

# Statistical Analysis of the Relationship between Gene Expression and Location

Sachiyo Aburatani<sup>1</sup>      Nobuyoshi Sugaya<sup>1</sup>      Hiroo Murakami<sup>1</sup>  
sachiyo@ims.u-tokyo.ac.jp      sugaya@ims.u-tokyo.ac.jp      hiroo@ims.u-tokyo.ac.jp

Makihiko Sato<sup>1,2</sup>      Katsuhisa Horimoto<sup>1</sup>  
makihiko@ims.u-tokyo.ac.jp      khorimot@ims.u-tokyo.ac.jp

- <sup>1</sup> Laboratory of Biostatistics, Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan  
<sup>2</sup> Makuhari Systems Laboratory, Fujitsu, 1-9-3 Nakase, Mihama-ku, Chiba-shi, Chiba 261-8588, Japan

**Keywords:** gene expression, gene location, correlation coefficient, *Escherichia Coli*

## 1 Introduction

DNA microarray data provide us with a large amount of information on gene expressions in the living cell. One of the important goals of such expression analysis is to estimate the relationships among genes in entire genome [4, 1]. The completion of bacterial genomes provided us the tendencies of genes with related functions to remain together across several genomes. Furthermore, the case of genes, whose protein products physically interact, conserved the gene order [3]. The observation implies the possibility of correlation between the gene expression level and the gene location on the genome.

In this study, we present a statistical approach to investigate the relationships between gene expression levels and their location on the genome.

## 2 Methods and Results

### 2.1 Data Set

We analyzed the data from 15 DNA microarray data performed on *E.coli* [2].

### 2.2 Statistical Analysis

We first identified the genes transcribed in the same direction of each replichore. Then, we calculated Pearson's correlation coefficients between the genes from their expression data according to their gene order on genome. Furthermore, we tested the correlation coefficients by t-test in the range from  $5 \times 10^{-3}$  to  $10^{-11}$  significance probability.

### 2.3 Relationship between Expression and Location

Relative distance between the genes with a significance probability is evaluated. Fig. 1 shows the relationship between the relative distance and the number of clusters with different probabilities. As easily seen, a large number of correlated clusters were found in the short relative distances with different probabilities. However, many correlated clusters were also found with the larger probabilities, regardless of relative distance.

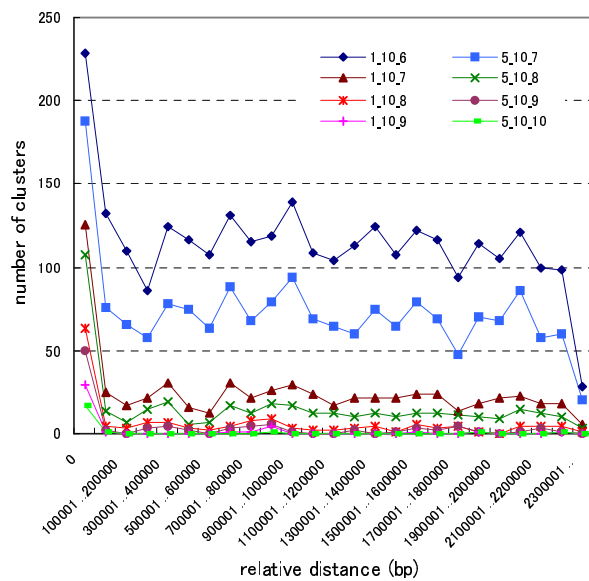


Figure 1: Relationship between relative distance and the number of correlated clusters.

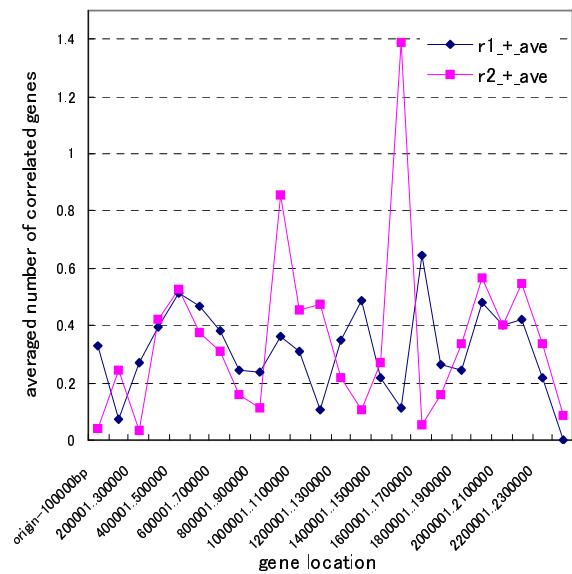


Figure 2: Pattern of the averaged numbers of correlated genes in replichore1 and replichore2.

Furthermore, we counted the number of correlated genes with a significance probability for each gene (criterion gene), and the numbers of correlated genes with the criterion genes were mapped on each replichore. In Fig. 2, the numbers of correlated genes located within each 100000bp are averaged for distinctive replichores. The two patterns of the averaged numbers of correlated genes are mutually similar in the same strand of replichore.

### 3 Discussion

The correlated gene clusters in the same strand were analyzed in terms of their biological function, and the well-known regulons were well found among them (data not shown). Similarity of correlated gene numbers between replichores might suggest the concerted transcription in the wider range on the genome.

### References

- [1] Bussemaker, H.J., Li, H., and Siggia, E.D., Regulatory element detection using correlation with expression, *Nature Genetics*, 27:167–171, 2001.
- [2] Courcelle, J., Khodursky, A., Peter, B., Brown, P.O., and Hanawalt, C., Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*, *Genetics*, 158:41–64, 2001.
- [3] Dandekar, T., Snel, B., Huynen, M., and Bork, P., Conservation of gene order: a fingerprint of proteins that physically interact, *TIBS*, 23:324–328, 1998.
- [4] Tong, A.H.Y., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Page, N., Robinson, M., Raghib izadeh, S., Hogue, C.W.V., Bussey, H., Andrews, B., Tyers, M., and Boone, C.M., Systematic genetic analysis with ordered arrays of yeast deletion mutants, *Science*, 294:2364–2368, 2001.