

## Correspondence

# MIRUMIR: an online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets

*Cell Death and Differentiation* (2013) 20, 367; doi:10.1038/cdd.2012.137; published online 23 November 2012

Dear Editor,

Although multiple clinical studies have correlated microRNA (miR) expression profiles with clinical outcome in cancer patients,<sup>1</sup> most data sets in the public domain have a complex interface, which is difficult to use by a non-specialist public and requires advanced informatics and statistical skills. Here, we present MIRUMIR, a tool which performs survival analyses and draws Kaplan–Meier (KM) plots for submitted ‘microRNA’ across several available data sets, which cover more than 800 patients. A robust statistical procedure is implemented to account for multiple testing. MIRUMIR is incorporated into BioProfiling.de, analytical portal for high-throughput cell biology,<sup>2</sup> and is freely available at <http://www.bioprofiling.de/MIRUMIR>.

Similar tools were developed recently to assess the relationship between expression levels of various genes with clinical outcome.<sup>3</sup> KM plots provides meta-analyses by combining multiple gene expression data sets. MIRUMIR is the first tool specifically designed for miRs. Unlike KM plots, MIRUMIR screens publicly available miR expression data sets one by one implementing robust statistical procedures to adjust *P*-value for multiple testing. Data sets have been downloaded from the GEO omnibus repository.<sup>4</sup> To be selected the data set must be a clinical data set (patients) and have at least 50 profiled samples annotated with survival information. Currently, MIRUMIR comprises nine data sets covering breast, prostate and ovarian cancers that we plan to expand shortly.

MIRUMIR exploits rank information from expression data sets. For each available data set samples are grouped with respect to expression rank of the user specified miR. The ‘low expression’ and ‘high expression’ groups are those where expression rank of miR is less or more than average expression rank across the data set, respectively. This separation of patients into ‘low’ and ‘high’ groups along with survival information is next used to find any statistical differences in survival outcome. The R statistical package is used to perform survival analyses<sup>5</sup> and to draw KM plots, see Supplementary Material for MIRUMIR practical analyses of several miR examples. There are sources of multiple testing that should be accounted for. Obviously, the miR hypothesis is tested across each available data set. Therefore, the final number of tested hypotheses is equal to the number of screened expression data sets. In addition, for some data sets, the miR can be mapped to several probes, which means several hypotheses tested for one data set. False discovery

rate<sup>6</sup> control procedure is implemented to automatically adjust *P*-values for multiple testing.

In summary, MIRUMIR supports the need of biomedical researchers to estimate the power of miR to serve as potential biomarker to predict survival of cancer patients. MIRUMIR provides such analyses based on several publicly available clinical miR data sets annotated with patient survival information. To our knowledge this is a first simple tool of this kind and, as shown by several examples (see Supplementary Materials), MIRUMIR is ideal for fast validation of the clinical relevance of a selected miR.

## Conflict of Interest

The authors declare no conflict of interest.

**Acknowledgements.** This work was supported by the UK Medical Research Council (MRC) and funding from RFBS (10-04-01234), MCB RAS, Russian Federal grants 16.740.11.036 (to NB) and 11.G34.31.0069 (to GM).

AV Antonov<sup>\*1</sup>, RA Knight<sup>1</sup>, G Melino<sup>1,2,3</sup>, NA Barlev<sup>3,4,5</sup> and PO Tsvetkov<sup>6</sup>

<sup>1</sup> Medical Research Council Toxicology Unit, Leicester University, Leicester LE1 9HN, UK;

<sup>2</sup> Department of Experimental Medicine and Surgery, University of Rome ‘Tor Vergata’, Rome 00133, Italy;

<sup>3</sup> Saint-Petersburg State Institute of Technology, Moscovskii avenue 26, Saint-Petersburg 190013, Russia;

<sup>4</sup> Department of Biochemistry, University of Leicester, Leicester LE1 9HN, UK;

<sup>5</sup> Institute of Cytology, 194064 Saint-Petersburg, Russia and

<sup>6</sup> Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Vavilov Street 32, Moscow 119991, Russia

\* Corresponding author: AV Antonov, Medical Research Council Toxicology Unit, Hodgkin Building, Lancaster Road, PO Box 138, Leicester LE1 9HN, UK. Tel: +44 (0) 116 252 5562; Fax: +44 (0) 116 252 5616; E-mail: [aa668@le.ac.uk](mailto:aa668@le.ac.uk)

1. Lovat F, Valeri N, Croce CM. *Semin Oncol* 2011; **38**: 724–733.

2. Antonov AV. *Nucleic Acids Res* 2011; **39**: W323–W327.

3. Gyorfy B *et al.* *Breast Cancer Res Treat* 2010; **123**: 725–731.

4. Barrett T *et al.* *Nucleic Acids Res* 2009; **37**: D885–D890.

5. Harrington, Fleming. *Biometrika* 1982; **69**: 553–566.

6. Hochberg Y, Benjamini Y. *Stat Med* 1990; **9**: 811–818.



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>