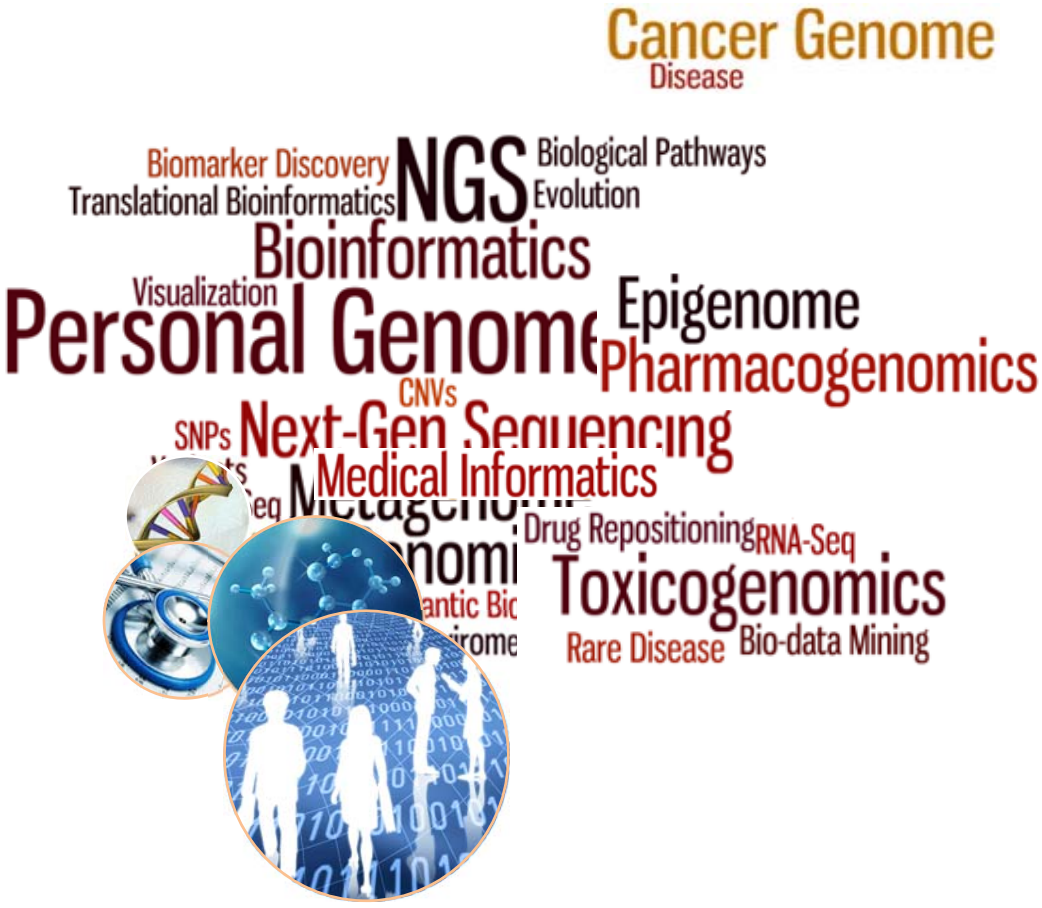
Translational Bioinformatics



2015.11.07 TBC

TIME and | 7th – 9th, November, 2015
LOCATION | BELLSALLE Nishi-Shinjuku, Tokyo, Japan
Conference | http://www.snubi.org/TBC2015/
The Korean Society of Medical Informatics
Systems Biomedical Informatics Research Center
Japan Association for Medical Informatics

Table of Contents

Welcome Messages

Translational Bioinformatics Conference (TBC) will aim to highlight the multi-disciplinary nature research field and provide an opportunity to bring together and exchange ideas between translational bioinformatics researchers. TBC puts its initial emphasis on promoting translational bioinformatics research activities initiated in Asia-Pacific region. Translational bioinformatics is a rapidly emerging field of biomedical data sciences and informatics technologies that efficiently translate basic molecular, genetic, cellular, and clinical data into clinical products or health implications. Translational bioinformaticians with a mix of computer scientists, engineers, epidemiologists, physicists, statisticians, physicians and biologists come together to create the unique intellectual environment of our meeting.

Learning Objectives

Major topic areas of this year are focused on infra-technological innovations from bench to bedside, with a particular emphasis on clinical implications

- To present and exchange the latest progresses in translational bioinformatics.
- To identify the current challenges, to find research and funding opportunities, and develop future perspectives.
- To demonstrate how genomic data-driven informatics approaches can facilitate clinical research, genomic medicine, and healthcare
- To facilitate trans-disciplinary interactions among computational biology, genomics, bio-data sciences, translational medicine, and healthcare.
- To provide educational opportunities for the rapidly growing new comers.
- To develop and deploy platform for resource and problem sharing among nation-wide biomedical informatics initiatives.

Greetings

Dear Colleagues,

On behalf of organizing committee chair, I am delighted to welcome you to Tokyo Japan, for the fifth annual Translational Bioinformatics Conference (TBC 2015). TBC 2015 provides a general forum for the latest research in genomics, bioinformatics, translational research, and biomedical informatics.

Past TBC conferences in Asia were successfully organized in Korea and Mainland China. Now it is coming to Japan. Japan Association for Medical Informatics is proud to host TBC 2015.

Thanks to the invited speakers and presenters from all around the world to this conference, I am sure that you will enjoy an exciting learning opportunities, exchanging of information, and endeavor at TBC 2015.

I wish all participants of the conference have pleasant and memorable experience. Please enjoy TBC 2015 and Tokyo city, pop-culture capital of the world.

With my best regards,

Michio Kimura, MD, PhD, FACMI

Chair, TBC 2015

Greetings

Dear Colleagues,

On behalf of the Japan Association for Medical Informatics (JAMI), I would like to extend a hearty welcome to the distinguished researchers from abroad on this special occasion of the Translational Biomedical Conference 2015 (TBC2015) in Tokyo, Japan. It is indeed a privilege for JAMI to have the opportunity of hosting the TBC2015.

Analysis of genomic information and translation of findings in basic research into clinical medicine are increasing. Genomic data already exists in hospital information systems, and biomedical data and lifestyle data are a part of healthcare information in clinical research to improve individual health and population health, and for preventive healthcare. We are keen to identify and resolve issues with the discipline of biomedical informatics in terms of sociology and ethics as well as clinical medicine and healthcare information technology.

We are delighted that scientists from more than 12 countries are getting together with a lively interest in a rapidly growing field of the Translational Bioinformatics. It is highly desirable that interested scientists get together to present and discuss new advances regarding translational bioinformatics, and to make plans for the future directions of research.

I wish the best success to the TBC2015, a success which I have no doubt and which will represent the most valued reward for those who devoted themselves to its organization with hard work and enthusiasm. I hope you all enjoy your stay in Tokyo and I wish your visit will prove pleasant and rewarding to you.

With my warmest regards,

President
Japan Association for Medical Informatics
Mihoko Okada, Ph.D.

Greetings

Dear Colleagues,

It is a great pleasure to welcome you all to the 5th annual Translational Bioinformatics Conference 2015. As the chair of science affairs in Japanese Association for Medical Informatics, I have been observing the domain of translational bioinformatics and expecting the enormous potential of the field as a bridge between bioinformatics research and clinical practice.

Over the past five years, the domain of the research is rapidly growing and significantly improving the quality. In TCB 2015, topics of the paper presentations are ranging from bioinformatics algorithms, data privacy, and biomedical big data to multi-omic applications, GWAS, phenotype linkage, cancer bioinformatics, and clinical informatics for translation. Premier keynote speakers around the world will present hot subjects relevant to the domain. One of the feature sessions is a panel discussion of journals and publication. The editor-in-chiefs or editorial board members of JAMIA, Journal of biomedical and Computer Methods and Programs in Biomedicine will join the panel discussion.

We are hoping that TBC2015 will serve to advance your research and will bridge to another prosperity of TBC2016 next year.

Best regards,

Chair, Division of Scientific Affairs,
Japan Association for Medical Informatics

Tomohiro Sawa, MD, PhD

TBC 2015 Chair

Michio Kimura, MD, PhD, FACMI

Professor of Medicine , and Medical Informatics

Hamamatsu University School of Medicine

Director, Medical Informatics Department

Hamamatsu University Hospital

Japan

Scientific Committee Members

Tomohiro Sawa, M.D., Ph.D.
Teikyo University (Japan)

Mihoko Okada, Ph.D.
Kawasaki University of Medical Welfare (Japan)

Naoki Nakashima, M.D., Ph.D.
Kyushu University (Japan)

Hideto Yokoi, M.D., Ph.D.
Kagawa University (Japan)

Ken Imai, Ph.D.
University of Tokyo (Japan)

Ju Han Kim, M.D., Ph.D.
Seoul National University College of Medicine (Korea)

Atul Butte, M.D., Ph.D.
University of California, San Francisco (U.S.A.)

Luonan Chen, Ph.D.
Shanghai Institute for Biological Sciences (China)

Indira Ghosh, Ph.D.
Jawaharlal Nehru University, New Delhi, India (India)

Maricel Kann, Ph.D.
University of Maryland Baltimore County (U.S.A.)

Yves A. Lussier, M.D.
University of Arizona (U.S.A.)

Lucila Ohno-Machado, M.D., Ph.D.
University of California, San Diego (U.S.A.)

Marylyn D. Ritchie, Ph.D.
Pennsylvania State University (U.S.A.)

Executive Committee Members

Japan Association for Medical Informatics

Tomohiro Sawa, M.D., Ph.D.
Teikyo University (Japan)

Mihoko Okada, Ph.D.
Kawasaki University of Medical Welfare (Japan)

Naoki Nakashima, M.D., Ph.D.
Kyushu University (Japan)

Hideto Yokoi, M.D., Ph.D.
Kagawa University (Japan)

Ken Imai, Ph.D.
University of Tokyo (Japan)

Yuichiro Gomi, Ph.D.
Nihon University (Japan)

Tetsuya Narukiyo, MHA, PMP
tokyo medical university (Japan)

Harukazu Tsuruta, Ph.D.
Kitasato University (Japan)

Organizing Committee Members

Ju Han Kim, M.D., Ph.D.
Seoul National University College of Medicine (Korea)

Atul Butte, M.D., Ph.D.
University of California, San Francisco (U.S.A.)

Luonan Chen, Ph.D.
Shanghai Institute for Biological Sciences (China)

Indira Ghosh, Ph.D.
Jawaharlal Nehru University, New Delhi, India (India)

Maricel Kann, Ph.D.
University of Maryland Baltimore County (U.S.A.)

Yves A. Lussier, M.D.
University of Arizona (U.S.A.)

Lucila Ohno-Machado, M.D., Ph.D.
University of California, San Diego (U.S.A.)

Marylyn D. Ritchie, Ph.D.
Pennsylvania State University (U.S.A.)

Tomohiro Sawa, M.D., Ph.D.
Teikyo University (Japan)

Program at a Glance

Day 1, Saturday, Nov. 7, 2015

	Hall A	Room 2
09:00-09:20	**Opening** Michio Kimura : Chair, TBC2015 Mihoko Okada : President, Japanese Association for Medical Informatics	
09:20-10:00	**Keynote 1. Michio Kimura** moderator: Mihoko Okada Japan's Available Large Clinical Databases - Reimbursement claim data, Specialist qualification registry, and HL7 based standardized storage	
10:00-10:30	**Keynote 2. Yu-Chuan (Jack) Li** moderator: Michio Kimura Mapping Long-term-use Drugs and Personal Cancer Risk	
10:30-11:00	Coffee Break	
	S1. Data Privacy	**S2. Clinical Informatics for Translation**
	moderator: Jack Li	moderator: Hideto Yokoi
11:00-11:25	Privacy-Preserving Mechanisms for Transmission Disequilibrium Test in Genome-Wide Association Studies	An innovative model for reducing medication errors in computerized physician order entry (CPOE) system
11:25-11:50	SECRET: Secure Edit-distance Computation over homomoRphic Encrypted daTa	BEST: Next-Generation Biomedical Entity Search Tool for Knowledge Discovery from Biomedical Literature
11:50-12:15	Secure Multi-pArty Computation Grid LOgistic REgression (SMAC-GLORE)	Establishing a program for a clinical informatics subspecialty in Korea and needs assessment survey in physicians.
12:15-13:20	Lunch Break	
13:20-14:20	**Keynote 3. Yves A. Lussier** moderator: Naoki Nakashima The personalome era: precision therapy with dysregulated molecular networks	
	S3. Cancer Bioinformatics	**S4. Multi-omic Applications**
	moderator: Indira Ghosh	moderator: Luonan Chen
14:20-14:45	SFRP1 is a possible candidate for epigenetic therapy in non-small cell lung cancer	Integrated analysis of omics data using microRNA-target mRNA network and PPI network reveals regulation of Gnai1 function in the spinal cord of EWS KO mice
14:45-15:10	Unraveling synthetic lethal interactions for therapeutics of lung adenocarcinoma	Integration of bioinformatics and imaging informatics for identifying rare PSEN1 variants in Alzheimer's disease
14:10-14:35	Frequent hypermethylation of orphan CpG islands with enhancer activity in cancer	Using knowledge-driven genomic interactions for multi-omics data analysis: meta-dimensional models for predicting clinical outcomes in ovarian carcinoma
15:35-16:00	Discovering gene expression signatures responding to tyrosine kinase inhibitor treatment in chronic myeloid leukemia	Alternative Polyadenylation in Cardiac Development and Hypertrophy
16:00-16:30	Coffee Break	
16:30-17:30	**Keynote 4. Lucila Ohno-Macado** moderator: Ken Imai Developing a large clinical data research network: the patient-centered SCAlable National Network for Effectiveness Research (pSCANNER)	
17:30-18:00	Break	
18:00-	Welcome Reception (Room 4)	

*Poster session in Room1

Program at a Glance

Day 2, Sunday, Nov. 8, 2015

	Hall A	Room 2
		S5. Linking Phenotypes
		moderator: Indira Ghosh
08:30-08:55	**Keynote 5:Hiroshi Tanaka** moderator: Takako Takai Japan Association for Omics-based Medicine Complex systems theory of cancer metastasis	IVIG administration in Kawasaki disease patients causes major methylation alterations on the CpG markers of inflammatory immune associated genes
08:55-09:20	**Journal and Publication Panel** JAMIA, JBI, CMPB Lucila Ohno-Machado (JAMIA: Editor-in-Chief) Maricel Kan (JBI: Editorial Board) Jack Li (CMPB: Editor-in-Chief) moderator: Yves A. Lussier	Semantics based approach for analyzing disease-target associations
09:20-09:45		Inferring Crohn's disease association from exome sequences by integrating biological knowledge
09:45-10:10	Coffee Break	
10:10-11:10	**Keynote 6: Marylyn D. Ritchie** moderator: Hideto Yokoi Electronic health records and genomics - a dynamic duo for precision medicine	
11:10-12:10	**Keynote 7: Toru Hosoda** moderator: Naoki Nakashima	
12:10-13:00	Lunch [Intel Special Session] Modular, Scalable Computing Platform for Precision Medicine Ketan Paranjape moderator: Tomohiro Sawa	
	Highlight Research	**S6. New Technologies**
	moderator: Marylyn D. Ritchie	moderator: Luonan Chen
13:00-13:25	Extraction of Pharmacokinetic Evidence of Drug–Drug Interactions from the Literature	EyeSee; an Assistive Device for Blind Navigation with Multi-Sensory Aid
13:25-13:50	Diagnostic Role of Exome Sequencing in Immune Deficiency Disorders	PATTERN: Pain Assessment for paTients who can't TEll using Restricted Boltzmann machiNe
13:50-14:15	hiHMM: Bayesian non-parametric joint inference of chromatin state maps.	Creative Activity Aid using Active Tremor Cancellation
14:15-14:25	Break	
14:25-15:25	**Keynote 8: Griffin Weber** moderator:Hideto Yokoi Analyzing, Visualizing, and Facilitating Scientific Collaboration	
	S7. Bionformatics Algorithms	**S8. GAWS and Post-GWAS**
	moderator: Ken Imai	moderator: Maricel Kan
15:25-15:50	Fast Comparison of Genomic and Meta-Genomic Reads with Alignment-Free Measures based on Quality Values	Genome-wide association study identifies novel susceptibility genes associated with coronary artery aneurysm formation in Kawasaki Disease
15:50-16:15	Parametric analysis of RNA-seq expression data	Integrative Regression Network for Genomic Association Study
16:15-16:40	Nearest Neighbor Imputation Algorithms: A Critical Evaluation.	eMERGE Phenome-Wide Association Study (PheWAS) Identifies Clinical Associations and Pleiotropy for Stop-Gain Variants
16:40-17:00	Coffee Break	
17:00-18:00	**Keynote 9: Rae Woong Park** moderator: Ken Imai Clinical Outcome Modeling & Simulation using EMR Big Data	

*Poster session in Room1

Program at a Glance

Day 3, Monday, Nov. 9, 2015

	Hall A	Room 2
Session	**Tohoku Scientific Session**	**S9. Biomedical Big Data**
		moderator: Ju Han Kim
08:30-08:55	**Tohoku Medical Megabank Organization** - Overview of Tohoku Medical Megabank Project - Deep whole genome sequencing Japanese healthy population - Strategy of Three-Generation Cohort Study and Community Based Cohort Study - Integrated Database Systems for personal genomic and healthcare information in Tohoku area	DataRank: A Framework for Ranking Biomedical Datasets
08:55-09:20		Many larger worlds in a small world
09:20-09:45		CLASH: Complementary Linkage with Anchoring and Scoring for Heterogeneous BioMolecular and Clinical Data
09:45-10:10	Coffee Break	
10:10-11:10	**Keynote 10: Maricel G. Kann** modelator: Luonan Chen	
11:10-11:40	Excellent Paper Award Excellent Poster Award	
11:40-12:00	Closing Tomohiro Sawa Ju Han Kim	

Keynote Speakers

09:20-10:00 (Saturday, Nov. 7)

Michio Kimura

Hamamatsu University School of Medicine, Japan

Japan's Available Large Clinical Databases

- Reimbursement claim data, Specialist qualification registry, and HL7 based standardized storage

In Japan, following four large healthcare databases are available. This presentation outlines these databases with personal evaluations.

