Published online 19 November 2015 *Nucleic Acids Research, 2016, Vol. 44, Database issue* ***D203–D208***
doi: 10.1093/nar/gkv1252

NONCODE 2016: an informative and valuable data source of long non-coding RNAs

Yi Zhao[1,†], Hui Li[2,3,†], Shuangsang Fang[2,3], Yue Kang[3,4], Wei wu[3,4], Yajing Hao[3,4], Ziyang Li[2], Dechao Bu[2], Ninghui Sun[2], Michael Q. Zhang[1,*] and Runsheng Chen[4,*]

[1]School of Medicine, MOE Key Laboratory of Bioinformatics and Bioinformatics Division, Center for Synthetic and System Biology, TNLIST/Department of Automation, Tsinghua University, Beijing 100084, China, [2]Key Laboratory of Intelligent Information Processing, Advanced Computer Research Center, State Key Laboratory of Computer Architecture, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China, [3]University of Chinese Academy of Sciences, Beijing 100049, China and [4]CAS Key Laboratory of RNA Biology, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

Received September 15, 2015; Revised October 29, 2015; Accepted October 31, 2015

ABSTRACT

NONCODE (http://www.bioinfo.org/noncode/) is an interactive database that aims to present the most complete collection and annotation of non-coding RNAs, especially long non-coding RNAs (lncRNAs). The recently reduced cost of RNA sequencing has produced an explosion of newly identified data. Revolutionary third-generation sequencing methods have also contributed to more accurate annotations. Accumulative experimental data also provides more comprehensive knowledge of lncRNA functions. In this update, NONCODE has added six new species, bringing the total to 16 species altogether. The lncRNAs in NONCODE have increased from 210 831 to 527,336. For human and mouse, the lncRNA numbers are 167,150 and 130,558, respectively. NONCODE 2016 has also introduced three important new features: (i) conservation annotation; (ii) the relationships between lncRNAs and diseases; and (iii) an interface to choose high-quality datasets through predicted scores, literature support and long-read sequencing method support. NONCODE is also accessible through http://www.noncode.org/.

INTRODUCTION

Recent whole transcriptome studies have revealed that about three quarters of the human genome is capable of being transcribed, while protein-coding regions account for just 2% of the genome (1–3). Therefore, the vast majority of transcribed sequences do not encode proteins, and are called non-coding RNA (ncRNA). Accumulating evidence shows that non-coding RNAs play key roles in various biological processes, such as imprinting control, the circuitry controlling pluripotency and differentiation, immune responses, and chromosome dynamics (4). ncRNAs are as important as protein-coding genes to cellular functions (5,6). Notably, a growing number of long ncRNAs (lncRNAs), which are considered to be >200 nt in length and are often multiexonic (7), have been implicated in disease etiology (8–10). It is therefore of great importance to collect lncRNA information and store this information in a one-stop knowledge gateway for lncRNAs, the NONCODE database.

The development of high-throughput sequencing methodologies has reduced the cost of RNA sequencing, and as a result there has been an explosive rise in the number of newly identified lncRNAs. For example, in 2015, Chinnaiyan *et al*. established a consensus set of 384 066 predicted transcripts from 7256 RNA-seq libraries, which were designated as the MiTranscriptome assembly (11). Since then the revolutionary advancement of sequencing methods, such as single-molecule long-read techniques, leads us closer to the real lncRNA transcriptome. Given sufficient material, amplification-free sequencing of full-length cDNA molecules provides a more direct view of RNA molecules (12). NONCODE has collected data from literature published since the last update and includes the latest versions of several public databases (Ensembl (13), RefSeq(14), lncRNAdb (15) and GENCODE (16)). After the removal of false and redundant lncRNAs, NONCODE contains a total of 527,336 transcripts.

In addition to the identification of new lncRNAs, data on the genetics and biochemical properties of lncRNAs has accumulated rapidly. Of the papers retrieved from Pubmed for lncRNAs, we found that the vast majority studied lncRNA

[*]To whom correspondence should be addressed. Tel: +86 10 6488 8543; Fax: +86 10 6487 7837; Email: chenrs@sun5.ibp.ac.cn
Correspondence may also be addressed to Michael Zhang. Tel: +86 10 6279 5993; Fax: +86 10 6278 6911; Email: michaelzhang@tsinghua.edu.cn
[†]These authors contributed equally to the paper as first authors.

function, especially the relationship between lncRNAs and disease (8–10). In large-scale searches for single-base differences between diseased and healthy individuals, about 40% of the disease-related differences show up in genomic regions outside of protein-coding genes. This implicates non-coding regions as vital for genetic risk factors of disease (2). In order to enable a systematic compilation and integration of this information, we added the relationships between lncRNAs and diseases to the annotations of NONCODE. The sources for these annotations were derived from literature mining, differential lncRNA analysis utilizing public RNA-seq data and microarray data and mutation analysis from public genome-wide association study (GWAS) data.

