The Estimation of Genes Dependency Relation from Transcriptional Region and Gene Disruption Data

Tooru Ogura (110028)

School of School of Information Science,
Japan Advanced Institute of Science and Technology

February 14, 2003

Keywords: gene, dependency relation of genes, DNA , gene disruption data, statistical method, transcriptional region.

Recently, Deoxyribonucleic acid (DNA) genome sequences of human and many animate beings was decided from many biological experiences. Therefore, the focal point of the biological research is changing to clear the vital function of a gene or a protein in the cell.

DNA has genetic code of the animate being, and it is in the nucleus inside the cell. A gene is a fraction of DNA and creates the vital functional protein. This phenomenon is called a gene expression. The gene expression has a transcription process and a translation process. After the translation process, the protein is created from gene expression. The transcription process of the gene expression starts from binding the transcriptional activate protein on the upstream region of the gene. This upstream region is called transcriptional regulatory region. In the transcription process, messenger ribonucleic acid (mRNA) is made from a counterpart of based sequences of the gene. The created mRNA moves the extranuclear, and the protein is made from bonded amino acids which are corresponded mRNA based sequences in the translation process.

If the transcriptional repression protein binds to another transcriptional regulator region in the upstream region of the gene, this gene is stopped its expression. The protein is called transcriptional factor which activates

or represses the gene expression. The gene is called regulator gene which expresses transcriptonal factor.

If the expression of some genes are effected by transcriptional factor by the regulator gene expression, this regulator gene has dependency relation for these gene.

In this paper, if the gene has known transcriptional regualatory region in its upstream region, the effect of this gene expression for the regulator gene is judged using statistics methods for the gene disruption data. The purpose of this paper was to demonstrate a significance of the transcription region reseach for genes dependency relation from this papers method. The gene disruption data means the DNA microarary data which was measured from the gene disruption by the biological methods. The purpose of this paper was based of the estimation of dependency relation of genes from using the transcription region reseach. The estimated dependency relation of genes are used for the qualitative analysis of the inner cell. This analyzed result is used for new drug and etc. The target of the animate being in this paper was Bacillus subtilis which was known all gene location and based sequences on its DNA.

The method of transcriptional regulator region in this paper was using pattern matching for known transcriptional region to bind transcriptional factor and the upstream of all genes from Bacillus subtilis DNA based sequences data. DNA is double helical constitution from two strands which are bond each base sequences. The transcriptional regulator region research was seached for upstream region from both strands upstream. The upstream region of the gene wa identified from using the data which has the location of the gene and its transcriptional direction. The reseach of trancriptional regulator region was seached all genes upsteram region and the transcription regular region from using patterm matching.

The influence research of the regulator gene and the gene of the transcriptional regulator region in its upstream region was searched from using statistics methods and the gene disruption data. After calculated the correlation coefficient the regulator gene and all gene, if the correlation coefficient of the regulator gene and the gene of the transcriptional region in its upstream region from the gene disruption data is high, they have the possible of their dependency relation. The sample was decided genes of the

transcriptional region in its upstream region, and the parent population is all genes. The difference of the sample average and population average was checked for genes depenency relation. If the sample average is higher than the parent population, this paper was to judge their different from using test for the mean for their different research. If its judgement is different average, the regulator gene has the possible of the dependency relation for genes of this sample. This research decided four samples for the genes of the transcription region in its upstream region. Those sample were known dependency relation genes of transcriptional regulator region, genes of the same influence of known dependency relation genes (operon), genes of all genes of transcirption region in these upstream region, unknown dependency relation genes of the transcriptional region in their upstream region. The research of four samples for the possible of the dependency relation for the regulator gene.

The results of judgement the sample of known dependency relation genes and its operon were matched known fact and the analysis of DNA microarray data in more than half regulator genes. However, the sample of all genes of transcripstional region in its upstream region was judged less than half regulation genes which have the possible of the dependency relation. The sample of unknown dependency relation genes of the transcriptional region in its upstream region was judged dependency relation for the regulator gene.

The research of this sample was shown it has some genes of high correlation coefficient. However, this sample has many genes of low correlation coefficient. Therefore, the average of this sample was low than the parent population or same average from test for the mean.

Two reasons of many genes of low correlation coefficient in this sample were the influence of other unknown trancriptional factor, or independent for the regulator gene.

If the reason of the gene of low correlation coefficient in this sample is affected by other unknown regulator gene, this regulator gene is needed to identity one. Additionally, using the partial correlation function for regulator genes and other genes is solved this problem of the dependency relation of genes. If the reason is the independence gene in this sample, to calculate the calculation of correlation coefficient for the known depen-

dency gene and the gene in this sample is needed to solve this problem. This correlation coefficient is called vartial correlation coefficient for the correlation coefficient of the regulator gene. If the vartial correlation coefficient is high and positive sign, values of known gene and unknown gene are same for some DNA microarray data. The method is significant for the expection of independent genes in the sample from using theshold value for the virtual correlation coefficient.