1. National reimbursement claim data

About 10 years behind Korea, Japan Ministry started making reimburse claim data available only for researchers.

They are thorough; almost all claim data are included. But they have no lab results, no clinical contents by doctors, bogus diagnosed disease names. Timeliness is poor, as they are available after more than half year.

2. DPC data and voluntary hospital data pool (DPC: disease-procedure, DRG of Japan)

Cases are only of 1500 large hospitals, and only hospitalized cases. Disease classifications are not bogus. And they have some standardized clinical profiles, such as heart failure rate by NYHA. Doctors complain that coding input takes 15 minutes. Timeliness is same as claim data, but some hospitals gather earlier.

3. Surgeons' clinical data

17 surgeon societies allied and made one standardized form for operation case registration to be used for data analysis, and specialist accreditations for submitters. This database has almost all major operations, and they have many clinical items. But surgeons complain that it takes 30 minutes to submit a case. It is only available after more than a year, and scale-wide analysis is only allowed to professional societies.

4. HL7 standardized SS-MIX storage data.

Recommended by Ministries, HL7 based standardized storages are now in operation at 518 hospitals (3/2015). They store prescriptions, lab results, and diagnoses. Some medical professional societies for, such as diabetes, renal diseases, hypertension, and hyperlipidemia, are using this nationwide infrastructure for their case data collections. PMDA (FDA in Japan) are starting to use this for early drug side effect detection. Because it is based on CPOE systems, last week epidemic can be detected.

Keynote Speakers

10:00-10:30 (Saturday, Nov. 7)

Yu-Chuan(Jack) Li

Taipei Medical University, Taiwan

D2W2C: Mapping the Cancer Risk of Long-term Use Drugs

Based on big data analytics from large national claim databases, we were able to conduct Observational Studies with Control on a massive scale. In this talk, we will describe how a translational process from Dry lab to Wet lab to Clinical trials (D2W2C) can be achieved in the scenario of evaluating cancer risks of each long-term use drug. Initial data on sedatives and hypnotics as well as anti-diabetics drugs will be presented with their individual impact on common cancer risks. Collaborations from data scientists around the world are welcome to help us crack the secret of cancer risks from our drugs.

Keynote Speakers

13:20-14:20 (Saturday, Nov. 7)

Yves A. Lussier

University of Arizona, USA

The Personalome Era: Precision Therapy with Dysregulated Molecular Networks

Keynote Speakers

16:30-17:30 (Saturday, Nov. 7)

Lucila Ohno-Machado

University of California, San Diego, US

Developing a large clinical data research network: the patient-centered SCAlable National Network for Effectiveness Research (pSCANNER).

Large numbers of study participants are needed to provide sufficient power for observational or interventional comparative effectiveness studies. There are multiple challenges and opportunities available to make better use of data collected for health care, and much has been discussed abo ut the development of a 'learning healthcare system.' pSCANNER is part of PCORnet, a PCORI -funded initiative to promote faster and leaner comparative effectiveness studies (randomized cli nical trials, pragmatic trials, and observational studies), through use of Electronic Health Record (EHR) data, on topics that matter most to patients. We developed governance and harmonized EHR data from over 23 million patients. Nine health systems, which are part of federal and stat e governments or of independent non-profit organizations, were involved in phase 1. I will desc ribe our achievements and lessons learned in developing the pSCANNER infrastructure, with pa rticular emphasis on distributed analytics, and describe our plans for phase 2.

Keynote Speakers

10:10-11:10 (Sunday, Nov. 8)

Marylyn D. Ritchie

Pennsylvania State University, USA

Electronic Health Records and Genomics - A Dynamic Duo for Precision Medicine.

Keynote Speakers

11:10-12:10 (Sunday, Nov. 8)

Toru Hosoda

Takai University, Japan

Translational Perspectives of Cardiac Regenerative Medicine.

It was believed that cardiomyocytes lose their proliferative potential soon after birth and survive as long as our lifespan without being replaced. However, the discovery of resident stem cells in the heart challenged this long-lasting dogma. Different laboratories have identified distinct primitive cell categories. Among them, c-kit-positive cardiac stem cells (CSCs) are found to be one of the most primitive populations in the heart. They hold fundamental properties of stem cells: self-renewing, clonogenic, and multipotent. Under physiological conditions, c-kit-positive CSCs are stored in a “niche”, an interstitial structure that exists among mature cardiac myocytes. Upon stimulation, CSCs divide, migrate, and differentiate into cardiomyocytes, smooth muscle cells, endothelial cells, and fibroblasts, slowly but continuously replacing old parenchymal cells in the myocardium. Moreover, in sharp contrast to the previous theory, recent evidence has revealed that myocytes in the human heart are completely renewed more than 10 times during our life. This paradigm shift not only affected our knowledge of biology but also provided a fundamental basis for the field of translational regenerative medicine. Following numerous preclinical studies employing various animal models, severe chronic heart failure patients were treated by autologous c-kit-positive CSCs. We found outstanding improvement of the left ventricular function, which lasted for at least 2 years after the cell administration. Most importantly, this intervention did not increase the major adverse events in the 20 patients in the cell-treated group. In the lecture, I would like to discuss the future prospects of this promising therapy.

Keynote Speakers

14:25-15:25 (Sunday, Nov. 8)

Griffin Weber

Harvard Medical School, USA

Analyzing, Visualization, and Facilitating Scientific Collaboration.

Profiles Research Networking Software (RNS) is an open source website used by several dozen universities, pharmaceutical companies, and government agencies to generate searchable online profiles of their investigators (http://profiles.catalyst.harvard.edu). We initially developed it for Harvard's Clinical and Translational Science Center to break-down silos in biomedical research and encourage collaboration. Interactive network visualizations enable users to explore the different ways people are connected, such as through co-authorship, having similar research interests, or working in physically nearby offices. Although Profiles RNS is primarily used to find experts in particular subject areas, it also provides a rich source of data for studying collaboration within an organization. This presentation will show how Profiles RNS has been used to analyze (1) collaboration patterns across different disciplines, (2) gender and race differences in scientific networks, (3) attributes of teams that predict whether they will be awarded funding, and (4) the role of interdisciplinary teams on translational science.

Keynote Speakers

17:00-18:00 (Sunday, Nov. 8)

Rae Woong Park

Ajou University School of Medicine, Korea

Clinical Outcome Modeling & Simulation using EMR Big Data.

Keynote Speakers

10:10-11:10 (Monday, Nov. 9)

Maricel G. Kann

University of Maryland, USA

The protein domain landscape of disease mutations

The body of disease mutations with known phenotypic relevance continues to increase and is expected to do so even faster with the advent of new experimental techniques such as whole-genome sequencing coupled with disease association studies. However, genomic association studies are limited by the molecular complexity of the phenotype being studied and the population size needed to have adequate statistical power. One way to circumvent this problem, which is critical for the study of rare diseases, is to study the molecular patterns emerging from functional studies of existing disease mutations. Current gene-centric analyses to study mutations in coding regions are limited by their inability to account for the functional modularity of the protein. Previous studies of the functional patterns of known human disease mutations have shown a significant tendency to cluster at protein domain positions, namely position-based domain hotspots of disease mutations. By leveraging phenotypically relevant data across different species we provide a framework for the study of all human mutations including those associated with Mendelian disease and cancer somatic mutations from individual genome. Our methods show good potential for identifying disease-relevant mutations as well as rare driver mutations in current, large-scale tumor sequencing projects. In addition, mapping mutations to specific domains provides the necessary functional context for understanding how the mutations contribute to the disease, and may reveal novel or more refined gene and domain target regions for drug development.

Special Journal / Publication Panel

08:55-09:45 (Sunday, Nov. 8)

Publishing Translational Bioinformatics Papers in Peer-reviewed Informatics Journals

Editors of three major informatics journals will introduce their journals and the opportunities for TBI researchers to publish papers with them. A TBI research leader will also provide some insights from the perspective of a frequent author, reviewer, and editorial board member. There will be time for discussion with attendees after the brief formal presentations.

Panelists:

Yves Lussier (Moderator)

Maricel Kann (former associate editor, Journal of Biomedical Informatics)

Lucila Ohno-Machado (editor-in-chief, Journal of the American Medical Informatics Association)

Yu-Chuan (Jack) Li (editor-in-chief, Computer Methods and Programs in Biomedicine)

Organizer:

Edward H. Shortliffe (editor-in-chief, Journal of Biomedical Informatics)

Organization:

Introduction (Lussier)

JBI (Kann)

JAMIA (Ohno-Machao)

CMPBM (Li)

Author/Reviewer Perspective (Lussier)

Discussion

Special Lunch Session --- Intel Presents

12:10-13:00 (Sunday, Nov. 8)

Enabling Technology. Leveraging Data. Transforming Precision Medicine.

Ketan Paranjape, General Manager Life Sciences, Intel Corp.

Through collaborations, research and innovation, Intel is supporting the advancement of processing, storage, networking, data security, sequencing efficiency, accelerated bioinformatics and advanced analytics—to push the boundaries of this new "precision medicine" and bring us closer than ever to truly making care personal. Listen to the talk discuss how Intel's Collaborative Cancer Cloud can lead the way.

Scientific Paper Sessions

S1. Data Privacy Algorithms

Room: Hall A
Date: Saturday, Nov. 7, 11:00 - 12:15

S1-1: Privacy-Preserving Mechanisms for Transmission Disequilibrium Test in Genome-Wide Association Studies.

Meng Wang[1], Zhanglong Ji[1], Xiaoqian Jiang[1], Shuang Wang[1], and Lucila Ohno-Machado[1]

[1]Department of Biomedical Informatics, UC San Diego, 92093, US

Abstract

Genome sequencing data have great potential to improve biomedical research (e.g., precision medicine). However, genome data are highly sensitive, and inappropriate disclosure may put patients and their family's privacy at risk if handled inappropriately. In Genome-Wide Association Studies (GWAS), most exisitng privacy-preserving mechanisms focus on case-control studies between two populations with and without a certain disease, but few on family-based methods. Transmission disequilibrium test (TDT) is a family based association test, which is widely used to measure the over-transmission of an allele from heterozygous parents to one affected offspring. In this paper, we developed and evaluated a suite of differentially private mechanisms for TDT, which include Laplace mechanisms using the TDT test statistic, · -value, projected · -values, and exponential mechanisms using the TDT test statistic and the shortest Hamming distance score. Simulation experiments demonstrated that the exponential mechanism based on the shortest Hamming distance score preserves relatively high utility and privacy.

S1-2: SECRET: Secure Edit-distance Computation over homomoRphic Encrypted data.

Yuchen Zhang[1], Wenrui Dai[1,2], Shuang Wang[2], Mira Kim[3], Kristin Lauter[4], Jun Sakuma[5], Hongkai Xiong[1], Xiaoqian Jiang[2]

[1] Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Department of Biomedical Informatics, University of California, San Diego, San Diego, CA, 92093, USA
[3]Department of Mathematical Sciences, Seoul National University, Seoul 151-747, Republic of Korea
[4]Microsoft Research San Diego, CA, 92122, USA
[5]Department of Computer science, University of Tsukuba, Tsukuba, Ibaraki 305-8577, Japan

Abstract The biomedical community benefits from the increasing availability of genomic data, which enables institutions and researchers to develop personalized treatment and discover precision medicine. However, privacy and confidentiality of genomic data have been becoming a major concern for both patients and researchers in the data storage, transfer and analysis phases. In this paper, we proposed a protocol for Secure Edit-distance Computation over homomoRphic Encrypted daTa (SECRET), which leverages Homomorphic Encryption (HME) to securely outsource both genomic data and Edit-distance computation in an untrusted cloud environment without sacrificing data privacy. The proposed SECRET protocol develops integer based HME comparison primitive for exact Edit distance computation. Furthermore, we improved the efficiency of secure Edit distance calculation by using an optimized pathfinding based approach. Parallel implementation is also investigated to improve the performance in concurrently calculating multiple sequence pairs. Experimental results demonstrate that the computational costs of the proposed protocol are significantly reduced in comparison to the existing state-of-the-art HME-based method.

Scientific Paper Sessions

S1-3: Secure Multi-pArty Computation Grid LOgistic REgression (SMAC-GLORE).

Haoyi Shi[1,2,*], Shuang Wang[2,*], Wenrui Dai[2], Yuzhe Tang[1], Xiaoqian Jiang[2], Lucila Ohno-Machado[2]

[1]Syracuse University, Syracuse, NY 13210, USA
[2] Department of Biomedical Informatics, University of California, San Diego, CA 92093, USA

Abstract

In biomedical research, data sharing and information exchange are very important in improving quality of care, accelerating discovery, and promoting the meaningful secondary use of clinical data. A big concern in biomedical data sharing is protection of patients' privacy because inappropriate information leakage can put patients' privacy at risk. In this study, we developed a secure multi-party computation based grid logistic regression (SMAC-GLORE) framework. Unlike our previous work in GLORE, SMAC-GLORE protects not only patient-level data, but also all the intermediary information exchanged during the model learning phase. The experimental results demonstrate the feasibility of secure distributed logistic regression across multiple institutions without sharing data.

S2. Clinical Informatics for Translation

Room: Room 2
Date: Saturday, Nov. 7, 11:00 - 12:15

S2-1: An innovative model for reducing medication errors in computerized physician order entry (CPOE) system.

Phung-Anh (Alex) Nguyen, Ph.D.[1,*], Chu-Ya Huang, MBA[1], Richard Lu, M.D., MS[1], Yu-Chuan (Jack) Li, M.D., Ph.D.[1,2,*]

[1]College of Medical Science & Technology, Taipei Medical University, Taipei, Taiwan
[2]Dermatology Department, Wan-Fang Hospital, Taipei, Taiwan

Abstract

Medication errors are common, life threatening, costly but preventable. The association rules mining techniques are utilized for 9.6 million prescriptions from 2007 to 2011 of three Taipei Medical University hospital claims data. The disease-medication (DM) and medication-medication (MM) associations were computed by their co- occurrence and associations' strength were measured by the Q values, which were derived from 68.6 million diagnoses with ICD9CM and 45.6 million medications with ATC codes. By considering the number of DMQs and MMQs, the AESOP model was developed to determine the appropriateness of a given prescription. Firstly, 3200 prescriptions were randomly selected to evaluate by 7 experts, subsequently, 1,056 prescriptions were evaluated by physicians through web-service in hospital practice. The results showed 96% accuracy for appropriate and 45% accuracy for inappropriate prescriptions. Thus, AESOP model be able to improve patient safety and quality of care as well as to enable better decision support and quality measurement.

Scientific Paper Sessions

S2-2: BEST: Next-Generation Biomedical Entity Search Tool for Knowledge Discovery from Biomedical Literature.

Donghyeon Kim[1,†], Sunwon Lee[1,†], Kyubum Lee[1], Jaehoon Choi[1], Seongsoon Kim[1], Minji Jeon[1], Sangrak Lim[1], Donghee Choi[1], Aik-Choon Tan[1, 2], and Jaewoo Kang[1,*]

[1]Department of Computer Science and Engineering, Korea University, Seoul, Korea
[2] Translational Bioinformatics and Cancer Systems Biology Laboratory, Division of Medical Oncology, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Abstract

As the volume of publications in biomedicine rapidly grows, searching and extracting relevant information from the literature becomes more challenging. Many tools aim to address this problem. However, the existing tools face at least one of the following three challenges: slow response time, outdated results, and limited answer types. Some of the existing tools pass a user's query to PubMed and process the retrieved abstracts to extract information at query time, resulting in slow response time. Others attempt to preprocess the PubMed corpus to speed up the query processing, but those tools are often out of date. The existing tools do not support sophisticated queries such as searches for mutations.

S2-3: Establishing a program for a clinical informatics subspecialty in Korea and needs assessment survey in physicians.

