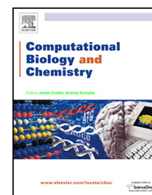
Computational Biology and Chemistry 50 (2014) 41–49

Contents lists available at ScienceDirect

Computational Biology and Chemistry

journal homepage: www.elsevier.com/locate/compbiolchem



Research Article

lncRNAMap: A map of putative regulatory functions in the long non-coding transcriptome

Wen-Ling Chan [a,b,c,d], Hsien-Da Huang [d,e,**], Jan-Gowth Chang [b,c,f,*]

[a] Biomedical Informatics, Asia University, Taichung, Taiwan
[b] Epigenome Research Center, China Medical University Hospital, Taichung, Taiwan
[c] School of Medicine, China Medical University, Taichung, Taiwan
[d] Institute of Bioinformatics and Systems Biology, National Chiao Tung University, Hsin-Chu, Taiwan
[e] Department of Biological Science and Technology, National Chiao Tung University, Hsin-Chu, Taiwan
[f] Department of Laboratory Medicine, China Medical University Hospital, Taichung, Taiwan

ARTICLE INFO

Article history:
Accepted 23 December 2013
Available online 23 January 2014

Keywords:
Long non-coding RNA (lncRNA)
Endogenous siRNA (esiRNA)
miRNA decoy
Function
Database

ABSTRACT

Background: Recent studies have demonstrated the importance of long non-coding RNAs (lncRNAs) in chromatin remodeling, and in transcriptional and post-transcriptional regulation. However, only a few specific lncRNAs are well understood, whereas others are completely uncharacterized. To address this, there is a need for user-friendly platform to studying the putative regulatory functions of human lncRNAs.
Description: lncRNAMap is an integrated and comprehensive database relating to exploration of the putative regulatory functions of human lncRNAs with two mechanisms of regulation, by encoding siRNAs and by acting as miRNA decoys. To investigate lncRNAs producing siRNAs that regulate protein-coding genes, lncRNAMap integrated small RNAs (sRNAs) that were supported by publicly available deep sequencing data from various sRNA libraries and constructed lncRNA-derived siRNA–target interactions. In addition, lncRNAMap demonstrated that lncRNAs can act as targets for miRNAs that would otherwise regulate protein-coding genes. Previously studies indicated that intergenic lncRNAs (lincRNAs) either positive or negative regulated neighboring genes, therefore, lncRNAMap surveyed neighboring genes within a 1 Mb distance from the genomic location of specific lncRNAs and provided the expression profiles of lncRNA and its neighboring genes. The gene expression profiles may supply the relationship between lncRNA and its neighboring genes.
Conclusions: lncRNAMap is a powerful user-friendly platform for the investigation of putative regulatory functions of human lncRNAs with producing siRNAs and acting as miRNA decoy. lncRNAMap is freely available on the web at http://lncRNAMap.mbc.nctu.edu.tw/.
© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

For over 50 years, the central dogma of molecular biology has been that DNA is transcribed into messenger RNA, which subsequently serves as the template for protein synthesis. However, only 2% of the human genome encodes proteins, whereas over 80% of the human genome produces non-protein coding RNAs (ncRNAs) (Alexander et al., 2010; Birney et al., 2007; Kapranov et al., 2007). These ncRNAs are divided into two major classes according to their nucleotide length, i.e. small (<200 bp) and long ncRNAs, each with distinct mechanisms. For the past 10 years, the small ncRNAs have been recognized as being involved in post-transcriptional regulation through translational repression (i.e. the RNAi pathway), while the much less well-studied long ncRNAs (lncRNAs) are implicated in epigenetic regulation (Deiters, 2010; Lee, 2012; Rinn and Chang, 2012; Snead and Rossi, 2010; Wang and Chang, 2011).

lncRNAs, transcripts >200 nt in length that are similar to protein-coding transcripts but are not translated, were initially considered to be transcriptional noise associated with low RNA polymerase fidelity and low sequence conservation (Marques and Ponting, 2009; Struhl, 2007). However, technological developments, especially of high-throughput sequencing methods, together with improvements in computational techniques, have led to an increasing number of functional lncRNAs being identified in mammals.

* Corresponding author at: Epigenome Research Center, China Medical University Hospital, No. 2, Yuh-Der Road, Taichaung, Taiwan. Tel.: +886 4 22052121x2008; fax: +886 4 22033295.
** Corresponding author at: Institute of Bioinformatics and Systems Biology, Department of Biological Science and Technology, National Chiao Tung University, 75 Bo-Ai Street, Hsin-Chu, Taiwan. Tel.: +886 3 5712121x56957; fax: +886 3 5739320.
E-mail addresses: bryan@mail.nctu.edu.tw (H.-D. Huang), d6781@mail.cmuh.org.tw (J.-G. Chang).

1476-9271/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiolchem.2014.01.003

These lncRNAs include not only antisense, intronic, and intergenic transcripts but also transcribed pseudogenes and retrotransposons. Most lncRNA functions still remain to be uncovered and specified through experiment. However, recent findings have revealed the involvement of lncRNAs in diverse cellular processes, such as chromatin modification, molecular scaffolding and cell cycle regulation, and lncRNAs have been found to play key roles in the circuitry controlling the embryonic stem cell state (Guttman et al., 2009, 2011; Khalil et al., 2009; Mercer et al., 2009; Rinn and Chang, 2012). In addition, lncRNAs may perform regulatory actions *in cis* (on neighboring genes) or *in trans* (on distantly located genes) (Hekimoglu and Ringrose, 2009; Hung and Chang, 2010; Khalil et al., 2009; Kim and Sung, 2012). Moreover, post-transcriptional regulation by lncRNA can be mediated through antisense RNA, with mechanisms involving siRNA, miRNA decoys or modulation of protein activation and localization (Chan et al., 2013b; Hung and Chang, 2010; Mattick et al., 2009; Poliseno et al., 2010; Wang and Chang, 2011).

Therefore, compilation and annotation of these data are crucial for a better understanding of lncRNAs. Indeed, a variety of public databases already provide some insight into lncRNA structure and function. For instance, lncRNAdb comprises lncRNA transcripts with well-documented functions in the literature (Amaral et al., 2011); LncRNADisease focuses on lncRNA-associated diseases (Chen et al., 2013); NRED provides probe-level expression profiles of lncRNAs (Dinger et al., 2009); ncRNAdb (Szymanski et al., 2007) and NONCODE (Bu et al., 2012) highlight RNA sequences and annotation from different sources; miRcode (Jeggari et al., 2012) and DIANA-LncBase (Paraskevopoulou et al., 2013) are tools for prediction of miRNA targets in lncRNA; and GENCODE (Derrien et al., 2012) and LNCipedia (Volders et al., 2013) reveal the structure, evolution and expression of human lncRNAs. Although each of these resources provides valuable information on lncRNAs, an integrated platform that explores the function of lncRNAs is currently lacking.

Small RNA-guided posttranscriptional regulation of gene expression is mainly covered by small RNA (sRNA) of 18–40 nt in size. More recently, deep sequencing approaches of various transcriptomes identified many classes of sRNA such as miRNAs, siRNAs, piRNAs, endogenous siRNAs (esiRNAs), mirtrons, miRNA-like sRNAs and others may derived from transcriptional read through of inverted repeat sequences, or from intermolecular double-strand RNA (dsRNA) precursors arising from transcription of either natural sense/antisense loci in the same (cis-nat siRNA) or different (trans-nat siRNA) position in the genome (gene/pseudogene) (Babiarz et al., 2008).

