International Journal of IT-based Public Health Management
Vol. 1, No. 1 (2014), pp. 1-10
http://dx.doi.org/10.21742/ijiphm.2014.1.1.01

A Study on Genes Information from Microarray Analysis of Cancer

V.Sujatha[1] and Shaheda Akthar[2]

[1]Department of CSE, Vignan'sNirula ITW Palakaluru, Guntur, AP, India[2]Department of CS, Government Women's College, Guntur, AP, India
[1]sujiekkurthi@gmail.com,[2]shaheda.akthar@yahoo.com

Abstract

A DNA microarray will track the expression levels of thousands of genes at the same time. Previous analysis has incontestable that this technology is helpful within the classification of cancers. Cancer microarray knowledge ordinarily contains a little range of samples that have an outsized range of organic phenomenon levels as options. To pick out relevant genes concerned in numerous kinds of cancer remains a challenge. So as to extract helpful gene information from cancer microarray knowledge and scale back spatiality, feature selection algorithms were consistently investigated during this study. Employing a correlation-based feature selector combined with machine learning algorithms like call trees and support vector machines, we tend to show that classification performance a minimum of nearly as good as printed results is obtained on cancer of the blood and diffuse massive B-cell cancer microarray data sets. During this paper, we tend to additionally demonstrate that a combined use of various classification and have choice approaches makes it potential to pick out relevant genes with high confidence.

Keywords: *Animal genomics, Immunogenomics, Microarray, Microbial genomics, Vaccines, Vaccinogenomics.*

1. Introduction to microbial genomics

The field of microbial genomics provides exciting new opportunities in the control and prevention of a wide range of veterinary diseases [8]. Genomics, and the functional analysis of genomic data, are leading to novel approaches for vaccine discovery, and improved methods for diagnosis and epidemiology. Genomes of several viral and bacterial pathogens that impact veterinary medicine have been sequenced. Each of these studies has provided new information and unique views into viral and bacterial pathogenesis [19]. In this introduction, the authors provide a brief overview of how bacterial genomic sequences are deduced and howgenesareidentifiedfrom thesedata.Inthefollowingsection,theyprovideabrief overview of some bacteria of importance to veterinary medicine for which genomic sequenceshavebeendeterminedandtheydescribehowgenomicdatacanbeexploitedto understand the ecology and epidemiology of pathogenic bacteria, a critical element for diseasecontrol andprevention.Currently,twobasicapproachesare usedfordeterminingthe sequence of bacterial genomes[14].Both methods use a 'shotgun' approach, where by random segments of the genome are sequenced. In the traditional method, plasmid libraries of cloned DNA fragments are constructed, and portions of the cloned DNA adjacent to plasmid vector sequences are

Article history:
Received (March, 2014),Review Result (May 07, 2014), Accepted (June 28, 2014)

Copyright ⓒ 2014 GV School Publication

determined by primer extension reactions. In an alternative pyrosequencing method, short DNA fragments are attached to microbeads, the fragments amplified, and a series of extension reactions are done that record the sequence of each fragment. Both techniques have strengths and weaknesses. The traditional method is labor intensive, but yields individual sequencing reads of 800 to 1000 bp in length that are paired with the complementary sequences obtained from the opposite end of the cloned fragment being physically linked. This facilitates construction of a scaffold on which the entire sequence assembly can be constructed. These contiguous segments of the genome are referred to as contigs, and the next step in the assembly process is to join adjacent contigs together until the chromosome or plasmid sequence (commonly referred to as a replicon because it is an autonomous replicating unit) is completely connected.

