Nucleic Acids Research Advance Access published October 20, 2014

Nucleic Acids Research, 2014 **1**
doi: 10.1093/nar/gku988

lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs

Xiu Cheng Quek[1,2], Daniel W. Thomson[1], Jesper L.V. Maag[1,2], Nenad Bartonicek[1], Bethany Signal[1], Michael B. Clark[1,3], Brian S. Gloss[1,2,*] and Marcel E. Dinger[1,2,*]

[1]Garvan Institute of Medical Research, 384 Victoria Street, Sydney, NSW 2010, Australia, [2]St Vincent's Clinical School, University of New South Wales, Sydney, NSW 2052, Australia and [3]MRC Functional Genomics Unit, Department of Physiology, Anatomy, and Genetics, University of Oxford, Oxford OX1 3PT, UK

Received September 15, 2014; Accepted October 5, 2014

ABSTRACT

Despite the prevalence of long noncoding RNA (lncRNA) genes in eukaryotic genomes, only a small proportion have been examined for biological function. lncRNAdb, available at http://lncrnadb.org, provides users with a comprehensive, manually curated reference database of 287 eukaryotic lncRNAs that have been described independently in the scientific literature. In addition to capturing a great proportion of the recent literature describing functions for individual lncRNAs, lncRNAdb now offers an improved user interface enabling greater accessibility to sequence information, expression data and the literature. The new features in lncRNAdb include the integration of Illumina Body Atlas expression profiles, nucleotide sequence information, a BLAST search tool and easy export of content via direct download or a REST API. lncRNAdb is now endorsed by RNAcentral and is in compliance with the International Nucleotide Sequence Database Collaboration.

INTRODUCTION

The last decade has provided compelling evidence for the function of RNA beyond its canonical role as a messenger for protein-coding genes. Long noncoding RNAs (lncRNAs) are transcripts greater than 200 nucleotides in length with little or no protein-coding potential (1–3). This arbitrary size threshold, which was incidentally defined by the characteristics of common nucleic acid purification protocols, pragmatically distinguishes lncRNAs from other distinct classes of small RNAs such as microRNAs, tRNAs and snoRNAs. From the earliest descriptions of biologically important lncRNAs such as H19 and XIST almost two decades ago, the last few years have seen rapid growth in the functional explorations of individual lncRNAs. Concomitant with this increased growth of characterized lncRNAs is an increasing understanding toward biological mechanisms, as well as a growing awareness and recognition of the importance of lncRNAs in virtually every cellular and regulatory process (4).

Although initially triggered by high-throughput cDNA cloning and tiling microarrays, discovery of lncRNAs is now largely driven by next-generation sequencing of whole transcriptomes and, more recently, target enrichment of rare or lowly expressed transcripts (5). Currently, GENCODE (v20) conservatively annotates 14 470 independent lncRNA genes in the human genome (6). The implication of widespread functionality of all these lncRNAs, based only on the confirmed expression of their transcripts, remains an area of some controversy. However, the evidence of generic hallmarks of functionality of lncRNAs, such as sequence conservation, highly specific and regulated expression, association with epigenetic control elements, alternate splicing and differential stability, are accumulating (7). This argues against the dismissal of lncRNAs as transcriptional noise or artifact.

The expanding list of lncRNAs and accumulating functional evidence has necessitated a coherently curated database to act as a data repository and a platform for lncRNA research. By updating lncRNAdb (8), version 2.0 aims to grow its momentum as the most cited and up-to-date reference database of lncRNAs. Other lncRNA databases that have been released since the inception of lncRNAdb focus less on providing manually curated literature evidence on lncRNA functionality, but offer complementary tools for the analysis of lncRNAs. For example, algorithms for finding microRNAs targeting lncRNAs can be accessed at DIANA-LncBase (9) and starBase v2.0 (10), chromatin state of lncRNAs can be investigated at ChIPBase (11) and the ability for lncRNAs to act as competitive endogenous RNA (ceRNA) can be investigated at lnCeDB (12). lncRNAdb remains the only expertly curated reference database of biologically investigated lncRNAs and accord-

*To whom correspondence should be addressed. Tel: +61293555860; Fax: +61293555871; Email: m.dinger@garvan.org.au.
Correspondence may also be addressed to Brian S. Gloss. Tel: +61293555768; Fax: +61293555871; Email: b.gloss@garvan.org.au

ingly serves as a source for other integrative databases, such as RNAcentral (13) and NONCODE (14).