Kye Hwa Lee[1], and Ju Han Kim[1, *]

[1]Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea

Abstract

Introduction: This study describes the three years of experiences of building an education program, Certified Physician in BioMedical Informatics (CPBMI) for physicians as a previous step in the process of certified bio-informatics training courses. We also conducted a needs assessment to trainees about the educational contents and future demand.
Methods: The length of the CPBMI educational program is designed to be an 18-month coursework, which is roughly equivalent to the coursework requirement of a 'weak' master's degree-seeking program level. Physicians are required to complete formal classes for introductory biomedical informatics courses including basic computer programing skills, biostatistics, database technologies, data structure and algorithms, artificial intelligence in biomedicine, genomics and translational bioinformatics for two semesters for six hours a week. Python, SQL and R programming are chosen for the efficient and practical computer skills. Requirements for clinical systems rotation, research project presentation, paper presentation to KOSMI conference and journal, with the final certification exam for CPBMI were introduced.
Results: There are 124 graduates for three years training courses from November 2012 to September 2015. We conducted a demand survey to these graduates and 58% of them fully responded. The following were the top three ranked interested in among clinical-informatics fields: statistics and research method; information/computer science core; domain-specific information system such as hospital or research laboratory. Most of the respondents (63%) identified their area of activity in bioinformatics as a genome data analysis. The second most active field is clinical research informatics (35.4%), and the third is public health informatics (15.4%). Regarding their self-reported experience with software and programming, R or SPSS are the most frequently used tools (81% and 72% respectively). Python is the most popular language, except for R, 60% using Python.
Conclusion: Biomedical informatics is multi-disciplinary in nature. There is a growing need of informatics concepts and information technologies for physicians who understand the healthcare system, care process, biomedical research and industries. CPBMI provided an immediately recognized credential for organizations seeking to hire physician informaticians. The future step to the establishment of certification process of CPBMI is clearly that of biomedical informatics subspecialty in Korea.

Scientific Paper Sessions

S3. Cancer Bioinformatics

Room: Hall A
Date: Saturday, Nov. 7, 13:20 - 15:00

S3-1: SFRP1 is a possible candidate for epigenetic therapy in non-small cell lung cancer.

Y-h. Taguchi[1], Mitsuo Iwadate[2], Hideaki Umeyama[2]

[1]Department of Physics/2Department of Biological Science, Chuo University, 1-13-27 Kasuga, Bunkyo-ku, 112-8551 Tokyo, Japan.

Abstract

Background: Non-small cell lung cancer (NSCLC) remains a lethal disease despite many proposed treatments. Recentstudies have indicated that epigenetic therapy, which targets epigenetic effects, might be a new therapeutic methodology for NSCLC. However, it is not clear which objects (e.g., genes) this treatment specifically targets. Secreted frizzled-related proteins (SFRPs) are promising candidates for epigenetic therapy in many cancers, but there have been no reports of SFRPs targeted by epigenetic therapy for NSCLC.
Methods: This study performed a meta-analysis of reprogrammed NSCLC cell lines instead of the direct examination of epigenetic therapy treatment to identify epigenetic therapy targets. In addition, mRNA expression/promoter methylation profiles were processed by recently proposed principal component analysis based unsupervised feature extraction and categorical regression analysis based feature extraction.
Results: The Wnt/β-catenin signalling pathway was extensively enriched among 32 genes identified by feature extraction. Among the genes identified, SFRP1 was specifically indicated to target β-catenin, and thus might be targeted by epigenetic therapy in NSCLC cell lines. A histone deacetylase inhibitor might reactivate SFRP1 based upon the re-analysis of a public domain data set. Numerical computation validated the binding of SFRP1 to WNT1 to suppress Wnt signalling pathway activation in NSCLC.

S3-2: Unraveling synthetic lethal interactions for therapeutics of lung adenocarcinoma.

Jan-Gowth Chang[1,*], Chia-Cheng Chen[2,*], Kun-Tu Yeh[3,4,*], Yu-Chin Hsu[2], Geng-Ming Chang[2] and Grace S. Shieh[2,§]

[1]Department of Laboratory Medicine, and Center of RNA Biology and Clinical Application, China Medical University Hospital, China Medical University, Taichung 404, Taiwan.
[2]Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan.
[3]Department of Pathology, Changhua Christian Hospital, Changhua 505, Taiwan.
[4]Department of Pathology, School of Medicine, Chung Shan Medical University, Taichung 402, Taiwan.

Abstract

Background: Two genes are called synthetic lethal (SL) if their simultaneous mutations lead to cell death, but each individual mutation does not. Targeting SL partners of mutated cancer genes can kill cancer cells specifically, but leave normal cells intact. Therefore, synthetic lethality strategy offers an elegant alternative to killing cancer cells with non-druggable mutant tumor suppressor genes and stability genes, for example, TP53, by targeting its SL partners. We present a computational approach to identifying SL gene pairs for novel therapeutics in lung adenocarcinoma (LADC).
Methods: We first identified functionally relevant (simultaneously differentially expressed) gene pairs by screening the collected 668 SL pairs, which were verified in various cancers, using microarray gene expression data of paired lung cancerous and non-cancerous tissues. From the top-ranked pairs, 21 genes were chosen for immunohistochemistry (IHC)

staining at multiple cellular locations using tissues dissected from 137 LADC patients in Taiwan. To find novel SL pairs, we combined the 24 IHC of individual proteins to result in 273 IHC pairs. Next, we tested each of these IHC pairs for the two proposed synergistic effects with clinical features to identify tumor-cell-dependent pairs, which are our predicted SL pairs.
Results: Of the 19 predicted SL pairs, FEN1-RAD54B and BRCA1-TP53 have been verified in colorectal cancer and breast cancer cells in literature, respectively. Furthermore, several predicted pairs with more significant p- values than these verified ones, e.g., EFGR-RB1, are promising.
Conclusions: Our method identified one validated SL pair of TP53 and promising EFGR-RB1. After future in vitro and in vivo validations, BRCA1 and RB1, may be promising targets for TP53-mutant and EGFR-mutant LADC patients, while to date safety regarding inhibition of TP53 is controversial and many Asian LADC patients have EGFR mutations. This indicates that the proposed approach is useful in revealing novel drug targets for LADC via SL interactions.

S3-3: Frequent hypermethylation of orphan CpG islands with enhancer activity in cancer.

Min Gyun Bae[1], Jeong Yeon Kim[1] and Jung Kyoon Choi[1,*]

[1]Department of Bio and Brain Engineering, KAIST, Daejeon 305-701, Republic of Korea

Abstract

CpG islands (CGIs) are interspersed DNA sequences that have unusually high CpG ratios and GC contents. CGIs are typically located in the promoter of protein-coding genes. They normally lack DNA methylation but become hypermethylated and induce repression of associated gene in cancer. However, biological functions of non-promoter CGIs (orphan CGIs) largely remain unclear. Here, we identify orphan CGIs that do not map to the promoter of any protein-coding or non-coding transcripts but possess chromatin and transcriptional marks that reflect enhancer activity (termed eCGIs). They exhibit three-dimensional chromatin looping toward multiple target genes with high affinity. Intriguingly, transcription regulators were frequently associated with such CGI-containing enhancers. Remarkably, our analyses in cell lines and clinical tissues showed that eCGIs have more dynamic DNA methylation changes in cancer than promoter CGIs. The observed eCGI hypermethylation was accompanied by the loss of enhancer marks and the transcriptional inactivation of the target genes. Our results suggest that CGIs may indicate a distinct class of enhancers and perform a more instrumental role in tumorigenesis than typical CGIs in gene promoters.

S3-4: Discovering gene expression signatures responding to tyrosine kinase inhibitor treatment in chronic myeloid leukemia.

Kihoon Cha[1], Yi Li[1], and Gwan-Su Yi[1,*]

[1]Dept. of Bio and Brain Engineering KAIST Daejeon, South Korea

AbstractTyrosine kinase inhibitor (TKI)-based therapy is a recommended treatment for patients with chronic myeloid leukemia (CML). However, a considerable group of CML patients do not respond well to the TKI therapy. Challenging to overcome this problem, we tried to discover molecular signatures in gene expression profiles to classify the responders and non- responders of TKI therapy. We collected three microarray datasets of CML patients having total 73 responders and 38 non- responders. Using meta-analysis of three datasets, we selected seven differentially expressed genes (DEGs). We tested the classification performance of these genes and further selected discriminator gene sets by using random forest and iterative backward variable selection methods. Finally, we could identify a set of genes including CTSG,

SERPINA1, MLLT10, and MAGEA4 showing the lowest classification error rate less than 14%. Interestingly, all of four genes are on the signaling pathway of cellular apoptosis. This set of genes showed much higher performance than the average performance of other genes in downstream signaling of TKI target, BCR-ABL. In this study, we could find a good companion diagnosis marker set for TKI treatment and, at the same time, the potential of gene expression analysis to enhance the coverage of companion diagnosis.

S4. Multi-Omic Applications

Room: Room 2
Date: Saturday, Nov. 7, 13:20 - 15:00

S4-1: Integrated analysis of omics data using microRNA-target mRNA network and PPI network reveals regulation of Gnai1 function in the spinal cord of EWS KO mice.

Chai-Jin Lee[1], Hongryul Ahn[2], Sean Bong Lee[4], Jong-Yeon Shin[5], Woong-Yang Park[6], Jong-Il Kim[5], Junghee Lee[7,8], Hoon Ryu[7,8,#] and Sun Kim[1,2,3,#]

[1]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, 151-747, Republic of Korea
[2]Department of Computer Science and Engineering, Seoul National University, Seoul, 151-744, Republic of Korea
[3]Bioinformatics Institute, Seoul National University, Seoul, 151-747, Republic of Korea
[4]Department of Pathology & Laboratory Medicine, Tulane University School of Medicine, New Orleans, LA 70112, USA
[5]Genome Medicine Institute and Department of Biochemistry, Seoul National University College of Medicine, Seoul 110-799, Republic of Korea
[6]Samsung Genome Institute, Samsung Medical Center and Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul 135-710, Republic of Korea
[7]VA Boston Healthcare System, Boston, MA 02130, USA
[8]Boston University Alzheimer's Disease Center and Department of Neurology, Boston University School of Medicine, Boston, MA 02118, USA
[#]double corresponding authors

Abstract

Multifunctional transcription factor (TF) gene EWS is involved in various cellular process, such as transcription regulation, noncoding RNA regulation, splicing regulation, genotoxic stress response and cancer generation. Role of a TF gene can be effectively studied by measuring genome-wide gene expression, i.e., transcriptome, in an animal model of EWS knockout (KO). However, when a TF gene has complex multi- function, conventional approaches such as differentially expressed genes (DEGs) are not successful to characterize the role of the EWS gene. In this regard, network-based analyses that consider associations among genes are the most promising approach. Networks are constructed and used to show associations among biological entities at various levels, thus different networks represent association at different levels. Taken together, in this paper, we report contributions on both computational and biological sides. Contribution on the computational side is to develop a novel computational framework that combines miRNA-gene network and protein-protein interaction network information to characterize the multifunctional role of EWS gene. On the biological side, we report that EWS regulates G-protein, Gnai1, in the spinal cord of EWS KO mice using the two biological network integrated analysis method. Neighbor proteins of Gnai1, G-protein complex subunits Gnb1, Gnb2 and Gnb4 were also down-regulated at their gene expression level. Interestingly, up-regulated genes, such as Rgs1 and Rgs19, are linked to the inhibition of Gnai1 activities. We further verified the altered expression of Gnai1 by qRT-PCR in EWS KO mice. Our integrated analysis of miRNA-transcriptome network and PPI network combined with qRT-PCR verifies that Gnai1 function is impaired in the spinal cord of EWS KO mice.

S4-2: Integration of bioinformatics and imaging informatics for identifying rare PSEN1 variants in Alzheimer's disease.

Kwangsik Nho[1,3,13,*], Emrin Horgusluoglu[2,13], Sungeun Kim[1,3,13], Shannon L. Risacher[1,13], Dokyoon Kim[4], Tatiana Foroud[1,2,3,13], Paul S. Aisen[6], Ronald C. Petersen[7], Clifford R. Jack, Jr.[8], Leslie M. Shaw[5], John Q. Trojanowski[5], Michael W. Weiner[9,10], Robert C. Green[11], Arthur W. Toga[12], and Andrew J. Saykin[1,2,3,13,*], for the Alzheimer's Disease Neuroimaging Initiative (ADNI)[**]

[1]Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN, USA
[2]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA
[3]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN, USA
[4]Department of Biochemistry and Molecular Biology, Pennsylvania State University, PA, USA
[5]Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
[6]Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA
[7]Department of Neurology, Mayo Clinic Minnesota, Rochester, MN, USA
[8]Department of Radiology, Mayo Clinic Minnesota, Rochester, MN, USA
[9]Departments of Radiology, Medicine, and Psychiatry, University of California-San Francisco, San Francisco, CA, USA
[10]Department of Veterans Affairs Medical Center, San Francisco, CA, USA
[11]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
[12]The Institute for Neuroimaging and Informatics and Laboratory of Neuro Imaging, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA
[13]Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN, USA

Abstract

Background: Pathogenic mutations in PSEN1 are known to cause familial early-onset Alzheimer's disease (EOAD) but common variants in PSEN1 have not been found to strongly influence late-onset AD (LOAD). The association of rare variants in PSEN1 with LOAD-related endophenotypes has received little attention. In this study, we performed a rare variant association analysis of PSEN1 with quantitative biomarkers of LOAD using whole genome sequencing (WGS) by integrating bioinformatics and imaging informatics.
Methods: A WGS data set (N=815) from the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort was used in this analysis. 757 non-Hispanic Caucasian participants underwent WGS from a blood sample and high resolution T1-weighted structural MRI at baseline. An automated MRI analysis technique (FreeSurfer) was used to measure cortical thickness and volume of neuroanatomical structures. We assessed imaging and cerebrospinal fluid (CSF) biomarkers as LOAD-related quantitative endophenotypes. Single variant analyses were performed using PLINK and gene-based analyses of rare variants were performed using the optimal Sequence Kernel Association Test (SKAT-O).
Results: A total of 839 rare variants (MAF $< 1/\sqrt{(2N)}=0.0257$) were found within a region of ±10kb from PSEN1. Among them, six exonic (three non-synonymous) variants were observed. A single variant association analysis showed that the PSEN1 p.E318G variant increases the risk of LOAD only in participants carrying APOE ε4 allele where individuals carrying the minor allele of this PSEN1 risk variant have lower CSF Aβ1-42 and higher CSF tau. A gene-based analysis resulted in a significant association of rare but not common (MAF ≥ 0.0257) PSEN1 variants with bilateral entorhinal cortical thickness.
Conclusions: This is the first study to show that PSEN1 rare variants collectively show a significant association with the brain atrophy in regions preferentially affected by LOAD, providing further support for a role of PSEN1 in LOAD. The PSEN1 p.E318G variant increases the risk of LOAD only in APOE ε4 carriers. Integrating bioinformatics with imaging informatics for identification of rare variants could help explain the missing heritability in LOAD.

Scientific Paper Sessions

S4-3: Using knowledge-driven genomic interactions for multi-omics data analysis: meta-dimensional models for predicting clinical outcomes in ovarian carcinoma.

Dokyoon Kim[1], Ruowang Li[1], Anastasia Lucas[1], Shefali S. Verma[1], Scott M. Dudek[1], Marylyn D. Ritchie[1,2,*]

[1]Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, Pennsylvania, USA
[2]Biomedical & Translational Informatics, Geisinger Health System, Danville, Pennsylvania, USA

Abstract

Precision medicine, an emerging approach for prevention and treatment strategies by taking into account patients' genetic variability, has new technologies and advances which are moving toward a new era of personalized medicine. It is common that cancer patients have different molecular signatures even though they have similar clinical features such as prognosis due to the heterogeneity of tumors. Thus, new data integration approaches by incorporating biological knowledge such as pathway information need to be developed to overcome the variability of diagnostic or prognostic predictors. Previously, we developed a new approach that identifies knowledge-driven genomic interactions, which are associated with outcomes of interest, based on gene expression data alone. However, no systematic approach has been proposed to identify interaction models between pathways based on multi-omics data, which is called meta-dimensional knowledge-driven genomic interactions. Here we have proposed such a novel methodological framework. To test the utility of the proposed framework, we applied it to multi-omics data in ovarian cancer from TCGA for predicting stage, grade, and survival outcomes. We found that each knowledge-driven genomic interaction model, based on different genomic data sets, contains different sets of pathway features, which suggests each genomic data type may contribute to outcomes in ovarian cancer via different pathways. In addition, meta-dimensional knowledge-driven genomic interaction models significantly outperformed the single knowledge-driven genomic interaction model. From the meta-dimensional knowledge-driven genomic interaction models, many interactions between pathways associated with outcomes were found such as MAPK signaling pathway and the GnRH signaling pathway, both of whichare known to play important roles in cancer pathogenesis. As demonstrated by this study, the beauty of incorporating biological knowledge into the model based on multi-omics data is the ability to improve diagnosis and prognosis and further provide better interpretability. Thus, cancer patients' variability in molecular signatures based on these interactions between pathways may lead to better diagnostic/treatment strategies for better precision medicine.

S4-4: Alternative Polyadenylation in Cardiac Development and Hypertrophy.