Polymerase chain reactions (PCR) and primer walking of selected templates are used to improve sequence quality, and should ultimately yield a single contig per replicon. Genomic data are processed by a variety of software programs that help identify individual genes, and translate them into the predicted protein products. Different proteins with a common function often share segments with a similar sequence of amino acids. Protein segments having shared sequences and presumably similar functions are referred to as motifs. The genomes of some veterinary bacterial pathogens have now been sequenced, including those of several important zoonotic agents. Development of an annotated genomic sequence establishes a framework through which targets for epidemiological analysis can be identified. Genomic sites that contain tandem repeated sequences often vary in the number of copies of the core repeating sequence among different strains due to errors that occur during DNA replication[7][8]. Because changes in copy numbers within different variable nucleotide tandem repeats (VNTR) accumulate at independent rates, simultaneous analysis of multiple VNTRs provides a powerful method for differentiating similar strains of bacteria. Development of VNTR-based tools using PCR for epidemiological studies of several bacterial pathogens has been made possible through access to genomic sequences, and is especially useful in characterizing organisms that are otherwise difficult to differentiate, Microarrays allow investigators to assess genetic variation between isolates and characterize global patterns of gene expression. For microarray analysis, RNA or DNA samples are differentially tagged with chemical labels and used to hybridize with DNA targets on the array. Unhybridised material is removed by washing and the retained, tagged samples are modified with a chemical that fluoresces when excited by lasers in a specialized instrument. The intensity of each spot, representing a hybridization target, usually a specific gene, is measured and compared to control samples to determine either genetic diversity (DNA input) or differential gene expression (RNA input). A key point in using microarrays to study gene expression is that many putative genes identified by genomic analysis encode proteins of unknown function. By identifying genes that respond to environmental stimuli rather than selecting genes based on a bias formed by presumed function, it may be possible to identify bacterial proteins essential for survival in the host.

This information is critical for rational selection of proteins for development as subunit vaccines.

Genomic analysis is also helping to develop a more comprehensive understanding of signals used to direct proteins to extra cytoplasmic locations, including the outer membrane. Outer membrane proteins (OMPs) are often considered ideal vaccine candidates, and improved methods for identifying protein motifs that direct proteins to the outer membrane are essential to assign presumptive locations of proteins with unknown function. This problem is of particular importance in identifying putative OMPs in spirochetes, a distinctgroupofbacteriawithanunusualcellwall/cellmembranestructure.Modelsforpredicting

OMPs based on other bacteria were potentially misleading when applied to spirochetes. Availability of the genomic sequences for several pathogenic spirochetes enabled [14]. to developanalgorithmwithimprovedpredictivepowertoidentifypotentialOMPsinthis group of bacteria. This information is being used to help select and analyze potential vaccine candidates for a wide variety of spirochete diseases, including leptospirosis. Viruses, due in part to their small size, are more easily compared using genomic approaches than bacteria, and new studies are providing useful information on strain variation. These analyses illustrate how genomics equencing is increasing our understanding of the interaction of important pathogenic microbes with their environment and facilitating the identification of relevant targets for designing vaccine that are effective under field conditions.

Accordingly, the genomic approach that can most readily be applied to vaccine discovery is the creation of subunit vaccines. In most cases, the production of killed or whole cell protein reparations and attenuated live vaccine strains does not need genomic technology, but rather a protein chemist or microbiologist .However, the identification of suitable anti gensis crucial to successful vaccine development based on subunit approaches[2][3].Using a combination of proteomics, genomics and bio informatics, investigators can quickly narrow the list from thousands of genes down to a few dozen vaccine candidates [13].The genomics/proteomics methods define the coding capacity, and then the bioinformatics analyses strim off pseudogenes (non functional or non coding) and sequences similar to human proteins and make predictions about secreted and surface-located proteins. Then it is back to the laboratory to recombinant produce these proteins and determine their immune gentility, although bioinformatics can make limited predictions along these lines as well. From these exercises emerge the short list of solid vaccine candidates to testin an animal model for protection.

Genomics also has the capability to make DNA vaccination studies much more efficient [17].Before the genome sequence was available for Mycobacteriumavium sub species paratuberculosis, DNA vaccination was attempted for this cattle pathogen using the expression library immunization procedure [4]. This study revealed two pools of DNA that were shown to be protective in mice and limited efforts were made to identify the relevant DNA in those pools. Random expression library immunization was used because the genome sequence was not available at the project’s inception. This random cloning method meant that the majority of clones would be in the opposite orientation relative to the coding strandor would be out of frame with the coding sequence. Therefore, many additional clones were needed to make the library truly representative of every coding sequence in the genome. An approximate total of 16,500 clones were used to immunize mice in that study. With the genome sequence now complete, a directed expression library immunization project, in which each clone faithfully represented a single coding sequence, could be initiated. This method has the advantage that fewer clones are needed, making resulting clone pools less complex and there is no ‘garbage’ or nonfunctional clones such as those in the opposite orientation or out of frame. For such a study, only 4,350 clones would be needed because that is the total number of genes present in the M. avium sub species par tuberculosis genome [16]. An added benefit is that fewer mice would be needed to test the clone pools. Genomic approaches can also identify the best targets for knock out mutations that enable engineering of attenuated vaccine strains. However, as yet, there are no published studies for bacterial pathogens that demonstrate a genome wide approach that can identify a target, knock out this target, and show both attenuation and protection in an animal host. Rather, the literature reports studies in which genomics has been used to genetically define a known vaccines train [8][13]. The Salmonellaty phimurium vaccine strain is protective in mice and lacks the transcriptional