AIMS OF THE DATABASE

In response to the need for a repository of lncRNA sequences and supporting data, lncRNAdb aims to summarize our knowledge of eukaryotic lncRNAs in an easily accessible and searchable format. lncRNAdb provides an interface to researchers that allows for easy access via a web browser and also for automated queries through a REST API. lncRNAdb includes a variety of annotations for eukaryotic lncRNAs, including gene expression data, evolutionary conservation, structural information, genomic context, subcellular localization, functional evidence, links to the primary literature and the transcript sequence.

Entries into lncRNAdb are curated from evidence supported by the literature. This distinguishes this database from other lncRNA databases that, for example, aggregate data from diverse (often uncredited) sources, or supply computational tools to display and interpret datasets or prediction algorithms.

LNCRNADB V2.0

Since its launch, lncRNAdb has been widely accepted as a valuable catalog of biologically validated lncRNAs. For instance, the HUGO Gene Nomenclature Committee (HGNC) has included lncRNAdb as part of their lncRNA specific resources. Seventy-six of the 110 lncRNA entries on HUGO cite lncRNAdb (http://www.genenames.org/rna/LNCRNA).

As of August 2014, lncRNAdb has been inducted into RNAcentral as a third party data specialist database. RNAcentral is a network of resources that provides unified access to noncoding RNA sequence data supplied by external expert databases (13). Inclusion of lncRNAdb in RNAcentral requires compliance with guidelines set by the International Nucleotide Sequence Database Collaboration (INSDC). lncRNAdb entries exported to RNAcentral have been given an ENA TPA accession ID and its content can be readily obtained on lncRNAdb or RNAcentral (Supplementary Data 1).

NEW FEATURES

New entries

We have added a total of 87 new entries to the database, and existing entries have been updated to reflect recent literature. These changes are based on information derived via manual curation from 382 new publications. In total lncRNAdb now holds 283 entries, informed by 921 references and 260 nucleotide sequences (Figure1). These cover entries across 71 different organisms.

New user interface

A new user interface with expanded features has been included to promote easily searchable and downloadable content, queried through lncRNA name, tissue or disease association (Figure 2). Entry pages are presented in an accordion-style format that allows users to expand or collapse various sections of the content. All records are available for download either as a useful printer-friendly summary or as an XML record for easy programmatic access.

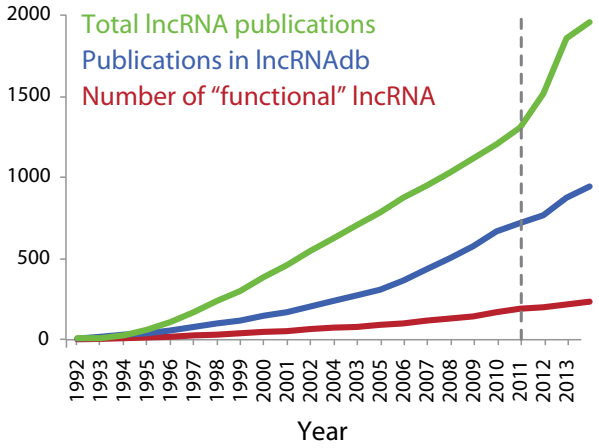

Figure 1. Coverage of the literature by lncRNAdb v2.0. Cumulative totals of all publications matching the search term 'long noncoding RNA' [MeSH] were extracted from PubMed from 1992–2013 (green) and the proportion incorporated into lncRNAdb as ascribing functional annotation to lncRNAs is shown (blue). Cumulative totals of the number of lncRNAs found in lncRNAdb described in literature (red).

Sequence search capabilities

In addition to word search tools, lncRNAdb v2.0 includes incorporation of a BLAST (Basic Local Alignment Search Tool) server for sequence-alignment search (15). On input of a query sequence by the user, lncRNAdb will return any entries that have significant similarity with the query sequence. The user also has the option to download the full-text result of the BLAST search (Supplementary Figure S1).

