Chapter 11

Chromosomal Spatial Correlation of Gene Expression in Plasmodium falciparum

J.B. Christian\textsuperscript{a}, C. Shaw\textsuperscript{c}, J. Noyola-Martinez\textsuperscript{a}, M.C. Gustin\textsuperscript{b}, D.W. Scott\textsuperscript{a} and R. Guerra\textsuperscript{a,\ast}

\textsuperscript{a}Department of Statistics, Rice University, Houston, TX 77005, USA
\textsuperscript{b}Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77005, USA
\textsuperscript{c}Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

Abstract

Malaria is responsible for half a billion infections and two million deaths each year. Understanding the biology of Plasmodium falciparum is critical if effective vaccines are to be developed to fight against this aggressive parasite. New information about the regulatory mechanisms of P. falciparum promotes the elucidation of the fundamental metabolic and transcriptional pathways which we must understand to design vaccines and better treatments. Of particular importance is the intraerythrocytic development cycle (IDC), the part of the P. falciparum life cycle spent in the blood stream of host mammals and that is responsible for the physical symptoms of malaria. The goal of this investigation is to examine spatially dependent co-regulation of gene expression over the 48-hour IDC. Correlation between gene expression and gene location over a few genes demonstrates evidence of co-regulated genes or operons, while correlation over many genes may provide evidence for some other transcriptional regulation mechanism such as chromatin remodeling or enhancers. We develop and apply a visualization and statistical testing methodology to examine expression–location correlation in a time-course microarray study of the IDC transcriptome. Contrary to the current paucity of evidence, our findings show evidence for spatial correlation. The biological implications of detected blocks of moderate but consistent spatial correlation provide novel insights into the transcriptome of P. falciparum.

Keywords: co-regulation, spatial correlation, DNA sequence data, microarray data, integration of data sources, visualization, permutation tests

\ast Corresponding author.
1. INTRODUCTION

Understanding the regulatory mechanisms in *P. falciparum* helps identify new targets for both preventing or stopping malaria infections. The study of transcriptional regulation is paramount to achieving these goals, and there are many interesting transcriptional phenomena in *Plasmodium*. With secondary, tertiary and quaternary levels of structure in the DNA, there is much speculation about the randomness of the ordering of genes. Operons, chromatin remodeling and enhancers can affect gene transcription over short, mid and long distances, respectively, along the chromosome. In addition, protozoa such as *Plasmodium* are capable of regulating gene expression by altering their chromosome structure. For example, expression of the *var* cell surface protein of *Plasmodium* is regulated by a silencing mechanism [7]. In other eukaryotes, gene silencing and related epigenetic phenomena are typically mediated by covalent modification of histones that can spread along chromosomes, altering the accessibility of genes to the transcription apparatus [14]. Whether this type of regulation extends beyond the *var* genes to other genetic loci remains to be determined. This investigation explores the basic properties of location dependent transcriptional regulation by searching for both small and large chromosomal areas with correlated gene expressions.

The data we analyzed were collected by Bozdech et al. [5] who also considered the problem of spatial correlation. They reported finding a few regions of 2–7 genes each showing spatial correlation among the 14 linear chromosomes of *P. falciparum*. One limitation of their approach is that the search was based on correlated expression of adjacent genes independent of the physical distance between them. Thus, neighboring genes may be close or far apart. This approach might lead to a loss of power in detecting spatial correlation. Aburanti et al. [1] found correlations in *E. coli* by similar methods. In both investigations the main results were largely descriptive without a formal framework for statistical significance, especially with respect to the multiple testing aspects. One important observation made in this problem was by Balázs et al. [2] who discuss the potential for spurious spatial correlation on the chromosome due to the spatial arrangement of probes on the microarrays themselves. In addition to suggesting detection methods for such artifacts, they also propose numerical methods to minimize this type of experimental bias. Kluger et al. [17] also report similar cautionary measures. Using signal processing methods, Jeong, Ahn, and Khodursky [16] report a thorough investigation of spatial chromosomal patterns in *E. coli* and provide convincing evidence of higher-order organization of transcription in bacteria.

**Analytical Objective.** The overall objective of our work was to develop a visual and statistical framework to examine the correlation between gene expres-