6th Association of Asian Societies for Bioinformatics Symposium (AASBi2007)

2nd December 2007

Creation Theatrette, Matrix, 30 Biopolis Street, Singapore

Association of Asian Societies for Bioinformatics (AASBi) is a collaboration formed by Association for Medical and Bio Informatics Singapore (AMBIS), Bioinformatics Australia, Bioinformatics Society Taiwan, Japanese Society for Bioinformatics (JSBi), and Korean Society for Bioinformatics (KSBi). The purpose of AASBi is to promote bioinformatics in Asia-Oceanian area. Each year, AASBi will have one or more board meeting on issues related to collaboration and promotion of bioinformatics research and applications in its member societies.

The 6th AASBi Symposium (AASBi2007) will be on 2nd December 2007 at the Creation Theatrette in the Matrix Building, Biopolis, Singapore. The 2007 AASBi Board Meeting will be held on 2nd December 2007 at the same venue after the symposium. The board meeting is limited to the delegates from member societies. However, the symposium is open to the public.

AASBi2007 will be followed by the 18th International Conference on Genome Informatics (GIW2007, 3rd-5th December 2007, http://www.comp.nus.edu.sg/~giw2007) and by the 2nd International Symposium on Languages in Biology and Medicine (LBM2007, 6th-7th December 2007, http://lbm2007.biopathway.org) at the same venue. AASBi2007 is free and open to the public. But attendance at GIW2007 and LBM2007 require a single registration at http://www.comp.nus.edu.sg/~giw2007/registration.html.

The programme for AASBi2007 is given below. If you have any query on AASBi2007, please do not hesitate to let us know (wongls@comp.nus.edu.sg). The School of Computing and the Bioinformatics Programme at the National University of Singapore are honoured to jointly host AASBi2007. We warmly welcome you to Singapore.

Professor Limsoon Wong
Conference Chair, AASBi2007.

AASBi2007, 2nd December 2007, Creation Theatrette

Programme

09.00 Onsite registration	
Session I (chair: Limsoon Wong, NUS)	
09.20 Welcome address	
09.30 Keynote address	**Applying Structural and Computational Biology and Cyber Tools to Metagenomics** John Wooley University of California San Diego, San Diego, USA
10.15 Tea break	
Session 2 (chair: Louxin Zhang, NUS)	
10.45	**Inferring Phylogenetic Trees from Unaligned Molecular Sequences** Mark Ragan ARC Centre of Excellence in Bioinformatics, and Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia
11.15	**The Computational Biology of Genetically Diverse Assemblages** Allen Rodrigo Bioinformatics Institute, University of Auckland, New Zealand
11.45	**A Computational Study of Genome-Wide SNP Analysis** Michael K. Ng Department of Mathematics, Hong Kong Baptist University, Hong Kong
12.15 Lunch break	
Session 3 (chair: Greg Tucker-Kellogg, Lilly Singapore Centre for Drug Discovery)	
13.00	**Systems-Level Analysis and Engineering of E. coli for L-Valine Production** Sang Yup Lee KAIST, Daejon, Korea
13.30	**An Integrative Approach for Gene Annotation**

	based on Spectral Clustering and Network Modularity Hiroshi Mamitsuka Bioinformatics Center, Institute for Chemical Research, Kyoto University, Kyoto, Japan
14.00	**Knowlegator: A Literature-Driven, Ontology-Centric Content Delivery Platform** Christopher J. O. Baker Institute for Infocomm Research, Singapore
14.30	**TBA** Zeti Hussein School of Biosciences & Biotechnology, Faculty of Science & Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia
15.00 Tea break	
Session 4 (chair: Chee Keong Kwoh, NTU)	
15.30	**Petascale Computing for Systems Biology** Satoru Miyano Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan
16.00	**From Basic Statistical Concepts to Advanced Bioinformatics Algorithms** Tien-Hao Chang National Cheng Kung University, Tainan, Taiwan
16.30	**The Statistical Power of Phylogenetic Motif Models** Tim Bailey Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia
17.00	**An Update from ISCB** Tan Tin Wee Department of Biochemistry, National University of Singapore, Singapore
17.15 Closing address	
AASBi Board Meeting	
17.30	AASBi Board Meeting
19.00	AASBi Board Dinner @ Summer Pavillion, Ritz Carlton

