BIOINFORMATICS APPLICATIONS NOTE Vol. 29 no. 15 2013, pages 1898–1899
doi:10.1093/bioinformatics/btt296

Sequence analysis

Advance Access publication May 24, 2013

miRTCat: a comprehensive map of human and mouse microRNA target sites including non-canonical nucleation bulges

Ka-Kyung Kim[1,2], Juyoung Ham[2] and Sung Wook Chi[1,2,3,*]

[1]Samsung Biomedical Research Institute, [2]Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University and [3]Samsung Research Institute for Future Medicine, Samsung Medical Center, 81 Irwon-ro, Gangnam-gu, Seoul 135-710, Korea

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: MicroRNAs (miRNAs) regulate various biological functions by binding hundreds of transcripts to impart post-transcriptional repression. Recently, by applying a transcriptome-wide experimental method for identifying miRNA target sites (Ago HITS-CLIP), a novel non-canonical target site, named 'nucleation bulge', was discovered as widespread, functional and evolutionally conserved. Although such non-canonical nucleation bulges have been proven to be predictive by using 'pivot pairing rule' and sequence conservation, this approach has not been applied yet. To facilitate the functional studies of non-canonical miRNA targets, we implement miRTCat: a comprehensive searchable map of miRNA target sites, including non-canonical nucleation bulges, not only mapped in experimentally verified miRNA-bound regions but also predicted in all 3′-untranslated regions (3′-UTRs) derived from human and mouse (~15.6% as expected false-positive results).

Availability: http://ion.skku.edu/mirtcat.

Contact: swchi@skku.edu

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

Received on February 5, 2013; revised on May 7, 2013; accepted on May 20, 2013

1 INTRODUCTION

MicroRNAs (miRNAs) are small RNA molecules (~22 nt) that induce post-transcriptional repression of messenger RNAs (mRNAs), affecting various biological phenotypes (He and Hannon, 2004). Animal miRNAs typically recognize mRNA targets through imperfect matches, making the problem of predicting miRNA target sites challenging. Among these partial pairings, it is well known that miRNAs bind canonical miRNA target sites through short consecutive pairings with their seed regions (position 2–8; called 'seed pairing rule') (Bartel, 2009). Besides sequence complementarity, evolutional conservation has been considered, improving the performance of miRNA target predictions (Lewis *et al.*, 2005). However, a critical caveat of all seed-centric approaches is its inability to identify non-canonical miRNA target sites (Didiano and Hobert, 2006), overlooking priori recognized non-canonical target sites validated as functionally important (Tay *et al.*, 2008; Vella *et al.*, 2004).

*To whom correspondence should be addressed.

Recently, by applying cross-linking immunoprecipitation method (CLIP) coupled with high-throughput sequencing (HITS) (Licatalosi, 2008) to a protein called Argonaute (Ago HITS-CLIP), all miRNA-binding regions in target mRNAs were experimentally mapped in high resolution (Chi *et al.*, 2009). However, not all miRNA-bound regions could be explained by the canonical seed pairing rule, implicating existence of non-canonical miRNA–target interactions. By performing bioinformatic analyses in such orphan regions, non-canonical sites called 'nucleation bulges' were identified and validated as widespread, functional and evolutionally conserved (Chi *et al.*, 2012). Non-canonical nucleation bulge sites are predictable, as the mode of their interactions is generalized by 'pivot pairing rule', dictating a nucleotide in bulge (position 5 and 6) to be competent to pair with a nucleotide in position six of a miRNA, named 'pivot' to confer the thermodynamic stability on consecutive five base pairings (position 2–6, Supplementary Fig. S1A). Such initial five base pairings are proposed to be served as nuclei of interactions (nucleation), followed by bulge formation where originally matched pivot nucleotide in position six becomes bulged-out and propagating hybridization toward 3′-end of the miRNA. However, such a rule based on a pivot pairing has not been applied for miRNA target prediction yet.

To offer opportunities of studying non-canonical miRNA targets, here we describe miRNA target maps specialized for nucleation bulge sites, named 'miRTCat'. miRTCat provides comprehensive maps searchable for miRNA target sites including non-canonical nucleation bulges, of which locations are not only predicted in all 3′-untranslated regions (3′-UTRs; annotated by RefSeq in human and mouse) but also mapped in miRNA-bound regions that are experimentally verified (Chi *et al.*, 2009, 2012; Hafner *et al.*, 2010; Kishore *et al.*, 2011; Leung *et al.*, 2011). Such functionalities will aid to generate hypothesis starting from interactions between miRNA and non-canonical targets.

2 IMPLEMENTATION

Based on the established principles: sequence complementarity and evolutional conservation (see Supplementary Material for details), miRTCat pipeline (Fig. 1A), implemented mainly by using Perl, PHP and MySQL, identifies miRNA target sites of both canonical and non-canonical types for microRNA families, which were previously defined as mostly conserved across vertebrates (Friedman *et al.*, 2009). For canonical seed sites, a seed pairing rule is used to find motifs >6mers that match to the 5′-end seed regions of miRNAs (position 2–8; Fig. 1A and Supplementary Fig. S1A). For non-canonical nucleation bulge sites, a pivot pairing rule is applied to

