

Databases and ontologies

## ***OncomiR*: an online resource for exploring pan-cancer microRNA dysregulation**

**Nathan W. Wong<sup>1,2</sup>, Yuhao Chen<sup>2,3</sup>, Shuai Chen<sup>2</sup> and Xiaowei Wang<sup>1,2,\*</sup>**

<sup>1</sup>Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA,

<sup>2</sup>Department of Radiation Oncology, Washington University School of Medicine, St. Louis, MO 63108, USA and

<sup>3</sup>Department of Electrical and Systems Engineering, Washington University in St. Louis, St. Louis, MO 63130, USA

\*To whom correspondence should be addressed

Associate Editor: Alfonso Valencia

Received on April 24, 2017; revised on August 28, 2017; editorial decision on September 28, 2017; accepted on September 30, 2017

### **Abstract**

**Summary:** Dysregulation of microRNAs (miRNAs) is extensively associated with cancer development and progression. miRNAs have been shown to be biomarkers for predicting tumor formation and outcome. However, identification of the relationships between miRNA expression and tumor characteristics can be difficult and time-consuming without appropriate bioinformatics expertise. To address this issue, we present *OncomiR*, an online resource for exploring miRNA dysregulation in cancer. Using combined miRNA-seq, RNA-seq and clinical data from The Cancer Genome Atlas, we systematically performed statistical analyses to identify dysregulated miRNAs that are associated with tumor development and progression in most major cancer types. Additional analyses further identified potential miRNA-gene target interactions in tumors. These results are stored in a backend database and presented through a web server interface. Moreover, through a backend bioinformatics pipeline, *OncomiR* can also perform dynamic analysis with custom miRNA selections for in-depth characterization of miRNAs in cancer.

**Availability and implementation:** The *OncomiR* website is freely accessible at [www.oncomir.org](http://www.oncomir.org).

**Contact:** [xiaowei.wang@wustl.edu](mailto:xiaowei.wang@wustl.edu)

**Supplementary information:** [Supplementary data](#) are available at *Bioinformatics* online.

### **1 Introduction**

MicroRNAs (miRNAs) are short single-stranded RNA molecules of approximately 22 nucleotides that function in post-transcriptional regulation of gene expression. By targeting RNA transcripts for degradation or inhibition of translation, miRNAs are actively involved in controlling downstream proteomic profiles. This phenomenon is observed in numerous biological processes, such as embryonic formation, carcinogenesis and tumor progression (Ambros, 2004). Furthermore, miRNAs can serve as biomarkers for various diseases, with particular clinical interest in predicting likelihood of cancer development and progression.

miRNA biomarkers have been discovered in nearly all cancer types. For example, in breast cancer, miR-21-5p, miR-155-5p and miR-29a-3p, among others, are observed as upregulated compared to

normal tissue, while miR-30-3p and miR-221-3p are examples of downregulated miRNAs. However, it has also been noted that miR-29a-3p is downregulated in other cancer types, such as neuroblastoma and sarcoma, which suggests that the mechanisms driving tumor formation and progression are not uniform across all cancer types (Cho, 2010). The majority of studies that identify miRNA biomarkers focus on a single or a subset of miRNAs; however, with the growth of high-throughput sequencing technologies, it is possible to analyse the complete miRNomes from many patients across multiple cancer types.

Analysis of high-throughput sequencing data is often difficult for researchers with little computational expertise. Multiple databases have been established previously to characterize miRNA functions in cancer. For example, HMDD, miR2Disease and OncomiRDB present experimentally validated relationships between miRNA expression

and cancer development based on literature reports (Jiang *et al.*, 2009; Li *et al.*, 2014; Wang *et al.*, 2014). Other resources, including cBioPortal and FireBrowse among others, provide results of statistical analyses on high-throughput sequencing studies of cancer genomics in general, but not specifically focused on miRNA analysis (Cerami *et al.*, 2012; Broad Institute TCGA Genome Data Analysis Center, 2016; Lee *et al.*, 2015; Yang *et al.*, 2017).

Incorporating all the facets of miRNA biology into a comprehensive user-friendly toolset is a daunting task, as it requires identifying potential miRNA biomarkers and their targets and establishing their functional relationships in the context of cancer physiology. To address this issue, we present *OncomiR*, which is an online pan-cancer resource for analysis of miRNA dysregulation. *OncomiR* contains three major features: (i) a database of statistically dysregulated miRNAs associated with clinical characteristics of cancer; (ii) miRNA-target expression correlation and prediction across cancer types; and (iii) tools for dynamic analysis of miRNA-derived survival signatures and clustering of cancer types. The diverse functionality of *OncomiR* would make it a valuable resource to the miRNA and cancer research community.

## 2 Results

*OncomiR* consists of a primary backend database and a dynamic web server. Aligned and normalized miRNA-seq and RNA-seq data were obtained from The Cancer Genome Atlas (TCGA), covering approximately 1200 mature miRNA transcripts and 30 000 mRNA transcripts (The Cancer Genome Atlas Research Network, *et al.* 2013). Patient clinical data were also obtained; patients were excluded from subsequent analysis if follow-up data was poor, resulting in a data set encompassing about 10 000 patients spanning 30 cancer types (Supplementary Table S1). The results from the statistical analyses are stored in a MySQL database accessible through Perl CGI and Perl DBI. The *OncomiR* web server implements Perl CGI in conjunction with the R statistical program ([www.r-project.org](http://www.r-project.org)) for *ad hoc* backend analysis.