Ji Yeon Park[1,2], Mainul Hoque[1], Bei You[1], Dinghai Zheng[1], Ghassan Yehia[3], Peiyong Zhai[4], Junichi Sadoshima[4], Ju Han Kim[2], Bin Tian[1]

[1]Department of Microbiology, Biochemistry and Molecular Genetics, Rutgers New Jersey Medical School, Newark, New Jersey, USA
[2]Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul, Republic of Korea
[3]Transgenic Core Facility, Rutgers New Jersey Medical School, Newark, New Jersey, USA. [4]Department of Cell Biology and Molecular Medicine, Rutgers New Jersey Medical School, Newark, New Jersey, USA

Abstract

Alternative polyadenylation (APA) is an important layer of gene regulation in development and disease. Here, using a special deep sequencing method to examine APA genome-wide, we systematically define APA profiles in murine heart development and cardiac hypertrophy (enlargement of the heart in response to increased workload). Our analysis reveals widespread APA changes in both development and hypertrophy. Distinct and dynamic APA changes take place in different phases of development, involving substantial 3'UTR-APA and CDS-APA events. During hypertrophy, 3'UTRs

of mRNAs are generally shortened through APA, some of which are related to those regulated in development. 3'UTR length changes impact RNA-binding protein-binding sites, including those of CELF and MBNL, two key regulators of RNA metabolism in muscle cells. In addition, we reveal APA impacts various functional pathways relevant to heart functions. Taken together, our results indicate that APA plays a key role in gene regulation in heart development and hypertrophy, opening a new venue for future therapeutic interventions of heart diseases.

S5. Linking Phenotypes

Room: Room 2
Date: Sunday, Nov. 8, 08:30 - 09:45

S5-1: IVIG administration in Kawasaki disease patients causes major methylation alterations on the CpG markers of inflammatory immune associated genes.

Sung-Chou Li[1], Wen-Ching Chan[1], Mindy Ming-Huey Guo[2, 3], Ying-Hsien Huang[3], and Ho-Chang Kuo[2, 3, #]

[1]Genomics and Proteomics Core Laboratory, Department of Medical Research, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Kaohsiung, Taiwan
[2]KD Center, Kaohsiung Chang Gung Memorial Hospital, Kaohsiung, Taiwan.
[3]Department of Pediatrics, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan

Abstract

Kawasaki disease (KD) is an autoimmune disease preferentially attacking children younger than five years worldwide. So far, the principal treatment to KD is the administration of Intravenous immunoglobulin (IVIG). Although DNA methylation plays important regulation roles in diseases, few studies investigated the regulation roles of DNA methylation in KD. In this study, we focused not only on the DNA methylation alterations resulted from KD onset but also on DNA methylation alterations resulted from IVIG administration. To do so, we investigated the effects of KD's onset and IVIG administration on CpG marker's methylation alterations. We first found that DNA methylation alterations reflecting disease onset or IVIG administration are contributed mainly by the CpG markers on autosomes. In addition, we also demonstrated that some CpG markers carry dynamic methylation alteration among samples, forcing the expression abundance of the downstream genes to be also dynamically altered and negatively correlated with methylation profile. Finally, compared with KD's onset, IVIG administration brings stronger impact on methylation alteration. And, such alterations were conducted mainly by hyper-methylating CpG markers, forcing the corresponding genes to keep lower expression levels. Moreover, the genes regulated by the altered CpG markers with IVIG administration are enriched in the pathways associated with inflammatory immune response. In summary, our result provides researchers with another way into the regulation mechanism through which IVIG represses excessive inflammatory responses.

S5-2: Semantics based approach for analyzing disease-target associations.

Rama Kaalia[1] and Indira Ghosh[1]

[1]Center for School of Computational & Integrative sciences, Jawaharlal Nehru University, New Delhi, 110067, India

Abstract

A complex disease is caused by heterogeneous biological interactions between genes and their products along with the influence of environmental factors. There have been many attempts for understanding the cause of these diseases using experimental, statistical and computational methods. In the present work we have used a semantics-based approach to address the challenge of representation and integration of information from heterogeneous biomedical aspects of a complex disease. We have developed a Disease Association Ontology for Diabetes (DAO-db) that provides a standard ontology-driven platform for describing genes, proteins, pathways involved in diabetes and for integrating functional associations from various interaction levels (gene-disease, gene-pathway, gene-function, gene-cellular component and protein-protein interactions). An automatic instance loader module is also developed in present work that helps in adding instances to DAO-db on a large scale.
Our ontology provides a framework for querying and analyzing the disease associated information in the form of RDF graphs. Three semantic web based scoring algorithms (PageRank, HITS (Hyperlink Induced Topic Search) and HITS with semantic weights) were used to score the gene nodes on the basis of their functional interactions in the graph. The above developed protocol was used to predict novel potential targets involved in diabetes disease from the long list of loose (statistically associated) gene-disease associations.

S5-3: Inferring Crohn's disease association from exome sequences by integrating biological knowledge.

Chan-Seok Jeong[1] and Dongsup Kim[1]

[1]Department of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-338, Republic of Korea

Abstract

Exome sequencing has been emerged as a primary method to iden- tify detailed sequence variants associated with complex diseases including Crohn's disease in the protein-coding regions of human genome. How- ever, constructing an interpretable model for exome sequencing data is challenging because of the huge diversity of genomic variation. In addi- tion, it has been known that utilizing biologically relevant information in a rigorous manner is essential for effectively extracting disease-associated information. In this paper, we incorporate three different types of biolog- ical knowledge such as predicted pathogenicity, disease gene annotation, and functional interaction network of human genes, and integrate them with exome sequence data in non-negative matrix tri-factorization frame- work. Based on the proposed method, we successfully identified Crohn's disease patients from exome sequencing data and achieved the area under the receiver operating characteristics curve (AUC) of 0.816, while other clustering methods not using biological information achieved the AUC of 0.786. Moreover, the disease association score derived from our method showed higher correlation with Crohn's disease genes than other unrelated genes. As a consequence, by integrating biological information across mul- tiple levels such as variant, gene, and systems, our method could be useful for identifying disease susceptibility and its associated genes from exome sequencing data.

Scientific Paper Sessions

S6. New Technologies

Room: Room 2
Date: Sunday, Nov. 8, 14:25 - 15:40

S6-1: EyeSee; an Assistive Device for Blind Navigation with Multi-Sensory Aid.

Catherine Todd[1,*], Mohamed Watfa[1], Mohammad Albatat[1] and Amish Suchak[1]

[1]University of Wollongong in Dubai, Block 15, Knowledge Village, Dubai, 20183, United Arab Emirates.

Abstract.

EyeSee is a sensory aid comprising two interlinked components including a pair of sensor-driven glasses and a hand glove that provide a blind user with audio and vibration feedback of distance and movement of surrounding objects, including humans. While the glasses provide feedback for outdoor use, the hand glove is intended for indoor navigation. In comparison to assistive devices that provide minimal surface-based tactile feedback, such as the white cane, EyeSee provides an enhanced sensory experience with audio feedback through bone conduction as well as a GPS locator to notify users' contacts in the event of an emergency.

S6-2: PATTERN: Pain Assessment for paTients who can't TEll using Restricted Boltzmann machine.

Lei Yang[1,*], Shuang Wang[2,*], Xiaoqian Jiang[2], Samuel Cheng[1] and Hyeon-Eui Kim[2]

[1]Department of Electrical and Computer Engineering, University of Oklahoma, Tulsa, OK 74135, USA
[2]Department of Biomedical Informatics, University of California, San Diego,CA 92093, USA

Abstract

Accurately assessing pain for those who cannot make self-report of pain, such as minimally responsive or severely brain-injured patients, is challenging. In this paper, we attempted to address this challenge by answering the following questions: (1) if the pain has dependency structures in electronic signals and if so, (2) how to apply this pattern in predicting the state of pain. To this end, we have been investigating and comparing the performance of several machine learning techniques. We first adopted different strategies, in which the collected original · - dimensional numerical data was converted into binary data. Pain states are represented in binary format and bound with above binary features to construct (· + 1)-dimensional data. We then modeled the joint distribution over all variables in this data using the Restricted Boltzmann Machine (RBM). The experimental results show that RBM is able to model the distribution of our binary pain data. In addition, we show that discriminant RBM can be used in classification task, and the initial result is competitive with respect to other classifier such as support vector machine (SVM) using PCA representation and LDA discriminant method.

Scientific Paper Sessions

S6-3: Creative Activity Aid using Active Tremor Cancellation.

Catherine Todd[1,*], Mohamed Watfa[1], Shaikha Albadi[1], Nour El Sayed[1], Mina Makary[1] and Omar Mahmoud[1]

[1]University of Wollongong in Dubai, Block 15, Knowledge Village, Dubai, 20183, United Arab Emirates.

Abstract

Involuntary human tremor due to clinical disorders such as in Parkinson's disease, Multiple Sclerosis (MS), effect of a stroke or a traumatic brain injury, inhibit daily activities and can severely restrict lifestyle. Such tremor is associated with rhythmic movement of one or more body parts, where muscles contract and relax, causing twitching or sudden movement. Symptoms are most commonly the result of a psychological illness that causes the muscles to act involuntarily (Hallet, 1998). Medical treatments to human tremor may pose harmful side effects, including hallucinations, and drugs used for treatment have varying degrees of success in reducing the symptoms of the illness (Lieberman and Dhanani, 2015). Our research provides a new, assistive technology that minimizes the effects of hand tremor, as a substitute for drug- related therapy. The device is intended for use in creative art; a paint brush attachment,

for application in painting. The engineering solution promises greater application for attachment to a variety of implements, such as writing or instructional implements, for stabilization of implements used by persons suffering from hand tremor, irrespective of the medical cause.

S7. Bioinformatics Algorithms

Room: Hall A
Date: Sunday, Nov. 8, 15:25 - 16:40

S7-1: Fast Comparison of Genomic and Meta-Genomic Reads with Alignment-Free Measures based on Quality Values.

Matteo Comin[1] and Michele Schimd[1]

[1]Department of Information Engineering, University of Padova, Padova, Italy,

Abstract

Sequencing technologies are generating enormous amounts of read data, however the assembly of genomes and metagenomes is still one of the most challenging task. In this paper we study the comparison of genomes and metagenomes based only on read data, thus without the need of reference genomes or assemblies, using word counts statistics called alignment-free.
Quality scores produced by sequencing platforms are fundamental for various analysis, moreover future-generation sequencing platforms, will produce long reads but with an error rate around 15%. In this context it will be fundamental to exploit quality value information within the framework of alignment-free measures.
In this paper we present a family of alignment-free measures, called dq-type, that are based on k- mer counts and quality values. These statistics can be used to compare genomes and metagenomes based on their read sets. Results show that the evolutionary relationship of genomes can be recon- structed based on the direct comparison of theirs reads sets. The use of quality values on average improves the classification accuracy, and its contribution increases when the reads are more noisy. Also the comparison of metagenomic microbial communities can be performed efficiently. Similar metagenomes

are quickly detected, just by processing their read data, without the need of costly alignments.

Scientific Paper Sessions

S7-2: Parametric analysis of RNA-seq expression data.

Tomokazu Konishi[1]

[1]*Faculty of Bioresource Sciences, Akita Prefectural University, Akita 0100195, Japan*

Abstract

As was expected, depth data of RNA-seq was lognormally distributed; however, the level of noise was higher than hybridization-based methodology such as microarrays. Although the effect of noise derived from the Bernoulli trial was rather limited, some of the transcripts might have been overlooked. The overlooking occurred with tendency toward lower range of data. The range that would be affected were found in the normalization process; to avoid false discoveries, such range of data should not be used in analyses. To find out overlooking, number of replications would be more important than read depth; as cause of overlooking might be in sample preparation, more depth would not improve the accuracy. The appropriate distribution model would improve certainty and accuracy of analyses.

S7-3: Nearest Neighbor Imputation Algorithms: A Critical Evaluation.

Lorenzo Beretta[1] **and Alessandro Santaniello**[1]

[1]*Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy*

Abstract

Nearest neighbor (NN) imputation algorithms are popular methods to fill in missing data where each missing value on some records is replaced by a value obtained from related cases in the whole set of records. Besides the capability to substitute the missing datum with plausible values that are as close as possible to the true value, imputation algorithms should preserve the original data structure and avoid to distort the distribution of the imputed variable. Despite the efficiency of NN algorithms little is known about the effect of these methods on data structure. Simulation on synthetic datasets with different patterns and degrees of missingness were conducted to evaluate the performance of NN with one single neighbor (1NN) and with k neighbors without (kNN) or with weighting (wkNN) in the context of different learning frameworks: plain set, reduced set after ReliefF filtering, bagging, random choice of attributes, bagging combined with random choice of attributes. Whatever the framework, kNN usually outperformed 1NN in terms of precision of imputation and reduced errors in inferential statistics, 1NN was however the only method capable of preserving the data structure and data were distorted even when small values of k neighbors were considered; distortion was more severe for resampling schemas. The use of 3 neighbors in conjunction with ReliefF seems to provide the best trade-off between imputation error and preservation of the data structure.

Scientific Paper Sessions

S8. GWAS and Post-GWAS

Room: Room 2
Date: Sunday, Nov. 8, 15:25 - 16:40

S8-1: Genome-wide association study identifies novel susceptibility genes associated with coronary artery aneurysm formation in Kawasaki Disease.

Ho-Chang Kuo[1], Sung-Chou Li[2], Mindy Ming-Huey Guo[1], Ying-Hsien Huang[1], Wen- Ching Chan[2,*]

[1]Department of Pediatrics and Kawasaki Disease Center, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan
[2]Genomics and Proteomics Core Laboratory, Department of Medical Research, Kaohsiung Chang Gung Memorial Hospital and Chang Gung University College of Medicine, Kaohsiung, Taiwan

Abstract

Kawasaki disease (KD) or Kawasaki syndrome is known as a vasculitis of small to medium-sized vessels, and coronary arteries are predominantly involved in childhood. Generally, 20-25% of untreated with IVIG and 3-5% of treated KD patients have been developed coronary artery lesions (CALs), such as dilatation and aneurysm. Understanding how coronary artery aneurysms (CAAs) are established and maintained in KD patients is therefore of great importance. Upon our previous genotyping data of 157 valid KD subjects, a genome-wide association study (GWAS) has been conducted among 11 (7%) CAA-developed KD patients to reveal five significant genetic variants passed pre-defined thresholds and resulted in two novel susceptibility protein-coding genes, which are NEBL (rs16921209 ($P = 7.44 \times 10{-}9$; OR = 32.22) and rs7922552 ($P = 8.43 \times 10{-}9$; OR = 32.0)) and TUBA3C (rs17076896 ($P = 8.04 \times 10{-}9$; OR = 21.03)). Their known functions have been reported to associate with cardiac muscle and tubulin, respectively. As a result, this might imply their putative roles of establishing CAAs during KD progression. Additionally, various model analyses have been utilized to determinant dominant and recessive inheritance patterns of identified susceptibility mutations. Finally, all susceptibility genes hit by significant genetic variants were further investigated and the top three representative gene-ontology (GO) clusters were regulation of cell projection organization, neuron recognition, and peptidyl- threonine phosphorylation. Our results help to depict the potential routes of the pathogenesis of CAAs in KD patients and will facilitate researchers to improve the diagnosis and prognosis of KD in personalized medicine.

S8-2: Integrative Regression Network for Genomic Association Study.

Reddy Rani Vangimalla[1], Hyun-hwan Jeong[1], Kyung-Ah Sohn[1,*]

[1]Department of Information and Computer Engineering, Ajou University, Suwon 443-749, Republic of Korea

Abstract

The increasing availability of multiple types of genomic profiles measured on the same cancer patients has been providing a lot of opportunities for investigating genomic mechanisms underlying cancer. This study is about identifying genomic associations on multiple high-dimensional genomic profiles by taking into account the association between as well as within profiles. The conventional correlation-based association tests have the limitation of being prone to indirect associations. In this work, we employ high-dimensional regression techniques to first identify associations between different genomic profiles and based on the resulting regression coefficients, a regression network is constructed within each profile. The main motivation is that two gene features having similar regression coefficients with respect to a number of traits are likely to be involved in the same biological process and to have equally prominent weightage for the traits. To extract more reliable associations, multiple sparse structured regression techniques are applied and the resulting multiple networks are merged together by similarity fusion technique.
Experiments were carried out on different regression methods and various cancer types. The pros and cons of each regression method are explored as well, which shows that considering structural information is important for the association study. Combining all method's regression coefficients can have a biasing effect, but fusing multiple regression networks by using similarity measurements led the study to discovery significant gene pairs and a resulting network with better topological properties.

S8-3: eMERGE Phenome-Wide Association Study (PheWAS) Identifies Clinical Associations and Pleiotropy for Stop-Gain Variants.