Along with the ever increasing number of lncRNAs and the amount and functional study data, genome-wide conservation information is required for biologists to study the mechanisms of lncRNA actions. In order to explore the conservation information of lncRNAs, NONCODE collected six new mammalian species (chimpanzee, gorilla, orangutan, rhesus macaque, opossum and platypus) (17). Conservation annotation is available on the information page of each NONCODE lncRNA gene. Users can browse the conserved counterparts of any human lncRNA gene in other species through a phylogenetic tree layout. This conservation information should greatly increase the convenience of studying lncRNA functions.

DATA COLLECTION AND PROCESSING

Similar to the former iterations of NONCODE (18–21), the source of NONCODE 2016 includes the previous versions of NONCODE, the collated literature and other public databases. We searched PubMed using the key words ‘ncrna’, ‘noncoding’, ‘non-coding’, ’no code’, ‘non-code’, ‘lncrna’ and ‘lincrna’, and found 6532 new articles since 1 June 2013 (the last collection date for NONCODE). We retrieved the newly identified lncRNAs and their annotations from the supplementary material or web site of these articles. Together with the newest data from Ensembl, RefSeq, lncRNAdb, GENCODE and the old versions of NONCODE data, literature data were processed through a standard pipeline for each species. The pipeline included the following six steps:

(i) Format normalization. All of the input data were processed into bed or gtf formats based on one assembly version, for example, hg38 for human and mm10 for mouse.
(ii) Combination. All of the normalized data files were combined together using the Cuffcompare program in the Cufflinks suite (22). After eliminating redundancy, every new transcript ID and the accompanying resources were extracted.
(iii) Filtering protein-coding RNA. We filtered out protein-coding RNA using two methods. Firstly, the RNA was compared with the coding RNA in RefSeq and Ensembl, and the ‘=’ and ‘c’ transcripts were excluded. Secondly, the RNA was filtered through the Coding-Non-Coding Index (CNCI) (23) program and only the RNAs considered non-coding by CNCI were kept.
(iv) Information retrieval. We assigned each transcript a name according to the criterion of NONCODE v4 and extracted basic information such as location (24), exons, length, assembly sequence, source, etc.
(v) Advanced annotation. Advanced annotations included expression profiles, predicted functions, conservation, disease information, etc. Human expression profiles were collected from 16 tissues of the Human BodyMap 2.0 data (ENA archive: ERP000546) and eight cell lines (GEO accession no. GSE30554), while mouse data was collected from six different tissues (ENA archive: ERP000591). Functions for the lncRNA genes were predicted by lnc-GFP (25), a coding–non-coding co-expression network (26,27) based global function predictor.
(vi) Web presence. The new NONCODE has provided completely new web pages. More annotation information has been added and a more user-friendly interface has been introduced.

STATISTICS OF NONCODE

NONCODE contains 527,336 lncRNA transcripts from 16 species (human, mouse, cow, rat, chimpanzee, gorilla, orangutan, rhesus macaque, opossum, platypus, chicken, zebrafish, fruitfly, *Caenorhabditis elegans*, yeast and Arabidopsis,). According to the definition of lncRNA genes (18), NONCODE collected 337,880 genes altogether. A total of 101,700 and 86,935 genes were generated from 167,150 and 130,558 lncRNAs from human and mouse (shown in Table 1), respectively. Following the nomenclature of NONCODE v4 (18), both lncRNA transcripts and genes were designated systematically: NON+ three characters (representing a species) +T (transcript) or G (gene) + six sequential numbers. NONCODE has annotated expression profiles from all the human and mouse transcripts and genes, and a large number of these genes were annotated with predicted functions.