Copyright ⓒ 2014 GV School Publication

regulator RfaH. Use of whole genome microarrays identified the RfaH-dependent genes giving investigators insight into the mechanism of attenuation for this vaccine strain. The most famous example of the use of genomics to define an attenuated bacterial vaccine strain is M. bovis BCG (Bacillus Calmette-Guérin – named after the French scientists Calmette and Guérin), which is the most widely used global vaccine to prevent human tuberculosis (TB) [4][16]. The initial step involves annotating the genome to define its coding capacity and hence all potential antigens [2][3]. Bioinformatics similarity searches should then be performed to discard pseudo genes and anything resembling human proteins. The remaining list of genes is then cloned and expressed and analyzed immunologically for vaccine candidates. Over 3 billion individuals have been vaccinated with BCG without major side effects. Whole genome analysis of pathogens enables the targeted selection of protective immunogens encoded by the disease-causing pathogen. This allows investigators to move away from empirical approaches in vaccine development towards a more focused, logical development and discovery of protective DNA segments and proteins. In the next section, the authors describe the applications of bioinformatics in the design of the ideal vaccine. Bioinformatics and computational vaccinologydesigning an ideal vaccine depends greatly on several factors associated with targeted pathogens and host responses, including knowledge at the molecular level of the immune response, pathogenesis, host–pathogen interaction, and genetic and physiological variation among animals and pathogens [19]. Recently discovered genome sequences of food animals and pathogens together with rapid advances in biotechnology will allow us to collect an unprecedented amount of information on hosts and pathogens that may have significant implications for vaccine discovery. However, transforming this information into practical understanding requires intensive data mining using sophisticated computational and bioinformatics tools. Highly intensive computation using high-speed central processing unit, multi-thread, and 64 bit technologies have greatly facilitated this process. Using computational approaches in vaccine design has become known as „computational vaccinology".

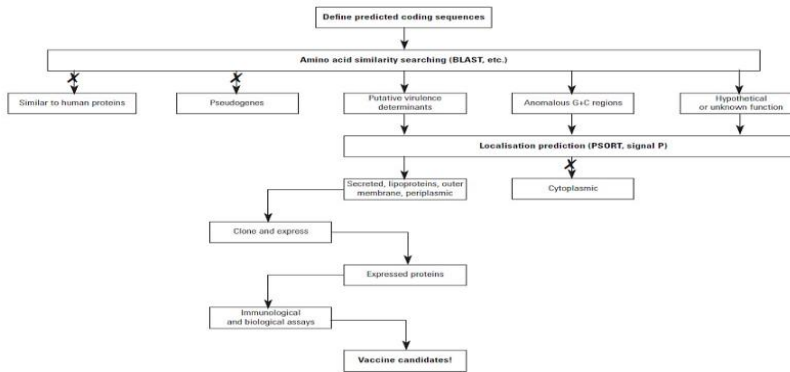

Figure 1. Schematic flow diagram showing the genomic approach to obtaining vaccine candidates for use in subunit vaccine approaches

In contrast to T-cell epitomes, B-cell epitopes remain much less predictable. Recently, using recurrent neural network, machine learning classifiers, and structural-energetic analysis improved the prediction of continuous B-cell epitopes [10][12], whereas the combination of protein 3D structures and statistics has been used to predict discontinuous B-cell epitopes. Although the technical difficulties of predicting B-cell epitopes remain to be overcome,

