Keynote Speeches

(1) Mapping genome-wide nucleosome dynamics

Steven Henikoff,
Fred Hutchinson Cancer Research Center,
Basic Sciences Division and Howard Hughes Medical Institute,
1100 Fairview Avenue North,
Seattle, WA, 98104, USA

Abstract

Eukaryotic gene expression occurs in the context of chromatin, and maintaining a region accessible to DNA-binding proteins for transcriptional regulation requires active processes that mobilize nucleosomes. Our approach to studying these processes has been to map nucleosome dynamics genome-wide, and we have introduced several different strategies to achieve this goal: (1) To measure relative levels of histone replacement across the genome, we have followed incorporation of the replication-independent histone variant, H3.3, which replaces replication-coupled H3 over the course of the cell cycle. (2) To map histone turnover kinetics directly we have developed a novel method based on metabolic labeling of proteins followed by affinity purification of newly synthesized histone core particles. (3) To map classical 'active' chromatin genome-wide we have applied salt fractionation to intact micrococcal nuclease-treated nuclei. (4) To effectively profile chromatin landscapes at single base-pair resolution, we have developed a simple sequencing library preparation protocol and data display method that we have applied to mapping transcription factors, nucleosome remodelers, nucleosomes and kinetochores from a single dataset. (5) To extend genome-wide chromatin profiling to tissues, we have introduced an affinity-based method for purification of nuclei expressing a nuclear envelope protein under control of a cell-type-specific promoter. Application of these methods to model organism genomes suggests that nucleosome turnover is crucial for epigenetic inheritance of gene activity and for maintaining a single centromere on a chromosome.

Biography

Steven Henikoff received a BS degree in Chemistry from the University of Chicago and a Ph.D. degree in Biochemistry and Molecular Biology from Harvard University, and carried out postdoctoral research at the University of Washington. He joined the Fred Hutchinson Cancer Research Center in Seattle in 1981, where he is a Member of the Basic Sciences Division and an Affiliate Professor of Genome Science at the University of Washington. He has been an Investigator of the Howard Hughes Medical Institute since 1990 and a Member of the US National Academy of Sciences since 2005. He is co-Editor-in-Chief of Epigenetics & Chromatin, a member of the Editorial Boards of Trends in Genetics, Current Opinion in Genetics and Development and Genome Biology, and a member of the Scientific Advisory Boards of Epizyme, Inc., the Center for Epigenomics of the Einstein School of Medicine and the Chicago Biomedical Consortium. His laboratory studies chromatin processes, epigenetic inheritance, centromere structure, function and evolution, and develops tools for epigenomics.

(2) On the interplay of protein physics and evolution in protein structure and function
Abstract

The intrinsic ability of protein structures to exhibit the geometric features required for molecular function in the absence of evolution is examined in the context of three systems: the reference set of real, single domain protein structures, a library of computationally generated, compact homopolypeptide, artificial structures with protein-like secondary structural elements, and a quasi-spherical random protein packed at the same density as proteins but lacking backbone secondary structure and hydrogen bonding. Without any evolutionary selection, the library of artificial structures has similar backbone hydrogen bonding, global shape, surface to volume ratio and statistically significant structural matches to real protein global structures. Moreover, these artificial structures have native like ligand binding cavities, and a tiny subset has interfacial geometries consistent with native-like protein-protein interactions and DNA binding. In contrast, the quasi-spherical random proteins, being devoid of secondary structure, have a lower surface to volume ratio and lack ligand binding pockets and intermolecular interaction interfaces. Surprisingly, these quasi-spherical random proteins exhibit protein like distributions of virtual bond angles and almost all have a statistically significant structural match to real protein structures. This implies that it is local chain stiffness, even without backbone hydrogen bonding, and compactness that give rise to the likely completeness of the library solved single domain protein structures. These studies also suggest that the packing of secondary structural elements generates the requisite geometry for intermolecular binding. Thus, backbone hydrogen bonding plays an important role not only in protein structure but also in protein function. Such ability to bind biological molecules is an inherent feature of protein structure; if combined with appropriate protein sequences, it could provide the non-zero background probability for low-level function that evolution requires for selection to occur.

