

# DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts

Maria D. Paraskevopoulou<sup>1,2,\*</sup>, Ioannis S. Vlachos<sup>1,2,3,†</sup>, Dimitra Karagkouni<sup>1,2,†</sup>, Georgios Georgakilas<sup>1,2</sup>, Ilias Kanellos<sup>4,5</sup>, Thanasis Vergoulis<sup>4</sup>, Konstantinos Zagganas<sup>4,6</sup>, Panayiotis Tsanakas<sup>5,7</sup>, Evangelos Floros<sup>7</sup>, Theodore Dalamagas<sup>4,\*</sup> and Artemis G. Hatzigeorgiou<sup>1,2,4,\*</sup>

<sup>1</sup>DIANA-Lab, Department of Computer & Communication Engineering, University of Thessaly, 382 21, Volos, Greece, <sup>2</sup>Hellenic Pasteur Institute, 127 Vasilissis Sofias Avenue, 11521, Athens, Greece, <sup>3</sup>Laboratory for Experimental Surgery and Surgical Research 'N.S. Christeas', Medical School of Athens, University of Athens, 11527, Greece, <sup>4</sup>'Athena' Research and Innovation Center, 11524, Athens, Greece, <sup>5</sup>School of Electrical and Computer Engineering, NTUA, 15773 Zografou, Greece, <sup>6</sup>University of Peloponnisos, Department of Informatics and Telecommunications, 22100, Tripoli, Greece and <sup>7</sup>Greek Research and Technology Network (GRNET), Athens 11527, Greece

Received September 30, 2015; Revised November 2, 2015; Accepted November 3, 2015

## ABSTRACT

microRNAs (miRNAs) are short non-coding RNAs (ncRNAs) that act as post-transcriptional regulators of coding gene expression. Long non-coding RNAs (lncRNAs) have been recently reported to interact with miRNAs. The sponge-like function of lncRNAs introduces an extra layer of complexity in the miRNA interactome. DIANA-LncBase v1 provided a database of experimentally supported and *in silico* predicted miRNA Recognition Elements (MREs) on lncRNAs. The second version of LncBase ([www.microrna.gr/LncBase](http://www.microrna.gr/LncBase)) presents an extensive collection of miRNA:lncRNA interactions. The significantly enhanced database includes more than 70 000 low and high-throughput, (in)direct miRNA:lncRNA experimentally supported interactions, derived from manually curated publications and the analysis of 153 AGO CLIP-Seq libraries. The new experimental module presents a 14-fold increase compared to the previous release. LncBase v2 hosts *in silico* predicted miRNA targets on lncRNAs, identified with the DIANA-microT algorithm. The relevant module provides millions of predicted miRNA binding sites, accompanied with detailed metadata and MRE conservation metrics. LncBase v2 caters information regarding cell type specific miRNA:lncRNA regulation and enables users to easily identify interactions in 66 different cell types, spanning 36 tissues for hu-

man and mouse. Database entries are also supported by accurate lncRNA expression information, derived from the analysis of more than 6 billion RNA-Seq reads.

## INTRODUCTION

microRNAs (miRNAs) are short (~23 nt) non coding RNAs (ncRNAs) present in hundreds of species and are considered as central gene expression regulators. miRNAs can regulate more than half of the annotated mRNAs in human through target degradation, cleavage and/or translational repression (1).

Recent transcriptome-wide next generation sequencing (NGS) studies unveiled the large number of long non-coding RNA (lncRNA) transcripts and introduced their regulatory roles in the cell (2). lncRNAs are typically longer than 200 nt and are characterized by compartmental, tissue, disease and developmental stage-specific expression. Even though they generally exhibit poor sequence conservation, recent studies have described lncRNA conserved function (3).

lncRNAs are found to interfere in every known level of gene regulation, through *cis* and/or *trans* interactions (4,5). They can also act as guide molecules and protein scaffolds; promoting chromatin remodeling and enabling the formation of cellular complexes, while they may also encode small non-coding RNAs. miRNAs and lncRNAs are currently vigorously studied for their physiological and pathological implications. miRNA:lncRNA interactions have been recently reported, in both nuclear and cytoplasmic com-

\*To whom correspondence should be addressed. Tel: +30 24210 74758; Fax: +30 24210 74997; Email: arhatzig@inf.uth.gr  
Correspondence may also be addressed to Theodore Dalamagas. Tel: +30 210 6875415; Fax: +30 6856804; Email: dalamag@imis.athena-innovation.gr  
Correspondence may also be addressed to Maria D. Paraskevopoulou. Tel: +30 24210 74758; Fax: +30 24210 74997; Email: mparaskevopoulou@inf.uth.gr  
†Equal Contribution.

partments with either specific or high-throughput experimental techniques. For instance, CDR1as/ciRS-7 circular antisense lncRNA harbors multiple miRNA binding sites and can act as a sponge for miR-7 (6); lincMD1 acts as a decoy for two muscle-specific miRNAs in the cytoplasm (7); PTENP1 sequesters miRNAs that target its coding counterpart (8), while distinct validated miRNA:H19 interactions have been found as conserved in human and mouse species (9–12). According to recent evidence, shared miRNA Recognition Elements (MREs) between lncRNAs and mRNAs in miRNA-mediated crosstalk are protected in species by purifying selection (13). Examples of specific miRNA:lncRNA interactions that are experimentally verified in different tissues are provided in Supplementary Table S1.

