

Transcriptogram as a diagnosis tool for gene expression data

Ricardo M. Ferreira², Jose Luiz Rybarczyk-Filho², Mauro A. A. Castro^{1,2}, Rodrigo J. S. Dalmolin³, Jose C. F. Moreira³, Leonardo G. Brunnet² and Rita M. C. de Almeida^{1,2}

Short Abstract-- Expression data gives access to the metabolic state of the cell. In principle, this knowledge can be used to determine cell state, as well its response to environment and drugs, therefore, expression data can be used to diagnosis. This information, however, has proved to be hard to obtain. In this work we apply the transcriptogram, an experiment-independent method, to access the metabolic state of the cell and use it as diagnosis tool.

Keywords — Genome-wide expression data, Transcriptogram, Diagnosis

I. INTRODUCTION

Genome-wide expression data consists of expression levels of thousands of genes, thus giving access to the metabolic state of the cell. This information can be used, for example, to test drug response, diagnosis, early detection, and predictions for clinical treatment. The usual approach, however, is not always capable of predict hard endpoints and is highly dependent of data processing techniques. In this work we aim to consolidate the transcriptogram as a diagnosis tool. The transcriptogram method consists of two major steps: first, ordering genes according with their interactions, as obtained from databases as STRING, that groups functional modules associated with gene ontology terms; second, averages of expression data over neighboring genes. The transcriptogram has advantages when compared to other methods to predict clinical endpoints: the gene ordering depends on protein-protein associations, thus being experiment independent; data processing and feature extraction is not an issue; as genes have their expression dynamics determined by a network of interactions, moderate alterations can together produce a important alteration when measured by averaging over neighboring genes.

II. OBJECTIVES

In this work we apply the transcriptogram to classify clinic endpoints as proposed in the MAQC-II project. We aim to compare our results to those in the project, and to

build a solid method for applying the transcriptogram as a simple diagnosis tool.

REFERENCES

Rybarczyk-Filho, J.L. *et al.*, Towards a genome-wide transcriptogram: the *Saccharomyces cerevisiae* case. *Nucleic Acids Research*, **39**, 3005-3016 (2011)

Shi, L. *et al.*, The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nature Biotechnology*, **28**, 827-838 (2010)

¹National Institute of Science and Technology for Complex Systems

²Instituto de Física, Universidade Federal do Rio Grande do Sul

³Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul