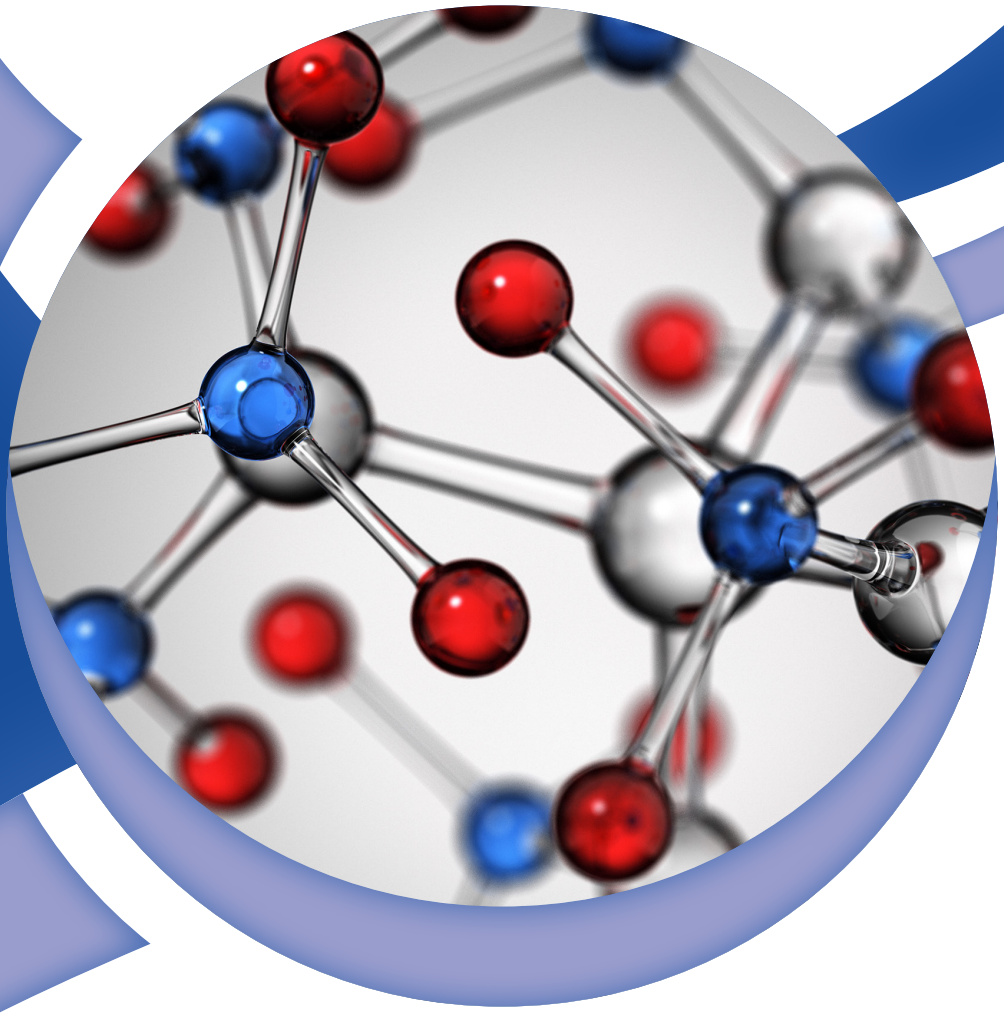


International Conference on

Computational Biology and Bioinformatics

September 05-06, 2018 Tokyo, Japan

Posters

Chun-Yip Tong et al., J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

CMEP: A database for circulating microRNA expression profiling

Chun-Yip Tong and Jian-Rong Li
National Chung Hsing University, Taiwan

In recent years, several experimental studies have revealed that the microRNAs (miRNAs) in serum, plasma, exosome and whole blood are dysregulated in various types of diseases, indicating that the circulating miRNAs may serve potential non-invasive biomarkers for disease diagnosis and prognosis. However, there is no database constructed to integrate the large-scale circulating miRNAs profiles, explore the functional pathway they involved and predict the potential biomarkers using feature selection between disease conditions. The Circulating MicroRNA Expression Profiling (CMEP) is a database for integrating, analyzing and visualizing the large-scale expression profiles of phenotype-specific circulating miRNAs. Although there have been several studies attempting to generate circulating miRNA database, they have not yet integrated the large-scale circulating miRNAs profiles and provided the biomarker-selection function using machine learning methods. To fill in this gap, we constructed the CMEP database for integrating, analyzing and visualizing the large-scale expression profiles of phenotype-specific circulating miRNAs. The CMEP database contains massive datasets manually curated from NCBI GEO including 61 datasets, 192 subsets and 9,444 samples. The CMEP provides the differential expression circulating miRNAs analysis and the KEGG functional pathway enrichment analysis. Furthermore, to provide the function of non-invasive biomarker discovery, implementation of several feature-selection methods including ridge regression, lasso regression, support vector machine and random forests. Finally, a user-friendly web interface was implemented to improve the user experience and visualize the data and results of CMEP.

Biography

Chun-Yip Tong is currently a Master's student at National Chung Hsing University, Institute of Genomic and Bioinformatics and currently researching on circulating micro RNA non-invasiave diagnosis

