

## Defining the pathogenesis of inflammatory and immune diseases through database mining

Fan Yang<sup>1</sup>, Irene Hwa Yang<sup>1</sup>, Daniel H. Chen<sup>1</sup>

<sup>1</sup>Temple University School of Medicine, Philadelphia, PA, 19140

### TABLE OF CONTENTS

1. Abstract
2. Introduction
3. What are the principles of database mining?
4. Example 1: stimulation-responsive alternative splicing is an important mechanism in generating self-antigen epitopes
5. Example 2: a three tier model for caspase-1 activation and inflammation privilege are important mechanisms underlying the differences in the inflammation initiation in tissues
6. Example 3: anti-inflammatory microRNAs may play critical roles in inhibiting the expression of pro-atherogenic molecules
7. Conclusion
8. Acknowledgements
9. References

## 1. ABSTRACT

Recent research in human and animal genomes, transcriptomes, proteomes, and antigen-omes has generated a large library of data and has led to the establishment of many experimental data-based searchable databases. Scientists now face new, unprecedented challenges to develop more systemic methods to analyze experimental data and generate new hypotheses. This review will briefly summarize our pioneering efforts in using new database mining methods to answer important questions in inflammatory and immune-related diseases. The new principles and basic methodologies of database mining developed in Dr. Yang's laboratory will be delineated in the following studies: 1) a stimulation-responsive alternative splicing model for generating untolerized autoantigen epitopes; 2) a three-tier model for caspase-1 activation and inflammation privileges of various organs; and 3) a group of anti-inflammatory microRNAs which inhibit proatherogenic gene expression during atherogenesis. With technological advances, database mining has provided important insight into new directions for experimental research.

## 2. INTRODUCTION

Cardiovascular disease (CVD) continues to be a leading cause of morbidity and mortality in developed countries(1, 2). Despite a vast amount of research that has characterized both traditional and non-traditional risk factors for CVD, some mechanisms for CVD onset have only recently been discovered. Atherosclerosis is a chronic, inflammatory, autoimmune disease and its progression involves both innate and adaptive immune systems. Improving our understanding of the molecular pathogenesis of the involved immune response may lead to the future development of novel therapies.

Through much experimental research, an immense amount of untapped resources is available in biomedical literature and databases (3, 4). Both traditional hypothesis-driven research and discovery-driven “-omics” research, including genomics, transcriptomics(5), proteinomics, metabolomics, glycomics, lipidomics, localizomics, protein-DNA interactome, protein-protein interactome, fluxomics, phenomics(6), and antigen-omics (<http://www.cancerimmunity.org/links/databases.htm>) (7-

10) have generated and established many experimental electronic (redundant)databases. These databases include PubMed and numerous protein and nucleotide databases generated by the National Institutes of Health (NIH)/National Center for Biotechnology Information (NCBI) (see the NCBI handbook at <http://www.ncbi.nlm.nih.gov/books/NBK21101/>) and other institutions. The development of these new resources holds many opportunities for biomedical scientists to develop more systemic approaches to analyze the data and generate new hypotheses. The discrepancy between the vast amount of experimental data in various databases and the fewer numbers of actual database-mining research papers (< 50 papers on database mining in inflammatory and immune responses) indicates the technical and methodological difficulties and out of date concepts that biomedical scientists face.

Traditionally, medical literature search using the Index Medicus was the only way to identify knowledge gaps and generate new hypotheses. Now, literature searches have been significantly enhanced using more systemic approaches such as 1) NCBI-PubMed search and Google Scholar search; 2) screening various arrays (nucleic acid arrays, protein arrays and metabolic arrays) (11-14); and 3) mining experimental databases(2, 15-19). When compared to microarray data screening, which requires bioinformatic algorithms, and expertise, database mining offers many advantages. First, when compared to the generation of algorithms, database mining requires less bioinformatic assistance since for easily searchable purpose databases are established by bioinformatic experts. (20). Secondly, it provides extensive insight on existing knowledge gaps and allows the user to generate new hypotheses for further experimental research. Also, database mining enables maximum value extraction from costly experimental data. Despite these advantages, database mining still requires scientists to have a full understanding of its capabilities and limitations. Database mining is used to analyze experimental data that has been generated from numerous research projects, and does not predict theoretical results based on pure theoretical bioinformatic studies. Due to its immense library of data, database minings are not limited to sequence comparisons of nucleic acids and proteins(21), sequence alignments, analysis of hydrophobicity indexes, and functional domain predictions of proteins. In addition, database mining is not usually a required course for graduate students or postdoctoral fellows, which poses a challenge to set up new course to train young investigators to use mining techniques for their future careers. Lastly, reviewers of database mining publications often incorrectly regard the electronic data found in databases as “non-experimental or theoretical” and demand costly, redundant laboratory experiments to be performed, sometimes even requiring the use of outdated experimental methods.

