Computational Approaches to the Analysis of Biomolecular Sequences, Structures and Their Functions and Applications to Biotechnology and Clinical Data Studies (23 - 25 Mar 2020)

Organizing Commettiee

Co-chairs

Igor N. Berezovsky (Bioinformatics Institute, A*STAR)
Lars Nordenskiöld (Nanyan Technological University)
Frank Eisenhaber (Bioinformatics Institute, A*STAR)

Members

Zhang Louxin (National University of Singapore)
Choi Kwok Pui (National University of Singapore)

- **Day 1. Monday, March 23, 2020**
  Workshop and Tutorial on Sequence-Based Analysis of Biomolecules and Molecular Modelling for Therapeutics

  Morning Session 9:30 – 12:30
  Chair: Frank Eisenhaber

  9:30 – 10:30

  Frank Eisenhaber, Bioinformatics Institute, A*STAR, Singapore

  Welcome by Organisers *and*

  Talk "About the limited potential, yet instrumental role of computational biology and the target identification bottleneck due to the decline in biomolecular mechanism discovery after 2000"

  ABSTRACT
  Computational biology is all about studying biomolecular sequences, structures and biological images as well as composite data types such as clinical data and the understanding of mechanisms that lead determine the phenotype for a given genome. The discovery of sequence homology was the prime reason for pushing bioinformatics into the center stage as it connects various areas of life science research via common biomolecular mechanisms. Although there is still no coherent biological theory, nevertheless, computational biology became the main driver for hypothesis generation in life sciences.

  It is generally believed that full human genome sequencing was a watershed event in human history that boosted biomedical research, biomolecular mechanism discovery and life science applications. At the same time, researchers in the field of genome annotation see that there is a persisting, substantial body of functionally insufficiently or completely not characterized genes (for example, ~10,000 protein-coding in the human genome) despite the availability of full genome sequences. A survey of the biomedical literature shows that the number of reported new protein functions had been steadily growing until 2000 but the trend reversed to a dramatic decline thereafter (1,2) when, at the same time, the annual amount of new life science publications doubled between 2000 and 2017.

  This reduction in the supply of newly characterized pathways has profound implications for the drug development pipelines in industry as well as for research. The example of the SUGCT function discovery (3) shows the significance that many uncharacterized genes will have in aging, metabolic diseases and their complications.

  1) Darkness in the human gene and protein function space: Widely modest or absent illumination by the life science literature and the trend for fewer protein function discoveries since 2000. Sinha S, Eisenhaber B, Jensen LJ, Kalbuaji B,

Eisenhaber F. Proteomics. 2018 Nov;18(21-22):e1800093. doi: 10.1002/pmic.201800093
2) A decade after the first full human genome sequencing: when will we understand our own genome? Eisenhaber F. J Bioinform Comput Biol. 2012 Oct;10(5):1271001. doi: 10.1142/S0219720012710011
3) Knockout of the non-essential gene SUGCT creates diet-linked, age-related microbiome disbalance with a diabetes-like metabolic syndrome phenotype. Niska-Blakie J, Gopinathan L, Low KN, Kien YL, Goh CMF, Caldez MJ, Pfeiffenberger E, Jones OS, Ong CB, Kurochkin IV, Coppola V, Tessarollo L, Choi H, Kanagasundaram Y, Eisenhaber F, Maurer-Stroh S, Kaldis P. Cell Mol Life Sci. 2019 Nov 13. doi: 10.1007/s00018-019-03359-z

11:00 – 11:45

Samuel Gan, Bioinformatics Institute, A*STAR, Singapore

Selected short talk "Antibody engineering, Scientific Phone Apps, Viral research, and Device prototyping in APD Lab"

ABSTRACT

Many significant biomedical discoveries of old were made in the private property of famous scientists e.g. Leeuwenhoek and Archimedes. Today, discoveries are made in brightly-lit, hi-tech, ergonomic buildings that house research institutes. While such development is advantageous in many aspects, the spatial restriction of research into well-organized structures may delay and limit the spontaneity necessary for discoveries. The smartphone and peripheral mobile devices have the potential to not only increase the productivity and mobility of biomedical research, but also restore some freedom from spatial constraints. One possible way this can occur is the development of a mobile biomedical lab that allows researchers to carry out core research processes 'on-the-go' without being spatially restrained within a building or availability of equipment. This talk introduces the world of Scientific Phone Apps and device prototyping in APD Lab, with a sneak preview into the unique antibody engineering and viral research by the lab.