Abstracts &

CV of Speakers

Session 1 (Keynote):

Applying Structural and Computational Biology and Cyber Tools to Metagenomics

John Wooley
University of California San Diego, San Diego, USA

CV

John Wooley is Associate Vice Chancellor for Research at the University of California San Diego, an adjunct Professor in Pharmacology, and in Chemistry and Biochemistry, and a Strategic Advisor and Senior Fellow of the San Diego Supercomputer Center. He received his Ph.D. degree in 1975 at The University of Chicago, working with Al Crewe and Robert Uretz in biological physics. Dr. Wooley created the first programs within the US federal government for funding research in bioinformatics and in computational biology, and has been involved in strengthening the interface between computing and biology for more than a decade. For the new UCSD California Institute for Telecommunication and Information Technology [Cal-(IT)2], Dr. Wooley directs the biology and biomedical layer or applications component, termed Digitally-enabled Genomic Medicine (DeGeM), a step in delivering personalized medicine in a wireless clinical setting. His current research involves bioinformatics and structural genomics, while his principle objectives at UCSD are to stimulate new research initiatives for large scale, multidisciplinary challenges. He also collaborates in developing scientific applications of information technology and high performance computing; creating industry-university collaborations; expanding applied life science opportunities, notably around drug discovery; establishing a biotechnology and pharmacology science park on UCSD's health sciences campus zone.

Session 2:

Inferring Phylogenetic Trees from Unaligned Molecular Sequences

Mark Ragan
ARC Centre of Excellence in Bioinformatics, and Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia

Abstract

It is commonly believed that to infer a phylogenetic tree that adequately represents the evolutionary history of a set of homologous molecular sequences, one must first align these sequences, *i.e.* arrange them relative to each other in a way that presents the best available hypothesis of homology at each and every sequence position (alignment column). Indeed, under a wide range of biologically relevant situations, sub-optimal alignment diminishes the accuracy of the resulting tree. Unfortunately, multiple sequence alignment is NP-hard. This raises the question of whether sufficient homology information can be inexpensively extracted from un-aligned sequences to support the inference of arbitrarily accurate phylogenetic trees. In this presentation I describe a range of approaches to alignment-free phylogenetic inference, describing the effects of inference method (distance-based or Bayesian), statistical variables (alphabet, word length and degeneracy), and one feature of evolving biomolecular sequences (across-sites rate

variation). The best alignment-free methods perform surprisingly well for biologically relevant problems, and can be superior to alignment-based approaches for certain problems. As a bonus, we find an apparent regularity in statistical parameterisation of alignment-free distances. (This talk is a based on joint work with , Michael Höhl.)

CV

Mark Ragan is founding head of Genomics and Computational Biology at the Institute for Molecular Bioscience at UQ, and Director of the ARC Centre of Excellence in Bioinformatics (2003-2010). He was recruited to UQ by Prof. John Mattick to contribute to the development of IMB. In Australia, he led the development of the ARC Centre in Bioinformatics and the Queensland Facility for Advanced Bioinformatics. His expertise spans comparative, computational and evolutionary genomics; molecular phylogenetics, including likelihood, Bayesian and experimental approaches; supertrees; automated pipelines and workflows for high-throughput bioinformatics; computational biology, especially involving large datasets; technologies for data management and knowledge integration in molecular bioscience; high-performance computing; and project management. His most exceptional research achievements in the past five years include the first quantitative map of lateral genetic transfer among prokaryotes (PNAS 2005) and the development of methods for phylogenetic inference and ortholog identification that do not require multiple sequence alignment (Systematic Biology 2007).