Here, we developed a systematic workflow to predict the siRNA-encoding or miRNA decoy functions of human lncRNAs. Accordingly, we have constructed an innovative and comprehensive database, lncRNAMap (http://lncRNAMap.mbc.nctu.edu.tw/), capturing information including lncRNA sequence data, homologous protein-coding gene and neighboring gene annotation, deep sequencing data, predicted RNA secondary structure, gene expression profiles, miRNA annotation and target prediction. To our knowledge, lncRNAMap is the first database to identify the neighboring protein-coding genes within a 1 Mb distance of known lncRNA and to focus on two mechanisms of action of human lncRNAs: (i) endogenous siRNA (esiRNA) generation from lncRNAs for the regulation of protein-coding genes; and (ii) lncRNA-mediated miRNA decoy.

2. Materials and methods

2.1. Data generation

A total of more than 20,000 human lncRNA transcripts were obtained from the Ensembl Genome Browser (Ensembl 65, GRCH37) (Flicek et al., 2012) using BioMart (http://www.ensembl.org/index.html). Functional small RNAs (fsRNAs) with sequence length from 18 to 40 nt were collected from the Functional RNA Database (fRNAdb) (Mituyama et al., 2009), which hosts a large collection of known/predicted non-coding RNA sequences from public databases. The public deep sequencing data from sRNA libraries were obtained using human embryo stem cells, liver tissues or hepatocellular carcinoma (HCC) tissues (Hou et al., 2011; Morin et al., 2008; Seila et al., 2008; Yeo et al., 2007). We summarize the statistics of the deep sequencing data from various sRNA libraries (Supplementary Table S1). The miRNA related information, including the accessions and miRNA sequences were obtained from miRBase release 18 (Griffiths-Jones et al., 2008; Kozomara and Griffiths-Jones, 2011). The genomic sequences and conserved 3′-UTR data were obtained from UCSC hg19.

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compbiolchem.2014.01.003.

2.2. Systematic workflow for exploring the regulatory function of lncRNAs

The workflow for this study is presented in Fig. 1, and mainly consists of the collection of datasets such as the lncRNAs themselves, protein-coding genes, fsRNAs, sRNA deep sequencing data, gene expression profiles, the integration of various tools, and the identification of human lncRNA functions, together with how lncRNAs are regulated. Table 1 lists the integrated databases and tools for mining potential regulatory functions of human lncRNAs. Based on a genome-wide computational pipeline of sequence-alignment approaches, this work highlights two major discoveries: lncRNA-derived esiRNA–target interaction, and the miRNA-decoy mechanism of action of lncRNAs. The detailed analyses are described below.

2.3. Identification of lncRNA-derived esiRNAs using publicly available next-generation sequencing (NGS) data

According to the previously studies, esiRNAs may derive from long non-coding RNAs to regulate protein-coding genes. To systematical explore that human lncRNAs may generate esiRNAs to regulate protein-coding genes, a computational pipeline was developed (Fig. 2). The sRNA deep sequencing reads were obtained from Gene Expression Omnibus (GEO) (Barrett and Edgar, 2006) and the Sequence Read Archive (SRA). The FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), SRA-Toolkit (http://www.ncbi.nlm.nih.gov/sra), and bowtie (http://bowtie-bio.sourceforge.net/) were incorporated into the identification of lncRNA-derived esiRNAs. The pipeline included adapter trimming, quality control and read alignment. Then these candidates of esiRNAs were mapped by lncRNAs with no mismatch. Finally, the hairpin structure predicted by Mfold (Zuker, 2003) was then determined by using the extended sequences of these candidate esiRNAs.

2.4. Identification of lncRNA-derived esiRNA–target interactions (eSTIs)

Our previously described approaches (Chan et al., 2013a,b) were modified to predict lncRNA-derived esiRNA targets. Briefly, esiRNA target sites within the conserved coding regions, 5′-UTRs and 3′-UTRs of genes were identified in 12 metazoan genomes (data were obtained from UCSC table) using three computational approaches, TargetScan (Friedman et al., 2009; Grimson et al., 2007; Lewis et al., 2005), miRanda (John et al., 2004) and RNAhybrid (Kruger and Rehmsmeier, 2006). The minimal free energy (MFE)

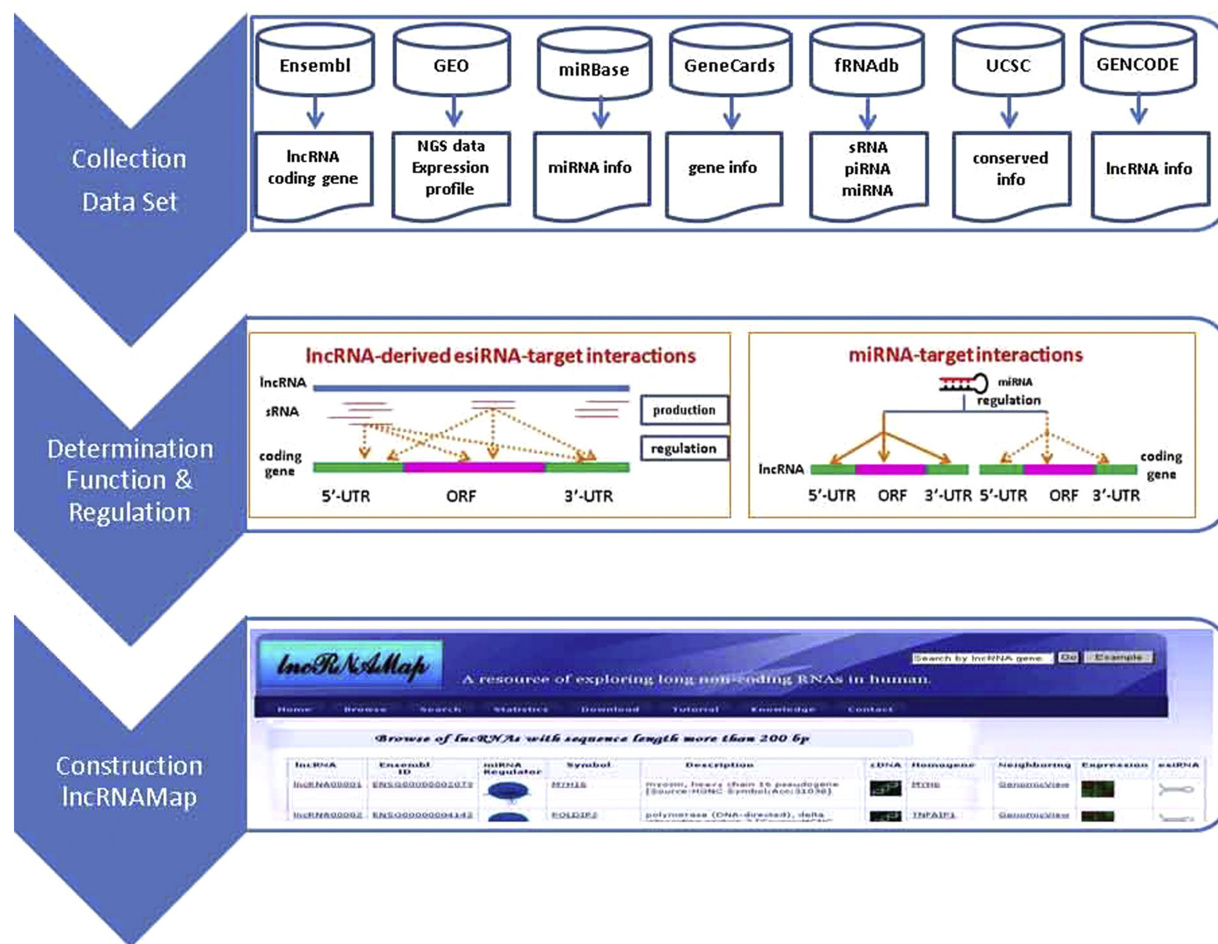


Fig. 1. Schematic of workflow for mining the putative regulatory functions of lncRNAs.

