Vos,J., Verloop,R., van de Wetering,M., Guryev,V., Takada,S. *et al.* (2006) Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. *Genome Res.*, **16**, 1289–1298.