Incorporation of gene expression data

For entries with a corresponding human Ensembl Gene ID, expression data from the Illumina Body Atlas is available (16). This feature provides an overview of the expression of the selected lncRNA in 16 human tissues. Data from the human body were generated via the Tuxedo suite (17) using the Gencode V15 Gene model and can be exported in XML format. For details of the analysis pipeline, see http://www.lncrnadb.org/help#BodyAltas.

Improved data accessibility

To enable easily downloadable content, lncRNAdb v2.0 includes a REST API for users to download raw data files programmatically. Content is available in XML, which is easily convertible to other formats, such as BED, FASTA and GTF. To ensure high integrity of nucleotide sequences, we provide corresponding International Nucleotide Sequence Database Collaboration (INSDC) IDs, and link out sequences to the European Nucleotide Archive (ENA). To ensure compliancy, the entries are now annotated with a corresponding ENA TPA. Content from pages can be exported in XML or printer-friendly format (Figure 2).

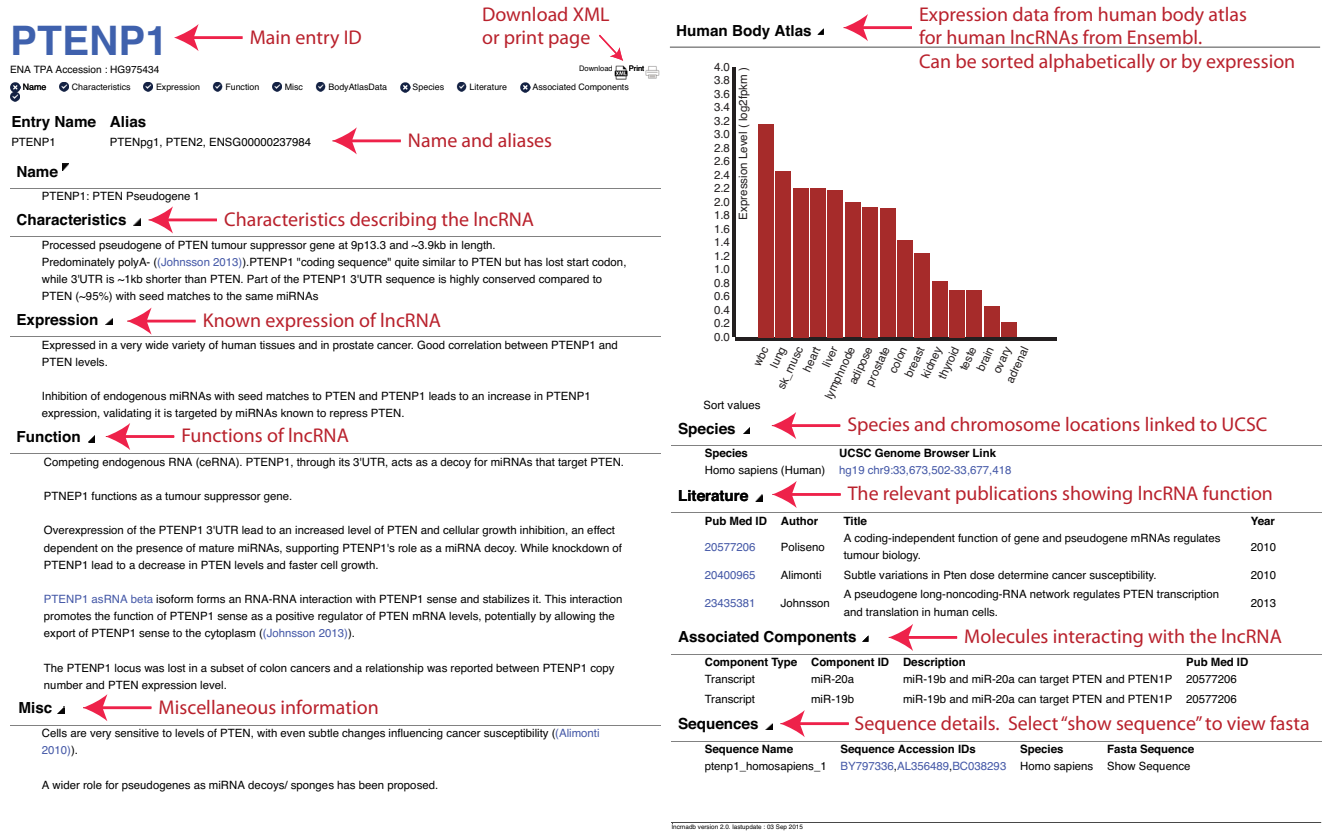

Figure 2. The lncRNAdb v2.0 user interface. A screenshot of an example profile highlighting new features.