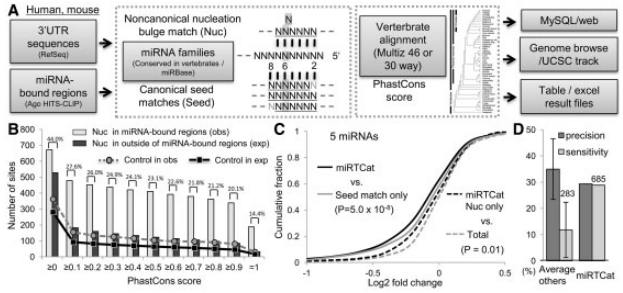

Fig. 1. miRTCat (**A**) Workflow for mapping of non-canonical nucleation bulge (Nuc) and canonical seed sites (Seed). A pivot nucleotide (position six) and pivot competent nucleotides are indicated as shade in the middle panel. (**B**) Number of nucleation bulges in miRNA-bound regions in the range of evolutional conservation (PhastCons cut-offs) compared with the expected number of sites in outside of miRNA-bound regions. (**C**) Cumulative fraction analysis for different types of targets based on the analysis of experiment datasets for miRNA-dependant repression (miR-1, miR-16, miR-30 a, miR-155 and let-7 b) (Alexiou *et al.*, 2009); *P*-values, Kolmogorov–Smirnov test. (**D**) Comparison with average performance of other programs based on the same datasets used in (C); error bars, SD

search 7mer motifs of which bulged sequence (position 5 and 6) is complement to pair with pivot (position six; Fig. 1A and Supplementary Fig. S1A). To consider evolutional conservation, phastCons score, derived from Multiz 46 or 30 way whole-genome alignments (Dreszer *et al.*, 2011; Siepel *et al.*, 2005) is used.

3 RESULTS

As previously reported (Chi *et al.*, 2012), nucleation bulge sites have propensity to be more conserved in miRNA-bound regions [mean phastCons score (PCS) = 0.59 ± 0.22] like seed sites (PCS = 0.82 ± 0.17), compared with the control sites (PCS = 0.32 ± 0.21) in human 3′-UTRs. Taking this notion, we examine the performance of miRTCat depending on the level of conservation by using experimentally identified miRNA-bound regions (Chi *et al.*, 2009), determine default cut-offs (PCS ≥ 0.9 for both sites) and estimate the false-positive rate (15.6%; 13.6% for seed matches and 20.1% for nucleation bulges) (Fig. 1B and Supplementary Fig. S1B–D). Validation analysis using experiment datasets of miRNA-dependant repression (Alexiou *et al.*, 2009) shows that transcripts identified by miRTCat (PCS ≥ 0.9) are more significantly downregulated than the targets found by simple seed matching ($P = 5.0 \times 10^{-8}$) (Fig. 1C and Supplementary Fig. S2). Furthermore, miRTCat is estimated to outperform in sensitivity (~2.6-fold increase) with comparable precision with other prediction methods (Fig. 1D and Supplementary Fig. S3). Overall, miRTCat offers comprehensive mappings with increased number of target sites by identifying non-canonical nucleation bulges, predicting *bona fide* 245 454 sites (33 461 nucleation bulges (7mers), comparable with 47 478 7mer seed sites) in human 3′-UTRs and mapping 263 361 sites (32 471 nucleation bulges, 44 665 7mer seed sites) in all compiled data from experimentally identified miRNA-bound regions (3272).

4 FUNCTIONALITY

The miRTCat web-interface offers search functionality of both canonical and non-canonical miRNA targets sites in two species, human and mouse. Users can retrieve two types of identification: *bona fide* prediction in 3′-UTR and experimentally mapping results in miRNA-bound regions, freely defining cut-off values of PCS (0.1–1.0) for both non-canonical nucleation bulge and canonical seed sites. Users are informed to decide cut-off values for their purpose of search considering experimentally estimated predictive values depending on the extent of evolutional conservation (Fig. 1B and Supplementary Fig. S1D). Centered on target mRNAs and/or regulator miRNAs, results of the target sites are returned in the form of lists as a table providing sorting functionality and links for UCSC genome browser (Dreszer *et al.*, 2011), enabling to browse with other genome information. In addition, all results are also available for download as an excel file format and easily used for other computational analysis.

In summary, here we provide a comprehensive map of miRNA target sites including non-canonical nucleation bulges in several formats to facilitate further biological studies of miRNA regulation through non-canonical targets, which has not been offered by current approaches based on seed pairings.

Funding: Korean Health Technology R&D Project, Ministry of Health & Welfare (A111989).

Conflict of Interest: none declared.

REFERENCES

Alexiou,P. *et al.* (2009) Lost in translation: an assessment and perspective for computational microRNA target identification. *Bioinformatics*, **25**, 3049–3055.
Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
Chi,S.W. *et al.* (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*, **460**, 479–486.
Chi,S.W. *et al.* (2012) An alternative mode of microRNA target recognition. *Nat. Struct. Mol. Biol.*, **19**, 321–327.
Didiano,D. and Hobert,O. (2006) Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions. *Nat. Struct. Mol. Biol.*, **13**, 849–851.
Dreszer,T.R. *et al.* (2011) The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res.*, **40**, D918–D923.
Friedman,R.C. *et al.* (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
Hafner,M. *et al.* (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*, **141**, 129–141.
He,L. and Hannon,G.J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.*, **5**, 522–531.
Kishore,S. *et al.* (2011) A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods*, **8**, 559–564.
Leung,A.K. *et al.* (2011) Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat. Struct. Mol. Biol.*, **18**, 237–244.
Lewis,B.P. *et al.* (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
Licatalosi,D.D. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. *Nature*, **456**, 464–469.
Siepel,A. *et al.* (2005) Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, **15**, 1034–1050.
Tay,Y. *et al.* (2008) MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature*, **455**, 1124–1128.
Vella,M.C. *et al.* (2004) The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3′UTR. *Genes Dev.*, **18**, 132–137.