threshold was −20 kcal/mol with a score ≥150 for miRanda and default parameters for TargetScan and RNAhybrid. The targets were identified using the following criteria: (i) potential target sites were determined by at least two approaches; (ii) multiple target sites were prioritized; and (iii) target sites must be located in accessible regions. Finally, we compared the expression profiles of lncRNA and its target gene to construct the esiRNA–target interactions (eSTIs).

2.5. Gene expression analysis

The expression levels of lncRNAs and mRNAs of protein-coding genes were obtained from Gene Expression Omnibus (Barrett and Edgar, 2006). These included GDS596, with data from 79 human physiologically normal tissues (Su et al., 2004); GSE2109, examined in 2158 samples from 61 tumor tissues; GSE3526, profiled in 353 samples from 65 normal tissues (Roth et al., 2006); and GSE5364, with data from primary human tumors and adjacent non-tumor tissues (Yu et al., 2008), representing 270 tumors and 71 normal-cancer pairs from patients with breast, colon, liver, lung, oesophagal and thyroid cancers. The Pearson correlation coefficient was computed for lncRNAs and their homologous protein-coding genes.

2.6. Neighboring genes analysis

Previously, studies have demonstrated either positive (Kim et al., 2010; Orom et al., 2010) or negative (Brockdorff et al.,

Table 1
Supported databases and tools in lncRNAMap.

Integrated database or tools	Dataset	Description
miRBase (Griffiths-Jones et al., 2008; Kozomara and Griffiths-Jones, 2011)	miRNA annotation	This database not only provides published miRNA sequences and annotations but also supplies known/predicted targets
fRNAdb (Mituyama et al., 2009)	sRNA annotation	A database to support mining and annotation of functional RNAs
Ensembl Genome Browser (Flicek et al., 2012)	Pseudogene, protein-coding gene	Produces genome databases for vertebrates and other eukaryotic species
UCSC Genome Browser (Kent et al., 2002)	Conserved region Genomic view of genes	This browser provides a rapid and reliable display of any requested portion of genomes at any scale, together with dozens of aligned annotation tracks
GeneCards (Stelzer et al., 2011)	Gene annotation	GeneCards is a searchable, integrated database of human genes that provides concise genome-related information on all known and predicted human genes
Mfold (Zuker, 2003)	RNA folding tool	Folded RNA structure
GEO (Barrett and Edgar, 2006)	Gene expression profiles and deep sequencing data	A public functional genomics data
BLAST (Altschul et al., 1990)	Sequence alignment tool	BLAST finds regions of similarity between biological sequences

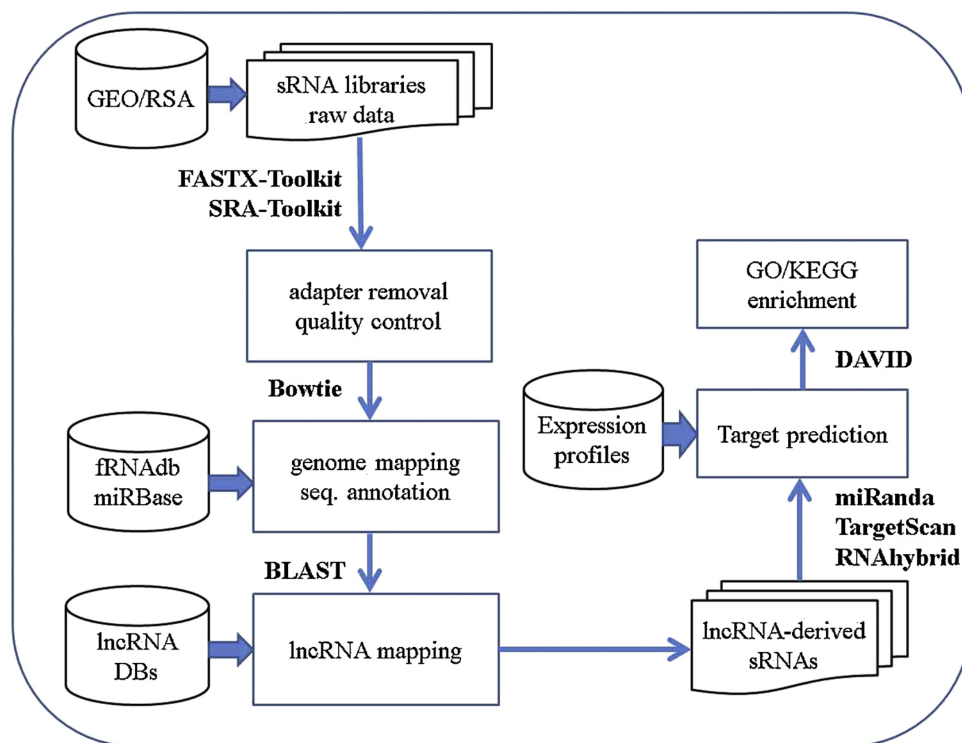


Fig. 2. A computational pipeline for exploring the lncRNAs may generate esiRNAs to regulate protein-coding genes.

1992; Nagano et al., 2008) regulation of neighboring protein-coding genes by intergenic lncRNAs (linRNAs). In this study, the number of protein-coding genes was detected within a 1 Mb distance of lncRNA as well as the expression profiles between lncRNA and its neighboring genes.

2.7. Determination of miRNA–target interactions (MTIs)

According to a previous study (Poliseno et al., 2010), pseudogenes *PTENP1* and *KRAS1P* act as a "miRNA decoy", binding to and thereby reducing the effective cellular concentration of miRNAs, resulting in their cognate genes escaping miRNA-mediated repression. In this work, we analyzed the potential for each specific lncRNA and its homologous protein-coding genes to be regulated by a miRNA decoy mechanism by examining miRNA–target interactions (MTIs) using the following pipeline. First, homologous genes were obtained by mapping the lncRNAs to cognate genomic sequences with the BLAST program (Altschul et al., 1990). The MTIs with lncRNAs and homologous protein-coding genes were then investigated as outlined previously (Chan et al., 2013a,b). The MFE threshold was −20 kcal/mol with a score ≥150 for miRanda and default parameters for TargetScan and RNAhybrid. Finally, the lncRNAs and their homologous protein-coding genes co-regulated by miRNAs were obtained. The miRNA and 3′UTR sequences were obtained from miRBase R18 (Griffiths-Jones et al., 2008; Kozomara and Griffiths-Jones, 2011) and Ensembl Genome Browser release 65, respectively.