Finally, a major improvement from the previous lncRNAdb release is the REST API. This feature was added due to a number of citations of the first edition of lncRNAdb from databases and publications that rely on programmatic data export from lncRNAdb. The API enables access to XML records in three levels, depending on the amount of requested content and the level of detail. In the simplest form, the user can select either the whole record (e.g. http://lncrnadb.org/rest/hotair) or specific content (e.g. http://lncrnadb.org/rest/hotair/sequence) for an individual lncRNA. The next level allows access to multiple entries at once. For example, the query http://lncrnadb.org/rest/search/brain+cancer/nomenclature/literature finds all the literature records for lncRNAs that are associated with brain cancer. Finally, the users can retrieve specific information for all entries, such as associated interacting components http://lncrnadb.org/rest/all/association. More information with examples can be found at http://lncrnadb.org/tools/.

User submission capacity

To assist in maintaining an informed repository of data, lncRNAdb provides an avenue for user submissions. New entries can be posted on the submission page with supporting information through a CAPTCHA-protected form (Supplementary Figure S2). All user-submitted data is processed by an expert human curator before incorporation into the database. A detailed description of the process and acceptance criteria for lncRNAdb contributions is available at http://lncrnadb.org/contribution. As the pace of lncRNA functional characterization continues to increase, we anticipate user-submitted data will become more crucial in keeping lncRNAdb up to date. We therefore encourage any researchers with newly published lncRNA data, or who find their discoveries are not included in the database, to submit their entry to lncRNAdb.

TOPICAL HIGHLIGHTS IN LNCRNA RESEARCH

Reported functions of lncRNAs

Reflective of the diversity of lncRNA size and structural characteristics, the numerous lncRNA functions described within lncRNAdb seldom fit into a discrete set of classifications. Among the heterogenous functions described, lncRNAs are capable of functioning as chromatin regulators (18,19), enhancer RNAs (20), nuclear scaffolds (21), snoRNA host genes (22), primary microRNA transcripts, pre-tRNAs, ceRNAs to sequester microRNAs (23) or the transcriptional machinery away from other genes. Even in terms of genomic context, lncRNAs evade ready categorization, with an individual lncRNA locus capable of comprising intergenic transcripts, overlapping transcripts, antisense transcripts and bidirectional transcripts.

To further confound easy categorization, individual lncRNA loci are not restricted to a single purpose. For example, the lncRNA *SNHG1* is a host to eight functional

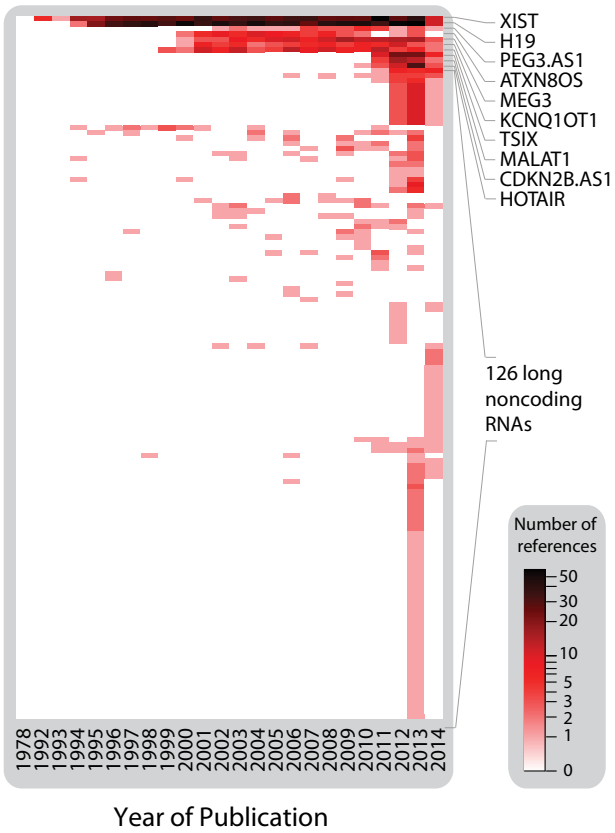

Figure 3. Heatmap showing the number of references within PubMed with the term 'RNA, long noncoding' [MeSH] for each given year. The top 10 most studied lncRNAs are named and color scale is logarithmic. Nomenclature information of all noncoding RNA from was obtained from HGNC. Entries without search results were removed. The remaining entries were visualized with a heatmap constructed using R Package 'gplots' (v.2.14.1).