combining laboratory and bioinformatic analysis, such as phage display and mimotope analyses, can increase the accuracy of predicting continuous and linear epitopes. Mimotopes were first described as peptides that mimic native epitopes of foot and mouth disease virus and can bind to the same antibody as native antigens [2][3]. Candidate vaccines can be identified based on mimotopes that can induce antibody capable of binding to native antigens of pathogens. This approach may be useful for developing multi-epitope vaccines to fight against pathogens with several serotypes, such as foot and mouth disease virus. One of the challenges of epitope-based vaccines is population coverage due to MHC polymorphism. Different MHC molecules display distinct peptide-binding specificity. However, it has been shown that certain MHC alleles share overlapping peptide-binding specificity and the alleles can be grouped into super types based on their common binding specificity. Predicting peptides that bind to MHC super types for vaccine development can avoid the complication of MHC polymorphism. MHC alleles can also be grouped into super types based on the bioinformatics analysis of MHC protein structures and sequences, and supertypic MHC legends can be predicted for multiepitope vaccine development to increase population coverage. It has been estimated that targeting only 3 to 6 class I HLA alleles should cover ~90% of the human population because of linkage disequilibrium in the MHC loci. MHC genes are also tightly linked in food animals. Another application of bioinformatics in vaccine development is the interpretation of data collected with functional genomics approaches to gain detailed understanding of the immune response, pathogenesis, and host–pathogen interaction [19]. The knowledge obtained can be implemented in vaccine design. DNA microarray and proteomic analyses are two common approaches used in the studies of functional genomics, measuring transcript and protein expression levels, respectively [9][18]. Because gene expression levels are collected in a genome-scale, the data must be stored in databases in order to be managed and analyzed effectively [6]. The data also contain a large portion of technical variation introduced by laboratory procedures. The variation must be removed or minimized by data normalization before statistical analysis. Because multiple statistical tests are used in the data analysis, significant thresholds must be adjusted to balance between false positives and false negatives in detecting differentially expressed genes. Differential gene expression can be further analyzed to infer biological conclusions based on known molecular pathways and gene functions [20]. Bioinformatics analysis will play a very important role in animal health by generating the detailed knowledge needed for rational vaccine development. In summary, bioinformatics has become an additional powerful approach in vaccine design. The impact of the, Application of bioinformatics on rational vaccine design will be very significant in the future as research in this field progresses. Short synthetic peptides have been considered to be the next generation vaccines; however, there are several technical difficulties in using peptides as vaccines. Many of the obstacles could be overcome by bioinformatics approaches. Currently, there are many challenges confronting animal health in the areas of disease prevention and eradication. Bioinformatics may allow us to take all relevant information into consideration, including the genetic diversity of hosts and pathogens, to formulate vaccines that have broader effects regardless of these variations. Combining genomics and biotechnologies, bioinformatics can provide us with the detailed knowledge needed for vaccine development [8][13][17]. However, the tools and infrastructure to facilitate these applications in animal health have yet to be fully developed. The next section provides an update on the animal genome initiatives. Animal genomics In the past two decades, molecular biology has changed the face of agricultural animal research, primarily in the arena of genomics and the relatively new offshoot areas of functional genomics, proteomics, transcriptomics, metabolomics and metagenomics [8][13]. We now

Copyright ⓒ 2014 GV School Publication

have in place a powerful toolbox for understanding the genetic variation underlying economically important and complex phenotypes.

Developedconcomitantlywiththesegenomeprojectshasbeenasuiteofassociatedtools, including:

- ESTlibraries
- Bacterialartificialchromosomemaps
- Integratedphysicalandlinkagemaps
- Full-length complementaryDNA(cDNA)libraries
- Microarraysorgenechips
- Identificationandvalidationofalargenumberofsinglenucleotidepolymorphism markers.

Currently, major efforts are underway to develop haplotype maps of these genomes in order to fine map QTL and enable whole genome selection for quantitative traits [17]. While the maturing field of livestock genomics has been largely centered on improvement of production traits up tothe present time, it is widely recognized that the highest potential of these technologies resides in difficult to measure and expensive traits such as efficiency of nutrient utilization and resistance to disease. In particular, genomics holds great promise for unraveling the interactions between various hosts and pathogens. Understanding host–pathogen interactions at the molecular level will increase our understanding of viral and bacterial pathogenesis and the mechanisms pathogens use to evade host immune responses, both of which are paramount to the discovery of the ideal vaccine for control and eradication of animal diseases [19]. The next section describes the role of functional genomics and the application of microarray technologies to understand host–pathogen interactions at the genomics level. Host–pathogen interactions at the genomics level Recent progress in sequencing the genomes of microbial pathogens and their hosts is providing sophisticated strategies for unraveling the biological complexity of host–pathogen interaction. Elucidating these interactions at the molecular level, however, remains largely unrealized because understanding of gene function lags behind gene expression analyses obtained through high throughput, large-scale functional genomics approaches [6]. Nonetheless, functional genomics is rapidly revolutionizing the analysis of whole genome responses of pathogens and hosts. This will lead to a better understanding of disease processes, the mechanisms through which pathogens evade host immunity and the genetic basis of host–pathogen interactions, which will ultimately result in the discovery of novel vaccines. Collectively, the integration of these approaches in vaccine research (vaccinogenomics) is likely to fundamentally change the way scientists approach the challenges of discovering safe and effective vaccines. DNA microarray technologies allow high-throughput measurement of global gene transcription patterns on a whole-genome or tissue-specific basis, thereby enabling the investigation of the transcriptional status of complex biological systems underlying host– pathogen interactions [9][18]. Specifically, genomic technologies combined with immunology (immunogenomics) permit in-depth analysis of complex immunologicalprocesses based on large-scale whole genome approaches. Unlike conventional methods of differential gene expression (e.g. SAGE [serial analysis of gene expression] and differential display) that enable functional annotation of sequenced genomes, DNA microarray hybridization analysis stands out for its simplicity, comprehensiveness, data consistency, speed, and high throughput methodologies [6]. Global profiling of host and pathogen gene expression is an attractive approach to identifying the novel genes involved in disease processes since, in general, genes are transcribed only when and where their function is required. Thus, determining the conditions under which a given gene is expressed allows inferences to be made about its function.

2. Gene expression profiling

A variety of human DNA and oligonucleotide microarrays are commercially available. The most commonly used host microarrays are largely composed of ESTs. DNA arrays have become popular because they are generally considered to be easier to use than other gene expression profiling methods, and they allow the simultaneous quantification of thousands of genes from multiple samples. DNA array technologies rely on nucleic acid hybridization between labeled free targets derived from a biologic sample, and an array of DNA fragments (the probes, representing genes of interest) tethered to a solid surface. The targets, often produced by reverse transcription of messenger RNA (mRNA) and simultaneous labeling of the corresponding cDNAs, form a complex mixture of fragments that hybridize with their cognate probes during the assay. The signal generated on each probe reflects the mRNA expression level of the corresponding gene in the sample [11]. After detection, quantification, and integration of signals with specialized software, intensities are normalized for technical deviations, providing a gene expression profile for each sample that may be compared with the profiles of other samples. Standard, robust statistical methods are required for assigning significance values to gene expression measurements and to infer meaningful information.

3. Microarray applications in host–pathogen interaction studies

Strategies to investigate host–pathogen interactions using high-throughput gene expression analysis have been described utilizing various in vitro and in vivo models with whole genomic or tissue-specific microarrays. The main objective of these studies is to identify groups of genes that are involved in the activation or repression of key regulatory pathways of interest. Additionally, high throughput gene expression arrays allow one to investigate the temporal sequences of induction or repression of transcription, a prerequisite for determining the order of Events following host–pathogen interaction. In most cases involving complex disease processes, it is difficult to investigate all of the interacting factors in vivo. Thus, in order to reduce the complexity of whole animals, and to facilitate the interpretation of genomic data, in vitro systems have been exploited (e.g. homogeneous cell lines that are relevant to the type of study), the results of which are compared to the results obtained with in vivo studies.

4. In vitro studies

The first reported application of whole genome expression arrays to analyze host–pathogen interactions used primary human fibroblast cells infected with human cytomegalovirus (CMV). RNA samples collected at 40 min, 8 h, and 24 h after CMV infection were used to interrogate gene chips containing oligonucleotides corresponding to > 6,600 human mRNAs

(GeneChip microarray, Affymetrix, Santa Clara, California, USA). At 40 min post infection, 27 mRNAs showed significant alterations in expression, and at 8 h and 24 h, the number of altered genes increased to 93 and 364, respectively.

These high numbers of genes were in contrast to previous results obtained by differential display that identified 15 interferon-inducible genes activated by CMV. Although CMV replicates in many different cell types and the response may be different from those seen in primary human fibroblasts, it can be speculated that many of the genes identified using the GeneChip array are involved in early response of host cells to this virus. Therefore, it is not surprising that data analysis using GeneChip software showed that substantial transcription changes began very early after infection involving the activation of many early transcription

Copyright ⓒ 2014 GV School Publication

factors and proinflammatory signaling molecules, including cytokines, chemokines, stress inducible proteins, and interferon-inducible proteins.

Macrophages are important cells of the host immune system and play an important role indictating the quantity and quality of immunity to microbial pathogens.