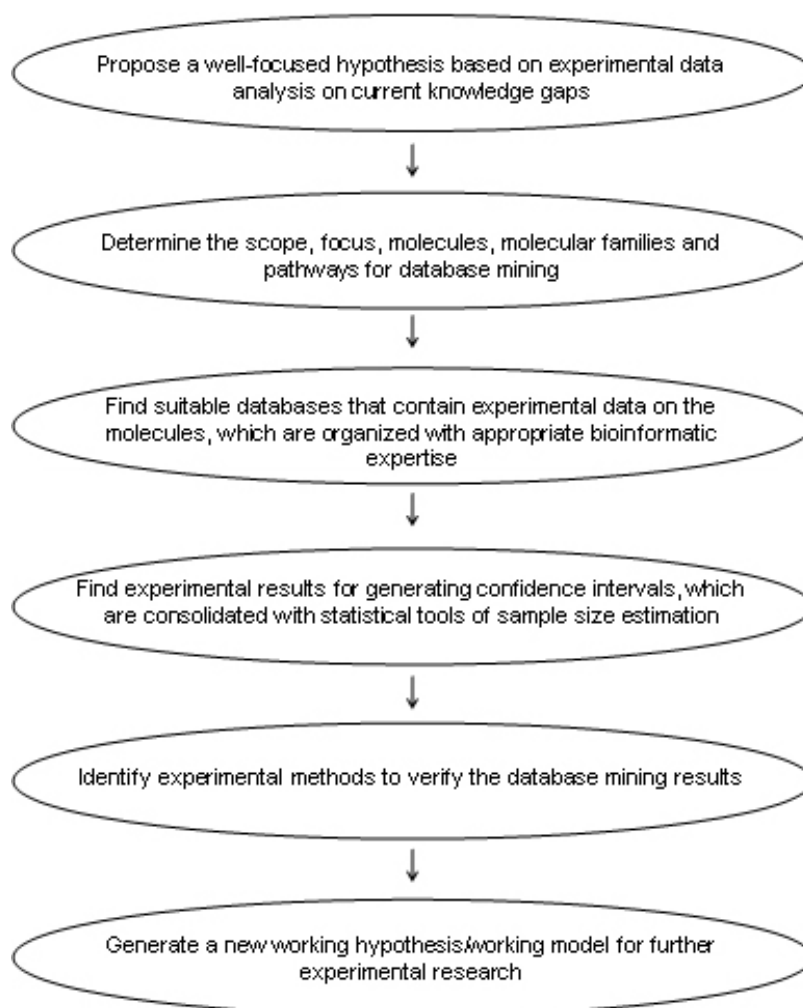
In the face of these challenges, bioinformatic scientists must work together with their colleagues to promote the significance of database mining projects in the biological sciences. It is encouraging to note that more database mining papers have already been published than in prior years. The 2011 (18<sup>th</sup>) database issue of the journal

“Nucleic Acid Research” features descriptions of 96 new and 83 updated online databases covering various areas of molecular biology(22). In addition to 32 databases of immunological interest that are now published(23), the Nucleic Acids Research online Database Collection, available at: <http://www.oxfordjournals.org/nar/database/a/>, now lists 1330 molecular biology databases. Moreover, our recently published review lists 11 B cell antigen epitope databases and 13 T cell antigen epitope analysis resources(2). All of the aforementioned suggests that a data mining approach has been accepted as an important mainstream tool used to analyze experimental data and generate new hypotheses (23).

Specifically, Dr. Yang’s laboratory has successfully pioneered major advances in the use of database mining in understanding adaptive and innate immune responses and inflammation (2, 15-19, 24). In this review, we will summarize the general approaches, principles, and databases used, as well as propose new working models for database mining research. Due to space limitations, we will not be able to discuss every database mining paper that Dr. Yang’s team has published. Our review will be particularly important and useful for biomedical scientists, since many are not involved in the generation of bioinformatic algorithms, but may desire to use database mining methods in their research either as parts of experimental studies or as free-standing projects. Of note, the database mining concept is not “brand new”. In fact, medical research has a long history of extracting data from costly experiments. For example, meta-analysis uses a statistical approach in order to combine the results of several epidemiological studies that address a set of related research hypotheses. In doing so, observations and conclusions may be made without using valuable funding. This practice of full value extraction began over 100 years ago and has been used across a wide variety of disease-related researches(25, 26). We believe that the use of database mining will become a routine part of experimental sciences used to generate new hypotheses from existing data.

### 3. WHAT ARE THE PRINCIPLES OF DATABASE MINING?

In recent years, many databases pertaining to biological immune responses and inflammation have been established(2, 16), which have not only expanded the scope and depth of publicly available online data, but have given birth to invaluable experimental analyses. Although research projects may vary in format, database mining approaches follow the same general principles (Figure 1): 1) Hypothesis: A clearly-defined hypothesis based on the current biomedical literature search in a given field is required to initiate a database mining project; 2) Scope: Database mining scope in terms of gene numbers are far more extensive than that examined using experimental methods. For example, our research examines the mRNA transcript expressions of about 30 genes including all the reported toll-like receptors, NOD-like receptors, and inflammatory caspases in more than ten different tissues. This broad scope allows us to obtain a panoramic view on a



**Figure 1.** Database mining flow-chart and principles.

complex pathway, and not limit ourselves to one gene or tissue; 3) Suitable databases: Databases that are suitable for examining the hypothesis must be available online; 4) Sizable experimentally verified data: In order to consolidate the results generated from database mining, a sizable amount of experimentally verified data published by various laboratories must be used to generate statistically significant confidence intervals (24)(15); 5) Verifiable methods: Experimental methods must be available to verify the data generated by database mining(27); and 6) A new working hypothesis: Using this approach, a new hypothesis will be proposed to test fewer, but much more-focused genes. In the next section, we will illustrate these principles with our own publications(2, 15-19, 24).