snoRNAs, at least one of which (SNORD25) is known to produce a miRNA (24). In principle, the same lncRNA transcript may also act as a ceRNA as an enhancer RNA and as a structural scaffold. The ability of a single locus to give rise to transcripts with multiple functions is not unique to lncRNAs (25). For example, the mRNA *KANSL2* is host to three snoRNAs, of which SNORA34 is a precursor to miR-1291.

The opinion has been put forward that evolutionary pressure to develop more sophisticated regulatory mechanisms has led to the requirement of a more complex transcriptome and consequently a greater number of lncRNAs. Evidence of a rapid expansion of lncRNA numbers and diversity over the recent period of primate evolution (26–28) supports this.

lncRNAs have minimal protein-coding capacity

LncRNAs were first described as a class in conjunction with early large-scale sequencing libraries of cDNA clones (29). At this time, assessment of coding potential was deduced mostly via assessment of open reading frames (ORFs). Because of the limitations of this approach the definition of 'noncoding' has remained ambiguous for many transcripts (2,30). More recent efforts to empirically determine the protein-coding ability add to this ambiguity by yielding reports that some annotated lnRNAs give rise to polypeptides (31). Counter to these observations is the growing body of evidence supporting that the protein-coding capacity for lncRNAs is minimal to absent. This includes data from bioinformatic assessment of ORFs and codon conservation frequency, as well as experimental assessment of ribosome occupancy using ribosome profiling (32) and mass spectrometry compared to RNAseq data (33,34).

Supporting evidence of noncoding RNA function

Assuming the absence of appreciable protein-coding capacity, any biological functionality held by lncRNAs is considered to be manifested at the RNA level. The majority of annotated lncRNAs do not have clearly defined functions. However, evidence from transcriptomic studies looking at lncRNAs as a class is highly suggestive of the functions of lncRNAs. This includes evidence surrounding evolutionary conservation, developmental- and tissue-specific expression, RNA structure and subcellular localization.

Evolutionary conservation

Although IncRNAs are under lower selective pressure than protein-coding genes, they are under higher selective pressure than repeat sequences that are considered to be under neutral selection (34). Interestingly, the promoters of IncRNAs display similar levels of conservation to that of coding genes (35).

RNA structure and sequence conservation

Due to the intrinsic differences in the encoding of structural information between protein-coding and noncoding genes, the associated primary sequences are subject to different evolutionary constraints. That is, in the case of protein-coding sequences, triplet nucleotides (codons) encode specific amino acids, where either single nucleotide polymorphism or insertions/deletions can drastically change or entirely prevent the production of a functional protein. In contrast, lncRNAs, which inherently encode RNA structures, may be considerably more resilient to sequence variation, where insertions/deletions may have little impact on structure and polymorphisms tolerated by complementary changes at partner folding sites. Therefore, if lncRNA function is dependent more on its structure than its primary sequence, significant conservation at the sequence level may be difficult to detect or entirely eroded through evolution, despite conservation of function. This hypothesis is supported by global investigations on the structure of lncRNAs, which indicate that it is evolutionarily conserved (36). The importance of secondary structure for function is exemplified by XIST, which maintains silencing of the inactive X chromosome by exploiting the three-dimensional conformation of the regions of the X-chromosome, not by specific sequences (37).