The target-mimetic lncRNA function introduces an extra layer of complexity in the miRNA interactome, modifying the components of competing endogenous RNA (ceRNA) regulatory networks (14).

## DATABASES INDEXING MIRNA:LNCRNA INTERACTIONS

Currently there are dozens of *in silico* implementations and databases providing miRNA:mRNA regulation. miRNA:lncRNA interactions, on the other hand, have not yet been adequately charted with experimental procedures or *in silico* computational approaches and there are only a few databases indexing the lncRNA sponge-like function.

DIANA-LncBase v1 (15) is considered as the first extensive database dedicated to cataloging miRNA:lncRNA interactions, providing the largest collection of experimentally supported and *in silico* predicted MREs on human and mouse lncRNAs.

miRcode hosts predicted miRNA binding sites on human lncRNA transcripts retrieved from GENCODE v11 (16), while LNCipedia (17) accompanies lncRNA entries with miRNA canonical predictions by using the MirTarget2 algorithm (18). Starbase (19) provides a collection of binding events for different RNA binding proteins. For Argonaute (AGO) binding sites, it can intersect miRanda/mirSVR-predicted (20) miRNA binding sites with the identified CLIP-Seq enriched regions spanning lncRNA transcripts. NPInter (21) integrates information from other repositories and literature regarding non coding regulation and interactions, including ncRNA:protein and ncRNA:miRNA binding events. It supports lncRNA annotation from different resources, while lncRNA:miRNA interactions are obtained from external databases such as Starbase. LncReg and LncNome (22,23), aim to catalog lncRNA-associated regulatory events. These databases also host a restricted number of miRNA binding sites on lncRNAs. These sites are either derived by text mining or are *in silico* inferred from high-throughput datasets.

In this manuscript we present LncBase v2, which has been significantly extended, compared to the previous release. A concise description of the updated database can be found in Table 1. LncBase v2 currently hosts ~70 000 experimentally supported and more than 10 million *in silico* predicted interactions for an integrative meticulously curated collection of lncRNA transcripts. The new database

enables the identification of miRNA:lncRNA regulatory interactions in numerous tissues, cell types and conditions, validated with low yield or high-throughput experimental methodologies. LncBase v2 facilitates the charting of tissue and cell-type-specific miRNA:lncRNA interactions with state-of-the-art experimental techniques. Database entries are escorted with rich metadata, including information on experimental methodologies, evolutionary conservation of miRNA targeted regions and lncRNA transcript expression profiles, assessed by analyzing in-house 58 raw RNA-Seq libraries comprising ~6.1 billion reads. The unique features of DIANA-LncBase are highlighted in Supplementary Table S2, along with a comprehensive summary of other leading repositories indexing experimentally supported miRNA:lncRNA interactions.

## Collected data

An extensive collection of manuscripts has been manually curated, while more than 150 raw NGS datasets harboring miRNA interactions with (non)coding transcripts were analyzed, in order to unveil and explore the lncRNA target-mimetic function.

## Experimental methodologies

Numerous experimental techniques can capture specific direct/indirect or a wider scale of miRNA:gene interactions with varying accuracy (24). Low yield techniques such as western blot or ELISA address interactions that induce reduction at protein level, while qPCR and northern blotting can be also applied to identify specific miRNA:lncRNA or miRNA:mRNA interactions that result in target cleavage and/or induced degradation (25).

Cell-based crosslinking experiments, followed by sequencing (CLIP-Seq) enable the identification of direct miRNA:transcript interactions on a transcriptome-wide scale. HITS-CLIP (high-throughput sequencing of RNA isolated by crosslinking immunoprecipitation) (26), PAR-CLIP (photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation) (27) and CLASH (28) (crosslinking, ligation and sequencing of hybrids), are considered as state-of-the-art CLIP-Seq approaches and can identify AGO:miRNA binding sites on targeted RNAs. The latter methodology can also reveal a fraction of ligated miRNA:target chimeric pairs.

LncBase v2 provides an extensive compendium of miRNA:lncRNA experimentally supported interactions from low yield and high-throughput methodologies extracted from manually curated publications and analyzed raw sequencing data (Table 1). LncBase v2 supports miRNA:lncRNA interactions derived from more than 150 CLIP-Seq (24 PAR-CLIP, 129 HITS-CLIP) libraries across a wide range of cell types, corresponding to the largest collection of AGO-CLIP data compared to any other relevant application. This compilation of high-throughput datasets corresponds to a 16-fold increase compared to the processed CLIP-Seq libraries available in LncBase v1.