Biography

Jeffrey Skolnick is the Director of the Center for the Study of Systems Biology in the School of Biology at the Georgia Institute of Technology and the Georgia Research Alliance Eminent Scholar in Computational Systems Biology. He attended graduate school in Chemistry at Yale University, receiving a Ph.D. in Chemistry in polymer statistical mechanics. He then held a postdoctoral fellowship at Bell Laboratories. Next, he joined the faculty of the Chemistry Department at Louisiana State University, Baton Rouge. Then, he moved to Washington University, where he was subsequently appointed Professor of Chemistry. There he was also Director of the Institute of Macromolecular Chemistry at Washington University. He joined the Department of Molecular Biology of the Scripps Research Institute, where he held the rank of Professor. Among his awards is an Alfred P. Sloan Research Fellowship and he is a Fellow of the American Association for the Advancement of Science, a Fellow of the Biophysical Society, a Fellow of the St. Louis Academy of Science. Recently, he moved to the Georgia Institute of Technology. He is the author of over 325 publications and has served on numerous editorial boards including Biophysical Journal, Biopolymers, Proteins, and the Journal of Chemical Physics. He is also a cofounder of an early stage structural proteomics company, GeneFormatics, and his software has been commercialized by Tripos. His research is in the area of Computational Systems Biology and has focused on the development of algorithms and their application to proteomes for the prediction of protein structure and function, the prediction of small molecule ligand-protein interactions with applications to drug discovery and the prediction of off-target uses of existing drugs, fundamental studies on the nature and completeness of protein structure space and the exploration of the interplay between protein physics and evolution in determining protein structure and function. He has also developed successful approaches for the prediction of protein-protein and protein-DNA interactions as well as a novel approach to cancer
metabolomics. Most recently, he has undertaken molecular based simulations designed to explore the basic physical principles underlying molecular motion within a cell.

(3) Integrative Multi-Scale Biomedical Informatics

Joel Saltz
Director, Center for Comprehensive Informatics, Emory University
Psychology Building 36 Eagle Row, Atlanta, GA, 30322

Abstract

Development of biomarkers that predict response to treatment and models that can direct development of new therapies requires integration of many complementary types of biomedical information captured at multiple scales. We are developing methodologies, information models, tools, and analytic pipelines that will make it feasible to systematically carry out large-scale integrative analyses of: 1) whole slide digital pathology and radiology based features, and 2) deep-sequencing data and patterns of protein and gene expression. The methods and tools will be designed to carry out the following closely interrelated tasks: 1) systematically manage, query and analyze results produced by data analyses composed of large numbers of interrelated algorithms, 2) compare results produced by workflows consisting of cascades of multiple algorithms, 3) efficiently manage result datasets that in aggregate will contain trillions of imaging derived features, 4) engage human neuropathologists and radiologists in validation of results and motivation of new analyses, and 5) support histological feature query and analysis patterns needed to link histological features with “omic” and outcome data.

Biography

Dr. Joel H. Saltz is Director of the Center for Comprehensive Informatics, Chair and Professor of Emory's Department of Biomedical Informatics, Professor in the Departments of Pathology, Biostatistics and Bioinformatics, and Mathematics and Computer Science at Emory University, Adjunct Professor of Computational Science and Engineering at Georgia Tech, Georgia Research Alliance Eminent Scholar in Biomedical Informatics, and Georgia Cancer Coalition Distinguished Cancer Scholar. Prior to joining Emory, Dr. Saltz was Professor and Chair of the Department of Biomedical Informatics at The Ohio State University (OSU) and Davis Endowed Chair of Cancer at OSU. He served on the faculty of Johns Hopkins Medical School, University of Maryland College Park and Yale University in departments of Pathology and Computer Science. He received his MD and PhD (computer science) degrees at Duke University and is a board certified Clinical Pathologist trained at Johns Hopkins University.

(4) The impact of Informatics and Information Management in MedImmune R&D

Mathew Woodwark
Lead, R&D informatics and Program Director of Knowledge and Information Management, MedImmune
Abstract
MedImmune, a subsidiary of AstraZeneca, manages a portfolio of over 150 Biologics drug projects across several major R&D sites worldwide. Biologics projects differ from small molecule projects in a number of ways, requiring different approaches to data and information management. In this talk, I will provide a review of how informatics and information management have been applied to support decision making in drug projects, using examples from my experience in both AZ and MedImmune.

Biography
With a PhD in Theoretical Population Genetics, Mathew became one of a very small number of Bioinformaticians in the UK in the early 1990s. After supporting the nascent use of sequence information at a network of Agricultural Research Institutes for several years, he went on to form part of the newly formed Bioinformatics group at Zeneca in 1996. Over a 9 year period at what became AstraZeneca, he came to lead a group responsible for the main "Omics" analysis platforms in the company, including Gene Catalogue, AZ's view of the Human Genome. Another strong focus at AZ was to foster a learning culture by introducing the concept of Training, Awareness and Utilisation (TAU) for informatics applications. In 2005, he moved to Cambridge Antibody Technology. Within a year he was working for AZ again, after their acquisition of CAT. In Cambridge, Mathew took on the role of Director of Knowledge and Information Management (KIM) with an emphasis on facilitating the creation, sharing, storage, use and critically re-use of drug project information. He held this role for 6 years, changing how MedImmune collaborates across the world in drug projects and functions. Very recently, he has come full circle, with a new role as Lead of R&D Informatics for MedImmune. This provides the opportunity to integrate data management, information management and knowledge management for drug project decision support in MedImmune.