#### 4. EXAMPLE 1: STIMULATION-RESPONSIVE ALTERNATIVE SPLICING IS AN IMPORTANT MECHANISM IN GENERATING SELF-ANTIGEN EPITOPES

As discussed in our review, the identification and molecular characterization of self-antigens expressed in

human malignancies are capable of eliciting an anti-tumor immune response in patients and thus, is an active field of research in tumor immunology(30). More than 2,000 tumor antigens have been identified to date, with most being self-antigens(30). Despite this research, how non-mutated self-protein antigens, generated from both normal and tumor cells, gain immunogenicity remains poorly understood (30). Elevated immunogenicity underlying some tumor-specific antigens may be a consequence of mutations such as those seen in tumor suppressor proteins p53 and Ras, and chromosome translocation abnormalities, such as the expression of fusion oncogene Bcr-Abl in chronic myelogenous leukemia (31-34). However, the mechanism underlying increased immunogenicity of most non-mutated self-tumor antigens is their abnormal overexpression in tumors(30). Zinkernagel *et al.*(35) suggested that the overexpression of self-antigens in tumors overcomes the threshold of antigen concentration at which an immune response is initiated(36). In untolerized regions of certain antigen epitopes, this threshold may be lower. Overexpressed genes often encode tumor antigens, identified by serological identification of self-antigens by screening a cDNA library with patients' sera (SEREX)(37).

This may reflect the inherent methodological bias for the detection of abundant transcripts (38). The overexpression of antigens seen in tumors may result from both transcriptional and post-transcriptional mechanisms. We recently demonstrated that the overexpression of tumor antigen CML66L in leukemia cells and tumor cells via alternative splicing is the mechanism for its immunogenicity in patients (27, 39). This discovery not only clearly illustrates overexpression, but also points to alternative splicing as the molecular mechanism by which antigen overexpression occurs (27). A large proportion of SEREX-defined self-tumor antigens are also autoantigens (40). Self-tumor antigen CML28, previously identified by our lab, is also known as autoantigen Rrp46p (41) in the library. Using this information from SEREX combined with the overexpression seen in tumor antigen CML66L, we hypothesized that alternative splicing is a general mechanism not only for the overexpression of untolerized self-antigen epitopes in tumors, but also in autoimmune diseases. To test this hypothesis, we used database mining techniques to search the NIH-NCBI AceView database for potential mechanisms of how non-mutated self-proteins gain new untolerized structures that trigger immune recognition (15). AceView provides an organized, non-redundant, and comprehensive sequence representation of all known public mRNA sequences (mRNAs from GenBank or RefSeq, and single pass cDNA sequences from dbEST and Trace (<http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/>)). Previous analyses of 9554 randomly selected human gene transcripts showed an alternative splicing rate of approximately 42% ( $p < 0.001$ ). In comparison, our results showed that alternative splicing occurs in 100% of autoantigen transcripts. Within isoform-specific regions of autoantigens, MHC class I and class II-restricted T-cell antigen epitopes were encoded 92% and 88% of the time respectively, and 70% encoded antibody binding domains. Alternative splicing may be canonical or noncanonical. Canonical splicing removes introns that have 5'GT and 3'AG consensus flanking sequences (GT-AG rule) (42). We found that 80% of the autoantigen transcripts undergo noncanonical alternative splicing, which is also significantly higher than the less than 1% rate observed in randomly selected gene transcripts ( $p < 0.001$ ). Thus, our studies suggest that noncanonical alternative splicing may be the mechanism that generates untolerized epitopes, which ultimately lead to autoimmunity. Furthermore, a transcript product that does not undergo alternative splicing is unlikely to be a target antigen in autoimmune diseases (15). To further evaluate this finding, we examined the effect of proinflammatory cytokine tumor necrosis factor- $\alpha$  (TNF- $\alpha$ ) on the prototypic alternative splicing factor (ASF)/SF2 in the splicing machinery. Our results showed that TNF- $\alpha$  downregulates ASF/SF2 expression in cultured muscle cells, which correlates with our previous finding of reduced expression of ASF/SF2 in inflamed muscle cells in patients with autoimmune myositis (28).

Based on our and others' experimental results, we recently proposed a new model of stimulation-responsive splicing for the selection of autoantigens and self-tumor antigens (16) (also see

<http://preview.ncbi.nlm.nih.gov/pubmed/16890493>). Our new model theorizes that the significantly higher rates of alternative splicing seen in autoantigen and self-tumor antigen transcripts that occur in response to external stimuli like proinflammatory cytokines induce the extra-thymic expression of untolerized antigen epitopes, which result in autoimmune and anti-tumor responses. Using B cell and T cell antigen epitope analysis databases listed in the tables in our recently published review (<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2858284/pdf/JBB2010-459798.pdf>) (2), we showed that the protein sequences encoded by alternatively spliced exons sufficiently allow antibody-binding antigen epitopes and MHC class I- and class II-restricted T cell antigen epitopes to stimulate B and T lymphocytes, respectively (15). Our newly proposed model not only applies to non-mutated self-tumor antigens associated with cancers and autoantigens associated with numerous autoimmune diseases, but also applies to the expansion of self antigen stem cells. By using database mining, we have generated a new model of differential epitope processing for MHC class I-restricted viral antigen and tumor antigen epitopes (17). Our reports have demonstrated the principles of database mining in adaptive immune responses (15, 16, 27-29).

