

Oral presentation

Open Access

rHVDM – a fast and user-friendly R package to predict transcription factor targets from microarray time series data

Martino Barenco^{*1,6}, Sonia Shah², Daniel Brewer^{3,6}, Robin Callard^{1,6}, Jaroslav Stark^{4,6}, Crispin Miller⁵ and Michael Hubank^{1,6}

Address: ¹Institute of Child Health, University College London, WC1N 1EH, UK, ²Bloomsbury Centre for Bioinformatics, University College London, WC1E 6BT, UK, ³Institute of Cancer Research, Sutton, SM2 5NG, UK, ⁴Department of Mathematics, Imperial College, London, SW7 2BZ, UK, ⁵Patterson Institute for Cancer Research, University of Manchester, M20 4BX, UK and ⁶CoMPLEX, University College London, NW1 2HE, UK

Email: Martino Barenco* - m.barenco@ucl.ac.uk

* Corresponding author

from BioSysBio 2007: Systems Biology, Bioinformatics and Synthetic Biology
Manchester, UK. 11–13 January 2007

Published: 8 May 2007

BMC Systems Biology 2007, 1(Suppl 1):S1 doi:10.1186/1752-0509-1-S1-S1

This abstract is available from: <http://www.biomedcentral.com/1752-0509/1?issue=S1>

© 2007 Barenco et al; licensee BioMed Central Ltd.

Background

Researchers dealing with gene microarray data are faced with daunting quantities of data in which lie hidden information that is important, including transcription factor activity profiles. We developed a model-based technique, HVDM (Hidden Variable Dynamic Modelling), which uses data from a small training set of known transcription factor (TF) targets plus a single anchoring degradation measurement to deduce the activity profile of the transcription factor – the hidden variable in the system. Using this activity profile, other targets of the same TF can then be identified by running the same model. Both stages rely on time course expression data obtained from microarrays. The sampling rate can be irregular, replicates are not required and measurement errors are explicitly taken into account so that results are ranked according to confidence, a must when dealing with noisy data. We tested HVDM on the DNA damage response network, focusing on p53, an important transcription factor. An independent experiment confirmed the accuracy of our predictions [1].

Results

We have now generated an R/Bioconductor version of HVDM, called rHVDM. The original implementation, written in C, used time consuming algorithms both for the optimization step (Nelder-Mead) and the confidence intervals determination (Markov Chain Monte Carlo). In

contrast, rHVDM uses a fast, gradient-based, optimisation step (Levenberg-Marquardt) from which accurate confidence intervals can also be obtained. As a result, a thousand genes can be screened in about five minutes on a standard current personal computer. Additionally, rHVDM includes an HTML report generator which allows visual quality assessment at each stage in the process.

Conclusion

rHVDM is applicable to large time course expression data sets, where identification and further exploitation of hidden variables can reveal critical information about network dynamics.

Availability

rHVDM can be downloaded from the bioconductor website <http://bioconductor.org/packages/2.0/bioc/html/rHVDM.html>.

References

1. Barenco M, Tomescu D, Brewer D, Callard R, Stark J, Hubank M: **Ranked predictions of p53 targets using Hidden Variable Dynamic Modelling (HVDM)**. *Genome Biology* 2006, **7(3):R25**.