Li Jian Rong is currently a postdoctoral fellow in the institute of Genomic and Bioinformatics, National Chung Hsing University, and current research topics are RNA-Seq of Oncidium and circulating micro RNA for non-invasiave diagnosis.

imanthony1993@hotmail.com.tw, fanicesiza@gmail.com

Notes:

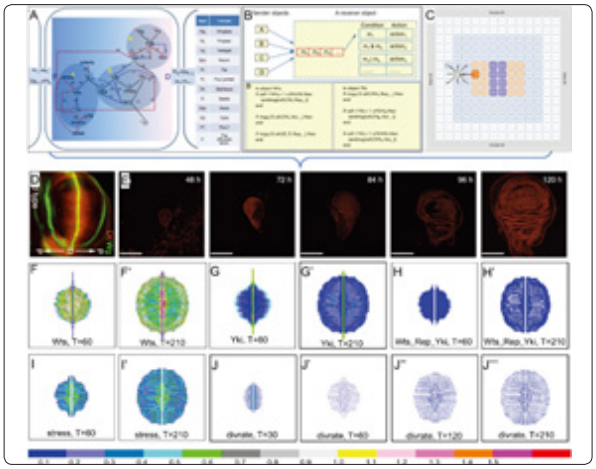
conferenceseries.com

Hao Zhu, J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

Reconstructed signaling events unveil how growth and growth arrest are controlled by PCP and Hippo signaling

Hao Zhu
Southern Medical University, China

How growth is controlled in normal tissues and organs but not in cancers is a question that has drawn wide attention but remained poorly understood. To uncover feedbacks and their properties in control mechanisms consisting of genes and their products, quantitative methods are needed but not enough, because gene expression and protein interaction are discrete events. A new model combining features of lattice models and vertex models and integrating differential equations, molecular signaling and mechanical force is developed to investigate growth and growth arrest of *Drosophila* wing. The model includes key elements in the Wnt, PCP and Hippo pathways, encapsulates proteins and their attributes and behaviors into objects, uses message-passing between objects to simulate signaling between proteins, simulates cell divisions in a 2D lattice, computes protein concentrations and cell polarity, captures the spatiotemporal distribution of signaling events and computes the spatiotemporal distribution of mechanical stress. The distributions of protein concentrations, cell polarity, cell division rates, and cell population agree well with experimental observations, justifying the unveiled control mechanism. Reconstructed signaling events uncover two intercellular feedbacks that jointly control growth, patterning and growth arrest. Specifically, the results indicate that the Warts repressing Yorkie event (*Wts_Rep_Yki*) distributes densely at the central in early stages but densely at the periphery in later stages. The spatiotemporal distribution of this critical event provides the unprecedented and most pertinent evidence suggesting that growth is gradually refrained from the periphery to the central of the wing pouch, which is a new and somewhat counter intuitive finding. The methods can be applied widely to systematically investigating signaling and patterning across the gene, molecular, cell and tissue levels.