LNCRNAS AND DISEASES

Definitive evidence has proven that transcription of the non-coding genome has produced functional RNAs (1). In particular, lncRNAs have been implicated in biological, developmental, and pathological processes, and acted through mechanisms such as chromatin reprogramming, cis regulation at enhancers, and post-transcriptional regulation of mRNA processing (28). lncRNAs are therefore considered to be important regulators of tissue physiology and disease processes including cancer (11). Although we have collected functional interactions between ncRNAs and biomolecules in NPInter (29–31), we think it is also necessary to include disease information into NONCODE.

The data retrieval pipeline is listed in Figure 1. Recent published papers have been explored, and the proven associations between NONCODE transcripts and diseases has been integrated into the latest version. The NONCODE assembly also assessed the overlaps of transcripts with the unique disease-associated Single-Nucleotide Polymorphisms (SNPs) from a catalog of GWASs (32) and the SNP database (dbSNP) (33). There were also a lot of rela-

Table 1. Transcript and gene statistics for NONCODE

Species		Number of lncRNA transcripts	Number of lncRNA genes
human	*Homo sapiens*	167,150	101,700
mouse	*Mus musculus*	130,558	86,935
cow	*Bos taurus*	23 599	18 189
rat	*Rattus rattus*	29 070	25 114
chimpanzee	*Pan troglodytes*	18 604	13 224
gorilla	*Gorilla gorilla*	20 785	17 140
orangutan	*Pongo pygmaeus*	15 601	13 432
rhesus macaque	*Macaca mulatta*	9325	6125
opossum	*Monodelphis domestica*	21 014	14 135
platypus	*Ornithorhynchus anatinus*	11 518	9394
chicken	*Gallus gallus*	13 085	9688
zebrafish	*Danio rerio*	5000	3635
fruitfly	*Drosophila melanogaster*	54 818	13 890
C. elegans	*Caenorhabditis elegans*	3269	2746
yeast	*Saccharomyces cerevisiae*	60	56
Arabidopsis	*Arabidopsis thaliana*	3853	2477
Total		527,336	337,880

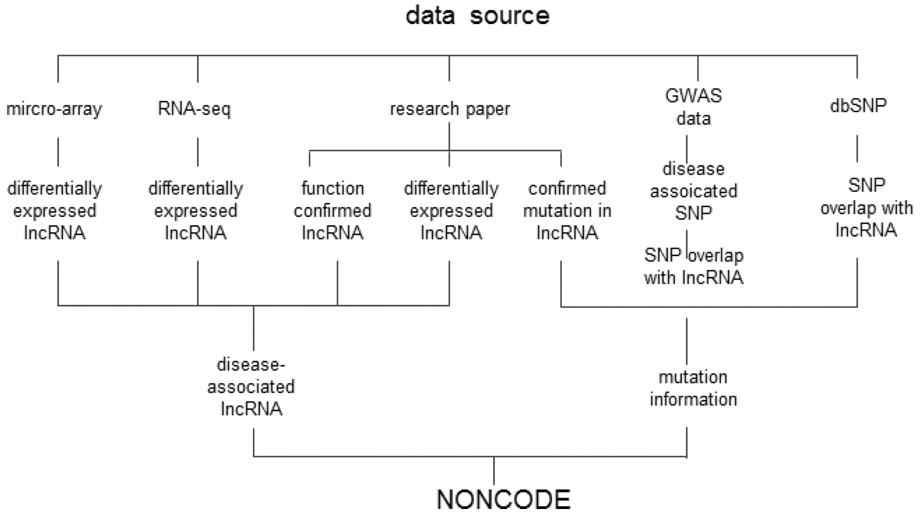

data source
mircro-array
RNA-seq
research paper
GWAS data
dbSNP
differentially expressed lncRNA
differentially expressed lncRNA
function confirmed lncRNA
differentially expressed lncRNA
confirmed mutation in lncRNA
disease assoicated SNP
SNP overlap with lncRNA
SNP overlap with lncRNA
disease-associated lncRNA
mutation information
NONCODE