The Computational Biology of Genetically Diverse Assemblages

Allen Rodrigo
Bioinformatics Institute, University of Auckland, New Zealand

Abstract

Just when we were getting used to dealing with single genomes, we now find that we have to develop tools to deal with genomic diversity. New sequencing and other bio-technologies mean that collections of molecular data from genetically diverse assemblages are fast becoming the norm. Our own work on genetically variable viruses has prompted us to look at the complexities of analysing such data. In this talk, I will discuss two aspects of our ongoing work: (1) a subsampling strategy for estimating population genetic parameters from large samples, and (2) computational metagenomic analyses. The results presented are "hot-off-the-press", but they signal the direction our research will take in the next few months/years.

CV

Dr. Allen Rodrigo is Professor of Computational Biology and Bioinformatics, and the Director of New Zealand's Bioinformatics Institute at the University of Auckland. He has over 70 international publications on bioinformatics and computational biology, phylogenetics and evolutionary genetics, and the molecular evolution of viruses. Prof. Rodrigo is an Associate Editor of *Evolutionary Bioinformatics*, sits on the Scientific Advisory Board of two bioinformatics companies, is a Partner Investigator of the ARC Centre of Research Excellence in Bioinformatics, an Associate Investigator of the Allan Wilson Centre for Molecular Ecology and Evolution, and is involved in several national and international collaborative projects on genomics and bioinformatics. His major research contributions are in the area of virus evolutionary genetics, where he has spearheaded the development of new methods to analyse time-series genetic data from viral populations.

A Computational Study of Genome-Wide SNP Analysis

Michael K. Ng
Department of Mathematics, Hong Kong Baptist University, Hong Kong, SAR

Abstract

The recent hapmap effort has placed focus on the application of genome-wide SNP analysis to assess the contribution of genetic variability, particularly SNPs, to traits such as disease. In this talk, we discuss some of my recent work in computational study of genome-wide SNP analysis including tagging, LD map, classification and clustering.

CV

Michael Ng is a Professor in the Department of Mathematics at the Hong Kong Baptist University. He obtained his B.Sc. degree in 1990 and M.Phil. degree in 1992 at the University of Hong Kong, and Ph.D. degree in 1995 at Chinese University of Hong Kong. As an applied mathematician, Michael's main research areas include Bioinformatics, Data Mining, Operations Research and Scientific Computing. Michael has published and edited 7 books, published more than 160 journal papers. He currently serves on the editorial boards of several international journals.

Session 3:

Systems-Level Analysis and Engineering of E. coli for L-Valine Production

Sang Yup Lee
KAIST, Daejon, Korea

CV

Sang Yup Lee is presently the LG Chem Chair Professor at the Department of Chemical and Biomolecular Engineering, KAIST. He also directs the BioProcess Engineering Research Center and the Bioinformatics Research Center at KAIST. Professor Sang Yup Lee has received numerous awards and honors including the First Young Scientist's Award from the President of Korea, the Scientist of the Month Award from the Ministry of Science and Technology, the Best Patent Award (SejongDaewang Award) from KIPO, the Citation Classic Award from ISI, USA, and the First Elmer Gaden Award (1999 Best Paper Award) from Biotechnology and Bioengineering (John Wiley & Sons, USA) at the ACS National meeting. In 2001, Prof. Lee has received the National Order of Merit from the President of Korea, for his contribution to the advancement of science and technology

An Integrative Approach for Gene Annotation based on Spectral Clustering and Network Modularity

Hiroshi Mamitsuka
Bioinformatics Center, Institute for Chemical Research, Kyoto University, Kyoto, Japan

Abstract

Predicting gene functions is a fundamental issue in the post-genome era. A lot of approaches have been proposed using different types of data, such as sequence information, microarray expressions, phylogenetic profiles etc. As each of the different data types has its own flaw, a reasonable approach is to integrate the different data types, which are categorized into two classes: structured and unstructured data. Our approach has two steps: clustering genes using differenty types of data and predicting functions of new genes using those of known genes in the same cluster. For the first step, we have developed a new approach based on spectral clustering and network modularity to combine different types of data, particularly structure data with unstructured data. We'll describe our original approach for gene clustering and experimental results obtained by examining the performance of our method.