Structure and results of the wing pouch growth model. (A) The model contains key elements in the Wnt, PCP and Hippo pathways. (B) Proteins are encapsulated into objects and signaling between proteins is captured as message-passing between objects. (C) In a 2D lattice each unit contains a set of PDE/ODE and represents a wing cell. Initial cells (the purple and pink ones) are at the central, new cells (the blue ones) generated by cell division make the cell population grow. (D) The wild-type of Wg and Dll distributions (note the two outer green ribbons are not wing pouch). (E) The wild-type of growth process. (FG) Wts and Yki concentrations in the lattice at early and late stages (T indicates non-dimensionalized time). (HI) The Wts_Rep_Yki event and cell stress in the lattice at early and late stages. (J) Cell division rates in cells decrease (the blue color darkens) gradually, making growth finally arrested. The color bar indicates non-dimensional protein concentrations.

Biography

Hao Zhu has obtained his MS in Computer Science from National University of Defense Technology, Changsha, China and PhD in Pathophysiology from Southern Medical University, Guangzhou, China. He has completed his Postdoctoral studies in Bioinformatics Institute of Singapore and School of Mathematical Sciences, University of Nottingham, UK. With a profound interest in Evo-Devo, currently he is working on mechanisms that make new genes (especially lncRNA genes) function coordinately with old ones to regulate gene expression and control developmental signaling and patterning. He developed a cellular automata style modeling tool to effectively address signaling events and lncRNA analysis tool/platform (Long Target) to analyze the evolution and function of lncRNAs.

zhuhao@smu.edu.cn

Yang-Zhan Huang et al., J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

Identification of the septic pathogen in the plasma by next generation sequencing

Yang-Zhan Huang[1], Li-Ching Hsieh[1] and Wen-Cheng Chao[2]
[1]National Chung Hsing University, Taiwan
[2]Taichung Veterans General Hospital, Taiwan

Sepsis is a leading cause of death worldwide and the most common cause of death in patients who admitted to intensive care units. Sepsis is characterized by the life-threatening host immune response against invading microorganisms. However, the accurate diagnosis of the microorganism for sepsis remains difficult given that the yield rate of blood culture, which is currently the golden standard in sepsis, is merely 10-15%. Advances in genome sequencing technologies have led to a drastically decreased cost of Next-Generation Sequencing (NGS) and a wide range of applications, including identification of the circulating pathogenic nucleic acid in sepsis. Nowadays several methods have been developed to identify the pathogens through NGS approaches, but a number of biases exist in bioinformatic analyses. In the present study, we used the plasma of patients with sepsis at Taichung Veterans General Hospital (TCVGH) and aimed to identify circulating pathogenic nucleic acid through the NGS approach. We used pathogen sequences assignment directly from filtered NGS reads by k-mer and this approach is characterized by the fast-screening although high-memory is required. We found that pathogen-specific sequences may be identified in short time consume under the low survive reads after filtering host reads.

Biography

Yang-Zhan Huang is currently pursuing his Masters in Bioinformatics and Metagenomics at Institute of Genomics and Bioinformatics, National Chung Hsing University, Taiwan. He has worked for Applied Microbiology Laboratory and is interested in microorganism research.

h7106019005@smail.nchu.edu.tw

Notes:

Yuki Sasahara et al., J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

Mathematical exploration of selective pressures that shaped the metabolic zonation of liver nitrogen metabolism

