BIOINFORMATICS APPLICATIONS NOTE Vol. 30 no. 11 2014, pages 1630–1632
doi:10.1093/bioinformatics/btu087

Gene expression

Advance Access publication February 11, 2014

SurvMicro: assessment of miRNA-based prognostic signatures for cancer clinical outcomes by multivariate survival analysis

Raul Aguirre-Gamboa and Victor Trevino*

Departamento de Investigación e Innovación, Escuela de Medicina, Tecnológico de Monterrey, 64849 México

Associate Editor: Ivo Hofacker

ABSTRACT

Summary: MicroRNAs (miRNAs) play a key role in post-transcriptional regulation of mRNA levels. Their function in cancer has been studied by high-throughput methods generating valuable sources of public information. Thus, miRNA signatures predicting cancer clinical outcomes are emerging. An important step to propose miRNA-based biomarkers before clinical validation is their evaluation in independent cohorts. Although it can be carried out using public data, such task is time-consuming and requires a specialized analysis. Therefore, to aid and simplify the evaluation of prognostic miRNA signatures in cancer, we developed SurvMicro, a free and easy-to-use web tool that assesses miRNA signatures from publicly available miRNA profiles using multivariate survival analysis. SurvMicro is composed of a wide and updated database of >40 cohorts in different tissues and a web tool where survival analysis can be done in minutes. We presented evaluations to portray the straightforward functionality of SurvMicro in liver and lung cancer. To our knowledge, SurvMicro is the only bioinformatic tool that aids the evaluation of multivariate prognostic miRNA signatures in cancer.

Availability and implementation: SurvMicro and its tutorial are freely available at http://bioinformatica.mty.itesm.mx/SurvMicro.

Contact: vtrevino@itesm.mx or victor.trevino@gmail.com

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

Received on October 10, 2013; revised on January 6, 2014; accepted on February 6, 2014

1 INTRODUCTION

MicroRNAs (miRNAs) are ~22-nt molecules that play an important post-transcriptional regulatory role (Reinhart *et al.*, 2000). This molecular mechanism has been widely studied in cancer. It is known that some miRNAs act like oncogenes or tumor suppressor genes (Esquela-Kerscher and Slack, 2006) and whose regulatory roles are linked to cancer hallmarks (Ruan *et al.*, 2009). MiRNAs have also been proposed as drug targets and biomarkers (Calin and Croce, 2006; Healy *et al.*, 2012).

The roles of miRNAs in the tumor cell have been explored using high-throughput techniques such as microarrays and sequencing. These data have also been used to associate miRNAs to clinical outcome in patients generating signatures for diagnostics and prognostics that are composed of several miRNAs (Calin and Croce, 2006). A miRNA signature trial has been recently announced (MRX34 trial). A key issue for miRNA signatures for cancer is their validation in diverse cohorts (Kern, 2012). Such task generates difficulties even though many public cohorts are available. These are related to time-consuming tasks, such as exploration of repositories, acquisition and processing and the modeling of miRNA levels, to assess their value as prognostic biomarkers.

The current approaches for miRNAs signature validation are limited. ProgMir (Goswami and Nakshatri, 2012) and MIRUMIR (Antonov *et al.*, 2013) implement only univariate analyses even though signatures are composed of several miRNAs (Raponi *et al.*, 2009; Wei *et al.*, 2013). Moreover, both tools are restricted regarding the available cohorts. These characteristics limit the assessment of multivariate miRNA biomarkers.

To provide assessment of multivariable prognostic miRNA biomarkers and to simplify the evaluation in several cohorts, we developed SurvMicro. SurvMicro is a curated and updated web tool of miRNA expression levels associated with clinical outcome that provides survival analysis and risk assessment in cancer.

2 METHODS

2.1 Datasets

Data were collected by searching for keywords related to cancer, clinical outcome, miRNA platforms and >30 samples. The searches were performed in GEO (http://www.ncbi.nlm.nih.gov/geo), GEOmetadb (Zhu *et al.*, 2008), ArrayExpress (https://www.ebi.ac.uk/arrayexpress/) and TCGA (https://tcga-data.nci.nih.gov/tcga). TCGA data were obtained at level 3 from the May 2013 run. TCGA read counts generated by RNA-seq were log2 transformed. All datasets were quantile-normalized and log2 transformed if needed. Datasets were grouped using disease ontologies (Schriml *et al.*, 2012).