CV

Hiroshi Mamtisuka is a Professor of Bioinformatics Center, Institute for Chemical Research of Kyoto University, being jointly appointed as a Professor of Graduate School of Pharmaceutical Sciences of the same university. He received his Ph.D. degree in Information Sciences from the University of Tokyo in 1999. His current research interests are to develop data mining/machine learning techniques and apply them to the problems in biology.

Knowlegator: A Literature-Driven, Ontology-Centric Content Delivery Platform

Christopher J. O. Baker
Institute for Infocomm Research, Singapore

Abstract

The knowledge navigation platform 'Knowlegator' is designed to derive actionable-knowledge from the web. It comprises of (i) a

content acquisition engine that drives the delivery of literature from distributed public repositories, (ii) a workflow of natural-language processing algorithms that deliver semantically indexed sentences as A-boxes to OWL-DL ontologies, and (iii) a drag 'n' drop visual query composer that facilitates query navigation over concept hierarchies, instantiated object properties and visualization of data type properties in the ontology. The platform has been deployed in series of life science application scenarios. This talk will showcase 'instantiated' ontologies constructed for knowledge navigation in the domains of lipidomics, disease epidemiology and mutated phosphatases.

CV

Christopher Baker is a Principal Investigator at the Institute for Infocomm Research, A*STAR, in Singapore. His current focus is on the application of semantic web technologies in knowledge-based life science information systems. Prior to his current appointment he held the following scientific leadership roles: Bioinformatics Project Manager of the Génome Québec funded project, 'Ontologies, the semantic web and intelligent systems for genomics' and Group Leader In-silico Discovery at Ecopia BioSciences Inc. Dr Baker received post doctoral training at Iogen Corporation and the University of Toronto after completing his Ph. D. studies in Environmental Microbiology and Enzymology at the University of Wales, Cardiff, UK. He is co-editor of the first book to illustrate deployment of the semantic web technologies in the life sciences.

TBA

Zeti Hussein
School of Biosciences & Biotechnology, Faculty of Science & Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

CV

Zeti Azura Mohamed Hussein is a lecturer in the School of Biosciences & Biotechnology of the Universiti Kebangsaan Malaysia. He received his PhD in Bioinformatics from the University of Edinburgh in 2005. His research interests include structural bioinformatics, biological pathways modeling, biological database development, and function prediction based on structural and interaction information. His recent projects include, for example, deciphering biological pathways in polycystic ovarian syndrome and metabolite pathways in *P. minus* and *M. speciosa*.

Session 4:

Petascale Computing for Systems Biology

Satoru Miyano
Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan

Abstract

RIKEN has started the project for RIKEN Next-Generation Supercomputer R&D Center in 2006 and it is developing a supercomputer system with 10 peta flops computing ability in Kobe. Simultaneously, it launched the “Grand Challenge” project which will develop software applications for life sciences and nanotechnology. The grand challenge project for life sciences called “The Next-Integrated Life Simulations” aims at developing software applications that will enable us to simulate and analyze the processes that take place within living organisms, from the molecular level to the level of the whole body (see Fig.1).

It takes two approaches. The first is an analytical approach, where biological/ physiological phenomena (molecular scale, cellular scale, organ and body scale) are studied through basic principles (formulas and models). The second is a data-based inductive approach, where we will attempt to discover new processes and laws by analyzing large quantities of experimental data. We are involved with the second approach. Obviously, biology and medicine are facing with large-scale high dimensional heterogeneous data, e.g. transcriptome, proteome, P-P interactions, SNPs, physiological data, diseases, phenotypes, etc. Moreover, currently investigated biological systems are incomplete and complex, and there are big gaps between models and data. These models and data may be divided into two categories. One is the “general” one as human organism. The other is the “personal” one as

individuals. Our aim is to bridge and fuse the gaps between "general" and "personal" by constructing a petascale computational strategy that will develop and employ the following technologies (see Fig.2): (1) Inferring and analyzing large-scale gene networks, (2) Large-scale protein-protein interaction predictions, (3) Large-scale SNP data analysis, (4) Data assimilation for biological systems. With this strategy, we will contribute to practical application to new drug target discovery/design and diagnostic/therapeutic methods.

