Competing endogenous RNA database

Aaron L Sarver[1] & Subbaya Subramanian[1, 2*]

[1]Masonic Cancer Center; [2]Division of Basic and Translational Research, Department of Surgery, University of Minnesota, MN 55455; Subbaya Subramanian – E mail: subree@umn.edu; Phone: 612-626-4330; Fax: 612-626-7031; *Corresponding author

Abstract:
A given mRNA can be regulated by interactions with miRNAs and in turn the availability of these miRNAs can be regulated by their interactions with alternate mRNAs. The concept of regulation of a given mRNA by alternate mRNA (competing endogenous mRNA) by virtue of interactions with miRNAs through shared miRNA response elements is poised to become a fundamental genetic regulatory mechanism. The molecular basis of the mRNA-mRNA cross talks is via miRNA response elements, which can be predicted based on both molecular interaction and evolutionary conservation. By examining the co-occurrence of miRNA response elements in the mRNAs on a genome-wide basis we predict competing endogenous RNA for specific mRNAs targeted by miRNAs. Comparison of the mRNAs predicted to regulate PTEN with recently published work, indicate that the results presented within the competing endogenous RNA database (*ceRDB*) have biological relevance.

Availability: http://www.oncomir.umn.edu/cefinder/

Key words: ceRNAs, MRE, microRNA response elements, database, competing endogenous RNAs database, ceRDB

Background:
MicroRNAs (miRNAs) play an important role in almost all biological functions [1]. Transcriptional deregulations in miRNAs have been implicated in disease processes including cancers and developmental disorders [2]. It has been well established that a single miRNA can regulate the expression of many mRNAs/ transcripts and an mRNA can be regulated by multiple miRNAs [1]. miRNA gene regulation is mediated by a complex set of proteins termed RNA induced silencing complex. The miRNAs are guided to the miRNA response elements (MRE) present in the target mRNAs, which may result in transcript degradation and/or translational inhibition [3]. Recently it has been established that miRNA activity on the target gene can be influenced by the presence or absence of other competing endogenous (ceRNA) mRNAs that contain shared MREs [4-7]. These miRNA activity modulators can act as a sponge, absorbing and releasing miRNA based on the level of the transcript. Several modulators of miRNA activity have been recently characterized [8]. Salmena et al proposed a hypothesis that these modulators can communicate with each other in a miRNA dependent manner mediated through MREs [9]. This complex miRNA-mRNA network and interactions opens up a new chapter in miRNA-mediated gene regulation. However, currently t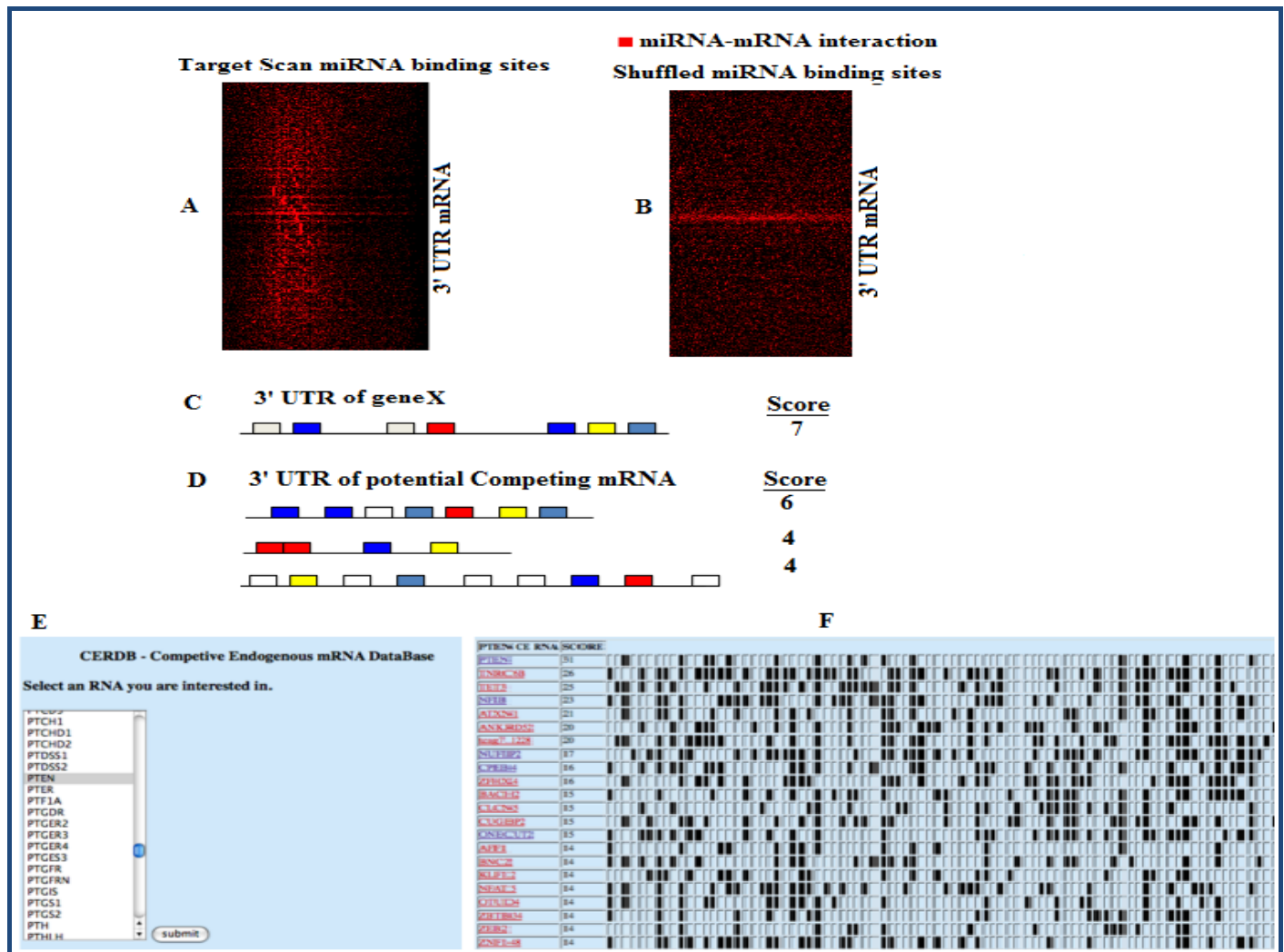here are no publicly available resources that identify and catalog the list of genes that can act as miRNA activity modulators or ceRNAs. Here we developed a comprehensive and easy to use resource named '*competing endogenous RNA database (ceRDB)*' that lists potential MRE containing genes that can act in a sponge like fashion for a given mRNA based on a set of scoring and ranking criteria.

