

# An Investigation into the Feasibility of Detecting Microscopic Disease Using Machine Learning

Plenary Keynote Lecture of IEEE 7<sup>th</sup> International Symposium on Bioinformatics and Bioengineering, October 14-17, 2007. The Conference Center at Harvard Medical School  
Boston, Massachusetts USA

**Mary Qu Yang**

National Human Genome Research Institute  
National Institutes of Health (NIH), Bethesda, MD 20852 U.S.A.  
And also Oak Ridge, DOE email: [yangma@mail.NIH.gov](mailto:yangma@mail.NIH.gov)

*Abstract*— The prognosis for many cancers could be improved dramatically if they could be detected while still at the microscopic disease stage. We are investigating the possibility of detecting microscopic disease using machine learning approaches based on features derived from gene expression levels and metabolic profiles. We use immunochemistry and QRT-PCR to measure the gene expression profiles from a number of antigens such as cyclin E, P27<sup>KIP1</sup>, FHIT, Ki-67, PCNA, Bax, Bcl-2, P53, Fas, FasL and hTERT in several particular types of neuroendocrine tumors such as pheochromocytomas, paragangliomas; and the adrenocortical carcinomas (ACC), adenomas (ACA), and hyperplasia (ACH) in Cushing's syndrome. We provide statistical evidence that, higher expression levels of hTERT, PCNA and Ki-67 etc. are associated with a higher risk that the tumors are malignant or borderline, as opposed to benign. We also investigated whether higher expression levels of the P27<sup>KIP1</sup> and FHIT etc. are associated with a decreased risk of adrenomedullary tumors. While no significant difference was found between cell-arrest antigens such as P27<sup>KIP1</sup> for malignant, borderline, and benign tumors, there was a significant difference between expression levels of such antigens in normal adrenal medulla samples and in adrenomedullary tumors.

It follows from a comprehensive statistical analysis that a number of antigens such as hTERT, PCNA and Ki-67 can be considered as cancer markers, while another set of antigens such as

P27<sup>KIP1</sup> and FHIT are possible markers for normal tissue. Because more than one marker must be considered to obtain a classification of cancer or no-cancer, and if cancer, to classify it as malignant, borderline, or benign, we must develop a intelligent decision system using machine learning techniques, including variants of support vector machines, neural networks, decision trees, self-organizing feature maps (SOFM) and recursive maximum contrast trees (RMCT). These variants and algorithms we developed tended to work very well, yielding an average accuracy that was generally in excess of 90%. Our frame work focused on not only different classification schemes and feature selection algorithms but also ensemble methods such as boosting and bagging in an effort to improve upon the accuracy of the individual classifiers. It is evident when all sorts of machine learning and statistically learning techniques are combined appropriately into one integrated intelligent medical decision system, the prediction power can be enhanced significantly.

This research has many potential applications, not only in providing an alternative diagnostic tool and a better understanding of the mechanisms involved in malignant transformation, but also in providing information that is useful for treatment planning and cancer prevention.

*Acknowledgements* - The research was jointly with Drs. Jack Y. Yang and Zuojie Luo of Harvard Medical School.

## *Biography*



Mary Qu Yang received her interdisciplinary Ph.D. degree in Biophysics and Computer Engineering with specialization in Computational Science and Engineering from Purdue and holds Master of Electrical and Computer Engineering Degree in Computer Engineering, and Master of Science Degree in Biophysics, also both from Purdue University, West Lafayette, Indiana, USA. She was the recipient of an outstanding Bilsland Dissertation Fellowship, a NIH Post Doctoral Fellowship for the National Human Genome Research and also a NIH - Oak Ridge, DOE research specialist fellowship. She was trained as a combined computational and experimental scientist with more than 15 years of teaching, research and engineering practice experience in the fields of software engineering and biomedical sciences. Dr. Yang has received more than a dozen distinguished awards, including outstanding achievement awards, best original research paper awards, best software development and application research paper awards, smart engineering system design awards, theoretical development of computational intelligence awards, and a number of plenary keynote speaker invitations from the IEEE and other international conferences. She has been an editor of more than a dozen journals and proceeding books. She is currently on the editorial boards of Journal of Supercomputing, International Journal of Bioinformatics Research and Applications, International Journal of Data Mining and Bioinformatics, Journal of Computational Intelligence in Bioinformatics, and Advances in Chemoinformatics and Computational Methods (ACCM) book series. She has published more than 40 peer-reviewed papers and a number of invited book chapters and keynote papers.