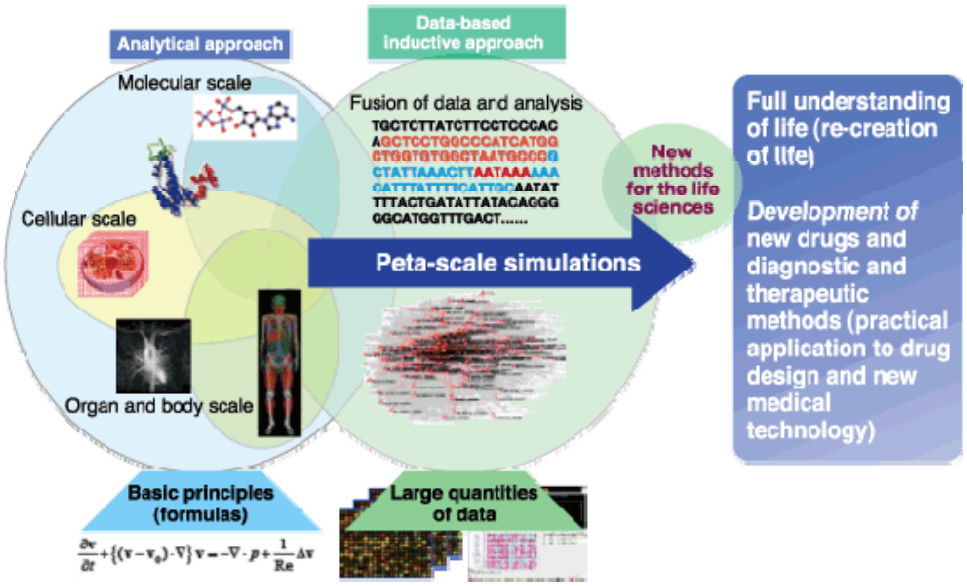
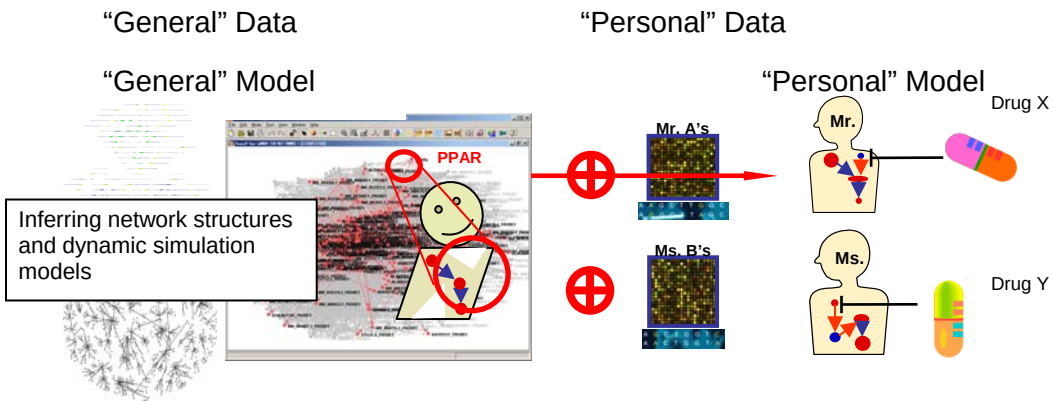


Figure 1



Predict large molecular networks, and use them as "maps" for searching drug/disease related genes and pathways.

Apply "data assimilation" to revise "general models" with "personal data".

