

Probabilistic Genetic Networks analysis of three *Plasmodium falciparum* strains from dynamical expression signals

Junior Barrera¹, Roberto M. Cesar-Jr¹, David C. Martins-Jr^{1*},
Ricardo Z. N. Vêncio¹, Carmen Fernandez-Becerra², Carlos A. B. Pereira¹,
Hernando A. del Portillo²

¹Institute of Mathematics and Statistics - University of São Paulo - Brazil,

²Institute of Biomedical Sciences - University of São Paulo - Brazil

The advent of genomics into malarial research is significantly accelerating the discovery of control strategies. Dynamical global gene expression measures of the intraerythrocytic developmental cycle (IDC) of the parasite at 1h-scale resolution were recently reported [1]. Moreover, by using Discrete Fourier Transform based techniques, it was demonstrated that many genes are regulated in a single periodic manner which allowed to order genes according to the phase of expression.

In this work we present a framework to construct genetic networks from dynamical expression signals [2]. The adopted model to represent these networks is the Probabilistic Genetic Network (PGN). This network is a Markov chain with some additional properties. This model mimics the properties of a gene as a non-linear stochastic gate and the systems are built by coupling of these gates. The PGN estimation is made through the mean conditional entropy minimization to discover subsets of genes which perform the best predictions of the target gene in the posterior time instant. Moreover, a tool that integrates mining of dynamical expression signals by PGN design techniques, different databases and biological knowledge, has been developed. The applicability of this tool for discovering gene networks of the malaria expression regulation system has been validated for simulated data and also for real *microarray* data using the glycolytic pathway as a “gold-standard”, as well as by creating an apicoplast as PGN network [2]. Also, a negative control between these two modules was confirmed through construction of PGN networks using four genes from glycolysis and four from apicoplast organelle as seed genes [2]. Together, this data demonstrates the value of the PGN model in generating biologically meaningful networks and which include genes not included by the Fourier approach [1]. Currently, we are applying the same technique for three malarial strains (3D7, Dd2, HB3) in order to analyze similarities and differences among them and to discover whether or not these three data sets may be joint, which would improve the PGN estimation.

- [1] M. Llinas, Z. Bozdech, E. D. Wong, A. T. Adai and J. L. DeRisi. *Comparative whole genome transcriptome analysis of three Plasmodium falciparum strains*. Nucleic Acids Research, 34(4):1166-1173 (2006).
- [2] J. Barrera, R. M. Cesar-Jr, D. C. Martins-Jr, R. Z. N. Vêncio, E. F. Merino, M. M. Yamamoto, F. G. Leonardi, C. A. B. Pereira and H. A. del Portillo. *Constructing probabilistic genetic networks of Plasmodium falciparum from dynamical expression signals of the intraerythrocytic development cycle*. In McConnel, Lin and Hurban, editors, Methods of Microarray Data Analysis V. Springer, 2005. (in press).

*Supported by grant 04/03967-0 from FAPESP - Brazil