BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 15 2012, pages 2062–2063
doi:10.1093/bioinformatics/bts344

Sequence analysis

Advance Access publication May 19, 2012

miRcode: a map of putative microRNA target sites in the long non-coding transcriptome

Ashwini Jeggari[1], Debora S Marks[2] and Erik Larsson[1,*]

[1]Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, SE-405 30, Sweden and [2]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

Associate Editor: Martin Bishop

ABSTRACT

Summary: Although small non-coding RNAs, such as microRNAs, have well-established functions in the cell, long non-coding RNAs (lncRNAs) have only recently started to emerge as abundant regulators of cell physiology, and their functions may be diverse. A small number of studies describe interactions between small and lncRNAs, with lncRNAs acting either as inhibitory decoys or as regulatory targets of microRNAs, but such interactions are still poorly explored. To facilitate the study of microRNA–lncRNA interactions, we implemented miRcode: a comprehensive searchable map of putative microRNA target sites across the complete GENCODE annotated transcriptome, including 10 419 lncRNA genes in the current version.

Availability: http://www.mircode.org

Contact: erik.larsson@gu.se

Supplementary Information: Supplementary data are available at *Bioinformatics* online.

Received on April 10, 2012; revised on May 18, 2012; accepted on June 10, 2012

1 INTRODUCTION

Large-scale studies in recent years have revealed that mammalian genomes encode thousands of long (>200 nt) transcripts that lack coding capacity, but are otherwise messenger RNA-like. These are collectively referred to as long non-coding RNAs (lncRNAs) (Mercer *et al.*, 2009). Although their overall biological importance has been debated, early functional examples were discovered more than 20 years ago, notably *H19* (Brannan *et al.*, 1990) and *XIST* (Brown *et al.*, 1991). Novel lncRNAs are now being uncovered at an increasing rate, with molecular functions that include recruitment of histone-modifying complexes to chromatin [e.g. *HOTAIR* and *HOTTIP* (Rinn *et al.*, 2007; Wang *et al.*, 2011)] and modulation of transcription and splicing by molecular interaction with relevant factors [e.g. *GAS5* and *MALAT1* (Bernard *et al.*, 2010; Kino *et al.*, 2010)].

A small number of studies describe interactions between small and lncRNAs, with lncRNAs acting either as inhibitory decoys of microRNAs (Ebert and Sharp, 2010) or as regulatory targets. In humans, miR-671 targets an antisense transcript of the human *CDR1* gene (Hansen *et al.*, 2011), and miR-29 can regulate the lncRNA *MEG3* in hepatocellular cancer, although only indirectly (Braconi *et al.*, 2011). Long non-coding transcripts that derive from ultra-conserved regions (T-UCRs) have also been suggested to be microRNA targets (Calin *et al.*, 2007). In plants, the IPS1 lncRNA inhibits miR-399 through a sponge/decoy effect (Franco-Zorrilla *et al.*, 2007). Herpesvirus-encoded RNAs can bind and inhibit human host miR-27 (Cazalla *et al.*, 2010), and the *HULC* lncRNA can bind and sequester miR-372 in liver cancer (Wang *et al.*, 2010). A pseudogene of the *PTEN* tumor suppressor can compete for microRNA binding with its coding counterpart (Poliseno *et al.*, 2010), and microRNA inhibition by lncRNAs is important during muscle differentiation (Cesana *et al.*, 2011). The decoy hypothesis is further supported by the observation that microRNAs with many targets in the cell tend to have a diluted effect on each individual target (Arvey *et al.*, 2010).

A recent study used lentiviral small hairpin RNAs to silence 147 lncRNAs at an average efficacy of 75% (Guttman *et al.*, 2011), demonstrating that lncRNAs in general are susceptible to regulation by Argonaute–small RNA complexes despite frequent nuclear localization. However, existing web-accessible microRNA target prediction databases, such as PicTar (Krek *et al.*, 2005), miRanda (Betel *et al.*, 2008) or TargetScan (Friedman *et al.*, 2009), are focused on 3′-untranslated region (UTR) of coding genes and fail to provide predictions for the long non-coding transcriptome.

To simplify the study of microRNA–lncRNA interactions, we here describe miRcode: a comprehensive map of putative microRNA target sites across the GENCODE long non-coding transcriptome (10 419/15 977 lncRNAs genes/transcripts in the current version based on GENCODE V11). miRcode is designed to be an easy to use, web-based tool, with search functionalities to aid hypothesis generation starting from a lncRNA or microRNA of interest. Custom genome browser views and downloadable tab-delimited files are also accessible through the miRcode web interface. miRcode additionally covers other GENCODE gene classes, including 12 549 pseudogenes and 19 999 coding genes both in typical (3′-UTR) and atypical (5′-UTR and CDS) positions.

2 IMPLEMENTATION

miRcode identifies putative target sites based on established principles: seed complementarity and evolutionary conservation (see Supplementary Material for detailed methods). The seed region, encompassing bases 2–8 from the 5′-end of the microRNA, is the major sequence determinant of microRNA targeting (Lewis *et al.*, 2003). The miRcode pipeline (Fig. 1), implemented using Perl, Matlab, PHP and MySQL, searches for complementary matches to established (Friedman *et al.*, 2009) microRNA seed families across GENCODE (Harrow *et al.*, 2006) transcripts. We consider 7-mers

*To whom correspondence should be addressed.