Yuki Sasahara, Yasuhiro Naito and Masaru Tomita
Keio University, Japan

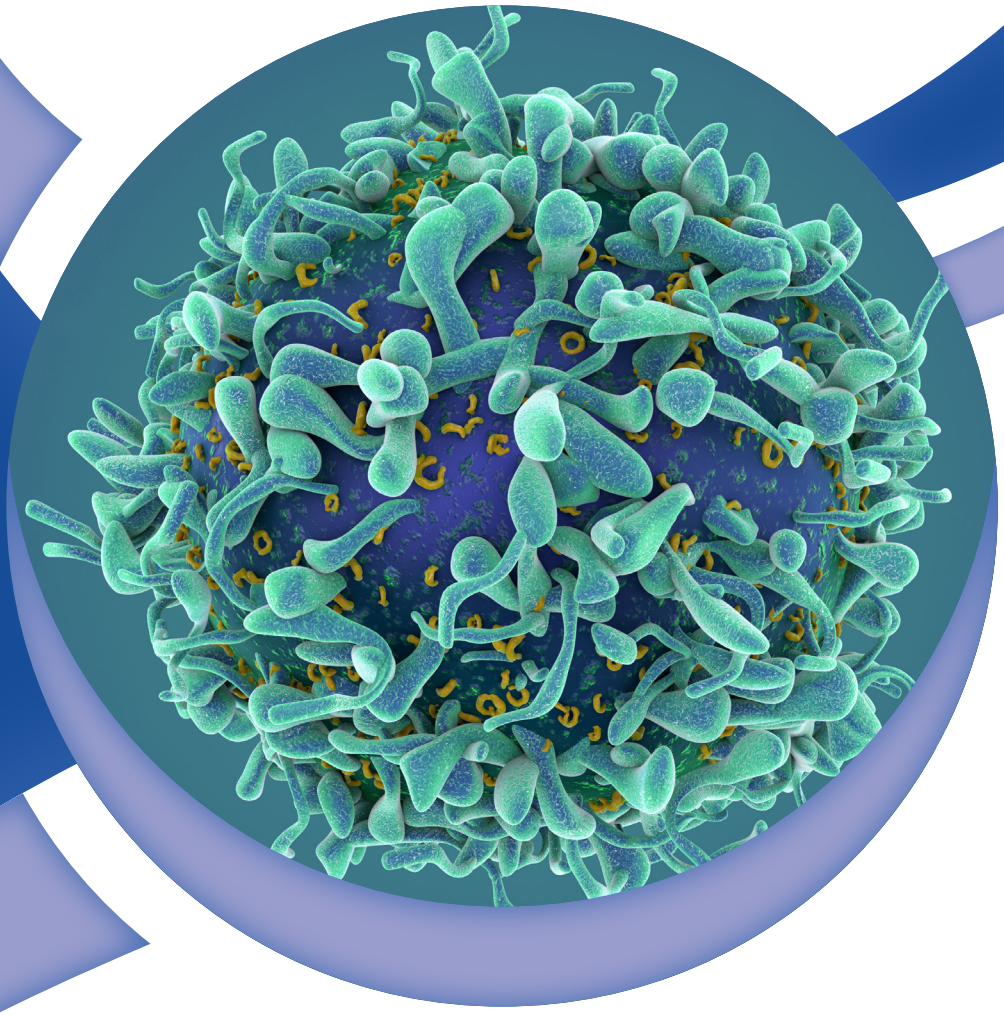
As ammonia which is one of the simplest nitrogen compounds is toxic for the central nervous system, its blood concentration should be maintained at low level. Human excrete most of nitrogen as urea in urine and urea synthesis exclusively occurs in liver. Ammonia is an inescapable metabolic intermediate during urea synthesis from various nitrogen compounds in hepatocytes. Most of ammonia is produced in hepatocytes and many of it is converted into urea and the rest is converted into reusable glutamine. The human liver is super-parallel metabolic filter with approximately 500,000 hepatic lobules that consist of approximately 500,000 hepatocytes. Blood flows into a hepatic lobule from fine branches of hepatic artery and portal vein and goes out from central vein. While upstream (periportal, PP) blood abundantly contains external molecules absorbed in the gastrointestinal tract, downstream (perivenous, PV) blood carries almost adjusted substances. Therefore, metabolic heterogeneity inevitably arises between PP and PV. Moreover, it is known that many enzymes heterogeneously express between PP and PV. Such heterogeneity is called metabolic zonation. For nitrogen metabolism, activity of urea cycle is dominant in PP and Glutamine Synthase (GS) activity is confined in PV. Recently, it is shown that hepatic GS deficient transgenic mice exhibit hyperammonemia and some organs other than liver affect its pathophysiology. In this study, we made the nitrogen homeostasis model of the whole body which incorporated metabolic zonation and tried to represent the systematic condition of nitrogen metabolism found in the GS deficient mice.