Anurag Verma[1,2,*], Shefali S. Verma[1,2,*], Sarah A. Pendergrass[2], Dana C. Crawford[4], David R. Crosslin[5], Helena Kuivaniemi[3,11], William S. Bush[4], Yuki Bradford[1], Iftikhar Kullo[8], Suzette J. Bielinski[8], Rongling Li[9], Joshua C. Denny[6], Peggy Peissig[7], Scott Hebbring[7], Mariza de Andrade[8], Marylyn D. Ritchie[1,2], Gerard Tromp[3,10]

[1]Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA
[2]Biomedical and Translational Informatics, Geisinger Health System, Danville, PA
[3]The Sigfried and Janet Weis Center for Research, Geisinger Health System, Danville, PA, USA
[4]Case Western Reserve University, Cleveland, OH
[5]Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA
[6]Vanderbilt University, Nashville, TN
[7]Marshfield Clinic, Marshfield, WI
[8]Mayo Clinic, Rochester, MN
[9]National Human Genome Research Institute, Bethesda, MD
[10]Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, 7505, South Africa

Abstract

We explored premature stop-gain variants to test the hypothesis that variants, which are likely to have a consequence on protein structure and function, will reveal important insights with respect to the phenotypes associated with them. We performed a phenome-wide association study (PheWAS) exploring the association between a selected list of functional stop-gain genetic variants (variation resulting in truncated proteins or in nonsense-mediated decay) and an extensive group of diagnoses to identify novel associations and uncover potential pleiotropy. In this study, we selected 25 stop-gain variants: 5 stop-gain variants with previously reported phenotypic associations, and a set of 20 putative stop-gain variants identified using dbSNP. For the PheWAS, we used data from the electronic MEdical Records and GEnomics (eMERGE)

Network across 9 sites with a total of 41,057 unrelated patients. We divided all these samples into two datasets by equal proportion of eMERGE site, sex, race, and genotyping platform. We calculated single effect associations between these 25 stop-gain variants and ICD-9 defined case-control diagnoses. We also performed stratified analyses for samples of European and African ancestry. Associations were adjusted for sex, site, genotyping platform and the first three principal components to account for global ancestry. We identified previously known associations, such as variants in LPL associated with hyperglyceridemia indicating that our approach was robust. We also found a total of three significant associations with p< 0.01 in both datasets, with the most significant replicating result being LPL SNP rs328 and ICD-9 code 272.1 “Disorder of Lipoid metabolism” (pdiscovery = 2.59x10-6, preplicating = 2.7x10-4). The other two significant replicated associations identified by this study are: variant rs1137617 in KCNH2 gene associated with ICD-9 code category 244 “Acquired Hypothyroidism” (pdiscovery = 5.31x103, preplicating = 1.15x10-3) and variant rs12060879 in DPT gene associated with ICD-9 code category 996 “Complications peculiar to certain specified procedures” (pdiscovery = 8.65x103, preplicating = 4.16x10-3) In conclusion, this PheWAS revealed novel associations of stop-gained variants with interesting phenotypes (ICD-9 codes) along with pleiotropic effects.

S9. Biomedical Big Data

Room: Room 2
Date: Monday, Nov. 9, 08:30 - 09:45

S9-1: DataRank: A Framework for Ranking Biomedical Datasets.

Arya Iranmehr[1], Huan Wang[2], Hannah Chen[2], and Xiaoqian Jiang[3]

[1]*Department of Electrical and Computer Engineering, UC San Diego, USA*
[2]*Department of Computer Science, UC San Diego, USA*
[3]*Department of Biomedical Informatics, UC San Diego, USA*

Abstract

Due to the advent of high-performance technologies for the generation, storage and processing of data, research in biomedical sciences is increasingly moving towards "data-driven research" via the manipulation and analysis of biomedical datasets. Unfortunately, the vast majority of these datasets are underutilized after their initial publications due to an increased rate of dataset generation and difficulty in searching for relevant datasets. In this paper, we develop the DataRank framework to rank biomedical datasets by incorporating three different criteria into our model: query-relevance, research-importance and user-preference. We implement query-relevance via a multi-label classification approach using binary relevance schemes. Research-importance of datasets is incorporated into ranking by utilizing a citation network. Preference ranking is implemented in an online setting where DataRank re-ranks searching results according to user feedback. Mean Average Precision (MAP) and Mean Reciprocal Rank (MRR) performance measures on DataRank query results are compared against search results from the Gene Expression Omnibus (GEO) data repository and Jaccard index method. The comparisons conducted on five different held-out validation set of queries and show that DataRank achieves higher MAP and MRR. Finally, we assess the effect of incorporating user feedback by measuring Regret for five different subjects and show that this measure is non-increasing.

S9-2: Many larger worlds in a small world.

Sunmin Yun[1,2], and Ju Han Kim[1, 2,*]

[1]Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea

[2]Systems Biomedical Informatics Research Center, Seoul National University, Seoul 110799, Korea

Abstract

Biological pathways is a biological system, is applied to biological network, consisting of a series of molecular actions in a cell and it can be classified into metabolism pathways, signal transduction pathways and genetic pathways as their functions. The members of each pathway are annotated to proteins, which control mostly biological systems in a cell. While many proteins perform their functions independently, the most of proteins interact with others for the biological activity. To understand the characteristics of pathways as biological network, we studied them by the classification of pathways as metabolism and the rest of pathways. The published genomic knowledge database like pathways, proteins and protein domains, which functional components constituting the protein downloaded and then the topology of each pathway and networks properties like density, average clustering coefficient and average path distance analyzed using their interactions. In addition, we investigated subcellular localization of components that make up the biological pathways. The results of this research show that there are differences between metabolism pathways and non-metabolism pathways in network topology, network properties and protein subcellular localization. Metabolism is relatively less complex than non-metabolism and not follows scale-free network. As a result, metabolism pathway is larger world than non-metabolism pathway, and biological network is constituted by many larger worlds in a small world.

S9-3: CLASH: Complementary Linkage with Anchoring and Scoring for Heterogeneous BioMolecular and Clinical Data

Yonghyun Nam[1], Myungjun Kim[1], and Hyunjung Shin[1,*]

[1]Department of Industrial Engineering, Ajou University, Wonchun-dong, Yeongtong-gu, Suwon 443-749, South Korea

Abstract

The study on disease-disease association has been increasingly viewed and analyzed as a network, in which the connections between diseases are configured using the source information on interactome maps of biomolecules such as genes, proteins, metabolites, etc. More abundance in source information leads to tighter connections between diseases in the network. However, for a certain group of diseases, e.g. metabolic diseases, the connections do not occur much because the source information is not sufficient; a large proportion of their associated genes are still unknown. One way to circumvent the difficulties in the lack of source information is to use available external information. In this study, we propose an algorithm that complement or strengthen connections between nodes in a disease network through jointly using outsourced information. When applied to the network of metabolic diseases that is sourced from protein-protein interaction data, the proposed algorithm recovered connections by 97%, and improved the AUC performance up to 0.71 (lifted from 0.55) by using the external information outsourced from text mining results on PubMed comorbidity literatures.

Highlight Research Tracks

Highlight Research

Room: Hall A
Date: Sunday, Nov. 8, 13:00 - 14:15

H1-1: Extraction of Pharmacokinetic Evidence of Drug–Drug Interactions from the Literature

Artemy Kolchinsky[1,2], Anália Lourenço[3,4], Heng-Yi Wu[5], Lang Li[5], Luis M. Rocha[1,2,*]

[1]School of Informatics and Computing, Indiana University, Bloomington, IN, USA
[2]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[3]ESEI: Escuela Superior de Ingeniería Informática, University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n 32004, Ourense, Spain
[4]CEB—Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[5]Center for Computational Biology and Bioinformatics and Department of Medical & Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA

Abstract

Drug-drug interaction (DDI) is a major cause of morbidity and mortality and a subject of in- tense scientific interest. Biomedical literature mining can aid DDI research by extracting evi- dence for large numbers of potential interactions from published literature and clinical databases. Though DDI is investigated in domains ranging in scale from intracellular bio- chemistry to human populations, literature mining has not been used to extract specific types of experimental evidence, which are reported differently for distinct experimental goals. We focus on pharmacokinetic evidence for DDI, essential for identifying causal mechanisms of putative interactions and as input for further pharmacological and pharma- coepidemiology investigations. We used manually curated corpora of PubMed abstracts and annotated sentences to evaluate the efficacy of literature mining on two tasks: first, identifying PubMed abstracts containing pharmacokinetic evidence of DDIs; second, ex- tracting sentences containing such evidence from abstracts. We implemented a text mining pipeline and evaluated it using several linear classifiers and a variety of feature transforms. The most important textual features in the abstract and sentence classification tasks were analyzed. We also investigated the performance benefits of using features derived from PubMed metadata fields, various publicly available named entity recognizers, and pharma- cokinetic dictionaries. Several classifiers performed very well in distinguishing relevant and irrelevant abstracts (reaching F1%0.93, MCC%0.74, iAUC%0.99) and sentences (F1%0.76, MCC%0.65, iAUC%0.83). We found that word bigram features were important for achieving optimal classifier performance and that features derived from Medical Subject Headings (MeSH) terms significantly improved abstract classification. We also found that some drug- related named entity recognition tools and dictionaries led to slight but significant improve- ments, especially in classification of evidence sentences. Based on our thorough analysis of classifiers and feature transforms and the high classification performance achieved, we demonstrate that literature mining can aid DDI discovery by supporting automatic extraction of specific types of experimental evidence.

H1-2: Diagnostic Role of Exome Sequencing in Immune Deficiency Disorders

Aashish N. Adhikari[1], Jay P. Patel[2], Alice Y. Chan[3], Divya Punwani[3], Haopeng Wang[3], Antonia Kwan[3], Theresa A. Kadlecek[3], Morton J.Cowan[3], Marianne Mollenauer[3], John Kuriyan[1], Shu Man Fu[4], Uma Sunderam[5], Sadhna Rana[5], Ajithavalli Chellappan[5], Kunal Kundu[5], Arend Mulder[6], Frans H J Claas[6], Joseph A Church[7], Arthur Weiss[3], Richard A Gatti[8], Jennifer M. Puck[3], Rajgopal Srinivasan[5], Steven E. Brenner[1,*]

[1]*University of California, Berkeley, CA, USA*
[2]*Children's Hospital of Los Angeles, Los Angeles, CA, USA*
[3]*University of California, San Francisco, CA, USA*
[4]*University of Virginia School of Medicine, Charlottesville, VA, USA*
[5]*Innovation Labs, Tata Consultancy Services Hyderabad, AP, India*
[6]*Leiden University Medical Centre, Leiden, The Netherlands*
[7]*University of Southern California, Los Angeles, CA, USA*
[8]*University of California Los Angeles, CA, USA*

We developed an analysis protocol for individual genome interpretation and used its distinctive features to diagnose numerous clinical cases. We applied the protocol to exomes from newborn patients with undiagnosed primary immune disorders. To yield high quality sets of possible causative variants, we used multiple callers with multisample calling and integrated variant annotation, variant filtering, and gene prioritization.
In two unrelated infant immunodeficient girls with no diagnoses, we discovered compound heterozygous variants in the ATM gene for both the infants offering a very early diagnosis of Ataxia Telangiectasia (AT) which allowed for avoidance of undue irradiation and live vaccinations.
In another case, the affected siblings had early onset bullous phempigoid, a chronic autoimmune disorder. Our analysis revealed compound heterozygous mutations in ZAP70, a gene associated with profound primary immunodeficiency, the opposite phenotype. Cellular immunological studies indicated that one variant was hypomorphic and the other was hyperactive. These combined to yield a novel presentation, adding to the existing phenotype repertoire of ZAP70 in humans.
Our analysis protocol focuses on genomic features that may be overlooked by other methods. In the case of a female with severe influenza pneumonia, our annotation tool, Varant, flagged variants in PRKDC apparently occurring after the genomically-encoded stop codon. This stop codon in the reference genome was correctly identified as premature due to a rare single base deletion. But accounting for the proband being normal at this position, we correctly annotated the proband's two variants as nonsynonymous mutations likely causative for the phenotype.
Our protocol has been similarly revealing in other SCID and CID cases including Nijmegen Breakage Syndrome, which highlight unique features of the analysis framework that facilitate genetic discovery. These help provide crucial information to offer prompt appropriate treatment, family genetic counseling, and avoidance of diagnostic odyssey.

H1-3: hiHMM: Bayesian non-parametric joint inference of chromatin state maps.

Kyung-Ah Sohn[1,2,3,†], Joshua W. K. Ho[4,5,6,7,†], Djordje Djordjevic[4,5], Hyun-hwan Jeong[1], Peter J. Park[6,7,*] and Ju Han Kim[2,3,*]

[1]Department of Information and Computer Engineering, Ajou University, Suwon 443-749, South Korea
[2]Seoul National University Biomedical Informatics (SNUBI), Division of Biomedical Informatics, Seoul National University College of Medicine, Seoul 110799, Korea
[3]Systems Biomedical Informatics Research Center, Seoul National University, Seoul 110799, Korea
[4]Victor Chang Cardiac Research Institute, Sydney, NSW 2010, Australia
[5]The University of New South Wales, Sydney, NSW 2052, Australia
[6]Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA
[7]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

Abstract

Motivation: Genome-wide mapping of chromatin states is essential for defining regulatory elem- ents and inferring their activities in eukaryotic genomes. A number of hidden Markov model (HMM)-based methods have been developed to infer chromatin state maps from genome-wide his- tone modification data for an individual genome. To perform a principled comparison of evolution- arily distant epigenomes, we must consider species-specific biases such as differences in genome size, strength of signal enrichment and co-occurrence patterns of histone modifications.
Results: Here, we present a new Bayesian non-parametric method called hierarchically linked infin- ite HMM (hiHMM) to jointly infer chromatin state maps in multiple genomes (different species, cell types and developmental stages) using genome-wide histone modification data. This flexible framework provides a new way to learn a consistent definition of chromatin states across multiple genomes, thus facilitating a direct comparison among them. We demonstrate the utility of this method using synthetic data as well as multiple modENCODE ChIP-seq datasets.
Conclusion: The hierarchical and Bayesian non-parametric formulation in our approach is an important extension to the current set of methodologies for comparative chromatin landscape analysis.
Availability and implementation: Source codes are available at https://github.com/kasohn/hiHMM. Chromatin data are available at http://encode-x.med.harvard.edu/data_sets/chromatin/.

Posters Session Room1

TBC-1: Diana Milena Gaitan-Vaca, Luisa Fernanda Daza Martinez, Juan David Henao-Sanchez, Daniel Camilo Osorio-Hurtado, Juan Lugo and Andres Mauricio Pinzon-Velasco

Computational model of Parkinson's disease suggests molecular mechanisms of nicotine, caffeine and ibuprofene neuroprotection on dopaminergic cells.

TBC-2: Hoim Jeong and Heui-Soo Kim

Biomarker for selecting the superior working dog using statistical microsatellite analysis.

TBC-3: Eun-Bi Kim, Bo-Kyeong Chang, Chang-Sik Pak and Chan-Yeong Heo

A U-health care solution about a chronic wound management with the mobile application.

TBC-4: Henrik Stranneheim, Valtteri Wirta and Anna Wedell

Rapid pulsed whole genome sequencing for comprehensive acute diagnostics of inborn errors of metabolism.

TBC-5: Ali Yousefian-Jazi and Jinwook Choi

FCM-EM: A Novel Approach for Discovering DNA Motifs

TBC-6: Hyesil Jung and Hyeoun-Ae Park

Evaluation of the content coverage of an adolescents' depression ontology.

TBC-7: Jeong-Eun Lee Lee, Ali Yousefian Jazi, Moon-Woo Seong and Jinwook Choi

Clinical Interpretation Pipeline of Targeted Next-Generation Sequencing Cardiovascular Panels.

TBC-8: Inge Seim, Pat Thomas, John Lai, Penny Jeffery and Lisa Chopin

Re-analysis of RNA-sequencing data reveals cancer-specific chimaeric transcripts.

TBC-9: Siriwon Taewijit and Tu Bao Ho

A Two-Stage Approach for Multivariate Infrequent Adverse Drug Events Analysis in Electronic Medical Records.

TBC-10: Shinuk Kim

Time dependent pathway regulation for cancer development.

TBC-11: Yuki Nakayama, Ryosuke Matsuo and Tu Bao Ho

Medical Synonym Extraction with Dual Space Model.

TBC-12: Gandhimathi Moharasan and Tu Bao Ho

Multilayered approach to extract temporal events and expressions from clinical narratives.

TBC-13: Ryosuke Matsuo and Tu Bao Ho

A Severity based Information Retrieval for Electronic Medical Records.

Posters Session Room1

TBC-14: Siming Ma and Vadim Gladyshev

Organization of the Mammalian Metabolome and Ionome according to Organ Function, Lineage Specialization, and Longevity.

TBC-15: Steven Brenner, John Moult and CAGI Participants

Findings from the Critical Assessment of Genome Interpretation, a community experiment to evaluate phenotype prediction.

TBC-16: Onkar Singh, Emily Chia-Yu Su and Hong Jie Dai

Building a Metastasis Gene Database using Principle-based Approach.

TBC-17: Joon Seol Bae

Comprehensive analysis of pediatric neuroblastoma using high depth cancer gene panel sequencing.

TBC-18: Je-Gun Joung

Tumor heterogeneity of advanced primary lung cancer evaluated by multiregion sequencing.

TBC-19: Chae-Gyun Lim, Hyo Jin Do and Ho-Jin Choi

Measuring Emotional Well-being Based on Physiological Indicators and Lifestyles.

TBC-20: Ben-Yang Liao and Meng-Pin Weng

Unraveling the association between mRNA expressions and mutant phenotypes in a genome-wide assessment of mice.

TBC-21: Hu Yu Hsuan

Prevalence and Outcome of Seizure related to Systemic Lupus Erythematosus.

TBC-22: Usman Iqbal, Phung-Anh Nguyen, Shabbir Syed-Abdul, Richard Lu, Hsuan-Chia Yang, Chih-Wei Huang, Wen-Shan Jian and Yu-Chuan Jack Li

Benzodiazepines use and breast cancer risk: A population-based study.

TBC-23: Shabbir Syed-Abdul, Hong-Jie Dai, Po-Ting Lai, Usman Iqbal, Phung-Anh Nguyen, Richard Lu, Hsuan-Chia Yang, Chih-Wei Huang, Wen-Shan Jian and Yu-Chuan Jack Li

A web browser extension to display phenome-wide association strengths observed in Taiwanese population: PWAS-PubMed.

