

Next Generation DNA Sequencing and Bioinformatics: Bottlenecks and Opportunities

Kwok-Wing Stephen Tsui

Professor, School of Biomedical Sciences
Director, Centre for Microbial Genomics and Proteomics
Director, Hong Kong Bioinformatics Centre
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
Tel: (852) 26096381; Email: kwtsui@cuhk.edu.hk

With the emergence of high-throughput genome sequencers from various companies including Applied Biosystems, Illumina and Roche, the conventional approach of genomic investigation has been revolutionized and new applications involving these sequencers have been designed and adopted not only in genome centres, but also in general laboratories working on biological and biomedical research. These genome sequencers are capable of generating terabytes of raw data daily, posing a big challenge to various aspects of bioinformatics and computational biology, e.g. data quality assurance, data storage, cross platform data analysis standards, data comparison, data annotation and gene network analysis. The problem will become more prominent when this technology is applicable in areas such as clinical diagnosis and personalized medicine. In foreseeable future, we believe bioinformatics will be one of the major bottlenecks of the further development of biological sciences. In this talk, some of the genome projects in the Centre for Microbial Genomics and Proteomics of the Chinese University of Hong Kong will be used as examples to unravel the difficulties researchers generally come across in the area of bioinformatics (references). Applications including genome sequencing, transcriptome analysis and methylome determination will be covered. Moreover, possibilities of the clinical use of high throughput sequencing will be explored. Next, based on our understanding of the recent development in computational biology, possible solutions to these difficulties will be discussed. Finally, the future opportunities in bioinformatics and next generation DNA sequencing technology will be presented. We believe this talk can provide useful insights and future research directions for audiences who are interested in bioinformatics and computation biology.

References:

- [1] Sung JJY, Tsui SKW, Tse CH, Ng YET, Leung KS, Lee KH, Mok TSK, Bartholomeusz A, Au TCC, Tsoi KKF, Locarnini S, Chan HLY. (2008). Genotype-specific genomic markers associated with primary hepatoma based on complete genomic sequencing of hepatitis B virus. *J Virol* 82:3604-3611.

- [2] Chim SSC*, Tsui SKW*, Chan KCA, Au TCC, Hung ECW, Tong YK, Chiu RWK, Ng EKO, Chan PKS, Chu CM, Sung JJY, Tam JS, Fung KP, Waye MMY, Lee CY, Yuen KY, Lo YMD. (2003). Genomic characterization of the SARS-coronavirus of the Amoy Gardens outbreak in Hong Kong. *Lancet* 362:1807-1808 (*co-first authors).
- [3] Tsui SKW, Chim SSC, Lo YMD and the Chinese University of Hong Kong Molecular SARS Research Group. (2003). Coronavirus genomic-sequence variations and the epidemiology of the severe acute respiratory syndrome. *N Engl J Med* 349:187-188.
- [4] Hwang DM, Dempsey AA, Wang RX, Rezvani M, Barrans JD, Dai KS, Wang HY, Ma H, Cukerman E, Liu YQ, Gu JR, Zhang JH, Tsui SKW, Waye MMY, Fung KP, Lee CY, Liew CC. (1997). A genome-based resource for molecular cardiovascular medicine - Towards a compendium of cardiovascular genes. *Circulation* 96:4146-4203.
- [5] Liew CC, Hwang DM, Fung YW, Laurensen C, Cukerman E, Tsui S, Lee CY. (1994). A catalogue of genes in the cardiovascular system identified by expressed sequence tags (ESTs). *Proc Natl Acad Sci USA* 91:10645-10649.

Biography

Stephen Kwok-Wing TSUI is currently a professor in the School of Biomedical Sciences and the directors of the Hong Kong Bioinformatics Centre and Centre for Microbial Genomics and Proteomics in the Chinese University of Hong Kong. Dr. Tsui received his B.Sc. and Ph.D. degrees at the Chinese University of Hong Kong in 1985 and 1995, respectively. He joined the Chinese University of Hong Kong as an Assistant Professor in the Biochemistry Department in 1997. In the last 15 years, his team had characterized and named more than ten novel human genes. He is also a member of the International HapMap Consortium and had been worked on the single nucleotide polymorphisms of the chromosome 3p. During the SARS outbreak in 2003, his team was one of the earliest teams that cracked the complete genome of the SARS-coronavirus. Totally, he has published over 70 scientific papers in international journals, including *Nature*, *New England Journal of Medicine*, *Lancet*, *PNAS* and *Circulation*.

His research interests are in bioinformatics, comparative genomics and molecular biology of clinical pathogens including human immunodeficiency virus, hepatitis B virus, influenza virus and *Mycobacterium tuberculosis*. Besides, his team is also working on the identification of human non-coding RNA and the changes in human methylome during cancer progression.