Figure 1. Disease related data acquisition pipeline

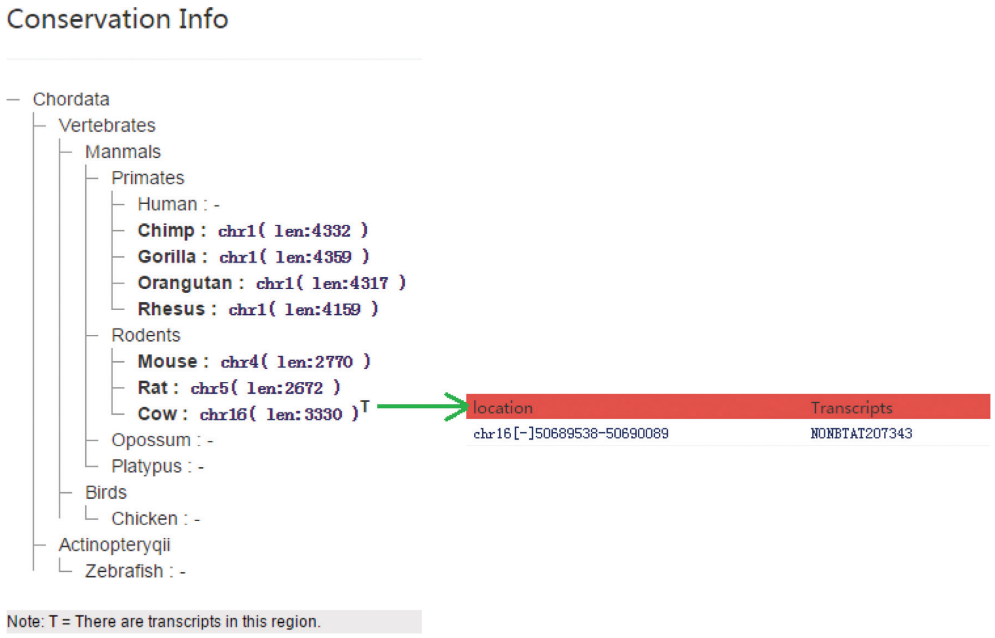
tional data between lncRNAs and diseases which were analyzed from RNA-seq and microarray data. After collecting the basic data, we compared it with the lncRNAs in NONCODE and retained data that overlapped with NONCODE lncRNAs. NONCODE 2016 contains 1110 lncRNAs which were related to 284 diseases. Among these associations, 153, 440, 101 and 429 lncRNAs were collected from 'literature', 'RNA-seq', 'microarray' and 'GWAS', respectively.

In the lncRNA gene description pages, users can retrieve the related diseases of the entry, and also get the source of the information, such as the PMID(s) of the reference paper(s). There is also mutational information retrieved from the literature, GWASs and the dbSNP database.

LNCRNA CONSERVATION

Compared to protein-coding genes and small RNAs (e.g. miRNAs and snoRNAs), several reports have suggested that lncRNAs are modestly conserved (11). Most lncRNAs are less conserved in sequence (34), but there are still many lncRNAs that are conserved in their genomic loci, exonic sequences and promoter regions (35). These are preserved across multiple species, attesting to their important functional potentials (36).

Benefiting from next-generation sequencing technologies, ncRNAs are now more easily identified via transcript sequencing. NONCODE has added six new species, mainly from multi-species RNA-Seq data (37,38). An evolutionary tree from 12 commonly studied species (human, mouse, cow, rat, chicken, zebrafish, chimpanzee, gorilla, orangutan, rhesus macaque, opossum and platypus) was constructed using methods introduced in phyloNONCODE (39). Each human lncRNA gene counterpart from the other listed species can be retrieved through browsing the evolutionary tree (shown in Figure 2). The counterpart of each lncRNA was computed using the UCSC LiftOver tool (40). In brief, LiftOver utilized BLASTZ (41), an independent implementation of the Gapped BLAST algorithm specifically designed for aligning two long genomic sequences, as a core algorithm to detect homologous regions in other genomes. After mapping to the second species, the counterpart region was intersected with the second species transcript. Users can browse the transcript



Conservation Info

- Chordata
  - Vertebrates
    - Manmals
      - Primates
        - Human : -
        - **Chimp :** chr1(len:4332)
        - **Gorilla :** chr1(len:4359)
        - **Orangutan :** chr1(len:4317)
        - **Rhesus :** chr1(len:4159)
      - Rodents
        - **Mouse :** chr4(len:2770)
        - **Rat :** chr5(len:2672)
        - **Cow :** chr16(len: 3330)T
      - Opossum : -
      - Platypus : -
    - Birds
      - Chicken : -
  - Actinopteryqii
    - Zebrafish : -

location	Transcripts
chr16[-]50689538-50690089	NONBTAT207343

Note: T = There are transcripts in this region.