**Table 1.** Comparison between LncBase v2 and LncBase v1

		LncBase v2	LncBase v1
Database entries	miRNA species	4	2
	miRNAs in interactions	~1400	127
	unique interacting miRNA:lncRNA pairs	~51 000	4982
	Cell lines	53	5
	Tissues	20	5
	Total interactions	>70 000	4994
Analyzed high-throughput datasets	Studies	22	2
	Conditions	67	6
	Libraries	153	9
Experimental methodologies	Number of methods	12	4
	Description	CLIP-Seq, AGO-IP, biotin miRNA tagging, RNA-Seq, microarrays, northern blot, qPCR, reporter assay	CLIP-Seq, qPCR, reporter assay, northern blot

The table summarizes the experimental module entries of the two databases, including the number of miRNAs targeting lncRNA transcripts, the unique miRNA:lncRNA interacting pairs, different cell lines and tissues supporting miRNA-related experimental methodologies, analyzed CLIP-Seq libraries and associated studies, experimental conditions, as well as the included low/high-throughput experimental methodologies.

### microRNA and lncRNA sequences

miRNA identifiers and sequences were obtained from miR-Base v21 (29). Annotation for lncRNA transcripts was derived from GENCODE v21 (16). GENCODE provides the largest available collection of high quality lncRNA transcripts, spatially classified into four main categories (sense intronic, sense overlapping, antisense and intergenic) according to their transcription orientation and locus of origin relative to protein coding genes. Transcripts annotated as 'processed transcripts' also clustered in the larger lncRNA family were included in LncBase v2. The finalized lncRNA collection includes all GENCODE indexed transcripts, as its main annotation, and also integrates lncRNAs contained in RefSeq (30) and the publication of Cabili *et al.* (31) presenting <90% sequence similarity with GENCODE entries. This integration was essential due to the highly dissimilar spliced transcripts that exist between different lncRNA resources. The final set of lncRNA transcripts comprised 1830 sense, 10 201 antisense, 18 029 long non-coding intergenic RNAs (lincRNAs) and 2163 processed transcripts for *Homo sapiens*. The respective set for *Mus musculus* consisted of 399 sense, 2642 antisense, 4542 lincRNA and 1689 processed transcripts.

## METHODS AND RESULTS

### Text-mining pipeline

Most identified miRNA:lncRNA interactions are fragmented in numerous manuscripts and raw NGS datasets. An in house semi-automated text mining pipeline (24) assisted the selection of publications presenting miRNA and/or lncRNA terms, as well as terms implying putative ncRNA regulation and lncRNA sponge decoy function. The text mining algorithm also detected all different experimental methodologies pertinent to miRNA function.

### miRNA:LncRNA experimentally supported interactions

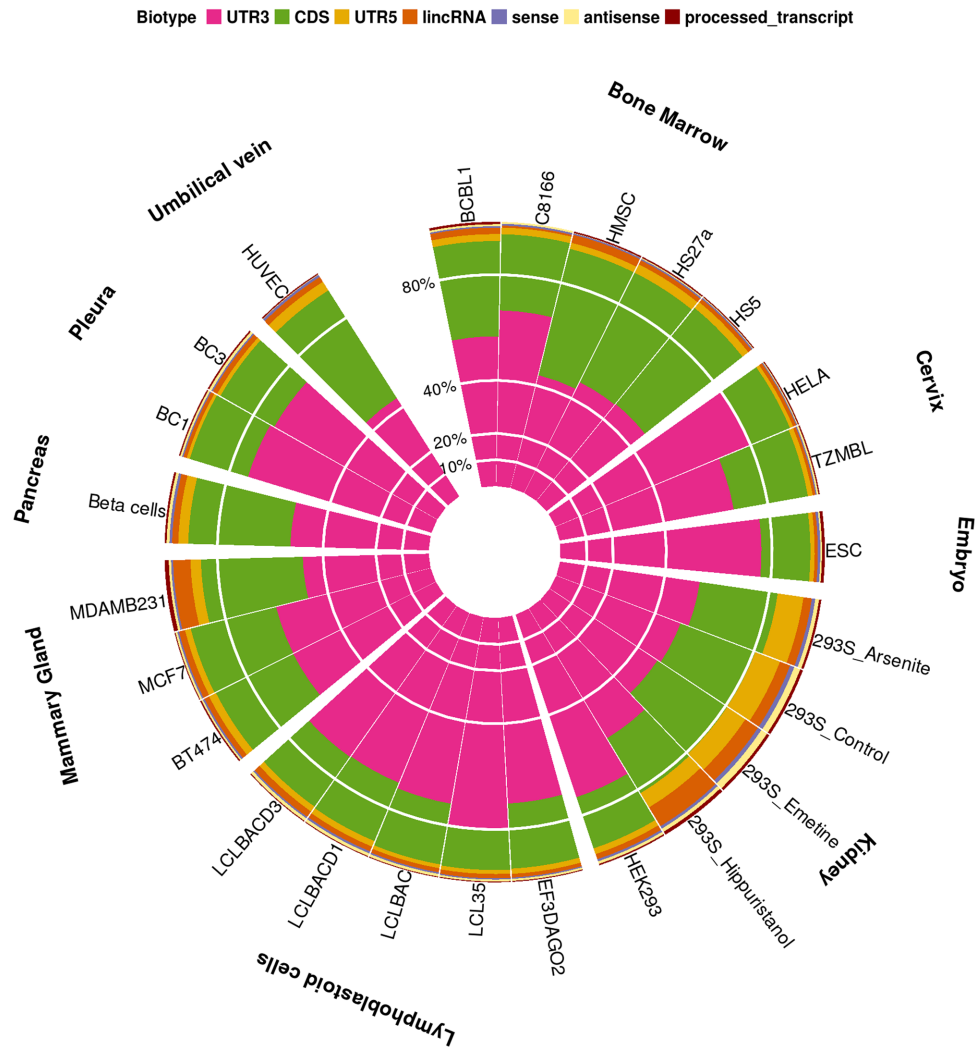
LncBase v2 hosts more than 140 miRNA:lncRNA interactions from specific low yield techniques, while thousands of