11:45 – 12:00

Chinh Trah To Su, Bioinformatics Institute, A*STAR, Singapore

Selected short talk "A holistic view of proteins: impact on antibody engineering and drug discovery"

ABSTRACT

The reductionist approach is prevalent in biomedical science. However, increasing evidence now shows that biological systems cannot be simply considered as the sum of its parts. With experimental, technological and computational advances, we can now do more than view parts in isolation, thus we propose that an increasing holistic view (where a protein is investigated as much as a whole as possible) is now timely. To further advocate this, we investigate and discuss several studies and applications involving allostery, where distant protein regions can cross-talk to influence functionality. As a result, we believe that an increasing big picture approach holds great promise, particularly in the areas of antibody engineering and drug discovery in rational drug design.

Afternoon Session 13:30 – 16:00
Chair: Frank Eisenhaber

13:30 – 14:00
Melvin Yin, Bioinformatics Institute, A*STAR, Singapore
*Selected short talk "*Deriving structure profiles corresponding to functional loops in proteins*"*

ABSTRACT
Functional segments in proteins are highly conserved across function-based families in both sequence and structure. We use sequence alignment and motif search techniques to automatically identify such segments in labelled proteins, and consolidate them into generalised structure profiles. These are manually verified to correspond to known functional loops. From the profiles, we build descriptors that contain the necessary structural information to replicate the environment that lead to the original protein performing the corresponding chemical reaction. These can be used in identification of current or previous functions in unlabelled proteins through a structure match scoring function, or in de novo protein design.

14:00 – 14:30

Swati Sinha, Bioinformatics Institute, A*STAR, Singapore

Selected short talk "Discovery of an Antifungal Compound BII-Rafflesfungin: A Computational Perspective"

ABSTRACT

Phomafungin is a recently reported broad spectrum antifungal compound but its biosynthetic pathway is unknown. We combed publicly available Phoma genomes but failed to find any putative biosynthetic gene cluster that could account for its biosynthesis. Therefore, we sequenced the genome of one of our Phoma strains (F3723) previously identified as having antifungal activity in a high-throughput screen. We found a biosynthetic gene cluster that was predicted to synthesize a cyclic lipodepsipeptide that differs in the amino acid composition compared to Phomafungin. Antifungal activity guided isolation yielded a new compound, BII-Rafflesfungin, the structure of which was determined. We describe the NRPS-t1PKS cluster 'BIIRfg' compatible with the synthesis of the cyclic lipodepsipeptide BII-Rafflesfungin [HMHDA-L-Ala-L-Glu-L-Asn-L-Ser-L-Ser-D-Ser-D-allo-Thr-Gly]. We propose a mechanism for BII-Rafflesfungin biosynthesis, which involves the formation of the lipid part by BIIRfg_PKS followed by activation and transfer of the lipid chain by a predicted AMP-ligase on to the first PCP domain of the BIIRfg_NRPS gene.

15:00 – 16:00

Marek Mutwil, School of Biological Sciences, Nanyang Technological University, Singapore

Talk "Evolutionary lessons learned from transcriptome data"

ABSTRACT

Marek Mutwil, SBS, NTU, Singapore

Evolutionary lessons learned from transcriptome data

To understand the evolution of various organs and metabolic pathways, it is necessary to understand the functions of the underlying gene products. Classical genomic approaches based solely on gene sequences are useful, but have shortcomings, as they cannot readily reveal which genes work together in a pathway, i.e., form a functional gene module. Consequently, to study the evolution of new traits, we need to integrate the classical genomic approaches with predicted functional gene modules, which can be identified by studying, for example, gene expression, protein-protein interactions or co-expression networks.