2.2 Web interface implementation

Two intuitive input pages and a results section were implemented using Java server pages, JavaScript, R, Ajax, Apache and MySQL. The first input page requires the list of miRNAs, the selection of the cohort and the treatment for multiprobe miRNAs. We annotate miRNAs using the miRBase nomenclature (Griffiths-Jones *et al.*, 2006).Owing to the presence of miRNA variants, we implemented two types of searches, a 'strict' for specific variant evaluation and a 'relaxed' one that reports every variant of the miRNA entered by the user (−3p, −5p and *). After web submission, the following analysis page obtains the rows corresponding to the entered miRNAs in the selected dataset as well as the available clinical characteristics. In this page, the user can customize the output of the analysis including risk group methodology, risk groups stratification, model fitting, heatmap and boxplot graphs, figure format, PDF report

*To whom correspondence should be addressed.

and others. Details of the implementation and options are thoughtfully documented in the tutorial. Finally, the results section displays the biomarker evaluation in publication-ready graphs and tables containing statistical parameters for easy exportation and manipulation.

2.3 MiRNA-based risk assessment

To perform the risk prediction from the list of miRNAs submitted by the user, SurvMicro uses the Cox multivariate fitting from R (Therneau, 2013). The risk groups are obtained by ranking the prognostic index, which is obtained by the sum of the products of the Cox coefficients and the expression values of corresponding miRNAs. The groups are generated splitting the prognostic index by the median, by the quantiles or by using a simple Log-Rank *P*-value optimization algorithm. The risk groups are used to generate the Kaplan–Meier plots and statistics (Log-Rank, concordance index and hazard ratios). Moreover, to analyze biomarkers computed with models other than Cox, the user can specify coefficients for each miRNA. Details are provided in the tutorial.

3 RESULTS

3.1 Database and web interface

To our knowledge, SurvMicro provides the biggest compilation of miRNA datasets related to clinical outcomes, comprising >40 datasets and 6300 samples grouped in 14 tissues. We also encourage any user to suggest datasets. As shown in Figure 1, SurvMicro is easily operated by providing only the list of miRNAs and selecting the target cohort in the input page. A link to a comprehensive tutorial and an example button for a quick example are located in the input page. The analysis page is generated in a few seconds. Finally, the results section is obtained in less than a minute (might vary if advance plots are selected) deploying the outcome plots and tables.

3.2 Validation and applications

To assess the capabilities of SurvMicro in the evaluation of multivariate prognostic biomarkers, we performed an analysis of a miRNA signature for liver hepatocellular carcinoma survival, an examination of a signature for lung squamous cell carcinoma (LSCC) and a performance comparison of two LSCC survival signatures. In liver, we evaluated a 20-miRNA signature for survival carcinoma (Wei *et al*., 2013) in two additional independent cohorts. The summary of the survival analyses is shown in Table 1 (Kaplan–Meier curves are shown in Section 5.1 of the Tutorial). The performance was highly significant in two datasets and acceptable in one cohort. Moreover, using the stratification option of SurvMicro, we examined the prognostic power of this signature in early lesions compared with advanced tumors (shown in the Tutorial). This explorative evidence suggests that the signature could be related to advanced tumors (Grade 3 or Stage III). For lung, a 20-miRNA signature proposed for survival and relapse in LSCC (Raponi *et al*., 2009) was evaluated in three LSCC datasets including the original dataset and a lung adenocarcinoma dataset. The results are shown in Table 1 (detailed in Tutorial Section 5.2). The Kaplan–Meier curves and statistical information in all datasets appear acceptable even though only two miRNAs were significant or marginally significant using a Cox model in the Raponi dataset. By using the fitting information option of SurvMicro in the Analysis Page, 10 of the 20 miRNAs are significant if fitted in the high-risk group suggesting that these miRNAs are important. For a comparison analysis of two signatures, we tested the 20 miRNAs of the Raponi signature and a 5-miRNA signature associated to survival in LSCC proposed by Yu *et al*.(2008). The summary shown in Table 1 (the details are included in the Tutorial Section 5.3) suggests that the Raponi signature is superior.