Figure 2. Conservation annotation for NONHSAG200087.

information by clicking the 'T' following the counterpart region (shown in Figure 2).

QUALITY OF LNCRNAS

Short-read sequencing technology allows high-throughput identification of lncRNAs. However, a proper method of *in silico* transcript reconstruction is an ongoing challenge. According to an assessment by the Paul Bertone group, <40% of known transcripts were well assembled from *Homo sapiens* RNA-seq data. The complexity of higher eukaryotic genomes imposes severe limitations on transcript recall and splice product discrimination that are likely to remain limiting factors for the analysis of current-generation RNA-seq data (42). Furthermore, multiple amplification steps during library preparation complicate the quantification of expression levels.

To some extent, third-generation sequencing technologies reduced the noise. This provides a more comprehensive assessment of the true complexity of the transcriptome. Given sufficient material, amplification-free and fragmentation-free sequencing of full-length cDNA molecules provides a more direct view of RNA molecules (12). NONCODE contacted the authors of the third-generation single-molecule long-read survey of the human transcriptome paper (12). After our analysis, the single-molecule lncRNA transcripts were included into NONCODE.

To meet the quality demands of researchers, NONCODE provides a subset searching interface. Users can choose the subset which is considered high quality. The quality controls include the source of the data, literature support, other database support and long-read sequencing method support. The controls also include selection of exon numbers, the lengths of the transcripts and prediction tools support. The web interface will return the subset according to the conditions users chose and allow users to download the data.

DISCUSSION

NONCODE 2016 contains 527,336 lncRNAs from 16 different species, this compares favorably with other lncRNA databases. For example, LNCipedia (human only) contains 111 685 transcripts (43), lncRNAtor (human, mouse, fly, zebrafish, worm and yeast) contains 34 605 transcripts (44), while the lncRNAWiki (human only) contains 105 255 transcripts (45). As mentioned above, technical limitations imposed by short-read sequencing lead to a number of computational challenges in transcript reconstruction and quantification. For many transcripts, automated methods failed to identify all of the constituent exons, and in cases in which all exons were reported, the protocols tested often failed to assemble the exons into complete isoforms (42). Considering this point, NONCODE filtered out some datasets. For example, although we have obtained all the data from MiTranscriptome (11), which contains 384 066 human lncRNAs from 7256 RNA-seq libraries, the detection of precise RefSeq splicing patterns from MiTranscriptome was only 31%, and the fraction of annotated genes within the entire MiTranscriptome was only 46%. Although it is reasonable to assume that unannotated transcription is unique to specific lineages, the low RefSeq detection rate was unusual. We therefore made a decision that NONCODE would not include MiTranscriptome data in the current version. In the future, we will attempt to make clear the real reason(s). Perhaps a more comprehensive construction tool is required to answer this question.

ACKNOWLEDGEMENT

We wish to thank Dr Greg Vatcher, Dr Jianjun Luo and Adistan Wahed for carefully reading our manuscript.

FUNDING

National High Technology Research and Development Program ('863' Program) of China [2014AA021103, 2014AA021502]; Training Program of the Major Research Plan of the National Natural Science Foundation of China [91229120]; National Natural Science Foundation of China [31371320, 31401119]; Chinese Academy of Science Strategic Project of Leading Science and Technology [XDA01020402]. Funding for open access charge: National High Technology Research and Development Program ('863' Program) of China [2014AA021103].

Conflict of interest statement. None declared.

