Application of Gene Chip and Data Processing

Baiyi An [1, a] and Wei Shen[2, b *]

[1]College of Horticulture, Jilin Agricultural University, Jilin, China

[2]College of Computer Science and Technology, Beihua University, Jilin, China

[a]swabyswaby@163.com, [b]abyswabysw@163.com

Keywords: Microarray; Data analysis; Gene expression; Data processing

Abstract. Microarray is a new experiment technique developed in 1990s. It can measure the expression levels of thousands of genes simultaneously which is very different from the traditional biological experiments. Because the data generated by microarray experiments are too large for the traditional biological data analysis methods, many new data analysis methods have been proposed with the development of microarray technology.

Introduction

Gene chip [1], also known as microarrays, how it is to use the technology, the modern probe solid-phase in situ synthesis technology [2], microelectronics technology such as photographic plate printing technology on the tiny solid phase carrier, such as glass, silicon, plastic, etc., and orderly integration of a series of addressable recognition of cDNA fragments or nucleotides [3], oligomeric fragments of measured after fluorescent tags or radioactive marker, the probe on the chip hybridization, by confocal laser scanners to chip array fluorescent signal or radiation signal strength testing, so that you can in one experiment study the expression of thousands of genes at the same time. Chip technology [4], the rapid, parallel, high-throughput detection efficiency makes it the genome era scientists research one of the most powerful tool of gene. It has been used in detection of gene expression [5], gene mutation and polymorphism analysis [6], gene sequencing and drug screening, etc. This review first review the practical application of gene chip [7], then on the basis of the processing method of gene expression data is discussed in this paper [8].

The Application of Gene Chip

In practical applications, biochip technology can be widely used in disease diagnosis and treatment, drug screening, crops the good example of optimization, judicial authentication, food hygiene supervision, environmental testing, defence, aerospace and many other fields. It will be for human to realize the origin of life, genetics, development and evolution, diagnosis, treatment and prevention of human disease in new ways[9], designed for new biological macromolecules and rapid screening of lead compounds in drug discovery and pharmacogenomics studies provide the technical support platform.

   Drug Screening and New Drug Development. The product function and effect, and chip technology has high flux, mass, and parallelism or protein (protein chip) analysis of gene expression ability, has the huge advantage in drug screening. Made a massive chip screening can omit a lot of animal experiment and clinical research, shorten the time used in drug screening, improve efficiency, reduce risk. As with the human genome, genetic engineering drugs will enter a period of great development, in the development and production of the genetic engineering drugs, biochip has a larger market. Insulin by genetic engineering, for example, when we put the human insulin gene transfer to e. coli cells, we need to use some way to analyze the gene engineering bacteria type, in order to confirm whether insulin gene transfer success. Method taken in the past, people is called "restriction fragment length polymorphism (hereinafter referred to as RELP)" [10], this method is very cumbersome and complex, in terms of cost and efficiency than gene chip, replaced by the chip

technology is the inevitable trend in the future. By using gene chip screening drugs has a huge advantage decides it will become a drug research trend in this century.

Disease Diagnosis. Gene chip as a kind of advanced, large-scale, high flux detection technology, applied to the diagnosis of the disease, its advantage has the following several aspects: one is the height of the sensitivity and accuracy;2 it is quick and easy; Three is can detect various diseases at the same time. As applied to prenatal genetic disease inspection, extract a little amniotic fluid can detect whether the fetus with genetic diseases, and at the same time identify disease may reach dozens or even hundreds of, this is irreplaceable by other methods, to improve "eugenics" the implementation of the national policy. Such as diagnosis of pathogenic microorganisms infection again, the longer for the laboratory diagnosis technology, the check is not comprehensive, doctors often can only according to the experience of clinical diagnosis, reduce diagnostic accuracy, if the application of gene chip technology in the examination, the doctor can know what the patient is in a short time pathogenic microorganism infection; And can determine whether the pathogen resistance, resistance to what kind of antibiotics, is sensitive to what kind of antibiotics, and so on, so the doctor can target setting for treatment; Again like to have a family history of high blood pressure, diabetes and other diseases of the high risk population census, contact poison material mass screening of malignant tumor, etc., such as using the gene chip technology, immediately can get reliable results, other cardiovascular diseases, diseases of the nervous system, endocrine system disease, autoimmune disease, metabolic disease, such as using the gene chip technology, the early diagnostic rate will rise greatly, and the misdiagnosis rate will be greatly reduced, at the same time help the doctor comprehensive understanding of the disease condition of each system [11].