2.8. Architecture of lncRNAMap

To enable the systematic compilation and updating of the results of analyses, as well as additional information, we constructed the database, lncRNAMap (http://lncRNAMap.mbc.nctu.edu.tw/), which is maintained with the MySQL (http://www.mysql.com/) relational database management system. Operating on an Apache HTTP server (http://www.apache.org/) with PHP (http://www.php.net/) on a Linux operation system (http://www.linux.com/), lncRNAMap was constructed using the Smarty template engine (http://www.smarty.net/). Based on PHP, JavaScript (http://www.javascriptsource.com/), CSS (http://www.w3schools.com/css/) and HTML languages (http://www.w3schools.com/html/), the web interface enables dynamic MySQL queries to be made with user-friendly graphics. All the above are open source technologies.

3. Results and discussion

3.1. Overview of lncRNAs

A total of 23,179 lncRNAs and 21,160 protein-coding genes were obtained from Ensembl 65. The number of lncRNAs and protein-coding genes per chromosome indicated that there are more lncRNAs than protein-coding genes on all chromosomes except chr14, 15, 16, 17, 18 and 19, which are not manually annotated by the HAVANA team (Supplementary Fig. S1). Of these lncRNAs, 4343 were detected by Affymetrix Human Genome U133A/U133Plus2 microarray. Subsequently, the gene expression profile was estimated based on $\log_2$ average intensity of lncRNAs in 79 human normal tissues. Hierarchical clustering results indicated that two fifths of lncRNAs were highly expressed in blood-related tissues (Supplementary Fig. S2 with red line genes). Expression profiles of six different tissues revealed that most lncRNAs were more highly expressed in tumors than paired non-tumor tissues (Supplementary Fig. S3 with red line genes).

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compbiolchem.2014.01.003.

3.2. A role for lncRNAs in the modulation of miRNA-mediated gene regulation

Given recent findings that pseudogenes *PTENP1* and *KRAS1P* can de-repress their cognate genes by a miRNA decoy

Table 2
A complete statistical analysis of lncRNAMap.

Dataset	Counts
No. of miRNA regulators	1886
No. of TPG-derived miRNAs	80
No. of TPG-derived piRNAs	4879
Deep sequencing data for profiling TPG-derived esiRNAs	
Human embryo stem cell-hB	1792
Human embryo stem cell-hESC	3847
Human embryo stem cell-hues6	761
Human embryo stem cell-hues6NP	489
Human embryo stem cell-hues6Neuron	216
HBV(+) Adjacent Tissue Sample 1	5424
HBV(+) Adjacent Tissue Sample 2	21,970
HBV(+) Distal Tissue Sample 1	5886
HBV(+) HCC Tissue Sample 1	6001
HBV(+) HCC Tissue Sample 2	15,790
HBV(+) Liver Tissue	13,190
HBV(+) Side Tissue Sample 1	6120
HCV(+) Adjacent Tissue Sample	71,912
HCV(+) HCC Tissue Sample	49,381
HBV(−) HCV(−) Adjacent Tissue Sample	13,298
HBV(−) HCV(−) HCC Tissue Sample	23,831
Human Normal Liver Tissue Sample 1	9690
Human Normal Liver Tissue Sample 2	7100
Human Normal Liver Tissue Sample 3	6587
Severe Chronic Hepatitis B Liver Tissue	1247

mechanism, we investigated MTIs using a similar approach. Analysis of the results indicated that 431 miRNAs with MFE threshold ≤ -20 kcal/mol and score ≥ 150 interact had many possible target sites in 21,164 lncRNAs and 5521 homologous protein-coding genes, and might potentially co-regulate specific pairs of lncRNAs and cognate protein-coding genes.

3.3. Possible role of lncRNA-derived esiRNAs in regulating protein-coding genes

Our survey of more than 20,000 human lncRNAs detected 4343 transcribed lncRNAs, 3280 of which were revealed in silico to have sequence features associated with the generation of endogenous siRNAs (esiRNAs). A total of 80 miRNAs, 4879 piRNAs and 264,532 Illumina-Solexa reads were represented in 310, 4533 and 3234 lncRNAs, respectively. Table 2 summarizes the complete statistical analysis of this research.

3.4. The number of lnRNAs across protein-coding genes within a 1 Mb distance

Previous studies have demonstrated linRNA regulation of neighboring protein-coding genes (Brockdorff et al., 1992; Kim et al., 2010; Nagano et al., 2008; Orom et al., 2010). Therefore, we examined the number of lncRNAs and neighboring protein-coding genes within a 1 Mb distance of the lncRNA-encoding locus. The results show that half of lncRNAs are across 1–10 protein-coding genes within a 1 Mb distance (Fig. 3).

3.5. Web interface and data visualization of lncRNAMap

To facilitate the systematic compilation and updating of datasets and any additional information, we constructed a web-based system called lncRNAMap by which we can comprehensively identify lncRNAs where (i) the lncRNA acts as a miRNA decoy; and/or (ii) the lncRNA produces esiRNAs to regulate protein-coding genes. There are two ways of accessing lncRNAMap: by browsing the database content, or by searching for a particular lncRNA. Fig. 4A displays the interface of output results of the browse gateway. The interface contains general information on lncRNAs, the involvement of lncRNA and its homologous protein-coding

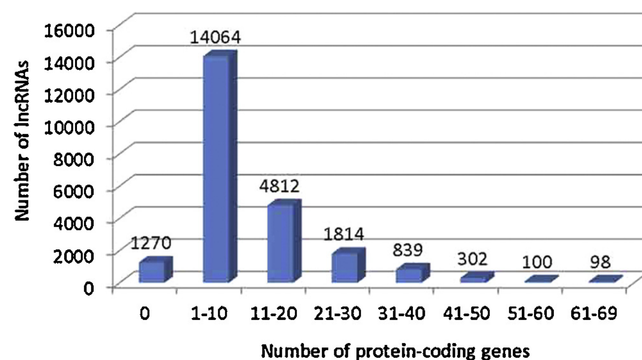


Fig. 3. The number of lncRNAs across protein-coding genes within a 1 Mb distance.

gene in miRNA-mediated repression, under the heading "miRNA Regulator"; lncRNA-derived esiRNA–target interactions, given as "esiRNA"; "Neighboring", indicating the genomic view of neighboring genes within a 1 Mb distance of a lncRNA; and "Expression", showing the gene expression profiles. The esiRNA view gives the lncRNA-derived esiRNAs themselves together with a graphical display of deep sequencing data (Fig. 4B). The red line represents the lncRNA, while the blue line refers to esiRNAs. We not only estimate the esiRNA–target interactions but also formulate the RNA folding structure of lncRNA-derived esiRNA. Fig. 4C provides a detailed view of the miRNA Regulator feature, which displays more fine-grained information. The "Expression" view presents the gene expression profiles not only of distinct lncRNAs, but also the corresponding homologous protein-coding genes under various experimental conditions (Fig. 4D). Moreover, all of the results and sequences can be downloaded for further analysis. In lncRNAMap, we also incorporate external sources, such as the UCSC genome browser (Kent et al., 2002) for a genomic view, GeneCards (Stelzer et al., 2011) for gene annotation, and miRBase (Kozomara and Griffiths-Jones, 2011) for miRNA annotation. Additionally, lncRNAMap also contains a tutorial and knowledge base for lncRNAs.