5. In vivo studies

Influenza A/Texas/36/91 virus causes a human-like influenza syndrome in pigtailed macaques and this animal model has been successfully used to study influenza virus infection at the genetic level. Transcriptional analysis of lung and tracheobronchial lymph nodes of pigtailed macaques infected with a genetically reconstructed strain of human influenza H1N2

A/Texas/36/91 virus was carried out to study host–virus interactions and to compare the antiviral response of macaques and humans. A commercially available human cDNA array (Agilent Technologies, Palo Alto, California, USA) containing duplicate spots of 13,026 unique clones was used in this study. Significant transcriptional activation of inflammatory cells with the activation of interferon, B cell, and apoptotic pathways accompanied by overt clinical signs was observed in the lungs of H1N2-infected macaques, which coincided with gross and histopathological signs of inflammation and tissue damage [5]. The results of this cDNA microarray study provided insights into the molecular and cellular mechanisms associated with local innate immunity to influenza virus which were consistent with clinical signs of disease [1][15]. Furthermore, gene expression profiling of influenza infected lungs revealed new views of the role of cytotoxic T cells and natural killer cells in clearing influenza virus from the lung.

6. Conclusion

Genomic-based approaches area unit driving elementary changes in our understanding of biology. Comparative analysis of microorganism strains is providing new insights into infective agent evolution, virulence mechanisms, and host varies specificity. most significantly, factor discovery and genetic variations will currently be employed in genotyping analyses and also the rational style of vaccines. New research strategies using high-throughput organic phenomenon analysis area unit providing novel platforms for additional comprehensive understanding of host–pathogen interactions. above all, genomics is chop-chop revolutionizing the analysis of whole ordination responses of host and pathogens, which can ultimately cause an improved understanding of malady processes and also the mechanisms through that pathogens evade host immunity; identification of the genetic basis of host–pathogen interactions; and discovery of novel vaccines, medicine and biotherapeutics. Ultimately, we'll be ready to monitor the 2 approach language between hosts and pathogens with the chop-chop developing public information of the utterly annotated genomic sequence datasets of the many hosts and pathogens, the utilization of sequence-based high-throughput expression identification technologies, associate degree integrated bioinformatics tools to research and interpret genomic knowledge. Through these multiple and combined approaches, we'll get an entire image of infectious diseases, microorganism pathologic process associate degreed protecting host immune mechanisms victimization an integrated systems biology that may be crucial in developing a brand new generation of intervention ways against pathogens infecting humans and animals. Microarray-based technologies for learning genome-wide transcriptional identification hold exceptional promise for infectious diseases studies, since transcriptional management plays a key role in host–pathogen interactions. chop-chop advancing microarray technology platforms (expression profiling) can permit larger flexibility by providing this technology with increasing array part densities, higher detection

sensitivities, and additional extremely efficient protocols.

References

[1] Baskin, C.R., Garcia-Sastre, A., Tumpey, T.M., Bielefeldt-Ohmann H., Carter, V.S., Nistal- Villan, E. and Katze, M.G., "Integration of Clinical Data, Pathology, and c DNA Microarrays". In: Influenza Virus-Infected Pigtailed Macaques (Macacanemestrina). Journal of Virology, Vol. 78, No. 19, pp. 10420-10432 **(2004)**

[2] Blake, D.P., Hesketh, P., Archer, A., Carroll, F., Smith, A.L. and Shirley, M.W., "Parasite Genetics and the Immune Host: Recombination between Antigenic Types of Eimeria Maxima as an Entree to the Identification of Protective Antigens". In: Molecural and biochemical Parasitology, Vol. 138, No. 1, pp. 143-152 **(2004)**

[3] Blake, D.P., Shirley, M.W. and Smith, A.L., "Genetic Identification of Antigens Protective against Coccidia". In: Parasite Immunology, Vol. 28, No. 7, pp. 305-314 **(2006)**

[4] Bloom, B.R., Murray and C.J.: Tuberculosis, "Commentary on a Reemergent Killer". In: Science, Vol. 257, No. 5073, pp. 1055-1064 **(1992)**

[5] Blythe, M.J. and Flower, D.R., "Benchmarking B cell Epitope Prediction: Underperformance of Existing Methods". In: Protein Science, Vol. 14, No. 1, pp. 246-248 **(2005)**