## 5. EXAMPLE 2: A THREE TIER MODEL FOR CASPASE-1 ACTIVATION AND INFLAMMATION PRIVILEGE ARE IMPORTANT MECHANISMS UNDERLYING THE DIFFERENCES IN THE INFLAMMATION INITIATION IN TISSUES

Atherosclerosis remains the leading cause of morbidity and mortality in the developed world. Several "traditional" risk factors have been identified for atherosclerosis including smoking, diabetes, hypertension, hyperlipidemia, obesity (43), oxidized low density lipoprotein, and hyperhomocysteinemia (HHcy). It is now known that chronic vascular inflammation plays an important role in the progression of atherosclerotic disease (44). Specifically, significant progress has been made in characterizing pathogen-associated molecular patterns' (PAMPs) receptor families (PAMP-Rs) and inflammasomes (the protein complex for activation of caspase-1), which further emphasizes the importance of proinflammatory cytokine interleukin-1 $\beta$  (IL-1 $\beta$ ) signaling in initiating inflammation (45). However, constitutive expression levels and readiness of PAMP-Rs, inflammasomes, and proinflammatory caspases seen in cardiovascular tissues continues to be ill-defined. Our study hypothesized that PAMP-Rs, inflammasome components, and proinflammatory cytokines like IL-1 and IL-18 are differentially expressed in cardiovascular tissues. To test our hypothesis, we searched the NCBI-UniGene database and analyzed cDNA cloning and DNA sequencing data from tissue cDNA libraries. In addition, we studied the expression profiles of Toll-like receptors (TLRs), cytosolic nucleotide binding and oligomerization domain (NOD)-like receptors (NLRs), inflammasome components, inflammatory caspases, and caspase-1 cleavable inflammatory cytokines. The UniGene database provides an organized view of the transcriptome, in which each

UniGene entry represents a set of transcript sequences based on information regarding protein similarities, gene expression, cDNA clone reagents, and genomic location (<http://www.ncbi.nlm.nih.gov/uniGene>). Upon analyzing our data obtained from UniGene, we made several important findings. Among the 11 tissues examined, only vascular and heart tissues express fewer types of TLRs and NLRs compared to immune system tissues such as blood, lymph nodes, thymus, and trachea. Additionally, brain, lymph nodes, and thymus tissue do not express proinflammatory cytokines IL-1 $\beta$  and IL-18 constitutively, which suggests that these two cytokines need to be upregulated when induced by inflammation. Finally, based on the expression data of three characterized inflammasomes (NALP1, NALP3 and IPAF), the examined tissues can be categorized into three tiers: the first tier tissues include brain, placenta, blood, and thymus and express inflammasome (s) constitutively; the second tier tissues have inflammasome (s) in a nearly-ready expression status requiring only the upregulation of one component; and the third tier tissues, like heart and bone marrow, require the upregulation of at least two components in order to activate functional inflammasomes. Given the expression readiness of inflammasomes in various tissues, we proposed a new working three tier model of inflammasome expression, which highlights the differences of tissues in initiating acute inflammation. Our model theorizes that (a) first tier tissues with constitutively expressed inflammasomes initiate inflammation quicker than second and third tier tissues; and (b) second tier tissues (requiring one component of upregulation) like the vasculature and third tier tissues (requiring more than one component of upregulation) like the heart have an inducible expression state of inflammasomes. Most likely, the inducible expressions of second and third tier inflammasomes are mediated through various signaling pathways and the interplay between these pathways must overcome a higher threshold than first tier tissues. Traditional concepts of immune privilege suggests that the lack of antigen-presenting self-MHC molecular expression protects against autoimmune destruction (30). Self MHC's lack of expression in immune privileged tissues, like testis, results in the failure of self-antigen presentation which stimulates the host's immune system, thus protecting the tissue from autoimmune destruction. Our lab proposed a new concept of tissues' immune privilege that focuses on a protective mechanism against tissue destruction which is mediated by inflammasome/IL-1 $\beta$ -based innate immune responses. In this new concept of immune privilege, vascular and heart tissue disproportionately express fewer types of TLRs and NLRs and may only inducibly express inflammasomes. In doing so, both heart and vascular tissues are protected against uncontrolled inflammatory destruction mediated by the inflammasome-based innate immune system (46). Our new model also explains the potential differences between cardiovascular tissues and other tissues with regards to acute inflammation initiation. First tier tissues have a higher percentage of experiencing acute inflammation compared to second and third tier tissues.

We and others showed that hyperhomocysteinemia (HHcy), elevated levels of plasma

homocysteine (Hcy), is an independent risk factor for cardiovascular diseases (CVD) including coronary heart disease and stroke (47-49). Recently, we performed an additional database mining study to examine the expression of more than 20 enzymes found in over 20 human and mice tissues that are involved in homocysteine metabolism and methylation (19). From the results, we proposed a new model of how hypomethylation (a post-translational protein modification) modulates the expressions of homocysteine-metabolizing enzymes (19). Taken together, our studies have demonstrated the usefulness of database mining in understanding innate immune reactions.

### 6. EXAMPLE 3: ANTI-INFLAMMATORY MICRORNAS MAY PLAY CRITICAL ROLES IN INHIBITING THE EXPRESSION OF PRO-ATHEROGENIC MOLECULES