REFERENCES

1. 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.
2. Pennisi,E. (2010) Shining a light on the genome's 'dark matter'. *Science*, **330**, 1614–1614.
3. Ponting,C.P., Oliver,P.L. and Reik,W. (2009) Evolution and functions of long noncoding RNAs. *Cell*, **136**, 629–641.
4. Guttman,M., Donaghey,J., Carey,B.W., Garber,M., Grenier,J.K., Munson,G., Young,G., Lucas,A.B., Ach,R., Bruhn,L. *et al.* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, **477**, U295–U260.
5. 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.
6. 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.
7. Wang,K.C. and Chang,H.Y. (2011) Molecular mechanisms of long noncoding RNAs. *Mol. Cell*, **43**, 904–914.
8. Gupta,R.A., Shah,N., Wang,K.C., Kim,J., Horlings,H.M., Wong,D.J., Tsai,M.C., Hung,T., Argani,P., Rinn,J.L. *et al.* (2010) Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, **464**, U1071–U1148.
9. Tsai,M.C., Manor,O., Wan,Y., Mosammaparast,N., Wang,J.K., Lan,F., Shi,Y., Segal,E. and Chang,H.Y. (2010) Long noncoding RNA as modular scaffold of histone modification complexes. *Science*, **329**, 689–693.
10. Weakley,S.M., Wang,H., Yao,Q.Z. and Chen,C.Y. (2011) Expression and function of a large non-coding RNA gene XIST in human cancer. *World J. Surg.*, **35**, 1751–1756.
11. Iyer,M.K., Niknafs,Y.S., Malik,R., Singhal,U., Sahu,A., Hosono,Y., Barrette,T.R., Prensner,J.R., Evans,J.R., Zhao,S. *et al.* (2015) The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.*, **47**, 199–208.
12. Sharon,D., Tilgner,H., Grubert,F. and Snyder,M. (2013) A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.*, **31**, 1009–1014.
13. Cunningham,F., Amode,M.R., Barrell,D., Beal,K., Billis,K., Brent,S., Carvalho-Silva,D., Clapham,P., Coates,G., Fitzgerald,S. *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.*, **43**, D662–D669.
14. Pruitt,K.D., Brown,G.R., Hiatt,S.M., Thibaud-Nissen,F., Astashyn,A., Ermolaeva,O., Farrell,C.M., Hart,J., Landrum,M.J., McGarvey,K.M. *et al.* (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.*, **42**, D756–D763.
15. Quek,X.C., Thomson,D.W., Maag,J.L., Bartonicek,N., Signal,B., Clark,M.B., Gloss,B.S. and Dinger,M.E. (2015) lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.*, **43**, D168–D173.
16. 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.
17. Merkin,J., Russell,C., Chen,P. and Burge,C.B. (2012) Evolutionary dynamics of gene and isoform regulation in mammalian tissues. *Science*, **338**, 1593–1599.
18. Xie,C.Y., Yuan,J., Li,H., Li,M., Zhao,G.G., Bu,D.C., Zhu,W.M., Wu,W., Chen,R.S. and Zhao,Y. (2014) NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res.*, **42**, D98–D103.
19. Bu,D.C., Yu,K.T., Sun,S.L., Xie,C.Y., Skogerbo,G., Miao,R.Y., Xiao,H., Liao,Q., Luo,H.T., Zhao,G.G. *et al.* (2012) NONCODE v3.0: integrative annotation of long noncoding RNAs. *Nucleic Acids Res.*, **40**, D210–D215.
20. He,S.M., Liu,C.N., Skogerbo,G., Zhao,H.T., Wang,J., Liu,T., Bai,B.Y., Zhao,Y. and Chen,R.S. (2008) NONCODE v2.0: decoding the non-coding. *Nucleic Acids Res.*, **36**, D170–D172.
21. Liu,C.N., Bai,B.Y., Skogerbo,G., Cai,L., Deng,W., Zhang,Y., Bu,D.B., Zhao,Y. and Chen,R.S. (2005) NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.*, **33**, D112–D115.
22. 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.
23. Sun,L., Luo,H., Bu,D., Zhao,G., Yu,K., Zhang,C., Liu,Y., Chen,R. and Zhao,Y. (2013) Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.*, **41**, e166.
24. Yin,Y., Zhao,Y., Wang,J., Liu,C., Chen,S., Chen,R. and Zhao,H. (2007) antiCODE: a natural sense-antisense transcripts database. *BMC Bioinformatics*, **8**, 319–321.
25. Guo,X., Gao,L., Liao,Q., Xiao,H., Ma,X., Yang,X., Luo,H., Zhao,G., Bu,D., Jiao,F. *et al.* (2013) Long non-coding RNAs function annotation: a global prediction method based on bi-colored networks. *Nucleic Acids Res.*, **41**, e35.
26. Liao,Q., Liu,C., Yuan,X., Kang,S., Miao,R., Xiao,H., Zhao,G., Luo,H., Bu,D., Zhao,H. *et al.* (2011) Large-scale prediction of long non-coding RNA functions in a coding-non-coding gene co-expression network. *Nucleic Acids Res.*, **39**, 3864–3878.
27. Liao,Q., Xiao,H., Bu,D., Xie,C., Miao,R., Luo,H., Zhao,G., Yu,K., Zhao,H., Skogerbo,G. *et al.* (2011) ncFANs: a web server for functional annotation of long non-coding RNAs. *Nucleic Acids Res.*, **39**, W118–W124.
28. Ulitsky,I. and Bartel,D.P. (2013) lincRNAs: genomics, evolution, and mechanisms. *Cell*, **154**, 26–46.
29. Wu,T., Wang,J., Liu,C., Zhang,Y., Shi,B., Zhu,X., Zhang,Z., Skogerbo,G., Chen,L., Lu,H. *et al.* (2006) NPInter: the noncoding RNAs and protein related biomacromolecules interaction database. *Nucleic Acids Res.*, **34**, D150–D152.
30. Yuan,J., Wu,W., Xie,C., Zhao,G., Zhao,Y. and Chen,R. (2014) NPInter v2.0: an updated database of ncRNA interactions. *Nucleic Acids Res.*, **42**, D104–D108.
31. Yuan,X., Liu,C., Yang,P., He,S., Liao,Q., Kang,S. and Zhao,Y. (2009) Clustered microRNAs' coordination in regulating protein-protein interaction network. *BMC Syst. Biol.*, **3**, 65–68.
32. Welter,D., MacArthur,J., Morales,J., Burdett,T., Hall,P., Junkins,H., Klemm,A., Flicek,P., Manolio,T., Hindorff,L. *et al.* (2014) The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.*, **42**, D1001–D1006.
33. Gong,J., Liu,W., Zhang,J., Miao,X. and Guo,A.Y. (2015) lncRNASNP: a database of SNPs in lncRNAs and their potential functions in human and mouse. *Nucleic Acids Res.*, **43**, D181–D186.
34. Marques,A.C. and Ponting,C.P. (2009) Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. *Genome Biol.*, **10**, 124–126.
35. Guttman,M., Amit,I., Garber,M., French,C., Lin,M.F., Feldser,D., Huarte,M., Zuk,O., Carey,B.W., Cassady,J.P. *et al.* (2009) Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature*, **458**, 223–227.
36. 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.
37. Brawand,D., Soumillon,M., Necsulea,A., Julien,P., Csardi,G., Harrigan,P., Weier,M., Liechti,A., Aximu-Petri,A., Kircher,M. *et al.*