Biography

Yuki Sasahara is a undergraduate student of department of environment and information studies of Keio University. She is expected to earn B.A.(Environment and Information Studies) in Mar.2019. She graduated Ferris Girls' Senior High School and entered Keio University in 2015. She has joined the E-Cell project and majors in computational biology at the institute for advanced biosciences of Keio University.

t15441ys@sfc.keio.ac.jp

Notes:



International Conference on

Computational Biology and Bioinformatics

September 05-06, 2018 Tokyo, Japan

Accepted Abstracts

J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

ETC: A toolkit for converting phenotype descriptions into computable data

Hong Cui[1], Thomas Rodenhausen[1], Bertram Ludäscher[2], James Macklin[3] and Nico Franz[4]
[1]University of Arizona, USA
[2]University of Illinois at Urbana-Champaign, USA
[3]Agriculture and Agri-Food, Canada
[4]Arizona State University, USA

The explorer of taxon concepts project has produced a web application that consists a set of five tools unlocking phenotype data from text narratives often found as taxonomic descriptions or character descriptions. Aside from the tools described below, the site supports division of labor by allowing users share their tasks. Text capture tool parses textual taxonomic descriptions and marks up anatomical entities and characters. The tool is powered by CharaParser and MicroPIE (for microbial taxonomic descriptions). Ontology building tool enables experts without ontology knowledge to organize a set of phenotypic terms (e.g. those discovered by using is a part of or synonym relationships. Resulting ontology can be used in the other tools to improve data quality. Matrix generation tool takes the output and assembles a raw taxon-character matrix for the user to edit and refine, with or without ontology. Key generation tool employs a novel algorithm that directly takes a taxon-character matrix with polymorphic characters as input and computes the information entropy scores for each character. The result is a multi-access key that order the characters based on their discrimination power within the pool of the taxa. Taxonomy comparison tool supports the task of taxon concept analysis by supporting both expert asserted RCC-5 (congruent, narrower/broader than, disjoint and overlap) relationships among taxa, as well as providing character data related to the taxa in question to aid expert decisions. These five tools can be used to construct several pipelines that generate a taxon-by-character matrix, create a multi-access identification tool or facilitate taxon concept comparisons. A related software application, matrix converter is public available for users to convert a raw matrix to a scored matrix for phylogenetic analysis.

hongcui@email.arizona.edu

conferenceseries.com

J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

In silico study of repositioning of drugs against a candidate drug target implicated in type-2 diabetes

Prateek Kumar
Indian Institute of Technology Mandi, India

Diabetes is the 7th major cause of deaths throughout the world. In 2015, 415 million people were living with diabetes and Type-2 Diabetes (T2D) consists about 90% of cases. T2D is characterized by hyperglycemia and caused due to improper production of insulin. Few years back, the genetic architecture of T2D was not clear. After 2007, several high throughput studies such as Genome Wide Association (GWA) studies and Next Generation Sequencing (NGS) have been conducted on the different populations. These studies have confirmed the association of several genes with T2D. GWA studies have proved the association of gene *KCNJ11*, potassium inwardly rectifying channel subfamily J member 11, in T2D signaling pathway. *KCNJ11* regulates the insulin secretion in pancreatic beta cells by inhibiting ATP sensitive potassium channel. Several drugs are available for the treatment of T2D but either due to their improper binding or their stability in the target protein they cause side effects. The crystal structure of human *KCNJ11* has not been solved yet, so structure modeling of *KCNJ11* was performed using computational approaches. To identify the interaction of drugs targeting *KCNJ11*, *in silico* docking was performed and the binding effects of these drugs were analyzed by molecular dynamic simulation. This study may provide a valuable insight on further structure-based drug design approaches for *KCNJ11* for the treatment of type-2 diabetes.

kumar.prateek3@yahoo.com

conferenceseries.com

J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

A Summary-Statistics-based Random Effect Model (SSREM) to estimate heritability and co-heritability in GWAS data analysis

Jin Liu, Can Yang, Mingwei Dai, Xiang Wan, Chao Yang, Yingying Wei, Mengjie Chen and Xiang Zhou
Centre for Quantitative Medicine-Duke-NUS Medical School, Singapore

