|  |
| --- |
| *Application Note*  **Circr: Identify and Prioritize circRNA-miRNA Associations through Integration of Bioinformatic Predictions and Experimental Data**  Martina Dori1✝, Jimmy Caroli1,2✝, Silvio Bicciato1\*  1 Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy  2 Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark  \*To whom correspondence should be addressed. ✝These authors contributed equally  Received on XXXXX; revised on XXXXX; accepted on XXXXX  Associate Editor: XXXXXXX |

Abstract

**Summary:** Here we present Circr, a computational tool for the prediction of circRNA-miRNA associations. It combines publicly available algorithms for *de novo* prediction of miRNA binding sites on target sequences with experimentally validated miRNA-AGO and miRNA-RNA sites in different organisms. Circr can be used with either the provided support files or with custom ones, allowing the analysis of novel and not previously annotated circRNAs in virtually any species of interest. Circr provides as output an annotated table that allows the user an easy selection of interesting circRNA-miRNA sites for validation and functional investigations.  
**Availability and implementation:** Circ is available at  
Circs is written in python 3.6 language and it is released under the GNU GPL3.0 License.

**Contact**: [silvio.bicciato@unimore.it](mailto:silvio.bicciato@unimore.it)

**Supplementary information:** Supplementary information is available at *Bioinformatics* online.

1. Introduction

Circular RNAs (circRNAs) are covalently closed RNA molecules characterized by a non-linear 3’-5’ junction resulting from an unusual splicing event and referred to as backsplice junction [Ref. **Lasda and Parker 2014**]. This particular splicing event causes circRNAs to lack both 3’ poly(A) tail and 5’ capping, conferring resistance to exonuclease activity [**Ref**.], therefore determining a general longer half-life as compared to linear RNAs [**Refs**.]. Despite the great interest towards this class of non-coding RNA, only few have been functionally characterized. Among the possible roles that have been proposed for circRNA, the one that is mainly implicated in numerous diseases is miRNA binding. To date computational resources for predicting these interactions are mainly focused on either already annotated human or mouse circRNAs, while software that would allow *de novo* identification are usually restricted to a fixed nucleotide sequence length. To overcome these limitations, some prediction tools also provide stand-alone versions of their algorithms, allowing a fully customizable analysis with the drawback of an extremely high rate of false positives. Following the approach proposed in [**Dori, Bicciato**] we developed Circr, a tool that automatically combines 3 different algorithms for miRNA:target prediction and integrates this information with validated miRNA-RNA interactions and AGO peaks data (the driver of miRNA mediated target interaction). The combination of multiple tool identification of a single miRNA recognition site together with experimental data will allow users to efficiently reduce the large pool of candidate miRNA-circRNAs interaction for functional studies.

1. Implementation

The core of Circr consists of three main steps: i) input file preparation from circRNA genomic coordinates; ii) prediction of miRNA-circRNA interactions with third party softwares; and iii) comparison of the predicted sites with a database of validated interactions and AGO peaks. Starting from the genomic coordinates of circRNAs, Circr splits transcripts overlapping in the sense strands with genes into their intron/exon coordinates, retrieving only the latter ones. Conversely, antisense circRNAs or intergenic ones are considered as a single exon. This preliminary step can be superseded by the user when exon/intron coordinates are provided as direct input. Given these genomic coordinates, Circr then reconstructs the fasta sequence for each of the provided circRNA, building an input database for the miRNA binding site prediction algorithms implemented, namely miRanda [**Ref**], RNAhybrid [**Ref**] and TargetScan [**Ref**]. Next, Circr converts the predicted interactions in genomic coordinates combining the output of the prediction algorithms, noting for each interaction the number of tools that were able to identify that specific miRNA binding site. Moreover, Circr annotates the miRNA seed sequence category according to [**Ref. Bartel 2009**] (e.g. 8mer seed sequence). Finally, in the last step Circr compares the database of predicted interactions with a collection of publicly available and experimentally validated miRNA-RNA pairs **[Take all Refs!]** as well as AGO peaks coordinates, reporting in the final table whether there is an overlap between the coordinates of validated seed regions and the predicted ones. The resulting database is saved in a csv format that allows users to easily explore and filter out the seed sequences of interest. For the sake of usability, we provide a detailed set of support files comprising genome and miRNA sequences in fasta format; gene annotation file as gtf; rRNA coordinates, AGO peaks and validated interactions in bed format for 4 different organisms (human, mouse, fruitfly and worm) and 2 genome builds each. Nevertheless, it is also possible to use custom files to perform dedicated analyses, as long as they maintain the same format as the ones provided.

1. Installation and usage

Circr is written in python and has been developed in version 3.6. It requires several packages to be run, listed in the Supplementary Materials. Circr is freely available at **Github/bicciatolab**, where it can be directly downloaded together with the necessary supporting files. Moreover, we provide an Anaconda yaml file, granting the possibility to create a simple yet complete working environment to perform analysis with Circr via the Anaconda platform.

1. Results

To test the performance of Circr, we selected 100 circRNAs from the developing mouse lateral cortex [**Dori 2019**]. Among this cohort of sequences, we included circRNAs overlapping genes in the sense strand, antisense and intergenic ones to cover all possible genomic features of circRNAs. Since we can decide whether to split genic circRNAs into exon/intron or to provide an already final set of coordinates, we run Circr on the same set of sequences, testing both options. When providing the genomic coordinates of circRNAs full sequences, Circr completed the analysis in 104 minutes and retrieved more than 590,000 circRNA:miRNA binding sites. On the other hand, by providing the same input file as a set of final coordinates, Circr took only 19 minutes for the prediction of 221,699 sites. By selecting only the sites that are overlapping either validated seed regions or AGO peaks, this number drops to 38 from the full sequence analysis and 96 for the coordinates set up, providing the user with a more manageable number of putative interacting miRNA-circRNA pairs for further validation. Tests were carried out on a Ubuntu 18.04 server as well as Unix based pc (running either MacOS 10.15 or Ubuntu 20.04).

1. Conclusions

Here we present Circr, a flexible tool for the identification and prioritization of miRNA:circRNA binding sites. Circr combines the output from three different standalone prediction algorithms with publicly available experimentally validated targets and AGO-peaks. It provides all necessary supporting files to perform the analysis in 4 different organisms and with 2 genome builds each, also allowing the user to provide custom reference files, virtually granting the analysis on any organism and any sequence. [Final sentence!!!]

Acknowledgments

*Funding*: This work was supported by the Italian Ministry of Education, University and Research through EPIGEN – the Italian Flagship Project on Epigenomics.

*Conflict of interest*: none declared.

References