Environmental Protection. On environmental protection, gene chip is widely used, on the one hand, can rapid detection of pollution microbes or organic compounds to the environment, the pollution and harm of human body, plants and animals, at the same time can also through mass screening for protection, preparation of genetic engineering drugs, prevention and control of hazards or to control pollution sources of gene products.

The Justice. Gene chip can also be used in judicial, at this stage can be identified by DNA fingerprinting criminals, the future can be established throughout the country and even the world of DNA fingerprint, then to direct at the scene of the crime may be the suspect to stay hair, saliva, blood, semen, etc. is analyzed, and immediately with the DNA fingerprint criminal system to store DNA "fingerprint" comparison [12], and accurate investigation as soon as possible.

Modern Agriculture. Gene chip technology can be used for screening of crop genetic mutations, and looking for high yield, resistance to diseases and pests, drought resistance, resistance to freezing of related genes, and genes can also be used to scan and gene library construction, commodity inspection and quarantine, and other fields. Now the market has yet to be developed.

Research Field. Including gene expression detection, looking for a new gene, hybrid sequencing, gene mutation and polymorphism analysis and gene library construction, and etc. [13].

Gene expression test. About 100000 different human genome encoding genes, only master gene sequence information data, to understand the gene function is not enough, therefore, it is very important to have substantial mrna experiment monitoring tools.On chip technology to detect gene expression and its sensitivity, specificity, research results show that the chip technology is easy to monitor very lots of mrnas and can sensitively reflect the tiny changes in gene expression. Using gene chip technology has been more successfully to a variety of organisms including arabidopsis, yeast and human genome expression was studied, and use the technology (a total of 157112 probe) one-time test several different strains of yeast gene expression patterns of thousands of differences.

Looking for a new gene. Related experiments show that under the condition of the lack of any sequence information and gene chip can be used to find, such as HME gene and melanoma growth stimulating factor is found by gene chip technology.

DNA sequencing. The implementation of the human genome project to promote the efficiency of higher, to the development of the automation operation sequencing method, hybrid sequencing technologies in chip technology and its adjacent pile hybridization technology is a new kind of high

efficient rapid sequencing method. Such as using Affymetrix companies in the United States in 1998 to produce the chips with 135000 gene probes can cause the human DNA decoding speed by 25 times.

DNA mutation detection and genome polymorphism analysis. The relevant experimental results have shown that DNA chip technology can quickly and accurately study specific genes in patients with a large number of samples all possible heterozygous mutation. For the identification, mapping of the human genome single nucleotide polymorphisms and classification, the 16.6 KB mitochondria genome polymorphism research, etc[14]. As the genetic disorders found that an increase in the number of genes involved in cancer, mutation and polymorphism analysis will be more and more important.

Gene Network Model

When we want to build a biological gene networks, in the first place to have a search.Though sometimes you need to do a lot of simplified, but still many details to consider.For example, while a random model more realistic dynamic situation, but a single gene network model is established.If no necessary to simplify the biological process, it is hard to the potential network model.The following introduces some commonly used control model.

At present, the model used in gene regulation networks modeling mainly has the following kinds: Directed graph (Directed graphs), Boolean networks (Boolean networks), Bayesian networks, Bayesian networks), Differential equations, Differential equations model, DEM), random equation (Stochastic equations), etc. [15]. Directed graph and Boolean network is relatively simple model, the simulation of the system is qualitative, relatively coarse; Bayesian network is a probability model can quantitatively describe regulation network, random, but the prediction accuracy is low, and exist the problem of computing resource consumption; Stochastic model because of the difficulty of calculation is bigger, often difficult to apply to practice. Differential equation model, the changes in gene expression are expressed as the function of gene expression level and environmental factors, the gene regulation networks the dynamic behavior of the quantitative dynamic model is established, the differential equation model as the advantages of the gene regulatory network reconstruction is a powerful and flexible, to describe the relationship between the gene regulation networks and complex. In 2000, Tominaga D S - system model is proposed for differential equation system with GA optimization of system parameters, but the structure of the differential equation of constant, only optimized parameters to identify gene regulatory network. In 2002, HitoshiIba proposed using genetic process (based program, GP) and minimum average error (further mean squre, LMS) optimizing the structure and parameters of the differential equation, to deduce the simpler gene regulatory network; In 2007, P.G ennemark and D.W edelin evolution ordinary differential equations using heuristic search algorithm, is used for gene regulation networks and the identification of metabolic pathways. In 2008, Qian Lijun etc. [16]. Put forward using genetic programming and the differential equation of kalman filter optimization model [17].