TBC-24: Kaavya A Mohanasundaram, Mani P Grover, Tamsyn M Crowley, Andrzej Goscinski, Merridee A Wouters

In silico prediction of pathogenicity of missense substitutions: a perspective of complex diseases.

Posters Session Room1

TBC-25: Chih-Wei Huang, Shen-Hsien Lin, Phung Anh Nguyen, Usman Iqbal, Wen-Shan Jian, Yu-Chuan (Jack) Li

Using Data Visualization Method to Estimate and Predict Chronic Kidney Disease Medical Expenditures.

TBC-26: Tony Kuo, Jun Sese, Martin Frith, and Paul Horton

A Statistical Model for the Refinement and Ranking of Variant Calls.

TBC-1: Computational model of Parkinson's disease suggests molecular mechanisms of nicotine, caffeine and ibuprofene neuroprotection on dopaminergic cells.

Gaitan-Vaca DM[1] , Daza-Martinez LF[1] , Henao-Sanchez JD[1] , Osorio-Hurtado DC, Lugo J & Pinzon-velasco AM.[1,*]

[1]Bioinformatics and Computational Systems Biology Laboratory, Institute for Genetics, National University of Colombia.

Abstract

Parkinson's disease (PD) is a neurodegenerative disease affecting approximately 2 of every 1000 older adults and characterized by loss of dopaminergic cells in the substantia nigra in the brain (Hernandez, et al. 2006). To date there are limited options for its treatment, therefore the characterization of neuroprotective agents that could provide options for modifying its progression, or to reduce the risk of disease onset has been an active field of research (Hernán, 2002). Caffeine (Ascherio, et al. 2000), nicotine (Sugita, et al. 2001) and ibuprofene (chen, et al. 2005) have been three common agents largely associated to neuroprotection in PD, although the molecular mechanisms underlying this protective characteristic are yet to be understood.

Animal models have been typically used for investigating the disease and some of its neuroprotection mechanims (Lee, et al. 2013), but not all results can be transferred to human physiology and not all experimental procedures can be easily perform on these biological models. Therefore, the use of mathematical/computational models have proven to be a valuable alternative for predicting specific molecular processes underlying this and other neurodegenerative diseases (Lewis, et al. 2010).

Based on previously reported works (Buchel, et al. 2013) we extended and refined a computational model of 150 biochemical reactions of a dopaminergic nerve cell and predicted potential molecular mechanisms of nicotine, caffeine and ibuprofene neuroprotection in PD.

All together, our results suggest that ibuprofene can be associated to neuroprotection on both, healthy and unhealthy cells through the activation of extracelular tyrosine transport to the cytosol and the synthesis and transport of dopamine. Our research have also suggested that on unhealthy cells, nicotine appears to have an strong positive effect enhancing the metabolic flux on all reactions related to dopamine synthesis. Also on unhealthy cells, caffeine intake led to a reduction on dopamine release from cytosol.

TBC-2: Biomarker for selecting the superior working dog using statistical microsatellite analysis.

Hoim Jeong[1] and Heui-Soo Kim[1,*]

[1]Department of Biological Sciences, College of Natural Sciences, Pusan National University, Busan 609-735, Republic of Korea

Abstract

Dogs (Canis familiaris) have been around humans for 12,000 years or more. Genetic background or external environment allow many dog breeds to be trained to perform highly specialized tasks such as detecting drugs and guiding visually impaired people. In this study, we selected 50 Sapsaree dogs, a Korean breed that has the potential as a working dog. They underwent training and were scored during the in-training examination; given their scores, they were divided into pass or fail groups. We analyzed genetic difference between pass and fail individuals using genotyping of 13 microsatellite loci. The mean number of alleles, allelic richness, and observed heterozygosity of the pass group were 4.077, 4.061, and 0.489, respectively, whereas those of the fail group were 4.154, 4.138, and 0.465, respectively. Using the genetic information, we constructed statistics modeling. The results of both logistic regression analysis and decision tree indicated that the superiority of Sapsaree was determined by 166 repeats of bases (166 allele) and 164 repeats (164 allele) at TAT locus and 198 repeats (198 allele) of DRPLA gene. The results presented herein indicate that these allelic differences between the pass and fail group can be a good biomarker for selection of superior Sapsaree individuals, and our statistical modeling could provide the standard for selection of working dogs.

TBC-3: A U-health Care Solution about a Chronic Wound Management with the Mobile Application.

Eun-Bi Kim[1], Bo-Kyeong Chang[1], Chang-Sik Pak[1] , Chan-Yeong Heo[1]

[1]Seoul National University Bundang Hospital, Republic of Korea

Abstract

Backgrounds: A chronic wound care needs many human resources and infrastructure that maybe affect spatial and temporal limit. In the present condition, breaking the spatial and temporal limit makes economic difficulties to majority of its patients and family of patients. This study starts from solving these problems. We think mobile application can be solution of these problems. This study investigated the effect of mobile application of wound care through in clinical study.

Methods: This study used the mobile application that perform to evaluate pressure sore status tool (PSST) and to select the recommended dressing type. The selected dressing type data and the value of evaluated PSST were obtained 4 times (0, 1, 2, 4week). Then, we compared the PSST total score of the telemedicine data using mobile application and the field data at 0week and 4week. Also, we compared data of dressing type telemedicine and field at those 4 times. We checked the accuracy of the accordance rate of these compared data by Kappa value.
Results: Dressing recommendation accordance rate was very high in the first and the second dressing both. Especially, the researchers conducted dressing directly and dressing formulated remotely results were consistent almost 100%. The recovery rate about a chronic wound was evaluated using the PSST score at 86.2%. And the PSST score was decreased up to -17 points.
Conclusion: We think our study shows feasibility of telemedicine in chronic wound care and helps patient who was participated in clinical study of this study to get over the various problems.

TBC-4: Rapid pulsed whole genome sequencing for comprehensive acute diagnostics of inborn errors of metabolism.

Henrik Stranneheim[1,2,*], Valtteri Wirta[3] and Anna Wedell[1,2]

[1]Department of Molecular Medicine and Surgery, Science for Life Laboratory, Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden.
[2]Centre for Inherited Metabolic Diseases, Karolinska University Hospital, Stockholm, Sweden.
[3]Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden.

Abstract

Background: Massively parallel DNA sequencing (MPS) has the potential to revolutionize diagnostics, in particular for monogenic disorders. Inborn errors of metabolism (IEM) constitute a large group of monogenic disorders with highly variable clinical presentation, often with acute, nonspecific initial symptoms. In many cases irreversible damage can be reduced by initiation of specific treatment, provided that a correct molecular diagnosis can be rapidly obtained. MPS thus has the potential to significantly improve both diagnostics and outcome for affected patients in this highly specialized area of medicine.
Results: We have developed a conceptually novel approach for acute MPS, by analysing pulsed whole genome sequence data in real time, using automated analysis combined with data reduction and parallelization. We applied this novel methodology to an in-house developed customized work flow enabling clinical-grade analysis of all IEM with a known genetic basis, represented by a database containing 474 disease genes which is continuously updated. As proof-of-concept, two patients were retrospectively analysed in whom diagnostics had previously been performed by conventional methods. The correct disease-causing mutations were identified and presented to the clinical team after 15 and 18 hours from start of sequencing, respectively. With this information available, correct treatment would have been possible significantly sooner, likely improving outcome.
Conclusions: We have adapted MPS to fit into the dynamic, multidisciplinary work-flow of acute metabolic medicine. As the extent of irreversible damage in patients with IEM often correlates with timing and accuracy of management in early, critical disease stages, our novel methodology is predicted to improve patient outcome.
All procedures have been designed such that they can be implemented in any technical setting and to any genetic disease area. The strategy conforms to international guidelines for clinical MPS, as only validated disease genes are investigated and as clinical specialists take responsibility for translation of results.

TBC-5: FCM-EM: A Novel Approach for Discovering DNA Motifs.

Ali Yousefian-Jazi[1], Jinwook Choi[2]

[1]Interdisciplinary Program, Bioengineering Major, Graduate School, Seoul National University, Seoul 151-742, Korea
[2]Department of Biomedical Engineering, College of Medicine, Seoul National University, Seoul 110-744, Korea

Abstract

Sequence motifs are important tools in molecular biology. A DNA motif is defined as a nucleic acid sequence pattern that has some biological significance such as being DNA binding sites for a regulatory protein, i.e., a transcription factor.
Most of the earlier literature categorized motif finding algorithms into two major groups (1) word-based (string-based) methods that mostly rely on exhaustive enumeration, (2) probabilistic sequence models where the model parameters are estimated using maximum-likelihood principle or Bayesian inference. The probabilistic algorithms are not guaranteed to find globally optimal solutions, since they employ some form of local search, such as Gibbs sampling or expectation maximization (EM) that may converge to a locally optimal solution. MEME is a popular probabilistic motif discovery program which uses the EM algorithm to infer position frequency matrices (PFMs). MEME scales poorly with large datasets and inability to find the motifs including insertions and deletions.
This research aims to develop a framework for discovering DNA motifs, where fuzzy C-means membership functions, and an EM technique are employed to extract putative motifs. We present a method, FCM-EM, for discovering motifs allowing Indels.

The motif discovery problem can be thought as searching an unknown number of l-length subsequences from an N-length DNA sequence. Thus, all of the l-length subsequences from the given DNA sequence should be extracted in order to be statistically analyzed for the probability of being a motif instance. This probability is inversely proportional with the likelihood ratio of the nucleotide frequencies of the subsequence to the background model. The method proposed in this study handles these issues by performing the following steps: (a) clustering subsequences into a certain number of clusters, i.e., PWMs, using FCM (b) utilization of fuzzy membership values of each subsequence as an initial posterior probability values in EM technique (c) testing each PWM to see whether it is statistically interesting or not.

TBC-6: Evaluation of the content coverage of an adolescents' depression ontology

Hyesil Jung, RN[1], Hyeoun-Ae Park, PhD, RN, FAAN[1,2]

[1]College of Nursing, Seoul National University, Seoul, Korea
[2]Research Institute of Nursing Science, Seoul National University, Seoul, Korea,

Abstract

Objectives: The purpose of this study is to evaluate the ability of an adolescents' depression ontology to represent the concepts of the counselling records about youth depression.
Methods: To evaluate the content coverage of ontology, we 1) extracted narratives from 66 youth depression counselling records of the Korean Youth Counselling and Welfare Institute, 2) performed natural language processing (NLP) in R with extracted narratives, 3) collected nominal words emerged from NLP, 4) extracted concepts by analysing the meaning of the words from context, 5) mapped extracted concepts to the concepts in the ontology, and 6) validated mapping results with a youth mental health expert. If there was any disagreement during the validation process, we had additional meeting to reach a consensus.
Results: In total, 1,574 narratives were extracted from counselling records. Also 1,028 nominal words emerged from NLP and 706 unique concepts were extracted. In total, 529 (74.93%) out of 706 concepts were lexically and semantically mapped to concepts in ontology. Out of 706 concepts, 136 concepts were partially mapped: 100 concepts mapped to broader ontology concepts, 16 mapped to narrower ontology concepts, and 20 mapped to more than one ontology concept. There were 41 counselling records concepts not mapped to ontology concepts. Reflecting the mapping results and experts' opinion, we revised the ontology by adding synonyms and missed concepts and modifying the level of structure.
Conclusion: we evaluated the content coverage of an adolescents' depression ontology by mapping to the concepts in counselling records. The ontology can represent most of the youth counselling records concepts. We revised the ontology based on the evaluation results. Improved ontology could be more effective tool for collecting and analysing social data.

TBC-7: Clinical Interpretation Pipeline of Targeted Next-Generation Sequencing Cardiovascular Panels.

Jeongeun Lee[1], Ali Yousefian Jazi[1], Moon-Woo Seong[2], Jinwook Choi[3]

[1]Interdisciplinary Program for Bioengineering, Graduate School, Seoul National University, Seoul 151-742 Korea
[2]Department of Laboratory Medicine, Seoul National University College of Medicine, Seoul 110-744, Korea
[3]Department of Biomedical Engineering, Seoul National University College of Medicine, Seoul 110-744, Korea

Abstract

As targeted next-generation sequencing (NGS) provides the potential to assess all the known disease genes in a single assay, there is increasing efforts to utilize targeted NGS panels in clinical laboratories. Although currently several companies are offering various cardiovascular genetic testing services to support these efforts, interpretations of the sequence data still need additional processes to meet clinical requirements because usually these services are ended up with variants calling and simple annotations. Here we developed a clinical interpretation pipeline of cardiology next-generation sequencing panel that annotate clinical significance of each variants and related clinical information, so to facilitate the diagnosis of cardiac diseases. Our pipeline provides information about in which cardiac disease category each variant may involve and how significant the variant is in clinical aspect. The system assesses clinical significance of each genetic alteration by using both an in-house mutation frequency database and public databases.

TBC-8: Re-analysis of RNA-sequencing data reveals cancer-specific chimaeric transcripts.

Inge Seim[1,2,3], Patrick B. Thomas[1,2,3], John Lai[3], Penny L. Jeffery[1,2,3], Lisa K. Chopin[1,2,3]

[1]Comparative and Endocrine Biology Laboratory, Translational Research Institute-Institute of Health and Biomedical Innovation (TRI-IHBI), Queensland University of Technology, 37 Kent St., Woolloongabba, Queensland, 4102, Australia
[2]Ghrelin Research Group, Translational Research Institute-Institute of Health and Biomedical Innovation (TRI-IHBI), Queensland University of Technology, 37 Kent St., Woolloongabba, Queensland, 4102, Australia

[3]*Australian Prostate Cancer Research Centre–Queensland, Queensland University of Technology and Princess Alexandra Hospital, Woolloongabba, Queensland 4102, Australia*

Abstract

The peptide hormone ghrelin is a potent appetite-regulating hormone produced predominantly in the stomach. It has a number of other biological actions, including roles in energy balance, the stimulation of growth hormone release and the regulation of cell proliferation. Recently, several ghrelin gene (GHRL) splice variants have been described. Here, we attempted to identify conserved alternative splicing of GHRL by cross-species sequence comparisons. GHRL was identified in diverse taxa, ranging from fish to human. Interestingly, GHRL is inactivated in Saker falcon and Peregrine falcon, apex predators, and egg-laying mammals, the platypus and echidna. This is suggestive of distinct metabolism in these species. Comparisons of GHRL orthologs revealed conservation of exons coding for a C-terminally truncated form of the ghrelin peptide which we have termed minighrelin. Minighrelin-encoding GHRL variants are expressed in human and mouse tissues, demonstrating conservation of alternative splicing spanning at least 90 million years. Minighrelin appears to have similar actions to canonical ghrelin in vitro and in vivo. This is the first study employing next-generation sequencing data to compare GHRL coding sequences of vertebrate taxa and identify novel derived transcripts. This work adds further impetus for studies into alternative splicing of the ghrelin gene and the function of novel ghrelin peptides in vertebrates.

TBC-9: A Two-Stage Approach for Multivariate Infrequent Adverse Drug Events Analysis in Electronic Medical Records.

Siriwon Taewijit[1,2], Tu Bao Ho[1]

[1]*Japan Advanced Institute of Science and Technology, Ishikawa 9231211, Japan*
[2]*Sirindhorn International Institute of Technology (SIIT), Thammasat University, Pathum Thani 12121, Thailand*

Abstract

Most of drug safety researches in text mining−based methods rely on bivariate analysis of relationship between a drug and an adverse event (AE). Our research is beyond such the first order problem by investigation of multivariate relationships among infrequent AEs induced by multiple drugs item (MD), which is often appeared in the reality situation.
There are several limitations of the existing methods. Adaptive of bivariate based-methods contributes to suffer from the exponential time-consuming issue due to the combination of drug−AE computation. Favoured co-occurrence distribution− and predefined rule−based methods are usually lacked of domain knowledge integrating to support research hypothesis and reliable results. On the other hand, most of the existing MD−AEs analysis deploys the well−known frequent pattern mining. Unfortunately, the method is not only to result the trivial and redundant patterns, which have been reported, but also to abandon a significant relationship, which is infrequent. According to the natural characteristic of prescribed drug combination, a physician, in general, prescribes safety multiple drugs item, thus results the conventional method failed into capture harm related low probability events.
To overcome the challenges, we propose a two-stage approach to disclose the novel infrequent MD- AEs from the unstructured text with high significant level as well as low time complexity. Our algorithm first visuals EMRs by semantic network, where a node represents the drug concept derived from UMLS, and an edge represents the distribution strength between each node-pair. The relationship information in this structural pattern representation leverages our multivariate relationships analysis. Then, we propose the semantic relatedness weighted index to enrich domain knowledge to the dependency graph. Finally, we deploy semantic subgraphs mining to derive the list of common MD that has high potential to cause an AE. The second stage, infrequent casual relation among such discovered MD and AEs are extracted based on the successful statistical measure of disproportionality.
In summary, we are not only to contribute the feature representation for MD−AEs signal detection, but also propose the novel framework for multivariate relationships analysis of MD−AEs. The further results are either automatically proven by our pipeline process, biological based-network of drug-AE, or manually proven by physician in our project to filter out the clinically interested events for drug safety awareness.

TBC-10: Time dependent pathway regulation for cancer development.