### 2.1 miRNA dysregulation in cancer development and progression

The use of miRNAs as molecular biomarkers in the clinical setting requires independent validation after initial discovery. With this in mind, we have conducted comprehensive statistical analysis on the relationships between miRNA expression and three primary clinical features: tumor development; tumor staging and grade; and patient survival outcome. Significance in tumor development was determined through paired Student's *t*-test between normal and tumor tissues; in staging and grade, ANOVA; and for survival, unpaired Student's *t*-test and univariate Cox proportional hazards analysis between living and deceased patients. The results of these analyses, including correction for multiple testing, are stored in a backend MySQL database and accessible through a web query interface.

Users can search by miRNA name or by cancer type to retrieve a list of significant miRNA-cancer relationships (Kozomara and Griffiths-Jones, 2014). Additionally, this query interface can be used to retrieve the RNA-seq data of one or more miRNAs across multiple cancer types for a detailed view of cancer-dependent miRNA expression profiles.

### 2.2 miRNA-target predictions in tumor samples

miRNAs function as post-transcriptional regulators of gene expression; dysregulation of miRNAs implies that downstream translation of target mRNA transcripts would be affected. Correlating predicted

targets with dysregulated miRNAs can provide insight into potential mechanisms for miRNA-driven clinical phenotypes. *OncomiR* can be queried for miRNA-target correlations, as calculated within the dataset from TCGA as well as predicted by miRDB (Wang, 2016; Wong and Wang, 2015). Pearson's correlation analysis was conducted using all paired tumor/non-tumor tissue samples, as well as all tumor miRNA samples within each cancer type. Each miRNA-target pair is presented with the correlation coefficient, the raw *P*- and FDR-values, and the target prediction score from miRDB. A combination of a significant negative correlation coefficient with a high prediction score suggests a likely miRNA-target effect in tumorigenesis for the selected cancer type.

### 2.3 Server interface for custom miRNA analysis

One of the most significant clinical applications of cancer biomarker research is the stratification of patients based on treatment outcome. To make this analysis more accessible to the clinical research community at large, *OncomiR* can analyse miRNA-derived survival outcome signatures dynamically for one or more cancer types. The performance of a submitted signature on predicting overall cancer survival is presented through a Kaplan–Meier survival curve. This feature would be useful for evaluation of new biomarker signatures discovered with TCGA data or validation of existing signatures derived from other independent studies.

Clustering of major cancer types based on sequencing results has been performed previously (Hoadley *et al.*, 2014), but dynamic clustering based on miRNA-seq data, however, has not been adequately evaluated. Through *OncomiR*, users can perform hierarchical or *k*-means clustering to identify how major cancer types can be grouped together by miRNA expression for a deeper understanding of the involvement of miRNAs in cancer progression within subgroups or across major cancer types.

## 3 Conclusions

*OncomiR* is a user-friendly resource for exploring miRNA dysregulation in cancer. We conducted statistical analysis on miRNomes from TCGA to provide a readily accessible repository of miRNA associations with cancer characteristics. Additionally, correlation and target analysis were conducted to provide insight into possible miRNA-driven mechanisms leading to cancer development and progression. Moreover, *OncomiR* also provides a set of dynamic tools for researchers to conduct custom miRNA analyses. In summary, *OncomiR* is a comprehensive tool to allow flexible miRNomic analysis across many cancer types.

## Acknowledgements

We would like to thank P. Daft, W. Liu, J. Lewis, A. Wong, H. Vande Krol, T. Sharma and D. Murphy for beta testing and feedback, and N. Rose for technical assistance.

## Funding

This work was supported by the National Institutes of Health [Grants R01GM089784 and R01DE026471].

*Conflict of Interest:* none declared.

## References

Ambros, V. (2004) The functions of animal microRNAs. *Nature*, **431**, 350–355.

- Broad Institute TCGA Genome Data Analysis Center (2016) *Analysis Overview for 28 January 2016*. Broad Institute of MIT and Harvard. <http://www.firebrowse.org> (6 October 2017, date last accessed).
- Cerami,E. *et al.* (2012) The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, **2**, 401–404.
- Cho,W.C. (2010) MicroRNAs: potential biomarkers for cancer diagnosis, prognosis and targets for therapy. *Int. J. Biochem. Cell Biol.*, **42**, 1273–1281.
- Hoadley,K.A. *et al.* (2014) Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*, **158**, 929–944.
- Jiang,Q. *et al.* (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–104.
- Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
- Lee,H. *et al.* (2015) The Cancer Genome Atlas Clinical Explorer: a web and mobile interface for identifying clinical-genomic driver associations. *Genome Med.*, **7**, 112.
- Li,Y. *et al.* (2014) HMDD v2.0: a database for experimentally supported human microRNA and disease associations. *Nucleic Acids Res.*, **42**, D1070–D1074.
- The Cancer Genome Atlas Research Network, *et al.* (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.*, **45**, 1113–1120.
- Wang,D. *et al.* (2014) OncomiRDB: a database for the experimentally verified oncogenic and tumor-suppressive microRNAs. *Bioinformatics*, **30**, 2237–2238.
- Wang,X. (2016) Improving microRNA target prediction by modeling with unambiguously identified microRNA-target pairs from CLIP-ligation studies. *Bioinformatics*, **32**, 1316–1322.
- Wong,N. and Wang,X. (2015) miRDB: an online resource for microRNA target prediction and functional annotations. *Nucleic Acids Res.*, **43**, D146–D152.
- Yang,Z. *et al.* (2017) dbDEMOC 2.0: updated database of differentially expressed miRNAs in human cancers. *Nucleic Acids Res.*, **45**, D812–D818.