Curriculum Vitae

Wen-Jun Shen

Personal Information

Room 4336, Academic Building 1
The Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong
Phone: (852) 64344370
E-Mail: wenjshen@ust.hk; wenjshen.my@gmail.com

Research Interests

Statistical machine learning and Bioinformatics

Education Background

City University of Hong Kong, Kowloon, Hong Kong
Ph.D. in Computer Science, Department of Computer Science
Advisor: Assistant Head & Associate Professor Hau-San Wong
My graduate research was also under the direction of Professor Stephen Smale.
3.97/4.3 GPA
July 2010-December 2013

Jinan University, Guangdong, China
M.S. in Applied Mathematics, Department of Mathematics
Advisor: Associate Head & Professor Chuan-Lin Zhang
86.66/100 GPA
September 2007-June 2009

Jinan University, Guangdong, China
B.S. in Mathematics and Applied Mathematics
88.4/100 GPA
September 2003-June 2007

Exchange Programe

Massachusetts Institute of Technology
Institute for Medical Engineering & Science
Laboratory for Computational Immunology
Principal Investigator: Director & Professor Arup K. Chakraborty
Whitehead Institute for Biomedical Research
Ploegh Lab
Principal Investigator: Professor Hidde Ploegh

February 2013- August 2013

Working Experiences

Division of Biomedical Engineering, The Hong Kong University of Science and Technology
Immuno-engineering group in cooperation with MIT
Postdoctoral Fellow
Principal Investigator: Head & Professor I-Ming Hsing
March 2014-present

Department of Computer Science, City University of Hong Kong
Research Assistant
Principal Investigator: Assistant Head & Associate Professor Hau-San Wong
October 2013-December 2013, December 2009-May 2010

Department of Computer Science, City University of Hong Kong
Teaching Assistant for the courses of CS5483 Data Warehousing and Data Mining; CS3483 Multimodal Interface Design; CS2204 Fundamentals of Internet Applications Development
September 2010-December 2012

Workshop and Conference Presentations

- HKUST Jockey Club Institute for Advanced Study, The Hong Kong University of Science and Technology, 2014
- IEEE International Conference on Bioinformatics & Biomedicine, Philadelphia, U.S.A., 2010

Awards and Honors

- A winner of Machine Learning in Immunology (MLI) Competition 2012 , the best overall predictor in Group 2 (Target MHC molecules: HLA-A*0201, -B*0702, -B*3501, -B*4403, -B*5301, and -B*5701; H2-Db and H2-Kb), 2012
  Organizers: Professor Vladimir Brusic, Boston University, MA, U.S.A.
  http://bio.dfci.harvard.edu/DFRMLI/HTML/natural.php, 2012
- Ph.D. Research Studentship, City University of Hong Kong, 2010-2013
- Excellent Master Thesis, Jinan University, 2009
- M.S. Research Studentship, Jinan University, 2007-2009
- Excellent Student Scholarship, Jinan University, 2003-2007

Publications

Journal Papers

1. Wen-Jun Shen, Hau-San Wong, Quan-Wu Xiao, Xin Guo and Stephen Smale, "Introduction to the Peptide Binding Problem of Computational Immunology: New Results," *Foundations of Computational Mathematics* (2013): 1-34. (This paper was first circulated with name "Towards a mathematical foundation of immunology and amino acid chains" on arXiv http://arxiv.org/abs/1205.6031v2.) .

2. Wen-Jun Shen, Shaohong Zhang and Hau-San Wong, "An effective and efficient peptide binding prediction approach for a broad set of HLA-DR molecules based on ordered weighted averaging of binding pocket profiles," *Proteome Science* 11.Suppl 1 (2013): S15.

3. Wen-Jun Shen, Yu Ting Wei, Xin Guo, Stephen Smale, Hau-San Wong and Shuai Cheng Li, "MHC Binding Prediction with KernelRLSpan and its Variations," *Journal of Immunological Methods* 406 (2014): 10-20.

Conference Papers

1. Wen-Jun Shen and Hau-San Wong, "OWA-PSSM - A Position Specific Scoring Matrix based Method Integrated with OWA Weights for HLA-DR Peptide Binding Prediction," *Proceedings of International Conference on Bioinformatics & Biomedicine* (BIBM 2012), Philadelphia, U.S.A.
   Published online: doi:10.1109/BIBM.2012.6392705.

2. Shaohong Zhang, Hau-San Wong, Wen-Jun Shen and Dongqing Xie, "AORS: Affinity-based Outlier Ranking Score," *The 2014 International Joint Conference on Neural Networks* (IJCNN 2014), Beijing, China.

Projects Experience

- *Prediction of influenza A antigenic evolution using cluster analysis of historical hemagglutinin sequence, April 2014 – present, Division of Biomedical Engineering, Hong Kong.*

Since 1968, Influenza A virus subtype H3N2 has been the most frequently occurring seasonal influenza. As the influenza virus undergoes gradual drift within subtypes (named antigenic drift), the flu vaccines need to be updated annually. Monitoring the major surface antigen hemagglutinin (HA) is essential for vaccine selection. The antigenic evolution of H3N2 viral strains tend to group in chronological clusters, and thus an effective vaccine should target an antigenic cluster that will be dominant in the coming season. In our study, we construct a cluster tree of H3N2 viruses from HA sequences to identify their antigenic evolution from 1968 to 2012. We aim at predicting the antigenic cluster that will become dominant in the coming season and suggesting a systematic way for vaccine selection.