(2011) The evolution of gene expression levels in mammalian organs. *Nature*, **478**, 343–348.

38. Necsulea,A., Soumillon,M., Warnefors,M., Liechti,A., Daish,T., Zeller,U., Baker,J.C., Grutzner,F. and Kaessmann,H. (2014) The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature*, **505**, 635–640.
39. Bu,D., Luo,H., Jiao,F., Fang,S., Tan,C., Liu,Z. and Zhao,Y. (2015) Evolutionary annotation of conserved long non-coding RNAs in major mammalian species. *Sci. China Life Sci.*, **58**, 787–798.
40. Dreszer,T.R., Karolchik,D., Zweig,A.S., Hinrichs,A.S., Raney,B.J., Kuhn,R.M., Meyer,L.R., Wong,M., Sloan,C.A., Rosenbloom,K.R. *et al.* (2012) The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.*, **40**, D918–D923.
41. Schwartz,S., Kent,W.J., Smit,A., Zhang,Z., Baertsch,R., Hardison,R.C., Haussler,D. and Miller,W. (2003) Human-mouse alignments with BLASTZ. *Genome Res.*, **13**, 103–107.
42. Steijger,T., Abril,J.F., Engstrom,P.G., Kokocinski,F., Hubbard,T.J., Guigo,R., Harrow,J., Bertone,P. and Consortium,R. (2013) Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods*, **10**, 1177–1184.
43. Volders,P.J., Verheggen,K., Menschaert,G., Vandepoele,K., Martens,L., Vandesompele,J. and Mestdagh,P. (2015) An update on LNCipedia: a database for annotated human lncRNA sequences (vol 43, pg D174, 2015). *Nucleic Acids Res.* **43**, 4363–4364.
44. Park,C., Yu,N., Choi,I., Kim,W. and Lee,S. (2014) lncRNAtor: a comprehensive resource for functional investigation of long non-coding RNAs. *Bioinformatics*, **30**, 2480–2485.
45. Ma,L., Li,A., Zou,D., Xu,X., Xia,L., Yu,J., Bajic,V.B. and Zhang,Z. (2015) LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Res.*, **43**, D187–D192.