Specific expression and subcellular localization

Multiple studies have shown that lncRNA expression is more cell type and developmentally specific than that of

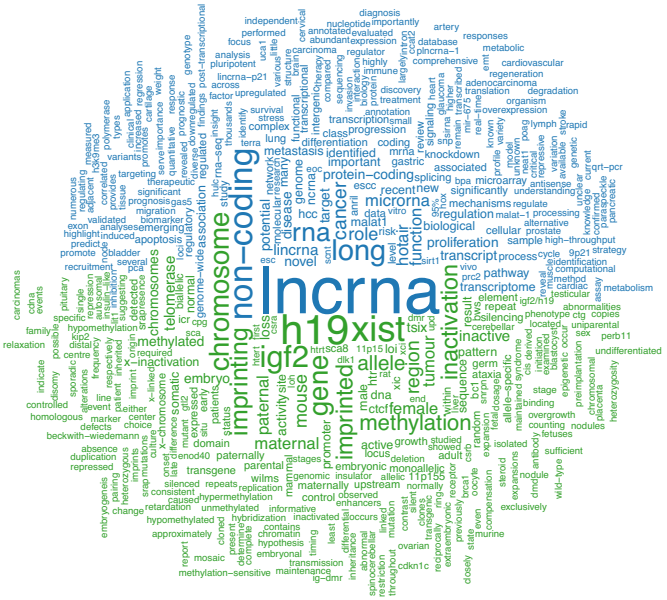

Figure 4. Word cloud collating the top 500 terms within abstracts found in PubMed, following a search with the term 'long noncoding RNA' [MeSH]. Terms appearing more frequent in publications before years 2011 are labeled green terms more frequent after in publications after year 2012 are labeled blue. The sizes of the words represent the difference in frequencies of the word appearing in the two sets of abstracts. Word frequencies are normalized against the number of publications in each set. Text preprocessing was conducted by an in-house script (available at http://lncrnadb.org/help) and the word cloud was created with R package 'tm' (v.0.6) and 'wordcloud' (v.2.5).

protein-coding genes (30,38–40). Moreover, as lncRNAs are more likely to be localized to the nucleus than coding transcripts it is suggestive of more regulatory roles of lncRNAs (41).

TRENDS IN LNCRNA RESEARCH

GENCODE (v20) annotated 14 470 independent lncRNAs in the human genome (6). However, other sources suggest numbers as high as 95 135 lncRNAs (42). Only a fraction of these have been functionally characterized. The majority of literature describing function of lncRNAs has been spearheaded by a few well-characterized lncRNAs, namely H19, XIST, HOTAIR, NEAT1 and MALAT1. This is evident by observing the number of articles for any given year that includes the search term 'lncRNA' (Figure 3). In recent years, the focus of lncRNA research expands to include a broader range of lncRNA genes. This trend is presumably due in large part to increased discovery rates driven by the availability of low-cost RNA-sequencing.

A switch of focus from the highly studied lncRNAs H19 and XIST can further be noticed by observing the search terms that accompany 'Long noncoding RNA' [MeSH] in literature via PubMed, comparing the terms from before the first release of lncRNAdb (August 2010) to those after. It can be observed that the term specific to lncRNAs H19 and XIST are replaced with terms describing functionality or mechanism of action such as 'function', 'cancer', 'differentiation', 'pathway', 'metastasis', 'role' and 'microRNA' (Figure 4). The increase of functionally annotated lncRNAs and their annotation within the lncRNAdb can also be seen (Figure 1).

CONCLUDING REMARKS

Our understanding of lncRNA function has largely been directed by the study of the first few discovered lncRNAs, the first of which was described almost two decades ago. The trend of lncRNA research is changing, presumably due to the increased availability of genomic and transcriptomic data. Research now includes a broader range of lncRNAs and a greater variety of mechanisms of action. Expecting this trend to continue, we anticipate many more lncRNAs to be supported with functional data, which in turn will prompt the continued demand for an easily accessible expert curated database. This remains the ambition behind lncRNAdb into the future.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The lncRNAdb team would like to acknowledge the following people for their contribution toward the new release of lncRNAdb. We would like to thank Dr Warren Kaplan and Derrick Lin from the Kinghorn Centre for Clinical Genomic at the Garvan Institute for provision of the required web infrastructure. Dr Richard Gibson and Dr Simon Kay from RNA Central, EBI for the inclusion of lncRNAdb as part of RNAcentral. Kenneth S. Sabir from the BioData Visualization at Garvan institute of Medical Research and Christina Stolte for their expert knowledge in web design and performance. Finally, we thank Dr Mark Cowley and Dr Mark McCabe for their helpful feedback on the project.

FUNDING

NHMRC Project Grant (APP1043971); CINSW Early Career Fellowship (to B.S.G.); NHMRC Early Career Fellowship (to M.B.C.).
Conflict of interest statement. None declared.