Conclusion

Microarray technology has been successfully applied in many fields and their data analysis, clinical diagnosis, such as tumor can be based on the analysis of gene mutation is early diagnosis for cancer patients; To infection medication guidance of resistance identification of bacteria, so as to guide the clinical medication; Drug screening, drug target genes and the mechanism of action, screening the effective ingredients of drugs; Basic medicine, used in the study of expression, gene mutation, gene type components and sequencing, etc. In addition, in preventive medicine, environmental protection, military, agriculture and other areas are used. But microarray technology is still in the stage of development, with the development of the technology and perfect, will generate huge amounts of data, for fast and reasonable analysis of these data, find out the hidden information, and is an important and difficult work, statistics, information science disciplines such as the development and combination of expectation for the extraction of microarray data provide new ideas and methods.

References

[1] D'haeseleer P, Liang S, Somogyi R. Genetic network inference: from co-expression clustering to reverse engineering[J]. Bioinformatics 2000, 16: 707−726.

[2] Dojer N, Gambin A, Mizera A, et al. Applying dynamic Bayesian networks to perturbed gene expression data[J]. BMC Bioinformatics, 2006, 7: 249.

[3] Beal M J, Falciani F, Ghahramani Z, et al. A Bayesian approach to reconstructing genetic regulatory networks with hidden factors[J]. Bioinformatics, 2005, 21: 349−356.

[4] Bansal M, Gatta G D, di Bernardo D. Inference of gene regulatory networks and compound mode of action from time course gene expression profiles[J]. Bioinformatics, 2006, 22: 815−822.

[5] Eugene N, Emmanuel B. Regulatory network reconstruction using an integral additive model with flexible kernel functions [J]. BMC Systems Biology, 2008, 2: 8.

[6] Nelson P A, Kahana Y. Spherical harmonics, singular-value decomposition and the head-related transfer function[J]. Journal of Sound and Vibration, 2001, 239(4): 607−638.

[7] Liang M L, Dai L F. The left and right inverse eigenvalue problems of generalized reflexive and anti-reflexive matrices [J]. Journal of Computational and Applied Mathematics, 2010, 234: 743−749.

[8] Brits R, Engelbrecht A P, van den Bergh F. Locating multiple optima using particle swarm optimization [J]. Appl Math Comput, 2007, 189(2): 1859−1883.

[9] Chen D Y, Chuang T R, Tsai S C. Jgap: A java-based graph algorithms platform [J]. Software Pract Exper, 2001, 31(7): 615−635.

[10] Tim F R, Philippe R S, Paul T S, et al. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB[J]. BMC Bioinformatics, 2006, 7: 489.

[11] Bansal M, Belcastro V, Ambesi-Impiombato A, et al. How to infer gene networks from expression profiles [J]. Mol Sys Biol, 2007, 3: 78.

[12] Barabái A L, Albert R. Emergence of scaling in random networks [J]. Science, 1999, 286: 509−512.

[13] Silva A C, da Silva J K L, Mendes J F F. Scale-free network with Boolean dynamics as a function of connectivity[J]. Phys Rev E, 2004, 70(6): 66140−66147.

[14] Wilhite S E, Barrett T. Strategies to explore functional genomics data sets in IVCBI's GEO database[J]. Methods Mol Biol, 2012, 802: 41−53.

[15] KEGG Cell cycle-yeast-Saccharomyces cerevisiae [EB/OL]. [2010−04−13].http://www.genome.jp/kegg/pathway/sce/sce04111.html

[16] Wang Z, Shaik S, Inuzuka H, et al. Systems Biology Approaches in Breast Cancer Studies. in Breast Cancer Metastasis and Drug Resistance. Ahmad AEd.: Springer New York, 2013. pp. 329-340.

[17] Penfold C A, Wild D L. How to infer gene networks from expression profiles, revisited. Interface Focus. 2011, 1(6). 857-870.