# Estimation of Probabilistic Genetic Networks of *Plasmodium falciparum* from Dynamical Expression Signals

Junior Barrera, Roberto M. Cesar-Jr, David C. Martins-Jr[*],
Ricardo Z. N. Vêncio, Carlos A. B. Pereira
Institute of Mathematics and Statistics - University of São Paulo - Brazil
Hernando A. del Portillo
Institute of Biomedical Sciences - University of São Paulo - Brazil

## Abstract

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 presents 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, was developed.

The applicability of this tool for discovering gene networks of the malaria expression regulation system has been validated for simulated data (`http://www.vision.ime.usp.br/CAMDA2004/simulations`) and also for real *microarray* data using the glycolytic pathway as a "gold-standard" (`http://www.vision.ime.usp.br/CAMDA2004/glycolysis.html`), as well as by creating an apicoplast as PGN network (`http://www.vision.ime.usp.br/CAMDA2004/apicoplast.html`) [2].

As our program creates PGN networks, a negative control was idealized to further validate the biological value of our findings. Thus, eight genes, four from glycolysis and four from the apicoplast organelle, were chosen randomly and used together as seed genes to create PGN networks based on single-gene and two-gene predictions. The results clearly demonstrated that the glycolysis and apicoplast PGN networks based on single-gene predictions were not interconnected (`http://www.vision.ime.usp.br/CAMDA2004/ga_c.html`). With the exception of two genes from the glycolytic PGN network that inter-connected with the apicoplast PGN network, remaining genes were not connected based on two-gene predictions (`http://www.vision.ime.usp.br/CAMDA2004/ga2_c.html`) [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].

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

## References

[1] Z. Bozdech, M. Llinas, B. L. Pulliam, E. D. Wong, J. Zhu and J. L. DeRisi. *The Transcriptome of the Intraerythrocytic Developmental Cycle of Plasmodium falciparum.* PLoS Biology, 1 (2003), 5.

[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 J. S. Shoemaker and S. M. Lin, editors, Methods of Microarray Data Analysis V. Simon, Springer, 2005. (in press).