Research has established that numerous genes are upregulated in atherogenesis through either epigenetic or genetic transcriptional mechanisms (50). However, transcription-independent mechanisms have received far less attention. Recent publications suggest that microRNAs, a newly characterized class of short (18-24 nucleotide long), endogenous, non-coding RNAs (51), contribute to the development of certain diseases by regulating biological processes such as cell growth, differentiation, proliferation, and apoptosis (52). More than 800 human microRNAs have been identified thus far, and up to 30% of human genes may be regulated by microRNAs (52, 53). Regulation is accomplished through post-transcriptional gene silencing (54) using Watson and Crick base-pairing predominately at the 3'-untranslated region (3'UTR) of messenger RNAs (mRNAs) (55, 56). Base-pairing can be further characterized as "perfect", "near perfect", leading to target mRNA cleavage and degradation, or "imperfect", leading to the inhibition of mRNA translation (54). Supporting evidence for microRNA involvement suggests that microRNAs function as key players during the critical stages of cell development, gene expression, and the maintenance of routine cellular functioning (57). Furthermore, microRNAs act on regulatory transcription factors, which lead to a broad indirect cellular effect as a result of their widespread gene modulating nature. Recent research has also demonstrated that changes in microRNA expression patterns are linked to several diseases including cardiovascular disease and subsequently, atherosclerosis. These studies have primarily focused on characterizing the elevated expression of microRNAs in disease models (58, 59). Current microRNA research has failed to answer two important questions, how microRNAs regulate atherogenic inflammatory genes, and whether the upregulation of atherogenic inflammatory genes is the result of anti-inflammatory microRNAs downregulation. To address these questions, our lab hypothesized that a group of anti-inflammatory microRNAs regulates the expressions of proatherogenic molecules (24). We developed a novel database mining approach using three databases including the online microRNA target prediction software TargetScan (<http://www.targetscan.org/>) (60-62), the Tarbase, an online database of experimentally verified microRNAs

(<http://diana.cslab.ece.ntua.gr/tarbase/>) (63, 64), and the online microRNA.org expression database (<http://www.microrna.org/microrna/home.do>) (65). By analyzing these databases using a statistical analysis strategy established in our previous database mining publications (15, 17-19, 66), our unique research yielded several key findings. First, we discovered that the expressions of 33 inflammatory genes (mRNAs) are upregulated in atherosclerotic lesions and second, that the mRNAs of those genes contain structural features in their 3'UTR for potential regulation by microRNAs. These structural features are statistically identical to previously experimentally verified 3'UTR microRNAs binding sites. We also found that 21 out of the 33 inflammatory genes (64%) are targets of highly expressed microRNAs, while the remaining 12 genes (36%) are targeted by normally expressed microRNAs. In addition, we established that 10 out of the 21 highly expressed microRNA-targeted inflammatory genes (48%) were targeted by a single microRNA, suggesting specificity of microRNA regulation. Meanwhile, 12 out of 25 highly expressed microRNAs (48%) targeted single inflammatory genes while the other 13 microRNAs targeted multiple inflammatory genes. Finally, microRNAs targeting atherosclerotic inflammatory genes use significantly higher binding interactions than microRNAs in the control group. Taken together, these results suggest that microRNAs regulating atherosclerotic inflammatory genes have unique features (24).

MicroRNAs play an integral role in modulating atherosclerosis-related processes including hypertension (microRNA-155), hyperlipidemia (microRNA-33, microRNA-125a-5p), plaque rupture (microRNA-222, microRNA-210), and atherosclerosis itself (microRNA-21, microRNA-126) (58). Given this, one must postulate whether certain microRNAs play a preventative role in disease development. One of the most interesting findings from our study is that the 25 microRNAs that are highly expressed under normal untreated conditions, target 21 out of the 33 atherosclerosis-upregulated inflammatory genes (64%). This finding suggests a novel mechanism by which a group of highly expressed anti-inflammatory microRNAs have the ability to suppress proatherogenic inflammatory gene upregulation under normal physiological conditions. While it is well established that microRNAs play important roles in the development of inflammation and cancer, our results are the first to suggest that microRNAs play a protective role by suppressing proatherogenic genes and by maintaining healthy arteries. Our conclusion is supported by other publications, which have shown that 7 out of the 20 microRNAs identified in this study were downregulated in studies using various proatherogenic factors (67-69). Together, our studies demonstrate the use and important of database mining in studying inflammation.

## 7. CONCLUSION

Active research in human and mouse “-omics” in the past decade has generated a tremendous amount of data and established many experimental data-based searchable databases. This offers unprecedented opportunities for investigators to develop more systemic and panoramic approaches to examine the databases and generate new

hypotheses. In this review, we summarize our pioneering efforts in using new database mining methods to address important questions in inflammations and immunological diseases. The new principles and basic methodologies of database mining developed in our laboratories are elucidated in several cases. With recent technological advances, database mining has provided significant new insights and hypotheses in defining the novel directions for experimental research.

## 8. ACKNOWLEDGEMENTS

Fan Yang and Irene Hwa Yang contributed equally to this work. We are very grateful to the Department of Pharmacology and the Dean's Offices in Temple University School of Medicine for the generous supports.

## 9. REFERENCES

1. Yang, X. F., Y. Yin & H. Wang: Vascular inflammation and atherogenesis are activated via receptors for pamps and suppressed by regulatory T cells. *Drug Discov Today Ther Strateg*, 5, 125-142 (2008)
2. Jan, M., S. Meng, N. C. Chen, J. Mai, H. Wang & X. F. Yang: Inflammatory and autoimmune reactions in atherosclerosis and vaccine design informatics. *J Biomed Biotechnol*, 2010, 459798 (2010)
3. Ishii, N., K. Nakahigashi, T. Baba, M. Robert, T. Soga, A. Kanai, T. Hirasawa, M. Naba, K. Hirai, A. Hoque, P. Y. Ho, Y. Kakazu, K. Sugawara, S. Igarashi, S. Harada, T. Masuda, N. Sugiyama, T. Togashi, M. Hasegawa, Y. Takai, K. Yugi, K. Arakawa, N. Iwata, Y. Toya, Y. Nakayama, T. Nishioka, K. Shimizu, H. Mori & M. Tomita: Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. *Science*, 316, 593-7 (2007)
4. Palakal, M., J. Bright, T. Sebastian & S. Hartanto: A comparative study of cells in inflammation, EAE and MS using biomedical literature data mining. *J Biomed Sci*, 14, 67-85 (2007)
5. Liang, M., A. W. Cowley, Jr., M. J. Hessner, J. Lazar, D. P. Basile & J. L. Pietrusz: Transcriptome analysis and kidney research: toward systems biology. *Kidney Int*, 67, 2114-22 (2005)
6. Joyce, A. R. & B. O. Palsson: The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol*, 7, 198-210 (2006)
7. Weinstein, J. N.: Fishing expeditions. *Science*, 282, 628-9 (1998)
8. Weinstein, J. N.: 'Omic' and hypothesis-driven research in the molecular pharmacology of cancer. *Curr Opin Pharmacol*, 2, 361-5 (2002)
9. Shimokawa, K., K. Mogushi, S. Shoji, A. Hiraishi, K. Ido, H. Mizushima & H. Tanaka: iCOD: an integrated