Shinuk Kim[1]

[1]*SangMyung University, Cheonan Chung-Nam 440-816, Republic of Korea*

Abstract

In this study, we developed an algorithm converting time independent data sets to time dependent data sets for inferring pathway regulatory networks along with cancer development. Glioblastoma multiforme (GBM) gene expression datasets from The Cancer Genome Atlas (TCGA) and KEGG pathway for pathway information were downloaded for the study materials. Gene Set Enrichment Analysis (GSEA, www.broadinstitute.org/gsea/) was used for gene sets

enrichment analysis for selecting meaningful pathways between normal and cancer patients groups. We stratified patients by using cancer development criteria such as survival time, grade and stage. The groups of stratified patients were arranged via given modules to describe cancer development. Comparing with normal data sets, the enriched pathways corresponding to each group were defined with core gene sets. We also developed mathematical model for pathway regulatory networks using those enriched pathways and core gene sets.

TBC-11: Medical Synonym Extraction with Dual Space Model.

Yuki Nakayama[1], Ryosuke Matsuo[1], and Tu Bao Ho[1]

[1]School of Knowledge Science, Japan Advanced Institute of Science and Technology, 1-1 Asahidai, Nomi, Ishikawa, Japan

Abstract

Information in narrative style document of electrical medical records (EMRs) is used in the immediate treatment of patients as well as the medical research. Because of advances in the medical language processing, it becomes possible to automatically extract information in EMRs. For information extraction in medical domain, a very important task is to extract synonyms in terms of medical terminology. By creating a dictionary that links concepts to their synonyms, it becomes possible to improve the accuracy of information extraction. In the medical domain, synonym extraction is often performed by medical lexicon such as UMLS (Unified Medical Language System). However, there are still a lot of medical expressions occurring within medical documents that are not included in such lexicon. Moreover, such medical language use changes with relatively high speed. Consequently, manual construction of medical lexicon that accurately reflects changes of terminology is costly, and may also result in low coverage. That is why automatic methods for extending synonym terminology are highly expected. The aim of this study is to apply the distributional hypothesis model with the assumption that words occurring in similar contexts likely have similar meaning- to the corpus of enormous medical documents in order to extract synonyms of preferred terms of SNOMED CT. In addition, we extract synonym phrases as well as synonym words. Although most of the approaches of synonym extraction uses technique that recognize phrases as terms when extracting synonym phrases, frequency of occurrence of terms decreases exponentially with the length of phrases. Then, it becomes difficult to express exact meaning of terms. For that reason, we not only applied Turney's model that uses similarity of topic and of function in order to compare each terms that include words and phrases, also developed model to extract synonymous phrases consisting of three or more words.

TBC-12: Multilayered approach to extract temporal events and expressions from clinical narratives.

Moharasan Gandhimathi[1], and Ho Tu Bao[1]

[1]School of Knowledge Science, Japan Advanced Institute of Science and Technology, Nomi, Ishikawa, 921-1211, Japan

Abstract

Implementation of electronic medical records (EMRs) produces a huge amount clinical data in the form of unstructured text. It is inevitable to convert the unstructured text to structured format, which makes easy to access the converted text and process by machine. One way is to transform clinical text is to detect and represent the Temporal Information (TI). Natural language processing (NLP) plays a significant role in discovering and transforming clinical text to structured representation in the TI form. To improve the accuracy of temporal information detection (medical terms and phrases) in clinical text, various external resources have been used with Conditional Random Fields (CRF) and hand-coded rules. However, existing methods have faced the difficulty in defining the long-range dependencies of sequential label prediction. To overcome this we are proposing the multilayered sequence labelling approach with significant external resources. In this layered approach, we detect the problem, treatment and test events with temporal expressions in the first stage. In the second stage, we use the first stage output as additional information to detect the rest of the evidential, clinical department and occurrence events. We used the partial I2B2 dataset, which has 190 records for training data and 120 records for evaluation in the original dataset. In this poster, we illustrate the problem formulation, introduce the key idea of our approach, describe the procedure and produce the initial stage results and ideas to further improvements of accuracy.

TBC-13: A Severity based Information Retrieval for Electronic Medical Records.

Ryosuke Matsuo[1], and Tu Bao Ho[1]

[1]Japan Advanced Institute of Science and Technology, Ishikawa, 923-1292, Japan

Abstract

The vector space model is the most popular one for information retrieval where term weighting (assigning a weight to each term in the document representation) plays an essential role. While the term weighting in vector space model often employs the term frequencies (TF-IDF and its variants), semantic term weighting is necessary in certain domains, in particular in medicine. In this work we propose a novel framework of semantic term

weighting for medical information retrieval, especially from electronic medical records (EMRs).
The term weighting in our proposed framework composes of three factors: TF-IDF, medical importance of the term and the patient status. The severity of patient is essential to identify the two new factors in addition to TF-IDF. Understanding of the patient severity is the critical element in medical domain as eminent medical ontology ICD that has been intimated with mortality. With two new semantic factors in term weighting, our method allows us to retrieval in EMRs databases with higher precision than method just uses TF-IDF.
We present the framework, the method of determining the medical importance of terms, and the experiment results from the TREC 2014 data.

TBC-14: Organization of the Mammalian Metabolome and Ionome According to Organ Function, Lineage Specialization, and Longevity.

Siming Ma[1,*], Sun Hee Yim[1,2,*], Clary B. Clish[2], Vadim N. Gladyshev[1,2]

[1]Harvard Medical School
[2]Broad Institute
[*]Equal contribution

Abstract

Biological diversity among mammals is remarkable. Mammalian body weights range seven orders of magnitude and lifespans differ more than 100-fold among species. While genetic, dietary, and pharmacological interventions can be used to modulate these traits in model organisms, it is unknown how they are determined by natural selection. By profiling metabolites and elements in brain, heart, kidney, and liver tissues of 26 mammalian species representing ten taxonomical orders, we report metabolite patterns characteristic of organs, lineages, and species longevity. Our data suggest different rates of metabolite divergence across organs and reveal patterns representing organ-specific functions and lineage-specific physiologies. We identified metabolites that correlated with species lifespan, some of which were previously implicated in longevity control. We also compared the results with metabolite changes in five long-lived mouse models and observed some similar patterns. Overall, this study describes adjustments of the mammalian metabolome according to lifespan, phylogeny, and organ and lineage specialization.

TBC-15: Findings from the Critical Assessment of Genome Interpretation, a community experiment to evaluate phenotype prediction.

Steven E. Brenner[1,*], John Moult[2], CAGI Participants[3]

[1]University of California, Berkeley, CA, USA
[2]IBBR, University of Maryland, Rockville, MD, USA
[3]The CAGI participants are: Predictors: Allison Abad, Ivan Adzhubey, Yesim Aydin Son, Benjamin Bachman, Violeta Beleva- Guthrie, Bonnie Berger, Brady Bernard, Marcus Breese, Yana Bromberg, Chen Cao, Emidio Capriotti, Rita Casadio, Chien-Yuan Chen, Shann-Ching Chen, Yun-Ching Chen, Melissa Cline, Andrea Corredor, Carla Davis, Greet De Baets, Mark Diekhans, Rezarta Islamaj Dogan, Christopher Douville, Roland Dunbrack, Andrea Eakin, Carlo Ferrari, Anna Flynn, Adam Frankish, Zaneta Franklin, Manuel Giollo, Nina Gonzaludo, Julian Gough, Jennifer Harrow, Ramin Homayouni, Raghavendra Hosur, Cheng Lai Victor Huang, Sohyun Hwang, Tadashi Imanishi, Chan-Seok Jeong, Yuxiang Jiang, Daniel Jordan, Rachel Karchin, Panagiotis Katsonis, Dongsup Kim, Eiru Kim, Jack Kirsch, Michael Kleyman, Pui-Yan Kwok, Ernest Lam, Insuk Lee, Pietro Di Lena, Emanuela Leonardi, Biao Li, Jun Li, Olivier Lichtarge, Chiao-Feng Lin, Rhonald Lua, Angel Mak, Pier Luigi Martelli, David Masica, Sean Mooney, Zev Medoff, Matthew Mort, John Moult, Steve Mount, Eliseos Mucaki, Jonathan Mudge, Katsuhiko Murakami, Yoko Nagai, Noushin Niknafs, Abhishek Niroula, Yanay Ofran, Ayodeji Olatubosun, Kymberleigh Pagel, Nathaniel Pearson, Vikas Pejaver, Jian Peng, Alexandra Piryatinska, Catherine Plotts, Predrag Radivojac, Aliz Rao, Lipika Ray, Graham Ritchie, Aharon Rodie, Peter Rogan, Frederic Rousseau, Jana Marie Schwarz, Joost Schymkowitz, George Shackelford, Maxim Shatsky, Jung Eun Shim, Junha Shin, Ilya Shmulevich, Brad Silver, Nathan Stitziel, Andrew Su, Paul Tang, Nuttinee Teerakulkittipong, Janita Thusberg, Silvio Tosatto, Yemliha Tuncel, Tychele Turner, Ron Unger, Gurkan Ustunkar, Jouni Valiaho, Joost Van Durme, Mauno Vihinen, Mary Wahl, Xinyuan Wang, Li-San Wang, Chunlei Wu, Qiong Wei, , Lijing Xu, Yuedong Yang, Christopher Yates, Yizhou Yin, Chen-Hsin Yu, Dejian Yuan, Maya Zuhl; Data providers: Adam P. Arkin, Madeleine Price Ball, Jason Bobe, George Church, Andre Franke, Nina Gonzaludo, Emma D'Andrea,Lisa Elefanti, Joe W. Gray, Linnea Jannson, John P. Kane , Pui-Yan Kwok, Rick Lathrop, Angel C. Y. Mak, Mary J. Malloy, Chiara Menin, John Moult, Robert Nussbaum, Lipika R. Pal, Clive R. Pullinger, Jasper Rine, Maria Chiara Scaini, Jeremy Sanford, Nicole Schmitt, Jay Shendure, Michale Snyder, Tim Sterne-Weiler, Paul L. F. Tang, Sean Tavtigian, Silvio Tosatto; Assessors: Rui Chen, Roland Dunbrack, Iddo Friedberg, Gad Getz, Rachel Karchin, Alexander Morgan, Sean Mooney, John Moult, Robert Nussbaum, Jeremy Sanford, David B. Searls, Artem Sokolov, Josh Stuart, Shamil Sunyaev, Sean Tavtigian, Silvio Tosatto; Organization and Management: Daniel Barsky, Navya Dabbiru, Gaurav Pandey, Sadhna Rana, Susanna Repo, Rajgopal Srinivasan, Sri Jyothsna Yeleswarapu, Stephen Yee, Maya Zuhl; Advisory Board: Russ Altman, George Church, Tim Hubbard, Scott Kahn, Sean Mooney, Pauline Ng, Susanna Repo; Scientific Council: Patricia Babbitt, Atul Butte, Garry Cutting, Laura Elnitski, Reece Hart, Ryan Hernandez, Rachel Karchin, Robert Nussbaum, Michael Snyder, Shamil Sunyaev, Joris Veltman, Liping Wei

Abstract

The Critical Assessment of Genome Interpretation (CAGI, 'kā-jē) is a community experiment to objectively assess computational methods for predicting the phenotypic impacts of genomic variation. In the experiment, participants are provided genetic variants and make predictions of resulting phenotype, for ten

challenges. These predictions are evaluated against experimental characterizations by independent assessors.
For example, in a challenge to predict Crohn's disease from exomes, several groups performed remarkably well, with one group achieving a ROC AUC of 0.94. The experiment also revealed important population structure to Crohn's disease in Germany. In another challenge, two groups were able to successfully map a significant number of Personal Genome Project complete genomes to corresponding trait profiles.
Other challenges were to predict which variants of BRCA1, BRCA2, and the MRN complex are associated with increased risk of breast cancer; to associate exomes, variants, and disease in lipid diseases; to predict how variants in p53 gene exons affect mRNA splicing; to predict how variants of p16 tumor suppressor protein inhibit cell proliferation; and to identify potential causative SNPs in disease- associated loci.
Overall, CAGI revealed that the phenotype prediction methods embody a rich representation of biological knowledge, making statistically significant predictions. However, the accuracy of prediction on the phenotypic impact of any specific variant was unsatisfactory and of questionable clinical utility. The most effective predictions came from methods honed to the precise challenge. Prediction methods are clearly growing in sophistication, yet there are extensive opportunities for further progress.
The fourth CAGI experiment is presently underway, with a prediction season running through November 2015 and a meeting planned for March 2016.
Complete information about CAGI may be found at http://genomeinterpretation.org.

TBC-16: Building a Metastasis Gene Database using Principle-based Approach.

Onkar Singh[1], Emily Chia Yu Su[1,*], and Hong-Jie Dai[2,*]

[1]Graduate Institute of Biomedical Informatics, Taipei Medical University, Taiwan.
[2]Department of Computer Science and Information Engineering, National Taitung University, Taiwan

Abstract

Metastasis is a major cause of mortality from cancer wherein cancer cells escape the primary tumor site and repopulate a different target organ. Despite its vast clinical importance and inherently fascinating underlying biology, metastatic tumors are often shows poor prognosis. Thus, the development of a metastasis gene database might be helpful for the manifestation of metastasis events. In this work, a flexible knowledge representation scheme, based on Infomap along with a partial matching algorithm that enables a single rule to match a lots of semantically similar expressions with high accuracy were used to extract relation among genes and metastasis-related concepts. We referred to it as the principle-based approach. In this approach, a collection of frames were developed to represent linguistic concepts. Each frame is a collection of slots with relations specified among them. A slot can be a word, phrase, semantic category, or another frame concept. One can specify position relations, collocation relations, agreement relations and others among its slots. These frames are composited by nine concepts including gene, microRNA, neoplasm metastasis, cytoskeleton, cell movement, cell adhesion neoplasms, organ, and tissues. The generated frames contain slots like [Gene]:[Positive event]: [Metastasis], which could be used to extract relations between genes and metastasis. To assess the performance of frames generated by the principle-based approach, the keyword "EMT and TGF-β [title/abstract]" was used to search abstracts on PubMed. 300 metastasis-related abstracts were randomly selected and applied to the constructed frames on the retrieved abstracts to extract different relations. The results were manually verified by our in-lab biologists. The developed frames achieved a Precision/Recall/F-score (PRF) of 68%/76%/72%. All the extracted relations were verified and store in a database, which might be helpful for biologists for understanding metastasis related pathways.

TBC-17: Comprehensive analysis of pediatric neuroblastoma using high depth cancer gene panel sequencing.

Joon Seol Bae[1]

[1]Samsung Medical Center, Seoul, Korea

Abstract

Background: Pediatric neuroblastoma (NB) is the most common solid tumor in children. The genetic landscape of the pediatric NB is unknown, although oncogenic variations have been identified by whole exome sequencing in Caucasian patients.
Methods: To investigate the genetic variations of pediatric NB, genomic DNA from sixty four pediatric NB patients were sequenced by targeted deep exome sequencing of CancerSCAN panel for 81 oncogenes.
Results: Ninety four percent of patients had at least one mutation, and total 256 somatic mutations (186 mutations and 6 Indels with freq. > 5%) and 107 copy number alterations (53 amplifications and 54 deletions) were identified. However, novel translocation of target genes was not found in this study. The most common mutations were found in the TP53, ROS1, ALK, ATM, EGFR, NF1 and PTCH1 genes (freq. > 10%). Compared with previous studies, we identified more recurrent variations in TP53 (25.4% in this study vs. 0.4% in previous study), ROS1 (15.9% vs. 1.3%), and ALK (14.3% vs. 9.2%). Interestingly, a deletion (rs3841650) at chromosome 15 was first discovered in IGF1R gene for our twenty three patients. It has been known that the expression of IGF1R

gene was related with pathogenesis of NB. The ALK R1275Q mutation and IGFR1R deletion were successfully validated by digital PCR.
Conclusions: In this study, we identified actionable mutations, considering for targeted treatment as ALK R1275Q, ALK F1741I, EGFR T790M, HRAS Q61R, and NRAS F12D, and a known mutation as TP53 G245D in COSMIC database. Our findings may help to understand the etiology of pediatric NB, and provide useful information in personalized anti-cancer drug treatment.

TBC-18: Tumor heterogeneity of advanced primary lung cancer evaluated by multiregion sequencing.

Je-Gun Joung[1], Joon Seol Bae[1], Su Yeon Lee[1], Jinha Park[2], Sang-Won Um[2], Woong- Yang Park[1]

1Samsung Genome Institute, Samsung Medical Center, Seoul, Korea
2Division of Pulmonary and Critical Care Medicine, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

Abstract

Cancers are composed of cells with distinct genetic and phenotypic characteristics. Tumor heterogeneity in each patient may influence on personalized medicine strategies that depend on results from single tumor-biopsy samples. However, tumor heterogeneity in advanced primary lung cancer has not been well studied. In order to investigate the tumor heterogeneity of lung cancer, we performed whole exome sequencing and RNA sequencing for multiple regions from four patients who were suspicious for primary lung cancer with multiple lymph node metastasis. Tumor samples were obtained from primary tumor and metastatic lymph node(s) or distant metastasis. The profiles of somatic mutations, copy number alterations and gene expression were compared among primary tumor and metastatic sites in each patient. There was profound tumor heterogeneity in terms of somatic mutation, copy number alteration, or gene expression in each patient. High frequency of private somatic mutation from spatially distinct regions was observed in two patients (59% in TH1, 44% in TH2) and known driver mutations were also heterogeneous. Analysis of the clonal architecture confirmed the presence of clonal heterogeneity in each patient. Phylogenetic reconstruction revealed the evidence of branched clonal evolution from advanced primary lung cancers to distinct regions. The single biopsy specimens may not represent the whole landscape of genetic alteration in advanced primary lung cancers.