miRNA binding sites on lncRNAs have been identified by in-house analyses of > 150 CLIP-Seq libraries.

*Analysis of high-throughput datasets.* Raw CLIP-Seq data were initially quality checked with FastQC (32) and further processed for contaminant removal with a combined use of Minion (33), Trimgalore (34) and Trimmomatic (35). Subsequently, obtained reads were aligned against the reference genome. An in-house developed CLIP-peak-guided MRE search algorithm was subsequently utilized to identify interactions of expressed miRNAs. The algorithm utilizes the search space that is defined by the AGO binding peaks for MRE identification. It takes into account CLIP-Seq-induced mutations and number of reads in peaks. In PAR-CLIP libraries only AGO-enriched peaks with adequate T-to-C or A-to-G (antisense strand) incorporation are retained. The AGO enriched regions are subjected to an MRE detection algorithm that includes features of miRNA binding type, miRNA:lncRNA duplex free energy, site accessibility, AU flanking content and conservation. The CLIP-Seq-based characteristics are used to pinpoint the MRE location, while the miRNA:lncRNA binding features are combined and scored by a general linear model classifier, as initially described by Rezeko *et al.* in microT-CDS algorithm (36), in order to identify the microRNA responsible for the binding. An interaction had to be present in at least two samples in experimental datasets that comprised biological replicates (Supplementary Methods).

The analysis of CLIP-Seq libraries resulted in a set of ~12 900 lncRNA transcripts harboring at least one MRE. More than half of the MREs identified on lncRNAs resided on intronic regions, which may be explained by the underestimation of their spliced length and number of exons. MREs detected on lncRNA introns are appropriately tagged and provided in the current release.

LncBase v2 also hosts 14 PAR-CLIP libraries derived from virus infected cells. For these datasets, host lncRNA transcripts were additionally searched for interactions with viral miRNAs. Expressed viral miRNAs were found to participate in more than 400 miRNA:lncRNA unique interacting pairs.



**Figure 1.** Spatial classification of miRNA-targeted regions as identified in human CLIP-Seq libraries. MREs are being distributed in 3'UTR, 5'UTR, CDS, lincRNA, (anti)sense and processed lincRNA transcript regions across different cell types, with  $5 \pm 2\%$  of the exonic MREs were annotated on lincRNAs.