clinical omics database based on the systems-pathology view of disease. *BMC Genomics*, 11 Suppl 4, S19 (2010)

10. Houle, D., D. R. Govindaraju & S. Omholt: Phenomics: the next challenge. *Nat Rev Genet*, 11, 855-66 (2010)

11. King, J. Y., R. Ferrara, R. Tabibiazar, J. M. Spin, M. M. Chen, A. Kuchinsky, A. Vailaya, R. Kincaid, A. Tsalenko, D. X. Deng, A. Connolly, P. Zhang, E. Yang, C. Watt, Z. Yakhini, A. Ben-Dor, A. Adler, L. Bruhn, P. Tsao, T. Quertermous & E. A. Ashley: Pathway analysis of coronary atherosclerosis. *Physiol Genomics*, 23, 103-18 (2005)

12. Warner, E. E. & B. K. Dieckgraefe: Application of genome-wide gene expression profiling by high-density DNA arrays to the treatment and study of inflammatory bowel disease. *Inflamm Bowel Dis*, 8, 140-57 (2002)

13. Loza, M. J., C. E. McCall, L. Li, W. B. Isaacs, J. Xu & B. L. Chang: Assembly of inflammation-related genes for pathway-focused genetic analysis. *PLoS One*, 2, e1035 (2007)

14. Pandey, R., R. K. Guru & D. W. Mount: Pathway Miner: extracting gene association networks from molecular pathways for predicting the biological significance of gene expression microarray data. *Bioinformatics*, 20, 2156-8 (2004)

15. Ng, B., F. Yang, D. P. Huston, Y. Yan, Y. Yang, Z. Xiong, L. E. Peterson, H. Wang & X. F. Yang: Increased noncanonical splicing of autoantigen transcripts provides the structural basis for expression of untolerized epitopes. *J Allergy Clin Immunol*, 114, 1463-70 (2004)

16. Yang, F., I. H. Chen, Z. Xiong, Y. Yan, H. Wang & X. F. Yang: Model of stimulation-responsive splicing and strategies in identification of immunogenic isoforms of tumor antigens and autoantigens. *Clin Immunol*, 121, 121-33 (2006)

17. Yang, X. F., D. Mirkovic, S. Zhang, Q. E. Zhang, Y. Yan, Z. Xiong, F. Yang, I. H. Chen, L. Li & H. Wang: Processing sites are different in the generation of HLA-A2.1-restricted, T cell reactive tumor antigen epitopes and viral epitopes. *Int J Immunopathol Pharmacol*, 19, 853-70 (2006)

18. Yin, Y., Y. Yan, X. Jiang, J. Mai, N. C. Chen, H. Wang & X. F. Yang: Inflammasomes are differentially expressed in cardiovascular and other tissues. *Int J Immunopathol Pharmacol*, 22, 311-322 (2009)

19. Chen, N. C., F. Yang, L. M. Capecci, Z. Gu, A. I. Schafer, W. Durante, X. F. Yang & H. Wang: Regulation of homocysteine metabolism and methylation in human and mouse tissues. *Faseb J*, 24, 2804-17 (2010)

20. Spasic, I., S. Ananiadou, J. McNaught & A. Kumar: Text mining and ontologies in biomedicine: making sense of raw text. *Brief Bioinform*, 6, 239-51 (2005)

21. Mount, D.: Historical Introduction and Overview. In: *Bioinformatics. Sequence and Genome Analysis*. Ed: D.

Mount. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (2004)

22. Galperin, M. Y. & G. R. Cochrane: The 2011 Nucleic Acids Research Database Issue and the online Molecular Biology Database Collection. *Nucleic Acids Res*, 39, D1-6 (2011)

23. Salimi, N., W. Fleri, B. Peters & A. Sette: Design and utilization of epitope-based databases and predictive tools. *Immunogenetics*, 62, 185-96 (2010)

24. Virtue, A., J. Mai, Y. Yin, S. Meng, T. Tran, X. Jiang, H. Wang, and X-F Yang: Structural evidence of anti-atherogenic microRNAs. *Frontiers in Bioscience* 16, in press (2011)

25. Egger, M. & G. D. Smith: Meta-Analysis. Potentials and promise. *Bmj*, 315, 1371-4 (1997)

26. Egger, M., G. D. Smith & A. N. Phillips: Meta-analysis: principles and procedures. *Bmj*, 315, 1533-7 (1997)

27. Yan, Y., L. Phan, F. Yang, M. Talpaz, Y. Yang, Z. Xiong, B. Ng, N. A. Timchenko, C. J. Wu, J. Ritz, H. Wang & X. F. Yang: A novel mechanism of alternative promoter and splicing regulates the epitope generation of tumor antigen CML66-L. *J Immunol*, 172, 651-60 (2004)