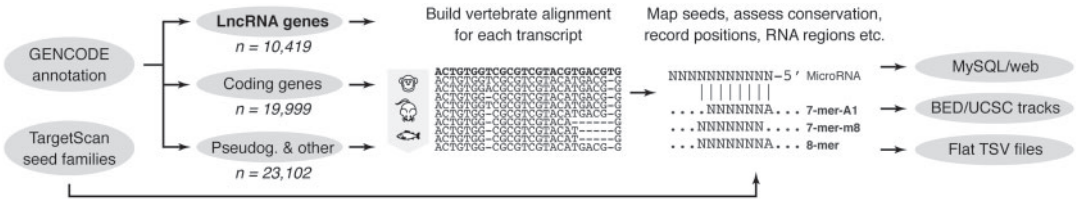

Fig. 1. Workflow for mapping of conserved putative microRNA target sites in lncRNAs

and adenosine-flanked 6-mer and 7-mer matches, but not 6-mers as these are only marginally effective (Grimson *et al.*, 2007; Selbach *et al.*, 2008).

GENCODE represents a comprehensive, high-quality description of the polyA+ transcriptome. It is updated on a regular basis and based largely on full-length or near full-length complementary DNA evidence and additionally contains many known RNA genes and microRNAs. Although all of GENCODE is analyzed and accessible in miRcode, we define a subset of lncRNA genes that produce only predicted non-coding transcripts with a mature (spliced) length of >200 nt. lncRNAs are further subdivided into intergenic (not overlapping with any coding gene) and non-intergenic.

To assess evolutionary conservation, a 46-way Multiz vertebrate genomic multiple alignment (Blanchette *et al.*, 2004; Fujita *et al.*, 2011) is remapped onto transcripts, and site conservation levels are determined based on site presence in primates, non-primate mammals and non-mammal vertebrates. Transcript regions (3′-UTR, CDS and 5′-UTR in case of coding genes) and possible overlaps with repeat sequences are recorded for each site. Sites are mapped first to transcripts to allow identification of splice-junction-spanning sites, and subsequently aggregated into non-redundant gene-level sets. Predictions are finally made accessible through a web interface.

3 FUNCTIONALITY

The miRcode interface provides basic search functionality for finding putative microRNA–target sites in lncRNAs of interest or predicted targets of specific microRNAs. Sites are returned in the form of lists, aggregated on genes, where conservation levels (fraction of species where site is present) are presented separately for primates, non-primate placental mammals and non-mammal vertebrates. In addition, custom UCSC browser views enable browsing of target sites in a genome context. Tab-delimited text files and BED files provide convenient access to whole-transcriptome target predictions for use in computational projects.

In summary, we provide, in several formats, a pan-GENCODE microRNA site map to facilitate further investigation into microRNA regulation of lncRNAs as well as other atypical target regions such as pseudogenes and 5′-UTRs.

Funding: Grants from the Swedish Medical Research Council; the Assar Gabrielsson Foundation; the Magnus Bergvall Foundation; and the Lars Hierta Memorial Foundation.

Conflict of Interest: None declared.

REFERENCES

Arvey,A. *et al.* (2010) Target mRNA abundance dilutes microRNA and siRNA activity. *Mol. Syst. Biol.*, **6**, 363.

Bernard,D. *et al.* (2010) A long nuclear-retained non-coding RNA regulates synaptogenesis by modulating gene expression. *EMBO J.*, **29**, 3082–3093.

Betel,D. *et al.* (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.

Blanchette,M. *et al.* (2004) Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.*, **14**, 708–715.

Braconi, C. *et al.* (2011) microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. *Oncogene*, **30**, 4750–4756.

Brannan,C.I. *et al.* (1990) The product of the H19 gene may function as an RNA. *Mol. Cell. Biol.*, **10**, 28–36.

Brown,C.J. *et al.* (1991) A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature*, **349**, 38–44.

Calin,G.A. *et al.* (2007) Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell*, **12**, 215–229.

Cazalla,D. *et al.* (2010) Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science*, **328**, 1563–1566.

Cesana,M. *et al.* (2011) A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, **147**, 358–369.

Ebert,M.S. and Sharp,P.A. (2010) Emerging roles for natural microRNA sponges. *Curr. Biol.*, **20**, R858–R861.

Franco-Zorrilla,J.M. *et al.* (2007) Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.*, **39**, 1033–1037.

Friedman,R.C. *et al.* (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.

Fujita,P.A. *et al.* (2011) The UCSC Genome Browser database: update 2011. *Nucleic Acids Res.*, **39**, D876–D882.

Grimson,A. *et al.* (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell.*, **27**, 91–105.

Guttman,M. *et al.* (2011) lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature*, **477**, 295–300.

Hansen,T.B. *et al.* (2011) miRNA-dependent gene silencing involving Ago2-mediated cleavage of a circular antisense RNA. *EMBO J.*, **30**, 4414–4422.

Harrow,J. *et al.* (2006) GENCODE: producing a reference annotation for ENCODE. *Genome Biol.*, **7** (Suppl 1), S4.1–4.9.

Kino,T. *et al.* (2010) Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci. Signal*, **3**, ra8.

Krek,A. *et al.* (2005) Combinatorial microRNA target predictions. *Nat. Genet.*, **37**, 495–500.

Lewis,B.P. *et al.* (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787–798.

Mercer,T.R. *et al.* (2009) Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.*, **10**, 155–159.

Poliseno,L. *et al.* (2010) A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature*, **465**, 1033–1038.

Rinn,J.L. *et al.* (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell*, **129**, 1311–1323.

Selbach,M. *et al.* (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.

Wang,J. *et al.* (2010) CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. *Nucleic Acids Res.*, **38**, 5366–5383.

Wang,K.C. *et al.* (2011) A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*, **472**, 120–124.