### miRNA:lincRNA *in silico* predicted interactions

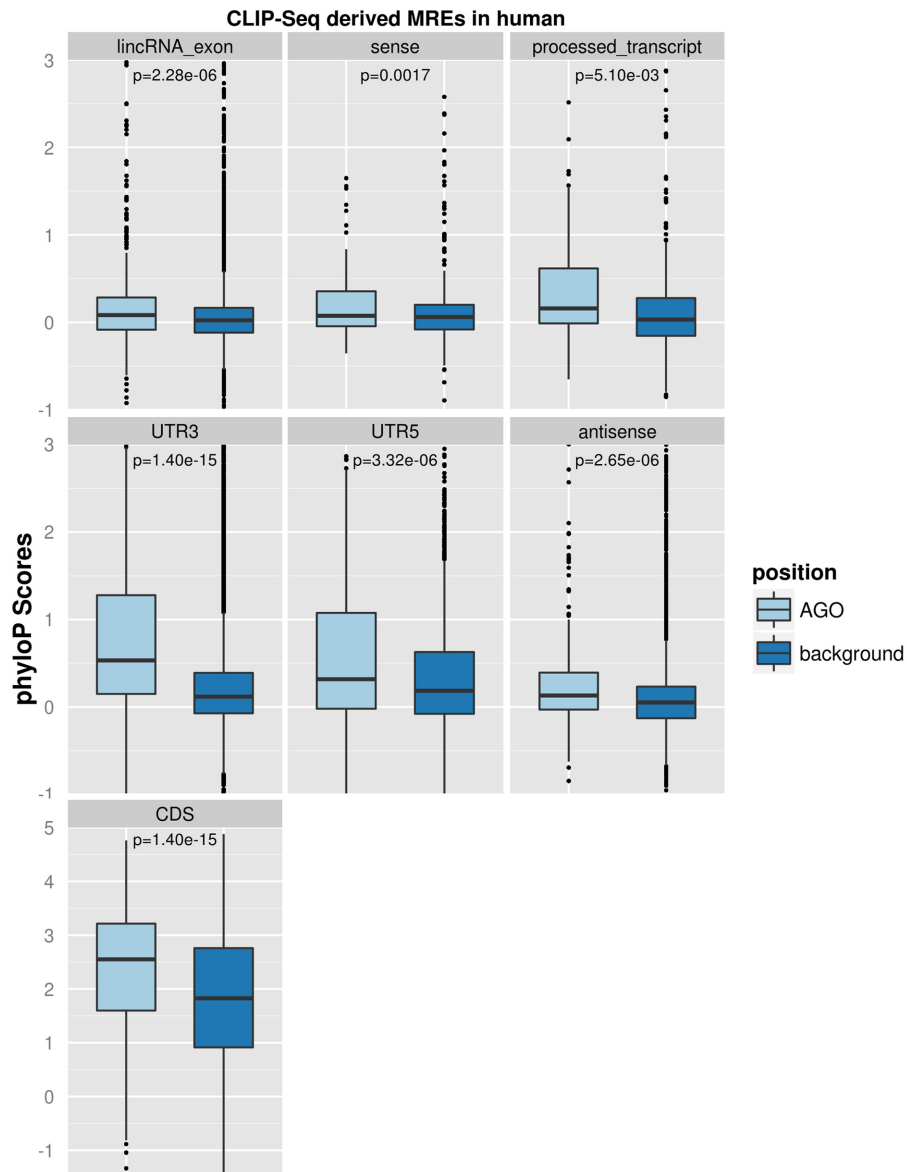
*In silico* target prediction for human and mouse spliced lincRNA sequences was performed with an appropriately adjusted DIANA-microT algorithm (37). MREs were scored separately and each miRNA:lincRNA interacting pair was characterized by a cumulative score which signifies the interaction strength.

Computationally predicted interactions exceed 10 million between 41 229 lincRNAs and 4503 miRNAs, for human and mouse. A subset of these interactions,  $\sim 5$  million, represent a set of highly scored predictions composed of 22 073 lincRNAs, 12 485 antisense, 14 681 sense, 3664 processed transcripts with at least one MRE.

### Tissue/cell type expression

*Collected expression data.* Raw RNA-Seq datasets were retrieved from ENCODE (2,38), UCSC (39) and Gene Expression Omnibus (GEO) (40) repositories in order to assess lincRNA transcript expression in a wide range of cell types for both human and mouse species. RNA-Seq data

corresponding to similar cell types with those in CLIP-Seq samples were preferentially selected. All RNA-Seq libraries were depleted of ribosomal RNAs. Whole transcriptome and poly-A selected libraries were analyzed. The analysis of deeply sequenced RNA samples enabled the extensive identification of expression patterns for targeted lincRNAs. Details concerning the accession codes of the processed RNA-Seq samples and each library specifications are provided in Supplementary Table S3. Raw datasets were quality checked and pre-processed to minimize contaminant sequences. Expression at transcript level was estimated using RSEM (41). Raw reads were aligned against human transcriptomes compiled from Ensembl 75 (GRCh37), RefSeq Release 106 (GRCh38) and Cabili *et al.* as well as mouse transcriptomes derived from Ensembl 81 (GRCm81) (42) and RefSeq Release 104 (GRCm38.p2). Transcript expression information, extracted from analyzed RNA-Seq data across 24 tissues and cell types in Cabili *et al.*, was also incorporated in LncBase v2.



**Figure 2.** Evaluation of CLIP-Seq-supported human MRE substitution rates. miRNA binding sites were spatially classified on CDS, 3'UTR, 5'UTR, lincRNA exons, processed transcripts and (anti)sense lincRNA regions. Random background regions retrieved from each spatially classified genomic group were additionally utilized as controls for the assessment of MRE evolutionary pressure. MRE and background region conservation were estimated using PhyloP pre-computed base-wise values from genome-wide multiple alignments of 46 vertebrate species. Pairwise comparisons revealed that MREs, even in lincRNA regions, are significantly more conserved than their background sequences, which is a phenomenon previously known to occur in MREs located in mRNA 3'UTRs. *P*-values derived from statistical analyses are marked in the relevant panels.

