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

Raul Aguirre-Gamboa1, and Victor Trevino1,*
1Cátedra de Bioinformática, Departamento de Investigación e Innovación, Escuela de Medicina, Tecnológico de Monterrey, México 64849.

ABSTRACT
Summary: MicroRNAs (miRNAs) play a key role in post-transcriptional regulation of RNAm 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 freely 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 more than 40 cohorts in different tissues and a web tool where survival analysis can be done in minutes. We presented evaluations to portrait 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.itesm.mx/SurvMicro.

1 INTRODUCTION
MicroRNAs (miRNAs) are ~22 nucleotide 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 has also been proposed as drug targets and biomarkers (Calin and Croce, 2006; Healy et al., 2012).

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

The roles of miRNAs in the tumor cell have been explored using high throughput techniques such as microarrays and sequencing. This data has also been used to associate miRNAs to clinical outcomes in patients generating signatures for diagnostics and prognostics that are composed of several miRNAs (Calin and Croce, 2006). Indeed, 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 signatures validation are limited. ProgMir (Goswami and Nakshatri, 2012) and MIRUMIR (Antonov et al., 2013) implements 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.
In order 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
Datasets. Data was collected by searching for keywords related to cancer, clinical outcome, miRNA platforms, and over 30 samples. The searches were performed in GEO, GEOMetadb (Zhu et al., 2008), ArrayExpress, and 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).

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 multi-probe miRNAs. We annotate miRNAs using the miRBase nomenclature (Griffiths-Jones et al., 2006). Due 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, 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.

To whom correspondence should be addressed.
miRNA-based risk assessment. In order 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 (PI), 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 PI by the median, quantiles or by using a simple Log-Rank (LR) p-value optimization algorithm. The risk groups are used to generate the Kaplan-Meier (KM) plots and statistics (LR, 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

Database and web interface. To our knowledge, SurvMicro provides the biggest compilation of miRNA datasets related to clinical outcomes, comprising more than 40 datasets and 6,300 samples grouped in 14 tissues. We also encourage any user to suggest datasets. As shown in Fig. 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.

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 (KM curves are shown in section 5.1 of the Tutorial). The performance was highly significant in both 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 to 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 miRNAs 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 KM 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 (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.

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.

![Fig. 1. An overview of three SurvMicro workflow and implementation.](http://bioinformatics.oxfordjournals.org/)

**Table 1. Assessment of three miRNA signatures.**

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

Columns mean 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.

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

**REFERENCES**