Discover genes related to phenotypes (diseases/drug responses), and predict them from SNPs and environmental data.

Figure 2

From Basic Statistical Concepts to Advanced Bioinformatics Algorithms

Tien-Hao Chang
National Cheng Kung University,
Tainan, Taiwan

Abstract

The idea to design a machine learning algorithm usually comes from some fundamental concepts. In this presentation, we will discuss alternative machine learning algorithms based on a simple statistical concept, density, and its applications in bioinformatics. The three bioinformatics problems addressed are: (1) protein sequence analysis; (2) protein structure analysis; and (3) protein-ligand docking. Here we regard the protein sequence analysis as a classification problem, the protein structure analysis as a geometric problem, and the protein-ligand docking as an optimization algorithm. Based on a novel density estimation algorithm, we can improve the performance of some existing approaches for all these problems.

CV

Dr. Tien-Hao Chang completed his Ph.D. degree in Computer Science and Information Engineering at National Taiwan University in 2006. He joined National Cheng Kung University in 2006 and is currently an assistant professor with the Department of Electrical Engineering. His research interests include machine learning, structure biology, and bioinformatics.

The statistical power of phylogenetic motif models

Tim Bailey
Institute for Molecular Bioscience,
University of Queensland, Brisbane,
Australia

Abstract

One component of the genomic program controlling the transcriptional regulation of genes are the locations and arrangement of transcription factors bound to the promoter and enhancer regions of a gene. Because the genomic locations of the functional binding sites of most transcription factors are not yet known, predicting them is of great importance. Unfortunately, it is well known that the low specificity of the binding of transcription factors to DNA makes such prediction, using position-specific probability matrices (motifs) alone, subject to huge numbers of false positives. One approach to alleviating this problem has been to use phylogenetic "shadowing" or "footprinting" to remove unconserved regions of the genome from consideration. Another approach has been to combine a phylogenetic model and the site-specificity model into a single, predictive model of conserved binding sites. I will describe a simplified, theoretical model that allows us to quantify the number of genomes required at varying evolutionary distances to achieve specified levels of accuracy (false positive and false negative prediction rates). I will show that this depends strongly on the information content of the position-specific probability matrix and on the evolutionary model. I will demonstrate the accuracy of the theoretical model by applying it to a transcription factor binding site prediction task in yeast, and show that it provides a reasonable estimate of the potential accuracy of phylogenetic motif search.

CV

Tim Bailey's research is about discovering knowledge from data using techniques from machine learning, data mining and statistics. The focus is on applications in computational biology and genomics where the need for fast and reliable ways to make sense of the flood of data from modern, high-throughput biological experiments is extremely acute. Tim has helped develop a number of tools for biological sequence analysis: MEME (for motif discovery) and MAST (for scanning sequences with motifs), Meta-MEME for protein and DNA sequence modeling, and various tools for predicting protein properties, structure and family membership. Current projects include developing static models of the transcriptional output of genes; studying the function of small RNA in human cells; exploring the regulation of genes involved in the development of red blood cells; predicting the targets of transcription factors using methods combining models DNA of evolution with models of DNA binding.