In the presence of individual-level data, Linear Mixed Model (LMM) is a commonly used powerful tool to conduct variance component analysis in Genome-Wide Association Studies (GWAS). Over the past few years, the methods to analyze summary statistics from GWAS become popular as (1) it is very difficult to fully access individual-level data, (2) the sample sizes are usually very large in meta-analysis across different studies, and (3) 1000 Genome Project data provides additional information to delineate linkage Disequilibrium (LD) over distinct populations. To maintain the estimation accuracy and efficiency of LMM, we proposed a unified approach, SSREM (Summary Statistics-based Random Effect Model), to explore genetic architecture of complex phenotypes using summary statistics, including both characterization of the overall genetic contribution to a phenotype (heritability estimation) and quantification of co-heritability, i.e. the genetic correlation between two phenotypes. The proposed method is based on the approximated likelihood of summary statistics uses samples from 1,000 genome project as the reference panel and implements an efficient parallel Gibbs sampling algorithm to ensure computational scalability for genome-wide data analysis. Results from empirical studies including both simulations and real data analysis (seven phenotypes from WTCCC and 25 more complex phenotypes) suggest that SSREM can be nearly as efficient as LMM that require the individual-level data and thus it outperforms other methods to estimate heritability and co-heritability using summary statistics. Majority of the findings are from real data analysis which is consistent with the results from previous studies.

jin.liu@duke-nus.edu.sg

J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

Genome Informatics: kmerHMM, SNPdryad and SignalSpider

Ka-Chun Wong
City University of Hong Kong, Hong Kong

There are three genome informatics methods are described here. The first kmerHMM is a pattern recognition method for discovering DNA motifs bound by proteins from Protein Binding Microarray (PBM) data. The novelty of kmerHMM lies in two aspects. First, it outperforms the existing methods in using Hidden Markov Models (HMMs) for modeling adjacent nucleotide dependency. Secondly, kmerHMM incorporates N-max-product algorithm and can derive multiple motifs. Comparisons of kmerHMM with other leading methods on several data sets demonstrated its effectiveness and uniqueness. Especially, it achieved the best performance on more than half of the data sets. In addition, the multiple binding modes derived by kmerHMM are biologically meaningful and will be useful in interpreting other genome-wide data. The second method named SNPdryad is a random forest method to predict the deleterious effect of non-synonymous SNPs on human proteins. It only includes protein orthologs in building a multiple sequence alignment. Among many other innovations, SNPdryad uses different conservation scoring schemes and uses Random Forest as a classifier. It has been demonstrated to have better performance than the existing methods (e.g. Harvard PolyPhen2 and JCVI SIFT) on well-studied datasets. It has been run on the complete human proteome, generating deleterious prediction scores for ALL possible non-synonymous SNPs in human. Lastly, the third method named SignalSpider will then be briefly introduced as a probabilistic graphical model for the integrative analysis of multiple ChIP-Seq (next generation sequencing) profiles from the ENCODE consortium. Comparing with similar existing methods, SignalSpider performs better in clustering promoter and enhancer regions. Notably, SignalSpider can learn higher-order combinatorial patterns from multiple ChIP-Seq profiles. The application of SignalSpider on the normalized ChIP-Seq profiles from the ENCODE consortium and learned model instances. We observed different higher-order enrichment and depletion patterns across sets of proteins. Those clustering patterns are supported by Gene Ontology (GO) enrichment, evolutionary conservation and chromatin interaction enrichment, offering biological insights for further focused studies. We also proposed a specific enrichment map visualization method to reveal the genome-wide transcription factor combinatorial patterns from the models built, which extend our existing fine-scale knowledge on gene regulation to a genome-wide level.

kc.w@cityu.edu.hk

J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

Developing a novel computational method for uncovering temporal correlation among chronic diseases using longitudinal medical records