28. Xiong, Z., A. Shaibani, Y. P. Li, Y. Yan, S. Zhang, Y. Yang, F. Yang, H. Wang & X. F. Yang: Alternative splicing factor ASF/SF2 is down regulated in inflamed muscle. *J Clin Pathol*, 59, 855-61 (2006)

29. Yang, X. F.: Immunology of stem cells and cancer stem cells. *Cell Mol Immunol*, 4, 161-71 (2007)

30. Yang, F. & X. F. Yang: New concepts in tumor antigens: their significance in future immunotherapies for tumors. *Cell Mol Immunol*, 2, 331-41 (2005)

31. Yotnda, P., H. Firat, F. Garcia-Pons, Z. Garcia, G. Gourru, J. P. Vernant, F. A. Lemonnier, V. Leblond & P. Langlade-Demoyen: Cytotoxic T cell response against the chimeric p210 BCR-ABL protein in patients with chronic myelogenous leukemia. *J Clin Invest*, 101, 2290-6. (1998)

32. Pinilla-Ibarz, J., K. Cathcart & D. A. Scheinberg: CML vaccines as a paradigm of the specific immunotherapy of cancer. *Blood Rev*, 14, 111-20. (2000)

33. Zorn, E., Orsini, E., Wu, C.J., Stein, B., Chillemi, A., Canning, C., Alyea, EP, Soiffer, R.J., and Ritz, J.: A CD4+ T cell clone selected from a CML patient after donor lymphocyte infusion recognizes BCR-ABL breakpoint peptides but not tumor cells. *Transplantation*, in press, (2001)

34. Clark, R. E., I. A. Dodi, S. C. Hill, J. R. Lill, G. Aubert, A. R. Macintyre, J. Rojas, A. Bourdon, P. L. Bonner, L. Wang, S. E. Christmas, P. J. Travers, C. S. Creaser, R. C. Rees & J. A. Madrigal: Direct evidence that leukemic cells present HLA-associated immunogenic peptides derived

- from the BCR-ABL b3a2 fusion protein. *Blood*, 98, 2887-93. (2001)
35. Zinkernagel, R. M. & H. Hengartner: Regulation of the immune response by antigen. *Science*, 293, 251-3. (2001)
36. Shlomchik, M. J., J. E. Craft & M. J. Mamula: From T to B and back again: positive feedback in systemic autoimmune disease. *Nat Rev Immunol*, 1, 147-53 (2001)
37. Sahin, U., O. Tureci, H. Schmitt, B. Cochlovius, T. Johannes, R. Schmits, F. Stenner, G. Luo, I. Schobert & M. Pfreundschuh: Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc Natl Acad Sci U S A*, 92, 11810-3 (1995)
38. Preuss, K. D., C. Zwick, C. Bormann, F. Neumann & M. Pfreundschuh: Analysis of the B-cell repertoire against antigens expressed by human neoplasms. *Immunol Rev*, 188, 43-50 (2002)
39. Yang, X. F., C. J. Wu, S. McLaughlin, A. Chillemi, K. S. Wang, C. Canning, E. P. Alyea, P. Kantoff, R. J. Soiffer, G. Dranoff & J. Ritz: CML66, a broadly immunogenic tumor antigen, elicits a humoral immune response associated with remission of chronic myelogenous leukemia. *Proc Natl Acad Sci U S A*, 98, 7492-7 (2001)
40. Chen, Y.: SEREX review. *Cancer Immunity*, <http://www.cancerimmunity.org/SEREX/>, (2004)
41. Yang, X. F., C. J. Wu, L. Chen, E. P. Alyea, C. Canning, P. Kantoff, R. J. Soiffer, G. Dranoff & J. Ritz: CML28 is a broadly immunogenic antigen, which is overexpressed in tumor cells. *Cancer Res*, 62, 5517-22 (2002)
42. Lewin, B.: Nuclear Splicing. In: Genes VII. Ed: B. Lewin. Oxford University Press Inc., New York, Cambridge (2000)
43. Ross, R.: Atherosclerosis. In: Cecil Textbook of Medicine. Ed: J. Wyngaarden, Smith, LH, Bennett, JC. W.B. Saunders Company, Philadelphia, London, Toronto, Montreal, Sydney, Tokyo (1992)
44. Hansson, G. K.: Inflammation, atherosclerosis, and coronary artery disease. *N Engl J Med*, 352, 1685-95 (2005)
45. Yang, X.-F., Ying Yin, Hong Wang: Vascular inflammation and atherosclerosis are activated via receptors for PAMPs and suppressed by regulatory T cells. *Drug Discovery Today Therapeutic Strategies*, 5, 125-142 (2008)
46. Streilein, J. W. & J. Stein-Streilein: Does innate immune privilege exist? *J Leukoc Biol*, 67, 479-87 (2000)
47. Maron, B. A. & J. Loscalzo: The treatment of hyperhomocysteinemia. *Annu Rev Med*, 60, 39-54 (2009)
48. Wang, H., X. Jiang, F. Yang, J. W. Gaubatz, L. Ma, M. J. Magera, X. Yang, P. B. Berger, W. Durante, H. J. Pownall & A. I. Schafer: Hyperhomocysteinemia accelerates atherosclerosis in cystathionine beta-synthase and apolipoprotein E double knock-out mice with and without dietary perturbation. *Blood*, 101, 3901-7 (2003)
49. Zhang, D., X. Jiang, P. Fang, Y. Yan, J. Song, S. Gupta, A. I. Schafer, W. Durante, W. D. Kruger, X. Yang & H. Wang: Hyperhomocysteinemia promotes inflammatory monocyte generation and accelerates atherosclerosis in transgenic cystathionine beta-synthase-deficient mice. *Circulation*, 120, 1893-902 (2009)
50. Turunen, M. P., E. Aavik & S. Yla-Herttuala: Epigenetics and atherosclerosis. *Biochim Biophys Acta*, 1790, 886-91 (2009)
51. Bartel, D. P.: MicroRNAs: target recognition and regulatory functions. *Cell*, 136, 215-33 (2009)
52. Zhang, C.: MicroRNAs: role in cardiovascular biology and disease. *Clin Sci (Lond)*, 114, 699-706 (2008)
53. Cheng, Y., N. Tan, J. Yang, X. Liu, X. Cao, P. He, X. Dong, S. Qin & C. Zhang: A translational study of circulating cell-free microRNA-1 in acute myocardial infarction. *Clin Sci (Lond)*, 119, 87-95 (2010)
54. Naeem, H., R. Kuffner, G. Csaba & R. Zimmer: miRSEL: automated extraction of associations between microRNAs and genes from the biomedical literature. *BMC Bioinformatics*, 11, 135 (2010)
55. Rasmussen, K. D., S. Simmini, C. Abreu-Goodger, N. Bartonicek, M. Di Giacomo, D. Bilbao-Cortes, R. Horos, M. Von Lindern, A. J. Enright & D. O'Carroll: The miR-144/451 locus is required for erythroid homeostasis. *J Exp Med*, 207, 1351-8 (2010)
56. Cordes, K. R., N. T. Sheehy, M. P. White, E. C. Berry, S. U. Morton, A. N. Muth, T. H. Lee, J. M. Miano, K. N. Ivey & D. Srivastava: miR-145 and miR-143 regulate smooth muscle cell fate and plasticity. *Nature*, 460, 705-10 (2009)
57. Baek, D., J. Villen, C. Shin, F. D. Camargo, S. P. Gygi & D. P. Bartel: The impact of microRNAs on protein output. *Nature*, 455, 64-71 (2008)
58. Rink, C. & S. Khanna: MicroRNA in Ischemic Stroke Etiology and Pathology. *Physiol Genomics* (2010)
59. Haver, V. G., R. H. Slart, C. J. Zeebregts, M. P. Peppelenbosch & R. A. Tio: Rupture of vulnerable atherosclerotic plaques: microRNAs conducting the orchestra? *Trends Cardiovasc Med*, 20, 65-71 (2010)
60. Vickers, K. C. & A. T. Remaley: MicroRNAs in atherosclerosis and lipoprotein metabolism. *Curr Opin Endocrinol Diabetes Obes*, 17, 150-5 (2010)
61. Rosero, S., V. Bravo-Egana, Z. Jiang, S. Khuri, N. Tsinoremas, D. Klein, E. Sabates, M. Correa-Medina, C. Ricordi, J. Dominguez-Bendala, J. Diez & R. L. Pastori: MicroRNA signature of the human developing pancreas. *BMC Genomics*, 11, 509 (2010)