I discuss how a variety of transcriptomic approaches can be used to answer evolutionary questions that cannot be easily addressed with genomics. More specifically, I will present how transcriptomes can be used to identify essential genes, discover metabolic pathways and study the evolution of new cell wall types and organs. I will also describe the existence of a gene expression program that is conserved over 1,500,000,000 years of evolution.

- **Day 2. Tuesday, March 24, 2020**
  Workshop and Tutorial on the Protein and Chromatin Structure and Function

  Morning Session 9:30 – 12:00
  Chair: Igor N. Berezovsky

  9:30 – 10:30

  Konstantin Pervushin, School of Biological Sciences, Nanyang Technological University, Singapore
  Talk "Abeta chaperones in Alzheimer's disease: friends or foes?"

  ABSTRACT

11:00 – 12:00

Zhen Wah Tan, Bioinformatics Institute, A*STAR, Singapore

Tutorial "Modeling allostery: global protein structural changes in response to local perturbations"

ABSTRACT

Characterizing the global structural effects of allosteric binding is a challenging task, often requiring extensive molecular dynamics simulations to reveal the change in the structural rigidity of different regions. To enable rapid screening of multiple potential allosteric sites, we have developed a structure-based statistical mechanical model that allows for quick estimates of changes in structural dynamics of ligand-bound or mutated proteins. In this talk, we will introduce two web resources we have developed: First, the AlloSigMA server (http://allosigma.bii.a-star.edu.sg) allows users to visualize and explore model predictions for user-defined protein structures, and screen for new --- or latent --- allosteric sites. Second, the AlloMAPS database (http://allomaps.bii.a-star.edu.sg) contains pre-computed data for a collection of (i) classical allosteric proteins, (ii) proteins associated with pathological SNPs, and (iii) protein chains representing a wide variety of structural folds.

Afternoon Session 13:30 – 16:00
Chair: Igor N. Berezovsky

13:30 – 14:30

Amartya Sanyal, School of Biological Sciences, Nanyang Technological University, Singapore
Talk "3D genome: from organization to function"

ABSTRACT
All genomic processes take place in the context of chromatin, the organization of genomic DNA with histones and hundreds of other proteins and RNAs inside the nucleus. The higher order folding of chromatin controls transcription and other nuclear processes, and is important for cellular identity and cell state transitions. At the beginning of the talk, I will provide a brief overview about the organization of higher-order structures of chromatin and their detection using chromosome conformation capture (3C)-based methods. Next, I will narrate our latest work on cancer drug resistance that involves interplay between genetic and epigenetic mechanisms, leading to transcriptional reprogramming. At the end, I will explain a simplified low-input 3C-seq protocol that we developed to interrogate chromatin organization in patient samples.

15:00 – 16:00

Zhen Wah Tan, Bioinformatics Institute, A*STAR, Singapore

Tutorial "Capturing communities in diffusive chromatin polymers via Markov State Modeling and Transition Path Theory"

ABSTRACT

Chromatin is a complex structure composed of densely compacted DNA and a myriad of molecular factors. Biological processes within chromatin remain highly organized despite spatial crowding of the system, which is believed to be due to the organization of chromatin into community structures. Capturing and understanding these community structures is thus the first step in studying how chromatin structure relates to its function and regulation. To achieve this, we modelled data from high-throughput chromosome conformation capture (Hi-C) using a Markov State Model defined on the interaction network. Various aspects of Transition Path Theory were used to develop a concrete definition of structural partitions and their interactions in chromatin, which enabled us to detect the structural hierarchy in chromatin organization, and characterize whole-genome structure for further analysis.