- *Assessing the antigenicity of human influenza viruses by using fitness landscape and statistical coupling analysis, April 2013-August 2013, Dept. Chemical Engineering, Massachusetts Institute of Technology, MA, U.S.A; March 2014 – present, Division of Biomedical Engineering, Hong Kong.*

The hemagglutinin is an antigenic glycoprotein existing on the surface of influenza viruses. It initiates the first stage of viral infection by binding to the sialic acid surface receptor of host target cells (erythrocytes or epithelial cells in the upper respiratory tract) and facilitates cellular and viral membrane fusion. In addition, the hemagglutinin is the major surface antigen as well as the major target of neutralizing antibodies, and thus it undergoes major changes in antigenic composition. Current influenza vaccines we get each year are noninfectious or killed virus vaccines. The inactivated vaccines would be incredibly powerful weapons for controlling infection when the hemagglutinin of the vaccine strain is highly similar to the one of the circulating virus strain.

In this project, we characterize the antigenicity of influenza virus by analyzing the hemagglutinin HA1 subunit using fitness landscape and statistical coupling analysis. In our current study, we have verified that the predicted antigenic difference correlated well with the logarithm of the antigenic relatedness derived from ferret antisera HI assays. And thus it is a promising tool to predict antigenic variants as well as predict vaccine induced cross-reactive antibody responses.

- *Develop a consensus program for MHC class I eluted peptide prediction, July 2012-February 2013, Dept. Computer Science, City University of Hong Kong, Hong Kong.*

This project is partially supported by a grant from the City University of Hong Kong [Project No. 7002771]; GRF/ECS Grant NO. 9041805 [CityU 124512]; and [Project No. 9380050].

In an attempt to further improve the prediction performance of naturally processed peptides, the Machine Learning in Immunology (MLI) Competition was held by Dr. Brusic and his colleagues associated with the InCoB 2012 conference. Our consensus method was awarded the best overall predictor in Group 2. We demonstrate that a high performance model for MHC class I-eluted peptide prediction can be constructed by combining three or four MHC class I-peptide binding prediction approaches by support vector machine (SVM). In this work, we verify that the combination of KernelRLSpanI with NetMHCpan and NetMHC achieves the best and robust performance.

- *Develop a new method for MHC class II/peptide binding prediction by using position specific scoring matrices approach, February 2012-December 2012, Dept. Computer Science, City University of Hong Kong, Hong Kong.*

This project is partially supported by a grant from the City University of Hong Kong [Project No. 7002771]; National Natural Science Foundation of China [Project No. 61202273]; and Natural Science Foundation of Guangdong Province, China [Project No. S2012040007206].

The binding of foreign peptides to MHC class II molecules plays a vital role in stimulating CD4+ helper T lymphocytes immune response. In this work, we present a pan-specific method, OWA-PSSM, which is a significantly extended version of a well known method called TEPITOPE. The TEPITOPE method is able to perform prediction for only 50 MHC alleles, while OWA-PSSM is able to perform prediction for much more, up to 879 HLA-DR molecules. We evaluate the method on five benchmark datasets. The method is demonstrated to be the best one in identifying binding cores compared with several other popular state-of-the-art approaches. Meanwhile, the method performs comparably to the TEPITOPE and NetMHCIIpan2.0 approaches in identifying HLA-DR epitopes and ligands, and it performs significantly better than TEPITOPEpan in the identification of HLA-DR ligands and MultiRTA in identifying HLA-DR T cell epitopes.

- *Develop a novel kernel with application to Major Histocompatibility Complex/peptide binding prediction and human leukocyte antigen DR (HLA-DR) molecules supertype classification, July 2010-February 2013, Dept. Computer Science, City University of Hong Kong, Hong Kong.*

This project is supported by the GRF grant [Project No. 9041544] and [Project No. CityU 103210] and [Project No. 9380050].

We attempt to set a mathematical foundation of immunology and amino acid chains. To measure the similarities of these chains, a novel correlation kernel (we call $\hat{K}^3$) on strings is defined using only the sequence of the chains and a good amino acid substitution matrix (e.g. BLOSUM62).

The kernel is used in learning machines to predict binding affinities of peptides to major histocompatibility complex molecules. On both fixed allele and pan-allele benchmark databases, our algorithm achieves the state-of-the-art performance.

In addition, we employ the tools of diffusion map and diffusion kernel to improve this sequence correlation kernel $\hat{K}^3$ for MHC II binding prediction. The idea is to develop a diffusion kernel K_{diff}, which inherits the sequential similarity information from $\hat{K}^3$, and makes full use of the distribution of the sequences.

The kernel is also used to define a distance on an HLA-DR allele set based on which a clustering analysis precisely recovers the serotype classifications assigned by the World Health Organization (WHO).

These results suggest that our kernel relates well the chain structure of both peptides and MHC molecules to their biological functions, and that it offers a simple, powerful and promising methodology to immunology and amino acid chain studies.