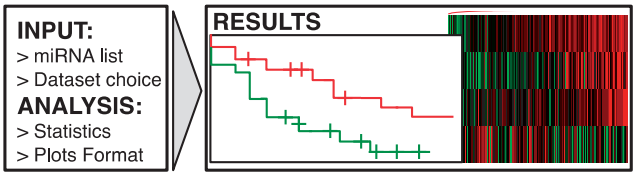

Fig. 1. An overview of SurvMicro workflow and implementation

Table 1. Assessment of three miRNA signatures

Database	Log-Rank	CI	HR	miRNAs	Samples
Wang liver biomarker					
Wei Wang	3.0e-7	70.02	3.52*	20	58
Li Gu	1.2e-2	58.91	1.59*	10	156
TCGA LIHC	3.1e-4	78.68	5.86*	14	166
Raponi lung biomarker					
Raponi	6.0e-6	77.41	5.18	20	71
TCGA LUAD	5.0e-5	66.86	3.58	14	112
TCGA LUSC	2.0e-3	63.51	2.43	14	147
Yu	1.0e-3	65.83	3.07	10	224
Yu lung biomarker					
Raponi	6.3e-1	60.62	1.19	5	71
TCGA LUAD	5.0e-2	66.08	1.54	4	112
TCGA LUSC	8.2e-2	56.06	1.61	4	147
Yu	1.0e-1	58.52	1.87	5	224

Columns means Log-Rank *P*-value, CI for concordance index, HR for hazard ratio, miRNAs for the number of miRNAs found in the dataset.
*$P<0.05$. LIHC, LUAD, and LUSC means Liver hepatocellular carcinoma, Lung adenocarcinoma, and Lung squamous cell carcinoma respectively.

4 CONCLUSION

The evaluations demonstrate that a curated database and systematic tool for biomarker evaluation can rapidly produce valuable information for the validation of miRNA signatures in various cohorts. We conclude that SurvMicro is a free and easy-to-use web tool to assess the performance of prognostic miRNA-based biomarkers.

Funding: The authors thank CONACyT grant 140601 and Cátedra Bioinformática CAT220 from ITESM for funding and scholarships.

Conflict of Interest: none declared.

REFERENCES

Antonov,A.V. *et al.* (2013) MIRUMIR: an online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets. *Cell Death Differ.*, **20**, 367.

Calin,G.A. and Croce,C.M. (2006) MicroRNA signatures in human cancers. *Nat. Rev. Cancer*, **6**, 857–866.
Esquela-Kerscher,A. and Slack,F.J. (2006) Oncomirs—microRNAs with a role in cancer. *Nat. Rev. Cancer*, **6**, 259–269.
Goswami,C.P. and Nakshatri,H. (2012) PROGmiR: a tool for identifying prognostic miRNA biomarkers in multiple cancers using publicly available data. *J. Clin. Bioinforma.*, **2**, 23.
Griffiths-Jones,S. *et al.* (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, **34**, D140–D144.
Healy,N.A. *et al.* (2012) Systemic mirnas as potential biomarkers for malignancy. *Int. J. Cancer*, **131**, 2215–2222.
Kern,S.E. (2012) Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures. *Cancer Res.*, **72**, 6097–6101.
Raponi,M. *et al.* (2009) MicroRNA classifiers for predicting prognosis of squamous cell lung cancer. *Cancer Res.*, **69**, 5776–5783.
Reinhart,B.J. *et al.* (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, **403**, 901–906.
Ruan,K. *et al.* (2009) MicroRNAs: novel regulators in the hallmarks of human cancer. *Cancer Lett.*, **285**, 116–126.
Schriml,L.M. *et al.* (2012) Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res.*, **40**, D940–D946.
Therneau,T.M. and Grambsch,P.M. (2000) *Modeling Survival Data: Extending the Cox Model*. Springer, New York.
Wei,R.R. *et al.* (2013) Clinical significance and prognostic value of microRNA expression signatures in hepatocellular carcinoma. *Clin. Cancer Res.*, **19**, 4780–4791.
Yu,S.L. *et al.* (2008) MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell*, **13**, 48–57.
Zhu,Y. *et al.* (2008) GEOmetadb: powerful alternative search engine for the Gene Expression Omnibus. *Bioinformatics*, **24**, 2798–800.