Invited Talks

(1) An In silico Approach to Target Discovery

James Cai
Hoffmann-La Roche in Nutley, New Jersey, USA

Abstract
Target discovery relies heavily on the intelligent use of scientific information. Recently high dimensional genomics data have become increasingly critical in discovering new targets and biomarkers. However it is difficult to combine findings from the large number of genomics studies and integrate across multiple data types. This presentation will describe a new in silico target discovery approach based on flexible pipelining and quantitative data integration using a meta-analysis framework. Examples will be given to show how results from multiple gene expression studies can be combined with mutation, disease association and competitor information to prioritize potential targets in oncology.

Biography
Dr. James Cai is the Head of Disease and Translational Informatics at Hoffmann-La Roche in Nutley, New Jersey, USA. His group develops scientific informatics solutions to support drug discovery in the Disease and Translational Areas (e.g., Oncology, Inflammation and Virology) at Roche. Dr. Cai received his Ph.D. in Molecular Biology from Cornell University and M.A. in Biomedical Informatics from Columbia University. He received his postdoc training in Biomedical Informatics at Columbia University through a fellowship provided by the National Library of Medicine. During the past 10 years at Roche, he has worked in various informatics areas including genomic data analysis, application development, knowledge management, and text analytics. He performed bioinformatics data analyses in several projects that have advanced into clinical trials. He and his team are responsible for the development of a number of scientific applications widely used at Roche.

(2) Phylogenetic analysis: the Tree of Life, whole genomes, and beyond.

Bernard Moret
Swiss Federal Institute of Technology in Lausanne, Switzerland

Abstract

The rapidly increasing number of sequenced genomes offers the chance to resolve longstanding questions about the evolutionary history of certain groups of organisms, to develop a better understanding of evolution, to make substantial advances in functional genomics, and to start bridging genomics and genetics. We are thus witnessing a significant increase in phylogenetic analysis using larger-scale features such as genomic rearrangements, duplications and losses of genomic regions, regulatory modules and networks, chromatin structure, etc. However, not only are the computational problems arising in such analyses much harder than those arising in sequence-based analyses, but we lack the vast collective experience and the detailed evolutionary models used in sequence-based analyses. In this talk, I will review the state of the art in phylogenetic analysis of whole genomes, explain what the main challenges are today, and describe recent progress, both in my group and elsewhere, that promises to resolve some of these challenges.

Biography

Bernard Moret is Professor of Computer Science, holding the chair of Bioinformatics, at the EPFL, the Swiss Federal Institute of Technology in Lausanne, Switzerland. He received his PhD in 1980 from the U. of Tennessee and was on the faculty of the Department of Computer Science at the U. of New Mexico until 2006, serving as chairman from 1991 till 1993. His research interests are in the area of algorithms and applications, particularly in computational molecular biology. He founded the ACM Journal of Experimental Algorithmics in 1995, serving as its editor-in-chief for 7 years. Since 2000, he has focused on the development of models and algorithms for evolutionary genomics, publishing around 80 peer-reviewed articles in the area and founding, in 2001, the annual Workshop on Algorithms in Bioinformatics (WABI).
(3) Analysis and Integration of Inconsistent and Unreliable Biomedical Prediction Models

Zoran Obradovic,  
Director, Data Analytics and Biomedical Informatics Center, Temple University

Abstract

In biomedical applications, multiple predictors are often developed for the same problem using multiple training datasets of various qualities. Selecting a single model from such a collection based on accuracy evaluation on a small biased set of annotated data is not necessarily the best strategy when the objective is large scale application of the model. In this talk we will discuss how to address this problem by uncertainty analysis in the reference models and in data. In addition, we will present an iterative algorithm for integrating predictions of multiple models without relying on any annotated data. The proposed solutions will be illustrated on the problem of predicting intrinsic disorder in proteins that lack a stable tertiary structure but still have important biological functions.

Biography

Zoran Obradovic, professor of Computer and Information Sciences and the director of the Data Analytics and Biomedical Informatics Center at Temple University in Philadelphia is an internationally recognized leader in data mining and bioinformatics. He has published about 240 articles addressing data mining challenges in health informatics, climate and ecological management, the social sciences, and other domains. Obradovic was the program chair at 6, track chair at 10 and program committee member at about 40 international conferences on data mining. He currently serves as an editorial board member at 8 journals and is the executive editor at the journal on Statistical Analysis and Data Mining which is the official publication of the American Statistical Association (ASA).