### Database statistics

*Distribution of MREs in (non)coding regions.* miRNA binding sites overlapping transcript exons were predominantly encountered in coding sequences (CDS) and 3'UTR regions of mRNAs, which was consistent in all cell types and tissues. The analyses of > 100 CLIP-Seq libraries in human revealed that  $91 \pm 5\%$  of the identified MREs were found on CDS and 3'UTR regions, and  $5 \pm 2\%$  on intergenic, sense, antisense and processed lincRNA transcripts (Figure 1). A similar distribution of miRNA targeted regions was observed in the HITS-CLIP datasets in mouse (Supplementary Figure S1).

*Clustering of cell types on targeted lincRNAs.* CLIP-Seq libraries from different cell types were hierarchically clustered based on the identified miRNA:lincRNA interactions. Specific cell type groups such as lymphoblastoid, HeLa and bone marrow-derived cell lines in human were found clustered together in the resulting dendrogram; depicting a high similarity in the identified interactions (Supplementary Figure S2). Similar clusters were also observed in targeted mouse lincRNAs of muscle cognate cell lines and thymocytes which are also densely grouped in the dendrogram (Supplementary Figure S3).

**1. Search Terms** → miRNA: hsa-miR-101-3p, hsa-miR-106a-5p; lncRNA: ENSG00000251562

**2. Coordinate Query** → Search by location

**3. Filters** → Filters

**4. Interaction Info**

Gene	miRNA	Pr. score	External Links	Methods
MALAT1	hsa-miR-106a-5p	0.797	mT TB InP mP	IP
MALAT1	hsa-miR-101-3p	0.686	mT TB InP mP	IP RS qP

**5. Gene Details**

**6. UCSC**

**7. miRNA Details**

**8. Disease Cloud**

**9. Experiment Details**

Publication	Tissue	Cell Type	Methods
Karginov FV et al. 2013	Kidney	293S	IP
Pillai MM et al. 2014	Mammary Gland	MDAMB231	IP

**10. mRE Details**

Location	Region	Method	Result	Validation Type	Source
11:65504368-65504396	exon	HITS-CLIP	+	DIRECT	LncBasev2
11:65505816-65505844	exon	HITS-CLIP	+	DIRECT	LncBasev2
11:65504368-65504396	exon	HITS-CLIP	+	DIRECT	LncBasev2
11:65502107-65502135	exon	HITS-CLIP	+	DIRECT	LncBasev2

**11. Change Module** → Go to Predicted module

**12. Help** → Help

**Figure 3.** Snapshot depicting the DIANA-LncBase v2 interface. Queries using one or more microRNAs and/or lncRNAs [1] or even the coordinates of a genomic location [2] are supported. Users can add and remove search terms or filter [3] their results based on cell/tissue type and experimental methodology, as well as the experimental outcome (positive/negative) or type of validation (direct/indirect). LncBase offers extensive information for each identified interaction, such as gene/miRNA details [4,5], as well as active links to UCSC graphical representation [6], Ensembl, miRBase and DIANA disease tag cloud [8]. LncBase also provides useful information for each performed experiment [9], including the methodology, cell or tissue that was utilized, as well as a link to the original publication. There are direct links to external applications such as microT, TarBase, miRPath, where the studied miRNA can be further examined. Interactions are also coupled with miRNA binding site details [10]. Users can navigate between the Experimental and Predicted LncBase v2 modules [11]. The Help button [12] leads to the LncBase Help section.

### Conservation of MRE regions

PhyloP (43) pre-computed scores from genome-wide multiple alignments of 46 and 60 vertebrate species for human and mouse respectively were utilized to assess evolutionary rates of miRNA targeted regions. PhyloP precompiled values were downloaded from the UCSC repository (39). Conservation signals of MRE regions were estimated as mean intensities of the overlapping PhyloP base-wise scores.

A non-redundant set of collapsed MREs collected from all analyzed CLIP-Seq datasets was defined and annotated accordingly to (non)coding exons. MREs with dual annotation due to overlapping transcript regions were excluded from the analysis. In all pairwise comparisons of conserva-

tion, binding sites positioned on lincRNA introns were considered as a separate category.

Stronger evolutionary pressure was observed in miRNA binding sites identified on coding and untranslated mRNA regions. MREs on lincRNA exons were significantly more conserved than those residing in introns, while no differences were observed in substitution rates of MREs on intergenic, sense, antisense and processed lincRNA transcripts (Supplementary Figure S4 and Supplementary Table S4). Statistical analysis of MRE conservation has also been performed for experimentally supported binding sites on mouse lincRNAs (Supplementary Figure S5 and Supplementary Table S5).

Random background regions retrieved from each spatially classified genomic group were additionally utilized as controls for the assessment of MRE evolutionary pressure. Pairwise comparisons revealed that CLIP-Seq-supported miRNA binding sites in human, even in lncRNA regions, are significantly more conserved than their background sequences (Figure 2), which is a phenomenon previously known to occur in MREs located in mRNA 3'UTRs (44). The evaluation of MRE evolutionary rates among different genomic classes compared to their background in mouse species produced similar results and is presented in Supplementary Figure S6.