REFERENCES

1. Wang,K.C. and Chang,H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell*, **43**, 904–914.
2. Mercer,T.R., Dinger,M.E. and Mattick,J.S. (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.
3. Kapranov,P., Cheng,J., Dike,S., Nix,D.A., Duttagupta,R., Willingham,A.T., Stadler,P.F., Hertel,J., Hackermuller,J., Hofacker,I.L. *et al.* (2007) RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, **316**, 1484–1488.
4. Amaral,P.P., Dinger,M.E., Mercer,T.R. and Mattick,J.S. (2008) The eukaryotic genome as an RNA machine. *Science*, **319**, 1787–1789.
5. Mercer,T.R., Gerhardt,D.J., Dinger,M.E., Crawford,J., Trapnell,C., Jeddeloh,J.A., Mattick,J.S. and Rinn,J.L. (2012) Targeted RNA sequencing reveals the deep complexity of the human transcriptome. *Nat. Biotechnol.*, **30**, 99–104.
6. Harrow,J., Frankish,A., Gonzalez,J.M., Tapanari,E., Diekhans,M., Kokocinski,F., Aken,B.L., Barrell,D., Zadissa,A., Searle,S. *et al.* (2012) GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.*, **22**, 1760–1774.

7. van Bakel,H., Nislow,C., Blencowe,B.J. and Hughes,T.R. (2010) Most 'dark matter' transcripts are associated with known genes. *PLoS Biol.*, **8**, e1000371.
8. Amaral,P.P., Clark,M.B., Gascoigne,D.K., Dinger,M.E. and Mattick,J.S. (2011) lncRNAdb: a reference database for long noncoding RNAs. *Nucleic Acids Res.*, **39**, D146–D151.
9. Paraskevopoulou,M.D., Georgakilas,G., Kostoulas,N., Reczko,M., Maragkakis,M., Dalamagas,T.M. and Hatzigeorgiou,A.G. (2013) DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res.*, **41**, D239–D245.
10. Li,J.H., Liu,S., Zhou,H., Qu,L.H. and Yang,J.H. (2014) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.*, **42**, D92–D97.
11. Yang,J.H., Li,J.H., Jiang,S., Zhou,H. and Qu,L.H. (2013) ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res.*, **41**, D177–D187.
12. Das,S., Ghosal,S., Sen,R. and Chakrabarti,J. (2014) lnCeDB: database of human long noncoding RNA acting as competing endogenous RNA. *PloS One*, **9**, e98965.
13. Bateman,A., Agrawal,S., Birney,E., Bruford,E.A., Bujnicki,J.M., Cochrane,G., Cole,J.R., Dinger,M.E., Enright,A.J., Gardner,P.P. *et al.* (2011) RNAcentral: a vision for an international database of RNA sequences. *RNA*, **17**, 1941–1946.
14. Bu,D., Yu,K., Sun,S., Xie,C., Skogerbo,G., Miao,R., Xiao,H., Liao,Q., Luo,H., Zhao,G. *et al.* (2012) NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res.*, **40**, D210–D215.
15. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
16. Wagner,F., Heidtke,K.R., Drescher,B. and Radelof,U. (2007) Development and perspectives of scientific services offered by genomic biological resource centres. *Brief. Funct. Genomic. Proteomic.*, **6**, 163–170.
17. Trapnell,C., Roberts,A., Goff,L., Pertea,G., Kim,D., Kelley,D.R., Pimentel,H., Salzberg,S.L., Rinn,J.L. and Pachter,L. (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.*, **7**, 562–578.
18. Rinn,J.L., Kertesz,M., Wang,J.K., Squazzo,S.L., Xu,X., Brugmann,S.A., Goodnough,L.H., Helms,J.A., Farnham,P.J., Segal,E. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323.
19. Mercer,T.R. and Mattick,J.S. (2013) Structure and function of long noncoding RNAs in epigenetic regulation. *Nat. Struct. Mol. Biol.*, **20**, 300–307.
20. Kim,T.K., Hemberg,M., Gray,J.M., Costa,A.M., Bear,D.M., Wu,J., Harmin,D.A., Laptewicz,M., Barbara-Haley,K., Kuersten,S. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182–187.
21. Bond,C.S. and Fox,A.H. (2009) Paraspeckles: nuclear bodies built on long noncoding RNA. *J. Cell Biol.*, **186**, 637–644.