In the search gateway, the ID or symbol of a particular lncRNA and/or its homologous protein-coding gene, can be deployed as search terms for further analysis. The search result screen is similar to the browse interface in that it contains general information on the searched lncRNA, its protein-coding gene, miRNA regulators, gene expression profiles and lncRNA-derived esiRNA–target interactions.

3.6. Comparison with other lncRNA databases

A variety of lncRNA databases have been constructed, including lncRNA Database, which provides comprehensive annotations of eukaryotic lncRNAs; Valadkhan Lab-lncRNA Database, a repository of mammalian long non-protein-coding transcripts with sequence and structure annotation; ncRNA expression database, a database integrating annotated expression data based on probe characteristics; and LncBase, a database providing predicted/experimentally verified miRNA–lncRNA interactions. However, these databases focus on automatic detection of lncRNAs using a variety of similarity-based approaches. Our database, lncRNAMap, aims to provide a comprehensive resource for genome-wide identification of the regulatory functions of human lncRNAs. In brief, there are four major features differentiating lncRNAMap from currently available public lncRNA databases. First, lncRNAMap elucidates the likelihood of lncRNA regulation of its homologous protein-coding gene by a miRNA decoy mechanism. Second, to explore the interaction of lncRNA and its homologous protein-coding gene, lncRNAMap provides the gene expression profiles of lncRNA and its homologous gene under various experimental conditions. Third,

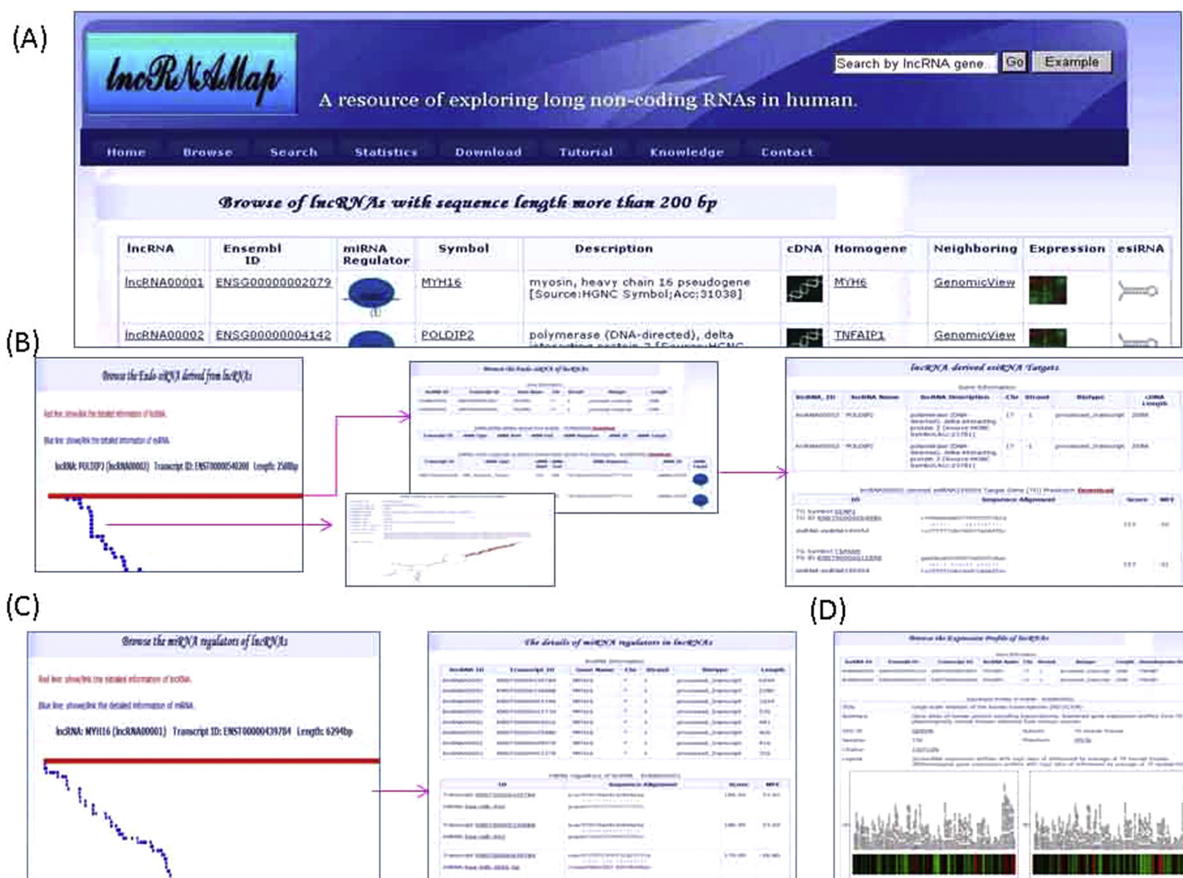


Fig. 4. lncRNAMap web interface.

Table 3
Comparison of lncRNAMap with currently available public databases of lncRNAs.

Supported features	lncRNAMap (our database)	lncRNA database	Valadkhan Lab-lncRNA database	NRED
Web interface	http://lncRNAMap.mbc.nctu.edu.tw/	http://www.lncrnadb.org/	http://www.valadkhanlab.org/database	http://nred.matticklab.com/cgi-bin/ncrnadb.pl
Description	lncRNAMap is a comprehensive resource for studying the functions of human lncRNAs	lncRNAdb is a database providing comprehensive annotations of eukaryotic lncRNAs	A repository of mammalian long non-protein-coding transcripts that have been experimentally shown to be both non-coding and functional	NRED integrates annotated expression data from various sources
Species supported	Human	Eukaryote	Human, mouse and rat	Human, mouse
Sequence download	Yes	Yes	–	–
lncRNA information	Yes	Yes	Yes	–
Homologous gene information	Yes	–	–	–
Knowledge of lncRNA	Yes	–	–	–
miRNA–lncRNA interactions	Yes	–	–	–
Gene expression profiles	Yes (both of lncRNA and its homologous gene)	–	–	Yes (probe-level)
lncRNA-derived esiRNAs	Yes	–	–	–
Deep sequencing data for profiling lncRNA-derived siRNAs	Yes	–	–	–
lncRNA's neighboring genes	Yes	–	–	–