[6] Boyce, J.D., Wilkie I., Harper, M., Paustian, M.L., Kapur, V. and Adler, B., "Genomic Scale Analysis of Pasteurella Multocida Gene Expression during Growth within the Natural Chicken Host". In: Infection Immunology., Vol. 70, No. 12, pp. 6871-6879 **(2002)**

[7] Bricker, B.J., Ewalt, D.R. and Halling, S.M., "Brucella 'HOOF-Prints': Strain Typing by Multi-Locus Analysis of Variable Number Tandem Repeats (VNTRs)". In: BMC Microbiology, Vol. 3, No. 15, **(2003)**

[8] Brosch, R., Gordon, S.V., Buchrieser, C., Pym, A.S., Garnier, T. and Cole, S.T., "Comparative Genomics uncovers Large Tandem Chromosomal Duplications in Mycobacterium Bovis BCG Pasteur". In: Yeast, Vol. 17, No. 2, pp. 111-123 **(2000)**

[9] Brown, P.O. and Botstein, D., "Exploring the New World of the Genome with DNA Microarrays". In: Nature Genetics, Vol. 21 (1 suppl.), pp. 33-37 **(1999)**

[10] Brusic, V., Bajic, V.B. andPetrovsky, N., "Computational Methods for Prediction of T-cell Epitopes – a Framework for Modeling, Testing, and Applications". In: Methods, Vol. 34, No. 4, pp. 436-443 **(2004)**

[11] Bulach, D.M., Zuerner, R.L., Wilson, P., Seemann, T., McGrath, A., Cullen, P.A., Davis, J., Johnson, M., Kuczek, E., Alt D.P.,Peterson-Burch, B., Coppel, R.L., Rood, J.I., Davies, J.K.. and Adler, B., "Genome Reduction in Leptospira Borgpetersenii Reflects Limited Transmission Potential". In: Proceedings of National Academy of Science, Vol. 103, No. 39, pp. 14560-14565 **(2006),** USA.

[12] Caoili, S.E., "A Structural-Energetic Basis for B-cell Epitope Prediction". In: Protein and Peptide Letters, Vol. 13, No. 7, pp. 743-751 **(2006)**

[13] Carrillo, C., Tulman, E.R., Delhon, G., Lu, Z., Carreno, A., Vagnozzi, A., Kutish, G.F. and Rock, D.L., "Comparative Genomics of Foot-and-Mouth Disease Virus. In: Journal of Virology, Vol. 79, No. 10, pp. 6487-6504 **(2005)**

[14] Chain, P.S., Comerci, D.J., Tolmasky, M.E., Larimer, F.W., Malfatti, S.A., Vergez, L.M., Aguero, F., Land, M.L., Ugalde, R.A. and Garcia, E., "Whole-Genome Analyses of Speciation Events in Pathogenic Brucellae". In: Infection Immunology, Vol. 73, No. 12, pp. 8353- 8361 **(2005)**

[15] Cohen, P., Bouaboula, M., Bellis, M., Baron, V., Jbilo, O., Poinot-Chazel, C., Galiegue, S., Hadibi, E.H. and Casellas, P., "Monitoring Cellular Responses to Listeria Monocytogenes with Oligonucleotide Arrays". In: Journal of biological Chemistry, Vol. 275, No. 15, pp. 11181-11190 **(2000)**

[16] Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E. 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A.,

McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S. andBarrell, B.G., “Deciphering the Biology of Mycobacterium Tuberculosis from the Complete Genome Sequence”. In: Nature, Vol. 393, No.6685, pp. 537-544 **(1998)**

[17] Collins, F.S., Green, E.D., Guttmacher, A.E. and Guyer, M.S., “A Vision for the Future of Genomics Research”. In: Nature, Vol. 422, No. 6934, pp. 835-847 **(2003)**

[18] Cummings, C.A. and Relman, D.A., “Using DNA Microarrays to Study Host-MicrobeInteractions”. In: Emerging Infectious Diseases, Vol. 6, No. 5, pp. 513-525 **(2000)**

[19] Cunningham, M.W., “Pathogenesis of Group A Streptococcal Infections”. In: Clinical Microbiology Reviews, Vol. 13, No. 3, pp. 470-511 **(2000)**

[20] Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C. and Conklin, B. R., “GenMAPP, A New Tool for Viewing and Analyzing Microarray Data on Biological Pathways”. In: Nature Genetics, Vol. 31, No. 1, pp. 19-20 **(2002)**