62. Dong, H., M. Paquette, A. Williams, R. T. Zoeller, M. Wade & C. Yauk: Thyroid hormone may regulate mRNA abundance in liver by acting on microRNAs. *PLoS One*, 5, (2010)
63. Papadopoulos, G. L., M. Reczko, V. A. Simossis, P. Sethupathy & A. G. Hatzigeorgiou: The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Res*, 37, D155-8 (2009)
64. Sethupathy, P., B. Corda & A. G. Hatzigeorgiou: TarBase: A comprehensive database of experimentally supported animal microRNA targets. *Rna*, 12, 192-7 (2006)
65. Betel, D., M. Wilson, A. Gabow, D. S. Marks & C. Sander: The microRNA.org resource: targets and expression. *Nucleic Acids Res*, 36, D149-53 (2008)
66. Shen, J., Y. Yin, J. Mai, X. Xiong, M. Pansuria, J. Liu, E. Maley, N. U. Saqib, H. Wang & X. F. Yang: Caspase-1 recognizes extended cleavage sites in its natural substrates. *Atherosclerosis*, 210, 422-429 (2010)
67. Ji, R., Y. Cheng, J. Yue, J. Yang, X. Liu, H. Chen, D. B. Dean & C. Zhang: MicroRNA expression signature and antisense-mediated depletion reveal an essential role of MicroRNA in vascular neointimal lesion formation. *Circ Res*, 100, 1579-88 (2007)
68. Elia, L., M. Quintavalle, J. Zhang, R. Contu, L. Cossu, M. V. Latronico, K. L. Peterson, C. Indolfi, D. Catalucci, J. Chen, S. A. Courtneidge & G. Condorelli: The knockout of miR-143 and -145 alters smooth muscle cell maintenance and vascular homeostasis in mice: correlates with human disease. *Cell Death Differ*, 16, 1590-8 (2009)
69. Chen, T., Z. Huang, L. Wang, Y. Wang, F. Wu, S. Meng & C. Wang: MicroRNA-125a-5p partly regulates the inflammatory response, lipid uptake, and ORP9 expression in oxLDL-stimulated monocyte/macrophages. *Cardiovasc Res*, 83, 131-9 (2009)

**Key Words:** Database mining, Inflammation, Immune disease, Antigen splicing, Inflammasome, microRNAs, Review

**Send correspondence to:** Fan Yang, Temple University School of Medicine, 3420 North Broad Street, MRB 300, Philadelphia, PA 19140, Tel: 215-285-4182, Fax: 215-707-7068, E-mail: fannibal@gmail.com