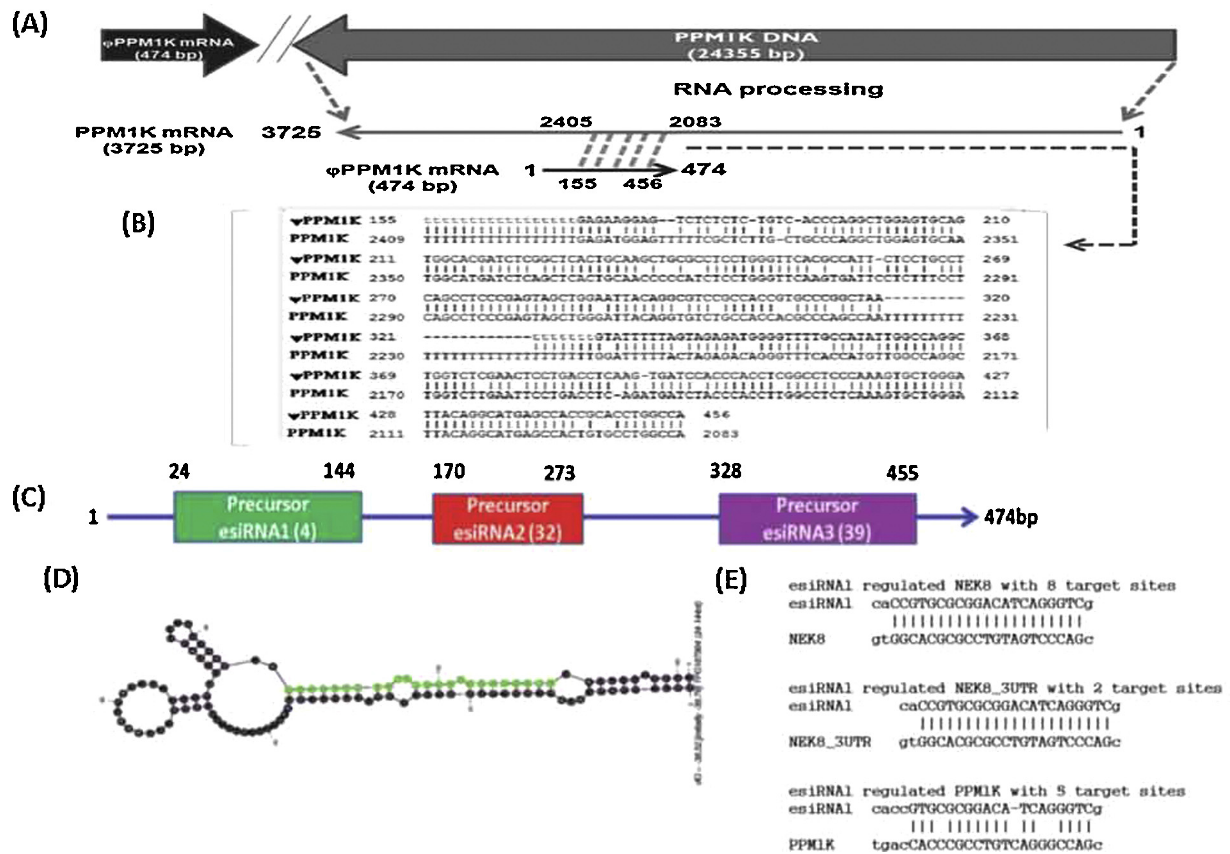


Fig. 5. Schematic representation of *ψPPM1K*-derived esiRNAs and their targets. (A) Location of *PPM1K* and *ψPPM1K*. (B) The mapping result between PPM1K and *ψPPM1K*. (C) Location and read counts of *ψPPM1K* from sRNA deep sequencing data. (D) Mfold was used to predict the hairpin structure of precursor esiRNA1. The predicted mature esiRNA1 sequence is depicted in green. (E) Matches of esiRNA1 sequences with target gene *NEK8* and parental gene *PPM1K*. Canonical pairings, solid lines; non-canonical pairings (G:U), dotted lines (modified from Chan et al. (2013b)). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

lncRNAMap curates lncRNA-derived esiRNAs, supported by deep sequencing data, as well as highlighting interacting gene targets in the human genome. Fourth, lncRNAmap surveys the neighboring genes within 1 Mb of a specific lncRNA. Table 3 lists detailed comparisons of lncRNAMap with other lncRNA databases.

3.7. Applications

There are two major applications of lncRNAMap, the first of which is the identification of non-coding RNA products of lncRNAs that may generate esiRNAs for the regulation of protein-coding genes in humans. In this case, lncRNAMap supplies NGS data from sRNA libraries to support the candidature of specific lncRNA-derived esiRNAs, together with gene expression profiles to verify the target interactions. The second application pertains when both the lncRNA and its cognate gene contain the same miRNA target sites, such that the lncRNA competes for miRNA repressor molecules, thereby alleviating miRNA-mediated gene repression. In other words, lncRNAMap helps to identify situations where the lncRNA may act as a miRNA decoy. lncRNAMap thus provides further insight into the lncRNA-mediated pathway of miRNA–target interactions.

3.8. Case study

Transcribed pseudogene is one class of lncRNAs. *PPM1K* (protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1K), located on chromosome 4q22.1, produces a 3725-nt mRNA encoding a mitochondrial matrix serine/threonine protein phosphatase shown to regulate the membrane permeability transition pore (MPTP) essential for cell survival and organ development (Lu et al., 2007). Pseudogene *PPM1K* (*ψPPM1K*), 474 bp in length, is partially retrotranscribed from *PPM1K* and is also located on chromosome 4 (Fig. 5A). According to our computational pipeline (Fig. 5B–E) and subsequent experimental tests, we found that *ψPPM1K* may generate esiRNAs to regulate protein-coding genes in HCC cell lines (Chan et al., 2013b).

4. Conclusions

Increasing studies have shown that lncRNAs are involved in a variety of biological functions. However, only a few specific lncRNAs are well understood. The lncRNAMap integrated various dataset, such as sequences, cognate and neighboring gene, NGS data, gene expression profiles and secondary structure. Moreover, we constructed a pipeline to investigate lncRNA producing esiRNAs to regulate protein-coding genes and lncRNA acting a miRNA decoy. We believe that lncRNAMap is useful for the studies of the potential functions and regulations of lncRNAs in human.

Author's contributions

WLC carried out all experimental concepts, wrote programs and the manuscript. HDH and JGC managed the study and assisted write and revise the manuscript. All authors read and approved the final manuscript.

Declarations

The publication costs for this article were funded by the corresponding author.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The authors thank the Asia University and the National Science Council of the Republic of China for financially supporting this research under Contract No. NSC 99-2320-B-037-006-MY3 (JGC), NSC 98-2311-B-009-004-MY3 (HDH), NSC 99-2627-B-009-003 (HDH), NSC 101-2311-B-009-003-MY3 (HDH) and NSC 100-2627-B-009-002 (HDH). This work was supported in part by the UST-UCSD International Center of Excellence in Advanced Bio-Engineering sponsored by the Taiwan National Science Council I-RiCE Program under Grant Number: NSC 102-2911-I-009-101 (HDH), and Veterans General Hospitals and University System of Taiwan (VGHUST) Joint Research Program under Grant Number: VGHUST101-G5-1-1 (HDH). This work was also partially supported by MOE ATU.

References

Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M., Gerstein, M.B., 2010. Annotating non-coding regions of the genome. Nat. Rev. Genet. 11, 559–571.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Amaral, P.P., Clark, M.B., Gascoigne, D.K., Dinger, M.E., Mattick, J.S., 2011. lncRNAdb: a reference database for long noncoding RNAs. Nucleic Acids Res. 39, D146–D151.

Babiarz, J.E., Ruby, J.G., Wang, Y., Bartel, D.P., Blelloch, R., 2008. Mouse ES cells express endogenous shRNAs, siRNAs, and other Microprocessor-independent, Dicer-dependent small RNAs. Genes Dev. 22, 2773–2785.