Non-parametric comparisons were performed with Kruskal–Wallis test in order to detect significant differences on substitution rates between multiple groups. Pairwise Mann–Whitney's U tests were adopted as a post-hoc non-parametric test. All *P*-values were FDR-adjusted to control family-wise error rates due to multiple comparisons (44). All tests were two-sided and *P*-values < 0.05 were considered as statistically significant.

## Interface

The database interface has been completely redesigned to provide an intuitive and easy to use application as well as high flexibility to different user queries (Figure 3). DIANA-LncBase v2 interface comprises two distinct modules for *in silico* predicted and experimentally supported miRNA:lncRNA interactions.

*Module for experimentally supported interactions.* Indexed interactions were enhanced with extensive metadata regarding the supporting publication, type of regulation, experimental methodologies used for miRNA:lncRNA interaction validation, experimental design (including treatment and conditions), as well as cell types and tissue information. Most of the experimentally supported interactions are now coupled with information regarding their genomic location. An advanced filtering/query panel for experimental methodologies, relevant cell types and species is also provided, in order to enable users to identify cell type and tissue-specific miRNA:lncRNA interactions.

*Module for in silico predicted interactions.* Predictions are enriched with information concerning MRE binding sites, structures and conservation. miRNA:lncRNA interactions can be visualized upon selection in an interactive UCSC genome browser (39) graphic, where the user is facilitated with all browser options and additional informative tracks. Prediction interaction score and lncRNA tissue/cell type expression can be utilized for filtering the displayed results.

LncBase v2 indexed interactions are seamlessly interconnected with other available tools in DIANA suite, including TarBase (24) and/or microT-CDS (37) for the identification of competing coding counterparts for miRNA binding and DIANA-miRPath (45) for functional characterization of microRNAs in molecular pathways.

## CONCLUSION

LncRNA functions remain widely uncovered, while others are still under debate. The recently introduced

sponge/decoy role of lncRNAs has been characterized for a few transcripts in specific tissue and/or disease conditions. LncBase v2 provides the first extensive compendium of miRNA:lncRNA *in silico* inferred and experimentally supported interactions covering a wide range of cell types and tissues for human and mouse. The analysis of extensive sequencing data unveiled thousands miRNA:lncRNA interactions, including lncRNAs harboring multiple miRNA binding sites and a set of ~400 unique viral-miRNA:lncRNA interacting pairs in virus infected cells. Spatial classification of miRNA-targeted regions in CLIP-Seq experiments, revealed similar percentages of targeted lncRNA transcripts across different cell types. A considerable amount of MREs residing on lncRNA transcript regions were highly conserved presenting stronger evolutionary pressure than their background regions, while miRNA sites located in lncRNA intronic regions presented accelerated evolutionary rates compared to those in lncRNA exons. AGO-CLIP-Seq cognate cell lines were densely grouped by targeted lncRNAs, possibly indicating a tissue specific miRNA:lncRNA regulation mechanism.

Accurate identification of miRNA coding and non-coding targets is crucial to the understanding of their function and to the detection of competing endogenous interactions. LncBase v2 can help toward this direction and become an indispensable tool for ncRNA regulation research.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

A significant part of the computations were performed on 'ARIS' National HPC Infrastructure of the Greek Research and Technology Network.

## FUNDING

'TOM' [2862], 'ARISTEIA' Action of the 'OPERATIONAL PROGRAMME EDUCATION AND LIFE-LONG LEARNING', General Secretariat for Research and Technology, Ministry of Education, Greece, European Social Fund (ESF) and National Resources; Fondation Santé grant to Artemis Hatzigeorgiou and John S. Latsis Public Benefit Foundation. Funding for open access charge: Fondation Santé grant to Artemis Hatzigeorgiou.

*Conflict of interest statement.* None declared.

## REFERENCES

- Huntzinger, E. and Izaurralde, E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.*, **12**, 99–110.
- Consortium, E.P. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
- Johnsson, P., Lipovich, L., Grander, D. and Morris, K.V. (2014) Evolutionary conservation of long non-coding RNAs; sequence, structure, function. *Biochim. Biophys. Acta*, **1840**, 1063–1071.
- Rinn, J.L. and Chang, H.Y. (2012) Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.*, **81**, 145–166.
- Gutschner, T. and Diederichs, S. (2012) The hallmarks of cancer: a long non-coding RNA point of view. *RNA Biol.*, **9**, 703–719.