Methodology:
MiRNA-mRNA target interactions were obtained from http://www.targetscan.org Release 5.2 June 2011. The predicted conserved target information file was parsed to obtain 54979 conserved human miRNA-mRNA interactions. To explore the structure of the dataset, the list of interactions was converted into a matrix containing 153 miRNA families on the X-axis and 9448 target mRNAs on the Y-axis. The presence of a predicted conserved miRNA-mRNA interaction is defined by the presence of a '1' at the defined gene row miRNA column corresponding to the interaction. The absence of an interaction is defined by the presence of a '0' at the corresponding interaction. To shuffle the matrix, interactions between each miRNA and mRNA were randomly assigned maintaining the total number of interactions for each mRNA. Both the real matrix and the shuffled matrix were filtered to only show genes with more than 5 miRNA binding sites and these were

clustered using Gene Cluster 3.0 hierarchical clustering of both the X and the Y-axis using Centroid linkage. The resulting clustered matrixes were visualized using Java Treeview. To score potential ceRNA interactions, the 54979 human interactions were loaded into a mySQL database and when the user selects a given mRNA all predicted miRNA targets for the given mRNA are obtained. These miRNA are then used to define all mRNAs that contain binding sites for the set of miRNAs. For each mRNA, an interaction score is then defined by adding up the total number of miRNA binding sites that overlap with the miRNA for a given mRNA. This interaction score is then used to sort the results and the top 50 predicted potential ceRNAs are returned. This process is carried out on the fly using PHP interactions with mySQL in a similar fashion as previously described in our publicly available databases such as sarcoma microRNA expression database (S-MED).

Figure 1: Visualization of co-occurrence in predicted miRNA-mRNA interactions. Heat map showing the presence of predicted miRNA-binding sites on the X-axis and the genes that contain the binding sites in the 3'UTR on the Y-axis. Only genes that show more than 5 binding sites are shown for **(A)** predicted interactions and **(B)** predicted interactions after shuffling. **(C)** Predicting competing mRNA via miRNA-mRNA interactions. miRNA binding site predictions in the 3'UTR are shown as colored boxes. The 'Score' is generated by counting the number of conserved predicted interactions. In this hypothetical case shown there are 7 predicted binding elements in the 3'UTR of the gene. **(D)** To predict potential competing RNA for the gene shown in A, binding sites for the predicted miRNA found in A are obtained and summed in all genes. The genes are then sorted by total number of overlapping binding sites and returned to the user. **(E)** Example of competing mRNA predictions from ceRDB for PTEN. The user selects an mRNA of interest from the list of available mRNA. In the case shown here the PTEN tumor suppressor is chosen. **(F)** Starting with the list of miRNA binding elements present in PTEN the tool predicts potential competing RNA and visualizes the extent of overlap between the miRNA binding sites. Only a representative subset of the matrix is shown, the full matrix is available online. Each predicted gene is linked back to the TargetScan database to visualize the position and total numbers of each miRNA element.