Barrett, T., Edgar, R., 2006. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods Enzymol. 411, 352–369.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., Kuehn, M.S., Taylor, C.M., Neph, S., Koch, C.M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J.A., Andrews, R.M., Flicek, P., Boyle, P.J., Cao, H., Carter, N.P., Clelland, G.K., Davis, S., Day, N., Dhami, P., Dillon, S.C., Dorschner, M.O., Fiegler, H., Giresi, P.G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K.D., Johnson, B.E., Johnson, E.M., Frum, T.T., Rosenzweig, E.R., Karnani, N., Lee, K., Lefebvre, G.C., Navas, P.A., Neri, F., Parker, S.C., Sabo, P.J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F.S., Dekker, J., Lieb, J.D., Tullius, T.D., Crawford, G.E., Sunyaev, S., Noble, W.S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I.L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H.A., Sekinger, E.A., Lagarde, J., Abril, J.F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J.S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M.C., Thomas, D.J., Weirauch, M.T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K.G., Sung, W.K., Ooi, H.S., Chiu, K.P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M.L., Valencia, A., Choo, S.W., Choo, C.Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T.G., Brown, J.B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C.N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J.S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R.M., Rogers, J., Stadler, P.F., Lowe, T.M., Wei, C.L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S.E., Fu, Y., Green, E.D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L.A., Wetterstrand, K.A., Good, P.J., Feingold, E.A., Guyer, M.S., Cooper, G.M., Asimenos, G., Dewey, C.N., Hou, M., Nikolaev, S., Montoya-Burgos, J.I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N.R., Holmes, I., Mullikin, J.C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W.J., Stone, E.A., Batzoglou, S., Goldman, N., Hardison, R.C., Haussler, D., Miller, W., Sidow, A., Trinklein, N.D., Zhang, Z.D., Barrera, L., Stuart, R., King, D.C., Ameur, A., Enroth, S., Bieda, M.C., Kim, J., Bhinge, A.A., Jiang, N., Liu, J., Yao, F., Vega, V.B., Lee, C.W., Ng, P., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M.J., Inman, D., Singer, M.A., Richmond, T.A., Munn, K.J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J.C., Couttet, P., Bruce, A.W., Dovey, O.M., Ellis, P.D., Langford, C.F., Nix, D.A., Euskirchen, G., Hartman, S., Urban, A.E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T.H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C.K., Rosenfeld, M.G., Aldred, S.F., Cooper, S.J., Halees, A., Lin, J.M., Shulha, H.P., Xu, M., Haidar, J.N., Yu, Y., Iyer, V.R., Green, R.D., Wadelius, C., Farnham, P.J., Ren, B., Harte, R.A., Hinrichs, A.S., Trumbower, H., Clawson, H., Hillman-Jackson, J., Zweig, A.S., Smith, K., Thakkapallayil, A., Barber, G., Kuhn, R.M., Karolchik, D., Armengol, L., Bird, C.P., de Bakker, P.I., Kern, A.D., Lopez-Bigas, N., Martin, J.D., Stranger, B.E., Woodroffe, A., Davydov, E., Dimas, A., Eyras, E., Hallgrimsdottir, I.B., Huppert, J., Zody, M.C., Abecasis, G.R., Estivill, X., Bouffard, G.G., Guan, X., Hansen, N.F., Idol, J.R., Maduro, V.V., Maskeri, B., McDowell, J.C., Park, M., Thomas, P.J., Young, A.C., Blakesley, R.W., Muzny, D.M., Sodergren, E., Wheeler, D.A., Worley, K.C., Jiang, H., Weinstock, G.M., Gibbs, R.A., Graves, T., Fulton, R., Mardis, E.R., Wilson, R.K., Clamp, M., Cuff, J., Gnerre, S., Jaffe, D.B., Chang, J.L., Lindblad-Toh, K., Lander, E.S., Koriabine, M., Nefedov, M., Osoegawa, K., Yoshinaga, Y., Zhu, B., de Jong, P.J., 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799–816.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., Rastan, S., 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71, 515–526.

Bu, D., Yu, K., Sun, S., Xie, C., Skogerbo, G., Miao, R., Xiao, H., Liao, Q., Luo, H., Zhao, G., Zhao, H., Liu, Z., Liu, C., Chen, R., Zhao, Y., 2012. NONCODE v3.0: integrative annotation of long noncoding RNAs. Nucleic Acids Res. 40, D210–D215.

Chan, W.L., Yang, W.K., Huang, H.D., Chang, J.G., 2013a. pseudoMap: an innovative and comprehensive resource for identification of siRNA-mediated mechanisms in human transcribed pseudogenes. Database (Oxford) 2013, bat001.

Chan, W.L., Yuo, C.Y., Yang, W.K., Hung, S.Y., Chang, Y.S., Chiu, C.C., Yeh, K.T., Huang, H.D., Chang, J.G., 2013b. Transcribed pseudogene psiPPM1K generates endogenous siRNA to suppress oncogenic cell growth in hepatocellular carcinoma. Nucleic Acids Res. 41 (6), 3734–3747.

Chen, G., Wang, Z., Wang, D., Qiu, C., Liu, M., Chen, X., Zhang, Q., Yan, G., Cui, Q., 2013. LncRNADisease: a database for long-non-coding RNA-associated diseases. Nucleic Acids Res. 41, D983–D986.

Deiters, A., 2010. Small molecule modifiers of the microRNA and RNA interference pathway. AAPS J. 12, 51–60.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J.B., Lipovich, L., Gonzalez, J.M., Thomas, M., Davis, C.A., Shiekhattar, R., Gingeras, T.R., Hubbard, T.J., Notredame, C., Harrow, J., Guigo, R., 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. 22, 1775–1789.

Dinger, M.E., Pang, K.C., Mercer, T.R., Crowe, M.L., Grimmond, S.M., Mattick, J.S., 2009. NRED: a database of long noncoding RNA expression. Nucleic Acids Res. 37, D122–D126.

Flicek, P., Amode, M.R., Barrell, D., Beal, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fairley, S., Fitzgerald, S., Gil, L., Gordon, L., Hendrix, M., Hourlier, T., Johnson, N., Kahari, A.K., Keefe, D., Keenan, S., Kinsella, R., Komorowska, M., Koscielny, G., Kulesha, E., Larsson, P., Longden, I., McLaren, W., Muffato, M., Overduin, B., Pignatelli, M., Pritchard, B., Riat, H.S., Ritchie, G.R., Ruffier, M., Schuster, M., Sobral, D., Tang, Y.A., Taylor, K., Trevanion, S., Vandrovcova, J., White, S., Wilson, M., Wilder, S.P., Aken, B.L., Birney, E., Cunningham, F., Dunham, I., Durbin, R., Fernandez-Suarez, X.M., Harrow, J., Herrero, J., Hubbard, T.J., Parker, A., Proctor, G., Spudich, G., Vogel, J., Yates, A., Zadissa, A., Searle, S.M., 2012. Ensembl 2012. Nucleic Acids Res. 40, D84–D90.

Friedman, R.C., Farh, K.K., Burge, C.B., Bartel, D.P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 19, 92–105.

Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J., 2008. miRBase: tools for microRNA genomics. Nucleic Acids Res. 36, D154–D158.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., Bartel, D.P., 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol. Cell 27, 91–105.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., Cabili, M.N., Jaenisch, R., Mikkelsen, T.S., Jacks, T., Hacohen, N., Bernstein, B.E., Kellis, M., Regev, A., Rinn, J.L., Lander, E.S., 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature 458, 223–227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., Yang, X., Amit, I., Meissner, A., Regev, A., Rinn, J.L., Root, D.E., Lander, E.S., 2011. lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature 477, 295–300.

Hekimoglu, B., Ringrose, L., 2009. Non-coding RNAs in polycomb/trithorax regulation. RNA Biol. 6, 129–137.

Hou, J., Lin, L., Zhou, W., Wang, Z., Ding, G., Dong, Q., Qin, L., Wu, X., Zheng, Y., Yang, Y., Tian, W., Zhang, Q., Wang, C., Zhuang, S.M., Zheng, L., Liang, A., Tao, W., Cao, X., 2011. Identification of miRNomes in human liver and hepatocellular carcinoma reveals miR-199a/b-3p as therapeutic target for hepatocellular carcinoma. Cancer Cell 19, 232–243.

Hung, T., Chang, H.Y., 2010. Long noncoding RNA in genome regulation: prospects and mechanisms. RNA Biol. 7, 582–585.

Jeggari, A., Marks, D.S., Larsson, E., 2012. miRcode: a map of putative microRNA target sites in the long non-coding transcriptome. Bioinformatics 28, 2062–2063.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., Marks, D.S., 2004. Human MicroRNA targets. PLoS Biol. 2, e363.

Kapranov, P., Willingham, A.T., Gingeras, T.R., 2007. Genome-wide transcription and the implications for genomic organization. Nat. Rev. Genet. 8, 413–423.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, D., 2002. The human genome browser at UCSC. Genome Res. 12, 996–1006.

Khalil, A.M., Guttman, M., Huarte, M., Garber, M., Raj, A., Rivea Morales, D., Thomas, K., Presser, A., Bernstein, B.E., van Oudenaarden, A., Regev, A., Lander, E.S., Rinn, J.L., 2009. Many human large intergenic noncoding RNAs associate with

chromatin-modifying complexes and affect gene expression. Proc. Natl. Acad. Sci. U. S. A. 106, 11667–11672.
Kim, E.D., Sung, S., 2012. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. Trends Plant Sci. 17, 16–21.
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., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P.F., Kreiman, G., Greenberg, M.E., 2010. Widespread transcription at neuronal activity-regulated enhancers. Nature 465, 182–187.
Kozomara, A., Griffiths-Jones, S., 2011. miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res. 39, D152–D157.
Kruger, J., Rehmsmeier, M., 2006. RNAhybrid: microRNA target prediction easy, fast and flexible. Nucleic Acids Res. 34, W451–W454.
Lee, J.T., 2012. Epigenetic regulation by long noncoding RNAs. Science 338, 1435–1439.
Lewis, B.P., Burge, C.B., Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 120, 15–20.
Lu, G., Ren, S., Korge, P., Choi, J., Dong, Y., Weiss, J., Koehler, C., Chen, J.N., Wang, Y., 2007. A novel mitochondrial matrix serine/threonine protein phosphatase regulates the mitochondria permeability transition pore and is essential for cellular survival and development. Genes Dev. 21, 784–796.
Marques, A.C., Ponting, C.P., 2009. Catalogues of mammalian long noncoding RNAs: modest conservation and incompleteness. Genome Biol. 10, R124.
Mattick, J.S., Amaral, P.P., Dinger, M.E., Mercer, T.R., Mehler, M.F., 2009. RNA regulation of epigenetic processes. Bioessays 31, 51–59.
Mercer, T.R., Dinger, M.E., Mattick, J.S., 2009. Long non-coding RNAs: insights into functions. Nat. Rev. Genet. 10, 155–159.
Mituyama, T., Yamada, K., Hattori, E., Okida, H., Ono, Y., Terai, G., Yoshizawa, A., Komori, T., Asai, K., 2009. The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs. Nucleic Acids Res. 37, D89–D92.
Morin, R.D., O'Connor, M.D., Griffith, M., Kuchenbauer, F., Delaney, A., Prabhu, A.L., Zhao, Y., McDonald, H., Zeng, T., Hirst, M., Eaves, C.J., Marra, M.A., 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 18, 610–621.
Nagano, T., Mitchell, J.A., Sanz, L.A., Pauler, F.M., Ferguson-Smith, A.C., Feil, R., Fraser, P., 2008. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. Science 322, 1717–1720.
Orom, U.A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q., Guigo, R., Shiekhattar, R., 2010. Long noncoding RNAs with enhancer-like function in human cells. Cell 143, 46–58.
Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Reczko, M., Maragkakis, M., Dalamagas, T.M., Hatzigeorgiou, A.G., 2013. DIANA-LncBase: experimentally verified and computationally predicted microRNA targets on long non-coding RNAs. Nucleic Acids Res. 41, D239–D245.
Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J., Pandolfi, P.P., 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465, 1033–1038.
Rinn, J.L., Chang, H.Y., 2012. Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81, 145–166.
Roth, R.B., Hevezi, P., Lee, J., Willhite, D., Lechner, S.M., Foster, A.C., Zlotnik, A., 2006. Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. Neurogenetics 7, 67–80.
Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., Sharp, P.A., 2008. Divergent transcription from active promoters. Science 322, 1849–1851.
Snead, N.M., Rossi, J.J., 2010. Biogenesis and function of endogenous and exogenous siRNAs. Wiley Interdiscip. Rev. RNA 1, 117–131.
Stelzer, G., Dalah, I., Stein, T.I., Satanower, Y., Rosen, N., Nativ, N., Oz-Levi, D., Olender, T., Belinky, F., Bahir, I., Krug, H., Perco, P., Mayer, B., Kolker, E., Safran, M., Lancet, D., 2011. In-silico human genomics with GeneCards. Hum. Genomics 5, 709–717.
Struhl, K., 2007. Transcriptional noise and the fidelity of initiation by RNA polymerase II. Nat. Struct. Mol. Biol. 14, 103–105.
Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R., Hogenesch, J.B., 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. U. S. A. 101, 6062–6067.
Szymanski, M., Erdmann, V.A., Barciszewski, J., 2007. Noncoding RNAs database (ncRNAdb). Nucleic Acids Res. 35, D162–D164.
Volders, P.J., Helsens, K., Wang, X., Menten, B., Martens, L., Gevaert, K., Vandesompele, J., Mestdagh, P., 2013. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. Nucleic Acids Res. 41, D246–D251.
Wang, K.C., Chang, H.Y., 2011. Molecular mechanisms of long noncoding RNAs. Mol. Cell 43, 904–914.
Yeo, G.W., Xu, X., Liang, T.Y., Muotri, A.R., Carson, C.T., Coufal, N.G., Gage, F.H., 2007. Alternative splicing events identified in human embryonic stem cells and neural progenitors. PLoS Comput. Biol. 3, 1951–1967.
Yu, K., Ganesan, K., Tan, L.K., Laban, M., Wu, J., Zhao, X.D., Li, H., Leung, C.H., Zhu, Y., Wei, C.L., Hooi, S.C., Miller, L., Tan, P., 2008. A precisely regulated gene expression cassette potently modulates metastasis and survival in multiple solid cancers. PLoS Genet. 4, e1000129.
Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.