TBC-19: Measuring Emotional Well-being Based on Physiological Indicators and Lifestyles.

Chae-Gyun Lim[1], Hyo Jin Do[1], and Ho-Jin Choi[1]

[1]School of Computing, Korea Advanced Institute of Science and Technology, 291, Daehak-ro, Yuseong-gu, Daejeon 305-338, Republic of Korea

Abstract

In modern society, maintaining people's mental health is important to ensure the high quality of life because there are various factors that make people stressful in their real life. However, it is difficult to measure current state of the mental health directly due to a characteristic of human affect which has subjective and conceptual quality. To sustain the mental health of people, new measurements targeting on their emotional well-being in daily life are required.
Many existing researches are particularly attuned to people with mental illness and psychiatric care. Some of the researches aimed to find implicative symptoms of mental disorders such as depression by analysing data from online social media. Other researches proposed methods integrating different types of features obtained from several sources to detect user's emotional states, but they targeted on patients who need psychiatric care. In this paper, we study estimates of mental health focused on daily lives of general people.
We propose a new measurement of human emotional well-being by integrating a user's daily lifestyles and physiological data such as heart rate, respiration, skin conductance, etc. By monitoring activities of the user, our system automatically collects lifestyle patterns which are repeated regularly. The physiological information obtained from wearable sensing devices may constitute a guide of user's affective reaction. In some cases, however, the information could not characterize the user's actual emotion due to its ambiguity. To avoid the issue that certain features may be unreliable, we combined multiple features extracted from lifestyles and physiological information by using machine learning techniques. Finally, we discuss our expectation about future implications of the proposed method for sustaining people's mental health.

TBC-20: Unraveling the association between mRNA expressions and mutant phenotypes in a genome-wide assessment of mice.

Meng-Pin Weng[1] and Ben-Yang Liao[1,*]

[1]Division of Biostatistics & Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County 350, Taiwan, R.O.C.

Abstract

High-throughput gene expression profiling has revealed substantial leaky and extraneous transcription of eukaryotic genes, challenging the perceptions that

transcription is strictly regulated and that changes in transcription have phenotypic consequences. To assess the functional implications of mRNA transcription directly, we analyzed mRNA expression data derived from microarrays, RNA-sequencing, and in situ hybridization, together with phenotype data of mouse mutants as a proxy of gene function at the tissue level. The results indicated that despite the presence of widespread ectopic transcription, mRNA expression and mutant phenotypes of mammalian genes or tissues remain associated. The expression-phenotype association at the gene level was particularly strong for tissue-specific genes, and the association could be underestimated due to data insufficiency and incomprehensive phenotyping of mouse mutants; the strength of expression-phenotype association at the tissue level depended on tissue functions. Mutations on genes expressed at higher levels or expressed at earlier embryonic stages more often result in abnormal phenotypes in the tissues where they are expressed. The mRNA expression profiles that have stronger associations with their phenotype profiles tend to be more evolutionarily conserved, indicating that the evolution of transcriptome and the evolution of phenome are coupled. Therefore, mutations resulting in phenotypic aberrations in expressed tissues are more likely to occur in highly transcribed genes, tissue-specific genes, genes expressed during early embryonic stages, or genes with evolutionarily conserved mRNA expression profiles.

TBC-21: Prevalence and Outcome of Seizure related to Systemic Lupus Erythematosus.

Yu-Hsuan Hu[1] and Kuang-Hui Yu[2]

[1]Chang Gung Memorial Hospital, No.5, Fu-Shin St., Kuei-Shan (333), Tao-Yuan County, Taiwan
[2]Division of Allergy, Immunology and Rheumatology, Chang Gung Memorial Hospital No.5, Fu-Shin St., Kuei-Shan (333), Tao-Yuan County, Taiwan

Abstract

Objective: To investigate the risk factors associated with systemic lupus erythematosus (SLE) related seizure and the long-term outcome of SLE related seizure. Methods: The medical records of 691 SLE hospitalized patients followed up in Chang Gung Medical Center in Taiwan from 2005 to 2014 were retrospectively reviewed. Results: 691 patients (11.3% men) who fulfilled at least four of the ACR criteria for SLE were studied. The mean onset age of SLE was 36.3 ± 16.4 years. 72 patients (10.4%) had SLE-related seizure. Univariate analysis indicated that independent predictors of seizure were photosensitivity, serositis, proteinuria, leukopenia, lymphopenia, thrombocytopenia, psychosis, hypertension, diabetic mellitus, and end stage renal disease. Multivariate Cox regression analysis indicated that independent predictors of mortality were onset age above 50 years at admission (HR 2.62, 95% CI 1.54–4.45, p < 0.001), male (HR 2.36, 95% CI 1.24-4.49, p =0.009), serositis (HR 2.86, 95% CI 1.72-4.76, p < 0.001), leukopenia (HR 2.06, 95% CI 1.14-3.70, p=0.016), thrombocytopenia (HR 3.84, 95% CI 2.08-7.09, p < 0.001), seizure (HR 1.94, 95% CI 1.12-3.35, p=0.018), and anti- dsDNA positivity (HR 0.43, 95% CI 0.25- 0.74, p=0.002). Conclusions: This study underlines the prevalence and outcome of seizure in SLE, which is significant related to patient survival rates. Survival was significantly lower in patients with seizure than that in patients without. Close follow up of SLE patients with seizure may be necessary.

TBC-22: Benzodiazepines use and breast cancer risk: A population-based study.

Usman Iqbal[1], Phung-Anh Nguyen[1], Shabbir Syed-Abdul[1], Richard Lu[1], Hsuan-Chia Yang[1], Chih-Wei Huang[1], Wen-Shan Jian[1,2,*], Yu-Chuan (Jack) Li[1,3,*]

[1]Graduate Institute of Biomedical Informatics, College of Medicine Science and Technology, Taipei Medical University
[2]School of Health Care Administration, Taipei Medical University
[3]Department of Dermatology, Taipei Medical University - Wan Fang Hospital, Taipei, Taiwan

Abstract

Introduction: Benzodiazepines (BZDs) are the most commonly used drugs in general population due to their anxiolytic and sedative effects and its use was observed as twice as much among females around the world. The carcinogenicity of BZDs is still unclear. We aimed to assess whether long-term benzodiazepines use is risk for breast cancer.
Methods: We used reimbursement data from the Bureau National Health Insurance (NHI) system in Taiwan and has registered all medical claims since 1996. We obtained the randomly selected two million sample population of NHI beneficiaries claim data from 1998 to 2009 year in Taiwan. We conducted a matched case-control study of the association between benzodiazepines and breast cancer. We identified cases with a first time breast cancer who were matched to cancer-free control (1:6) by using propensity score during Jan. 2001 to Dec. 2008. We also observed the outcomes according to the length of exposure and defined daily dose. To estimate the risk for breast cancer, we adjusted with potential confounding factors such as comorbid disease and other medications. The conditional logistic regression was used to analyse the results with 95% confidence intervals (CI).
Results: The adjusted odd ratio (OR) for breast cancer associated with overall BZDs use was 1.14 (95% CI, 1.05 to 1.24). While we looked at BZDs classes and found anxiolytics unsafe 1.12 (95% CI, 1.03 to 1.22) among other BZDs classesforbreastcancerrisk.
Conclusion: The BZDs use was associated with an overall increase breast cancer risk. Antiepileptics, hypnotics and sedatives, BZDs related found safer except anxiolytics. Our findings might provide evidence on the carcinogenic

effects of benzodiazepines. This could provide information in order to help physicians to select BZDs in future treatment.

TBC-23: A web browser extension to display phenome-wide association strengths observed in Taiwanese population: PWAS-PubMed.

Shabbir Syed-Abdul[1], Hong-Jie Dai[1], Po-Ting Lai[2], Usman Iqbal[1], Phung-Anh Nguyen[1], Richard Lu[1], Hsuan-Chia Yang[1], Chih-Wei Huang[1], Wen-Shan Jian[1,3], Yu-Chuan (Jack) Li[1,4,*]

[1]Graduate Institute of Biomedical Informatics, College of Medicine Science and Technology, Taipei Medical University
[2]Institute of Information Science, Academia Sinica, Taiwan
[3]School of Health Care Administration, Taipei Medical University
[4]Department of Dermatology, Taipei Medical University - Wan Fang Hospital, Taipei, Taiwan

Abstract

Introduction: Despite the rapid advancement in medical sciences, yet researchers are unclear about the holistic picture of disease associations. The study utilize augmented browsing technique to combined the quantified phenome-wide associations strengths (PWAS) observed in Taiwanese population with the literature searching results on PubMed by developing a browser extension.
Methods: We applied an informatics approach towards utilizing the patient phenotypes information from Taiwan's national health insurance research database (NHIRDB) to explore the disease-wide associations. PWAS database was created with a total of about 60.71 million of phenome-wide association values for 9.73 million male and 10.38 million female patients. These associations can be accessed online at http://associations.phr.tmu.edu.tw. This work intends to integrate the information of PWAS database on PubMed by developing a browser extension for PubMed website. The client side of the PWAS-PubMed extension was implemented by using the browser extension application programming interface (API). Because the coding system used by NHIRDB is ICD9CM, the recognized disease term mentions must be mapped to the ICD9CM codes. The maximum matching algorithm is then employed to combine the results from the two methods.
Results: This study develops a web browser extension to augment PubMed search results with phenome-wide association strengths. The users of web browser extension not only see the number of articles in the PubMed database discussing the two diseases but also will know the association strengths as shown in figure 1.
Conclusion: This unique PWAS database, when combined with existing online published literature of phenotypic, genetic and proteomic datasets could make an immense impact on our understanding of disease pathogenesis and their associations.

TBC-24: In silico prediction of pathogenicity of missense substitutions: a perspective of complex diseases

Kaavya A Mohanasundaram[1], Mani P Grover[1], Tamsyn M Crowley[1,2], Andrzej Goscinski[3], Merridee A Wouters[1,4]

[1]School of Medicine, Deakin University, Waurn Ponds, Geelong, Victoria, Australia
[2]Australian Animal Health Laboratory, CSIRO Biosecurity Flagship, Geelong, Victoria, Australia
[3]School of Information Technology, Faculty of Science Engineering and Built Environment, Deakin University, Waurn Ponds, Geelong, Victoria, Australia
[4]Olivia Newton-John Cancer Research Institute, Heidelberg, Victoria, Australia

Abstract

Non-synonymous single nucleotide polymorphisms (nsSNPs) are single nucleotide variations which occur in protein-coding regions of genes, substituting the wild-type amino acid with a new one, thus altering the protein sequence. These substitutions may affect the protein's structure and function and are thus of great interest to the study of human diseases. While cheap and quick sequencing of DNA, particularly exome sequencing, has allowed easy identification of variations in the genes of individuals with a common phenotype, distinguishing which variations are causal of the disease, and which are harmless, is not straightforward. Here, we review computational methods for distinguishing pathogenic and harmless variations. These amino acid substitution (AAS) methods can be classified roughly as "site sensitivity" or "distance" methods. The techniques use bioinformatic and/or physicochemical criteria derived from protein sequence and structure information to make predictions. We also look at their comparative performance on loss-of-function and gain-of-function mutations; and discuss the employment of these methods for diseases of Mendelian inheritance, cancer and also complex diseases, under both the rare variant and common variant scenarios. We suggest different methods are likely to be needed to understand complex diseases, particularly under the common variant scenario, and review efforts in this direction.

TBC-25: Using Data Visualization Method to Estimate and Predict Chronic Kidney Disease Medical Expenditures

Chih-Wei Huang[1,2], Shen-Hsien Lin[2], Phung Anh Nguyen[2], Usman Iqbal[1,2], Wen-Shan Jian[3], Yu-Chuan (Jack) Li[1,2]

[1]*Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University*
[2]*College of Medical Science and Technology, Taipei Medical University*
[3]*School of Health Care Administration, Taipei Medical University, Taipei 11031, Taiwan*

Abstract

At present, the chronic diseases are major cause of death for older people. Among chronic diseases, kidney disease was observed the main leading cause of death in which is ranked sixth in men, fifth in women, while over 65 with Chronic Kidney Disease (CKD) and all other ages. The end-stage renal disease (ESRD), it takes about 36% of the health care budget. Individuals with chronic kidney disease entering in the country with renal dialysis stage is growing 6% annually. Therefore, it has a heavy burden on society with deep finances influence. The retrospective cohort study research method were applied by using the National Health Insurance Database (NHIRD) of one million population (1997-2011 years) and selected patients with chronic kidney disease information to conduct a secondary data analysis. The aim of this study was to visualize patients with chronic kidney disease and calculate their medical expense in Taiwan. Because the Taiwanese medical expenses are under global budget and we calculated it for chronic kidney disease patients. This allows to evaluate the patients suffering from chronic kidney disease and their medical expenses prior to medical cost of chronic kidney disease. After linear relationship between the prognosis, to explore different intervals before and after chronic kidney disease from the timeline and calculated the medical costs associated with the consumption of their prognosis by applying paired t test to compare the medical expenses before and after CKD. Finally, the medical expenses along with different factors and disease information evaluated. By adding visual graphics, it offers physician's intuitive and efficient interactive query interface information to help patients make the best treatment decisions, to reduce the cost of medical institutions and enhance the quality of medical care of patients with chronic kidney disease.

TBC-26: A Statistical Model for the Refinement and Ranking of Variant Calls

Tony Kuo[1], Jun Sese[1], Martin Frith[1], and Paul Horton[1,*]

[1]*National Institute of Advance Industrial Science and Technology, AIST, Tokyo, Japan*

Abstract

Many human diseases have a genetic component. Thus, genome sequencing has become an important method in the study of genetic disorders. In particular, there is a great amount of interest in somatic mutations associated with cancer, as well as mutations associated with rare inherited diseases. As such, variant discovery has become a routine analysis method in research and in clinical settings. Currently, many tools and methods have been developed for the purpose of discovering variants. However, the results from different tools do not often agree and are very sensitive to small changes in parameters, implying the methods are not robust. It is known that uncertainties and ambiguities exist at all stages of the variant calling process. From alignment, to constructing a pileup of sequencing reads, and the variant calling process itself. This is likely due to the repetitive nature of the human genome. The calling of insertions and deletions in particular, tend to have difficult to resolve ambiguities and remains unreliable. This is especially true for variant calling on a single sample or individual.

Here, we present a statistical model to evaluate whether putative variants are significant. Given a set of variant calls, we construct the putative mutant genome. From which, we calculate the likelihood of the sequencing data given the reference genome versus the likelihood of the sequencing data given the mutant genome. Our method works under the principle of reciprocality and includes an analysis of the data with respect to the predicted mutant genome, something that current callers do not perform. Thus, our statistical model has explicit alternative hypotheses which we test against in order to refine and rank a set of variant calls.

We have begun to try our method on simulated and real data (targeted, exome and whole genome sequencing). Based on our manual inspection of numerous example of called variants, our ranking method appears promising as a way to reduce false positives without losing many true positives.

Venue

Belle Salle Nishi-Shinjuku (ベルサール西新宿)

- Address: 4-15-3 Nishi-Shinjuku, Shinjuku-ku, Tokyo, Japan, 160-0023
- Telephone: +81-3-3346-1396
- Homepage: http://bellesalle.co.jp/

Map

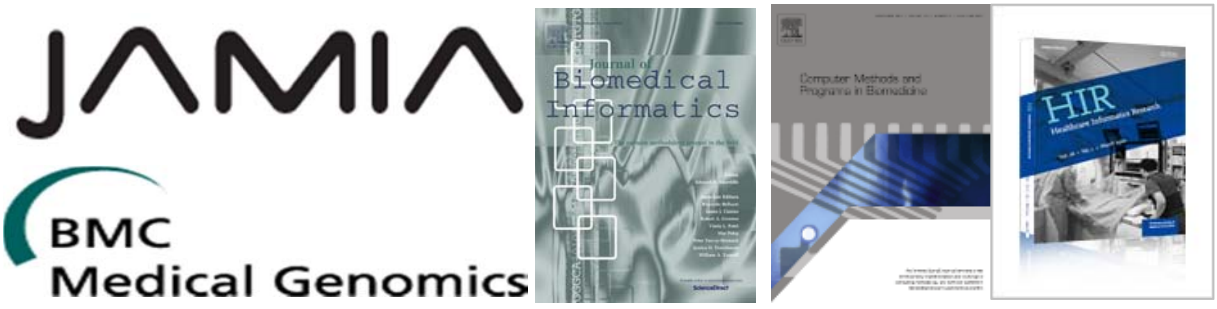
Conference App

Download now the TBC 2015 app for iOS and Android.

http://tbc2015.jp/tbc2015apps.html

Informatics Journals Supporting TBC

- **JAMIA** (Journal of American Medical Informatics Association)
- **JBI** (Journal of Biomedical Informatics)
- **BMC Medical Genomics**
- **CPBM** (Computer Methods and Programs in Biomedicine)
- **Healthcare Informatics Research**



Sponsors