6. Hansen, T.B., Wiklund, E.D., Bramsen, J.B., Villadsen, S.B., Statham, A.L., Clark, S.J. and Kjems, J. (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.*, **30**, 4414–4422.
7. Cesana, M., Cacchiarelli, D., Legnini, T., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A. and Bozzoni, I. (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, **147**, 358–369.
8. Poliseno, L., Salmena, L., Zhang, J., Carver, B., Haveman, W.J. and Pandolfi, P.P. (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, **465**, 1033–1038.
9. Imig, J., Brunschweiler, A., Brummer, A., Guennebig, B., Mittal, N., Kishore, S., Tsikrika, P., Gerber, A.P., Zavolan, M. and Hall, J. (2015) miR-CLIP capture of a miRNA targetome uncovers a lincRNA H19-miR-106a interaction. *Nat. Chem. Biol.*, **11**, 107–114.
10. Zhang, L., Yang, F., Yuan, J.H., Yuan, S.X., Zhou, W.P., Huo, X.S., Xu, D., Bi, H.S., Wang, F. and Sun, S.H. (2013) Epigenetic activation of the MiR-200 family contributes to H19-mediated metastasis suppression in hepatocellular carcinoma. *Carcinogenesis*, **34**, 577–586.
11. Shi, Y., Wang, Y., Luan, W., Wang, P., Tao, T., Zhang, J., Qian, J., Liu, N. and You, Y. (2014) Long non-coding RNA H19 promotes glioma cell invasion by deriving miR-675. *PLoS one*, **9**, e86295.
12. Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., Chen, N., Sun, F. and Fan, Q. (2010) CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.*, **38**, 5366–5383.
13. Tan, J.Y., Sirey, T., Honti, F., Graham, B., Piovesan, A., Merkenschlager, M., Webber, C., Ponting, C.P. and Marques, A.C. (2015) Corrigendum: Extensive microRNA-mediated cross-talk between lincRNAs and mRNAs in mouse embryonic stem cells. *Genome Res.*, **25**, 1410–1411.
14. 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.
15. Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Reczko, M., Maragkakis, M., Dalamagas, T.M. and Hatzigeorgiou, A.G. (2013) DIANA-LncBase: Extensive microRNA-mediated and computationally predicted microRNA targets on long non-coding RNAs. *Nucleic Acids Res.*, **41**, D239–D245.
16. 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.
17. 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 lincRNA sequences. *Nucleic Acids Res.*, **43**, 4363–4364.
18. Wang, X. and El Naqa, I.M. (2008) Prediction of both conserved and nonconserved microRNA targets in animals. *Bioinformatics*, **24**, 325–332.
19. 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.
20. Betel, D., Koppal, A., Agius, P., Sander, C. and Leslie, C. (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.*, **11**, R90.
21. 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.
22. Bhartiya, D., Pal, K., Ghosh, S., Kapoor, S., Jalali, S., Panwar, B., Jain, S., Sati, S., Sengupta, S., Sachidanandan, C. *et al.* (2013) lincRNome: a comprehensive knowledgebase of human long noncoding RNAs. *Database (Oxford)*, **2013**, bat034.
23. Zhou, Z., Shen, Y., Khan, M.R. and Li, A. (2015) LncReg: a reference resource for lincRNA-associated regulatory networks. *Database (Oxford)*, **2015**, doi:10.1093/database/bav083.
24. Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos, I.L., Manioudis, S., Karathanou, K., Kalfakakou, D. *et al.* (2015) DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.*, **43**, D153–D159.
25. Hansen, T.B., Jensen, T.I., Clausen, B.H., Bramsen, J.B., Finsen, B., Damgaard, C.K. and Kjems, J. (2013) Natural RNA circles function as efficient microRNA sponges. *Nature*, **495**, 384–388.
26. Chi, S.W., Zang, J.B., Mele, A. and Darnell, R.B. (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
27. Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M. Jr, Jungkamp, A.C., Munschauer, M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
28. Helwak, A., Kudla, G., Dudnakova, T. and Tollervey, D. (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell*, **153**, 654–665.
29. Kozomara, A. and Griffiths-Jones, S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
30. 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.
31. 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.
32. Andrews, S. (2015) FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
33. Davis, M.P., van Dongen, S., Abreu-Goodger, C., Bartonicek, N. and Enright, A.J. (2013) Kraken: a set of tools for quality control and analysis of high-throughput sequence data. *Methods*, **63**, 41–49.
34. Krueger, F. (2015) Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. [http://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).
35. Bolger, A.M., Lohse, M. and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
36. Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I. and Hatzigeorgiou, A.G. (2012) Functional microRNA targets in protein coding sequences. *Bioinformatics*, **28**, 771–776.
37. Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Reczko, M., Filipidis, C., Dalamagas, T. and Hatzigeorgiou, A.G. (2013) DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Res.*, **41**, W169–W173.
38. Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D. *et al.* (2014) A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, **515**, 355–364.
39. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M. *et al.* (2015) The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.*, **43**, D670–D681.
40. Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., Marshall, K.A., Phillippy, K.H., Sherman, P.M., Holko, M. *et al.* (2013) NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.*, **41**, D991–D995.
41. Li, B. and Dewey, C. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**, 323.
42. 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.
43. Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R. and Siepel, A. (2010) Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, **20**, 110–121.
44. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.
45. Vlachos, I.S., Zagganas, K., Paraskevopoulou, M.D., Georgakilas, G., Karagkouni, D., Vergoulis, T., Dalamagas, T. and Hatzigeorgiou, A.G. (2015) DIANA-miRPath v3.0: deciphering microRNA function with experimental support. *Nucleic Acids Res.*, **43**, W460–W466.