22. Askarian-Amiri,M.E., Crawford,J., French,J.D., Smart,C.E., Smith,M.A., Clark,M.B., Ru,K., Mercer,T.R., Thompson,E.R., Lakhani,S.R. *et al.* (2011) SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *RNA*, **17**, 878–891.
23. Salmena,L., Poliseno,L., Tay,Y., Kats,L. and Pandolfi,P.P. (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, **146**, 353–358.
24. Xia,J., Joyce,C.E., Bowcock,A.M. and Zhang,W. (2013) Noncanonical microRNAs and endogenous siRNAs in normal and psoriatic human skin. *Hum. Mol. Genet.*, **22**, 737–748.
25. Dinger,M.E., Gascoigne,D.K. and Mattick,J.S. (2011) The evolution of RNAs with multiple functions. *Biochimie*, **93**, 2013–2018.
26. Taft,R.J., Pheasant,M. and Mattick,J.S. (2007) The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays*, **29**, 288–299.
27. Liu,G., Mattick,J.S. and Taft,R.J. (2013) A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle*, **12**, 2061–2072.
28. Guennewig,B. and Cooper,A.A. (2014) The central role of noncoding RNA in the brain. *Int. Rev. Neurobiol.*, **116**, 153–194.
29. Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature*, **420**, 563–573.
30. Dinger,M.E., Pang,K.C., Mercer,T.R. and Mattick,J.S. (2008) Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput. Biol.*, **4**, e1000176.
31. Gascoigne,D.K., Cheetham,S.W., Cattenoz,P.B., Clark,M.B., Amaral,P.P., Taft,R.J., Wilhelm,D., Dinger,M.E. and Mattick,J.S. (2012) Pinstripe: a suite of programs for integrating transcriptomic and proteomic datasets identifies novel proteins and improves differentiation of protein-coding and non-coding genes. *Bioinformatics*, **28**, 3042–3050.
32. Ingolia,N.T., Lareau,L.F. and Weissman,J.S. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**, 789–802.
33. Banfai,B., Jia,H., Khatun,J., Wood,E., Risk,B., Gundling,W.E. Jr, Kundaje,A., Gunawardena,H.P., Yu,Y., Xie,L. *et al.* (2012) Long noncoding RNAs are rarely translated in two human cell lines. *Genome Res.*, **22**, 1646–1657.
34. Derrien,T., Johnson,R., Bussotti,G., Tanzer,A., Djebali,S., Tilgner,H., Guernec,G., Martin,D., Merkel,A., Knowles,D.G. *et al.* (2012) The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.*, **22**, 1775–1789.
35. Ponjavic,J., Ponting,C.P. and Lunter,G. (2007) Functionality or transcriptional noise? Evidence for selection within long noncoding RNAs. *Genome Res.*, **17**, 556–565.
36. Smith,M.A., Gesell,T., Stadler,P.F. and Mattick,J.S. (2013) Widespread purifying selection on RNA structure in mammals. *Nucleic Acids Res.*, **41**, 8220–8236.
37. Engreitz,J.M., Pandya-Jones,A., McDonel,P., Shishkin,A., Sirokman,K., Surka,C., Kadri,S., Xing,J., Goren,A., Lander,E.S. *et al.* (2013) The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science*, **341**, 720–721.
38. Cabili,M.N., Trapnell,C., Goff,L., Koziol,M., Tazon-Vega,B., Regev,A. and Rinn,J.L. (2011) Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.*, **25**, 1915–1927.
39. 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.
40. Mercer,T.R., Dinger,M.E., Sunkin,S.M., Mehler,M.F. and Mattick,J.S. (2008) Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl Acad. Sci. U.S.A.*, **105**, 716–721.
41. Djebali,S., Davis,C.A., Merkel,A., Dobin,A., Lassmann,T., Mortazavi,A., Tanzer,A., Lagarde,J., Lin,W., Schlesinger,F. *et al.* (2012) Landscape of transcription in human cells. *Nature*, **489**, 101–108.
42. Xie,C., Yuan,J., Li,H., Li,M., Zhao,G., Bu,D., Zhu,W., Wu,W., Chen,R. and Zhao,Y. (2014) NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.*, **42**, D98–D103.