An Update from ISCB

Tan Tin Wee

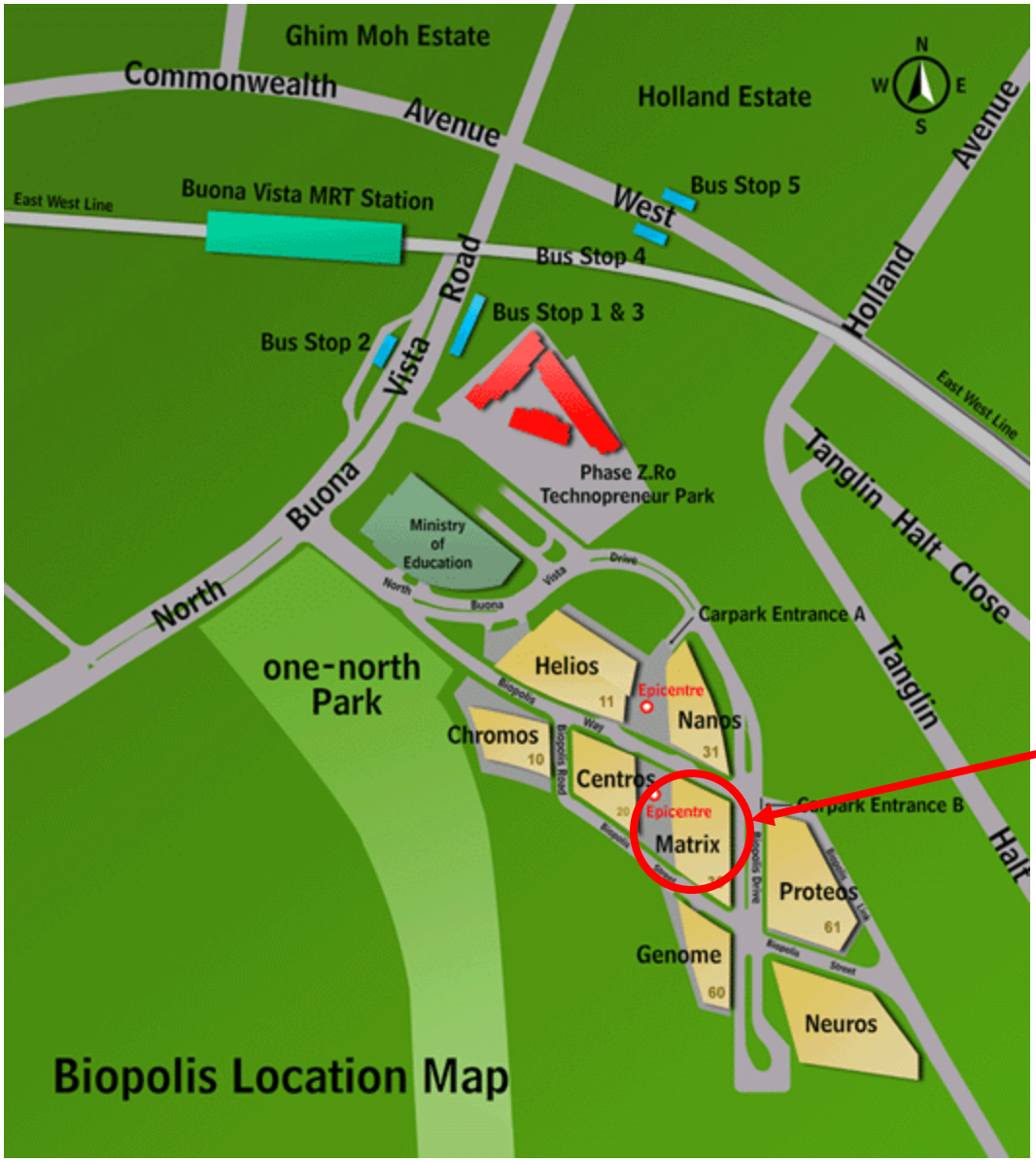
Department of Biochemistry, National University of Singapore, Singapore

Venue and Location

AABi2007 will be held in the Creation Theatrette at Matrix, Biopolis, Singapore on Sunday 2nd December 2007.

Address of Biopolis: Matrix, 30 Biopolis Street, Singapore 138671

Getting to Biopolis: By MRT: Alight at Buona Vista MRT station and walk 8 minutes; or take the free Biopolis Shuttle Bus service. **By Taxi**: Instruct driver to take you to "Biopolis" or to "MOE" (MOE = Ministry of Education, about 2 minutes walk to Biopolis).



AASBi 2007 conference venue

Sponsor:

Lilly

Answers That Matter.

Organizers

NUS National University of Singapore

School of Computing

NUS Bioinformatics Programme

Supporting Organizers

A*STAR Institute for Infocomm Research

A*STAR Bioinformatics Institute