Results:

In order to define the information content present within miRNA-mRNA predicted interactions we clustered a matrix containing miRNA families on the Y-axis with genes on the X-axis. Predicted binding interactions are labeled with a '1' and the lack of an interaction is labeled with a '0' at corresponding

points in the matrix. A heatmap of the clustered images as well as the branch structure indicate that miRNA binding sites coexist within 3'UTR at a much higher rate than would be expected at random. To visually show this we randomized the interaction matrix and clustered the results **(Figure 1A & B)**. Within the cell, each miRNA has many mRNA targets and each mRNA has potentially many miRNAs capable of regulation leading to a complex and dynamic regulatory system. One heretofore overlooked consequence of this system is that manipulation of the transcript level of a given mRNA may lead to changes in the concentration of available miRNAs leading to changes in alternate mRNA regulation via miRNA-mRNA interactions. In order to predict these interactions for a given target mRNA we determined all possible miRNA binding to this target mRNA and then found mRNAs (ceRNAs) that contained binding sites to these miRNAs. The potential for competition was ranked for each mRNA by counting the number of overlapping miRNA binding sites shared between the given mRNA and the potential ceRNA **(Figure 1C & D)**. Competing endogenous mRNA rankings were generated using the conserved mRNA-miRNA interactions. To access the data, we built a simple to use web interface and have made it available at http://www.oncomir.umn.edu/cefinder/. The user enters an mRNA they are interested in finding potential competing mRNAs that can regulate the gene of interest, and the tool returns potential ceRNA regulators. The list is sorted based on the overlap of the miRNA binding sites in the each of the pairwise relationships **(Figure 1E & F)**. Additionally the miRNA interactions present within the 3'UTR of the primary mRNA and all potential regulators are visualized in the final table**.**

Discussion:

In recent years miRNAs have taken center stage in many aspects of post transcriptional gene regulation. The complexity of miRNA-mediated gene regulation is compounded by the presence of multiple mechanisms that modulate either the levels of miRNAs and/or its target gene. Recently, the Pandolfi group proposed a novel concept in which mRNAs can regulate each other via common miRNA response elements **[4, 8]**. Through this cross talk novel mRNA-mRNA interactions have been identified in multiple cancer types. These findings suggest that modulation of miRNA activity by changing the levels of competing endogenous RNA is a key fundamental mechanism of gene regulation that will be applicable for many biological functions. Here we present a general and straightforward tool for identifying competing endogenous RNAs (ceRNAs) for a given gene of interest. Starting with the conserved set of miRNA-mRNA interactions, we observe that there is high degree of co-occurrence of miRNA binding sites within the miRNA-mRNA interaction dataset. This is consistent with the reports of Shalgi et al **[10].** We then use the co-occurrence of miRNA binding sites to predict and rank potential ceRNAs for all mRNAs. Our predictions are experimentally validated for PTEN and likely very relevant for a large number of additional genes **[5].** Several recent articles have described ceRNAs that are capable of regulating *PTEN* via competing reactions **[4, 5].** In these cases, loss of a competing mRNA releases miRNAs for

interaction with the tumor suppressor *PTEN* leading to decreased *PTEN* expression. Our database predicts many of the biologically validated interactions previously reported and uses a very straightforward algorithm in identifying these competing endogenous RNAs. Our search for ceRNAs for many established tumor suppressors in our database revealed some interesting observations. For example, genes such as *ONECUT2*, *NFIB* and *TNRC6B* appeared in many of the ceRNAs gene lists, these genes contains long 3'UTRs of up to 14kb in length and are predicted to contain many MREs that can potentially act as a sponge for multiple miRNAs. We are tempted to speculate that these ceRNAs with long 3'UTR can act as a 'master' MRE containing gene whose regulation may be affected in multiple disease conditions. Recently, TNRC6B was predicted to function as a ceRNA for *PTEN* and the downregulation of *TNRC6B* reduced the expression of *PTEN* **[5].**

In conclusion, we have developed the ceRDB resource to in the future accommodate multiple species such as model organisms and other types of sequences such as long non-coding RNAs and pseudogenes that can potentially also function as ceRNAs. We believe that the concept of competing endogenous RNA is likely to become a canonical central theme of gene regulation and having the ceRDB resource will significantly enhance our understanding of this fundamental gene regulatory mechanism.

Conflict of Interest:
We declare no conflict of interests

Author contributions:
AS and SS developed the idea. AS wrote the code and implemented. AS and SS wrote the manuscript.

References:
[1] Bartel DP, *Cell.* 2009 **136**: 215 [PMID: 19167326]
[2] Van Kouwenhove M *et al. Nat Rev Cancer.* 2011 **11**: 644 [PMCID: 21822212]
[3] Guo H *et al. Nature.* 2010 **466**: 835 [PMID: 20703300]
[4] Tay Y *et al. Cell.* 2011 **147**: 344 [PMID: 22000013]
[5] Karreth FA *et al. Cell.* 2011 **147:** 382 [PMID: 22000016]
[6] Cesana M *et al. Cell.* 2011 **147**: 358 [PMID: 22000014]
[7] Sumazin P *et al. Cell.* 2011 **147**: 370 [PMID: 22000015]
[8] Poliseno L *et al. Nature.* 2010 **465**: 1033 [PMID: 20577206]
[9] Salmena L *et al. Cell.* 2011 **146**: 353 [PMID: 21802130]
[10] Shalgi R *et al. PLoS Comput Biol.* 2007 **3**: e131 [PMID: 17630826]