Qun Feng Dong
Loyola University Chicago, USA

The availability of large-scale electronic medical records provides an unprecedented opportunity to investigate potential temporal correlations among different diseases using computational approaches. In this study, we present a novel computational method, implemented in R, which automatically builds retrospective matched cohorts from longitudinal electronic medical records to examine whether the occurrences of any diseases show significant temporal correlations using both cox proportional hazards regression and random forest survival analysis. In the method, time is correctly modeled as a continuous variable, accurately accounting for the temporal space between the onsets of different diseases. In addition, our method is flexible for incorporating relevant confounding factors such as age, gender and other demographic and medical information in the analysis. The output of our method is a disease correlation network, which is displayed using our web-based visualization tool, implemented in JavaScript, for users to interactively explore the correlations among diseases of interest based on statistical significance of the correlations and graph-theory-based network topology. We have successfully applied the method to a longitudinal electronic medical record dataset at Loyola with 10,832,319 distinct encounters detailing 92 diseases from 425,122 patients. Many uncovered disease correlations are strongly supported by in-depth literature reviews. Our computational package is freely available for download.

qdong@luc.edu

J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

Statistical machine learning in big data analytics

S Ejaz Ahmed
Brock University, Canada

Nowadays a large amount of data is available and the need for novel statistical strategies to analyze such data sets is pressing. This talk focuses on the development of statistical and computational strategies for a sparse regression model in the presence of mixed signals. The existing estimation methods have often ignored contributions from weak signals. However, many predictors altogether provide useful information for prediction, although the amount of such useful information in a single predictor might be modest. The search for such signals, sometimes called networks or pathways, is for instance an important topic for those working on personalized medicine. We discuss a new post selection shrinkage estimation strategy that considers the joint impact of both strong and weak signals to improve the prediction accuracy and opens pathways for further research in such scenarios.

sahmed5@brocku.ca

J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

A robust statistical method to integrate genotype and gene expression data identifies novel associated loci

Sarmistha Das and Indranil Mukhopadhyay
Indian Statistical Institute, India

Genome wide association studies identify many SNPs that are associated with disease traits. However, single marker test might miss SNPs with moderate effect. Moreover, gene expression contains information about the deregulation of genes when compared between cases and controls. Due to multiple testing or other issues some signals may remain unidentified especially when the sample size is not too large. Moreover, RNA being unstable than DNA, high cost is involved in RNA analysis leading to smaller sample for expression data than genotype data. No standard statistical procedure is available that integrates data from various sources to decode biologically sound interpretation on heritable traits. This motivates us to propose a novel method that tests for multi-loci association in the existing scenario. Based on a two-stage regression method our method essentially concatenates genome-wide expression data and disease-associated SNP data, when sample size for expression data is much smaller than genotype data. We integrated the information contained in both data sources into a latent variable based model. Our simple yet powerful multi-loci association test integrates two databases that broadcasts more of the deep-seated features comprehensively in a single test, which might be lost when datasets are considered in singularity. We also developed asymptotic distribution of our test statistic for fast calculation of p-value for real data set. Extensive simulation confirms that our method is powerful and robust to many genetic models. We have received promising result and identified few novel markers at genome-wide level even with a small gene expression dataset related to psoriasis.

sarmisthadascu@gmail.com

conferenceseries.com

J Proteomics Bioinform 2018, Volume 11
DOI: 10.4172/0974-276X-C1-113

Structural prediction and comparative studies of psychrophilic α-galactosidase from *Glaciozyma antarctica* against mesophilic and thermophilic enzymes

Shuhaila Mat-Sharani
National University of Malaysia, Malaysia

α-galactosidase is an essential enzyme that catalyses' the hydrolysis of α-1-6 linked terminal galactosyl residues from oligosaccharides, galacto-polysaccharides and glycol-conjugates. *Glaciozyma antarctica* is an obligate psychrophilic yeast with an optimal growth temperature of 12 °C and able to tolerate high temperature up to 20 °C. Since this yeast able to grow in an extremely cold temperature the enzyme must have special characteristics in adaptation to cold. The objectives of this study are to predict the structure and function of *G. antarctica* cold adapted α-galactosidase and to characterize its adaptation strategies towards the extremely cold environment based on homology modeling and molecular dynamics. It is commonly stated that there is a relationship between the flexibility of an enzyme and its catalytic activity at low temperature. Comparison between the structure of *G. antarctica* α-galactosidase enzymes with their homologs from the mesophilic and thermophilic fungus showed that *G. antarctica* α-galactosidase enzyme is more flexible because it has more loop structure with a lower number of hydrogen and disulfide bonds.

shuhaila129@gmail.com