(4) Statistical Challenges of multiple SNP analysis in Genome-wide Association Studies

Taesung Park  
Seoul National University, Seoul, Korea

Abstract

In recent years, genome-wide association (GWA) studies have successfully led to many discoveries of genetic variants affecting common complex traits, including height, blood pressure, and diabetes. Although GWA studies have made much progress in finding single nucleotide polymorphisms (SNPs) associated with many complex traits, such SNPs have been shown to explain only a very small proportion of the underlying genetic variance of complex traits. This is partly due to that fact that most current GWA studies have relied on single-marker approaches that identify single genetic factors individually and have limitations in considering the joint effects of multiple genetic factors on complex traits. Joint identification of multiple genetic factors would be more powerful and provide a better prediction of complex traits, since it utilizes combined information across variants. In this talk, we overview some
statistical methods for the joint identification of genetic variants for common complex traits and discuss challenges in application of GWA studies. We applied joint identification approaches to a large-scale GWA dataset (i.e., 8842 samples and 327,872 SNPs) in order to identify genetic variants of several phenotypes of interest for the Korean population. In addition, in order to test for the biological significance of the jointly identified SNPs, gene ontology and pathway enrichment analyses were further conducted.

Biography

Taesung Park received his M.S. degree in Statistics from Seoul National University, Korea in 1986, and received his Ph.D. degree in Biostatistics from the School of Public Health in the University of Michigan in 1990. From August 1991 to August 1992, he worked as a visiting research scientist at the National Institutes of Health, USA. From September 2002 to August 2003, he was a Visiting Professor in Department of Biostatistics at the University of Pittsburgh, USA. He served as the Chair of the Bioinformatics Program, Seoul National University (SNU) from April 2005 to March 2008 and the Chair of Department of Statistics of SNU from September 2007 and August 2009. He is currently the Professor and Director for the National Research Laboratory of Bioinformatics and Biostatistics at the Department of Statistics, SNU, Korea. His research areas include DNA microarray data analysis, genetic network, gene-gene interactions, and genome-wide association studies.

(5) Mathematical modeling of regulatory networks based on omics datasets

Tianhai Tian, 
Monash University, 
Australia

Abstract

Systems biology is the science of discovering, modeling, understanding and ultimately engineering the dynamic relationships between the biological molecules that define living organisms at the molecular level. There are two major approaches in systems biology. Top-down systems biology identifies molecular interaction networks on the basis of correlated molecular behavior observed in genome-wide ‘omics’ studies. Bottom-up systems biology examines the mechanisms through which functional properties arise in the interactions of known components. In this talk I will use two biological systems to demonstrate the great potential of mathematical models in systems biology. A nonlinear model will be used to reverse-engineering the dynamic regulation of p53 protein by using the microarray gene expression data and other ‘omics’ datasets. The second mathematical model is designed to analyze the signal transmission capacity of the mitogen-activated protein (MAP) kinase pathway.

Biography

Tianhai Tian received his PhD in Computational Mathematics in 2001 from the University of Queensland in Australia. He was a Research Fellow at the same University after submitting his PhD thesis. In 2003 he began his research in computational biology and bioinformatics. He obtained the Australian Research Fellowship from the Australian Research Council (ARC) in 2006 and then studied systems biology at the Institute for Molecular Bioscience in Queensland. He joined the University of Glasgow in Scotland as a
Lord Kelvin Fellow in 2007 and became a Reader in 2009. In 2011 he returned to Australia and now is an Associate Professor and ARC Future Fellow at the School of Mathematical Sciences, Monash University. His research interests include stochastic modeling and simulation of gene networks, mathematical modeling of cell signaling pathways, reverse-engineering of regulatory networks from omics datasets, as well as cancer therapy and drug resistance.

(6) The Third-generation Sequencing and the Upcoming Challenges in Bioinformatics

Stephen Kwok-Wing Tsui,
Chinese University of Hong Kong

Abstract

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. In the near future, the dream of USD1000 per genome will be achieved and affordable third-generation sequencers will be appeared in uncountable academic and diagnostic laboratories, or even hospitals. We anticipate that many problems will be encountered when genomics platforms migrate from the bench to the bed. In this talk, the future development and applications of DNA sequencing technology will be introduced. Also, the upcoming challenges in Bioinformatics will be presented and discussed.

Biography

Stephen Kwok-Wing TSUI is currently a professor in the School of Biomedical Sciences, Head of the Division of Genomics and Bioinformatics, Director of Hong Kong Bioinformatics Centre, Director of Microbial Genomics and Proteomics and the Associate Director of the newly formed CUHK-BGI Innovative Centre for Trans-omics in the Chinese University of Hong Kong. In the last 15 years, his team had named and characterized more than ten novel human genes. He was also a former member of the International HapMap Consortium. 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 110 scientific papers in international journals, including Nature, New England Journal of Medicine, Lancet, PNAS, Nucleic Acids Research, Bioinformatics and BMC Bioinformatics. 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.